Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **469,** 1691–1712 (2017)
Advance Access publication 2017 April 21

doi:10.1093/mnras/stx955

Comparing semi-analytic particle tagging and hydrodynamical simulations of the Milky Way's stellar halo

Andrew P. Cooper,[1]★ Shaun Cole,[1] Carlos S. Frenk,[1] Theo Le Bret[2,3] and Andrew Pontzen[2]

[1] *Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[2] *Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[3] *Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford OX1 3NP, UK*

Accepted 2017 April 14. Received 2017 March 17; in original form 2016 November 10

ABSTRACT

Particle tagging is an efficient, but approximate, technique for using cosmological *N*-body simulations to model the phase-space evolution of the stellar populations predicted, for example, by a semi-analytic model of galaxy formation. We test the technique developed by Cooper et al. (which we call STINGS here) by comparing particle tags with stars in a smooth particle hydrodynamic (SPH) simulation. We focus on the spherically averaged density profile of stars accreted from satellite galaxies in a Milky Way (MW)-like system. The stellar profile in the SPH simulation can be recovered accurately by tagging dark matter (DM) particles in the same simulation according to a prescription based on the rank order of particle binding energy. Applying the same prescription to an *N*-body version of this simulation produces a density profile differing from that of the SPH simulation by $\lesssim$10 per cent on average between 1 and 200 kpc. This confirms that particle tagging can provide a faithful and robust approximation to a self-consistent hydrodynamical simulation in this regime (in contradiction to previous claims in the literature). We find only one systematic effect, likely due to the collisionless approximation, namely that massive satellites in the SPH simulation are disrupted somewhat earlier than their collisionless counterparts. In most cases, this makes remarkably little difference to the spherically averaged distribution of their stellar debris. We conclude that, for galaxy formation models that do not predict strong baryonic effects on the present-day DM distribution of MW-like galaxies or their satellites, differences in stellar halo predictions associated with the treatment of star formation and feedback are much more important than those associated with the dynamical limitations of collisionless particle tagging.

Key words: methods: numerical – galaxies: dwarf – galaxies: haloes – galaxies: structure.

1 INTRODUCTION

A number of studies have used so-called particle tagging techniques to predict the distribution and kinematics of Milky Way (MW) halo stars accreted from tidally disrupted dwarf satellite galaxies (Bullock, Kravtsov & Weinberg 2001b; Bullock & Johnston 2005; De Lucia & Helmi 2008; Cooper et al. 2010; Tumlinson 2010; Libeskind et al. 2011; Rashkov et al. 2012). These techniques attempt to model both stars and dark matter (DM) with a single collisionless particle species in a cosmological *N*-body simulation by 'painting' subsets of the particles with stellar mass, according to a weighting function, without changing the mass of the particle used in the gravitational calculation. This is intended as an approximation to the more self-consistent approach of hydrodynamical simulations, in which a separate species of collisionless 'star particles' is inserted into the calculation to replace gas particles that become sufficiently cold and dense. This replacement is usually done according to a 'subgrid' model of star formation describing the state of the interstellar medium represented by the gas particle (e.g. Schaye et al. 2015). In particle tagging models, all the baryonic physics of galaxy formation (including dissipative cooling, star formation and feedback of mass and energy to the interstellar medium) are modelled semi-analytically[1] on the scale of DM haloes (De Lucia

★ E-mail: a.p.cooper@durham.ac.uk

[1] An alternative approach ignores the physics of galaxy formation and instead assigns stellar mass directly to DM haloes using theoretical or empirical scaling relations (e.g. Bullock & Johnston 2005; Rashkov et al. 2012; Laporte et al. 2013).

& Helmi 2008; Cooper et al. 2010). The assignment of stellar mass to DM is usually expressed as a function of the binding energy of the DM particles, in order to account for the prior dissipation of energy by the star-forming gas (Bullock et al. 2001a; Bullock & Johnston 2005; Peñarrubia, McConnachie & Navarro 2008).

The most significant differences between tagging schemes concern those two aspects of the approach – how the star formation histories (SFHs) of DM haloes are computed, and the algorithm used to associate stellar mass with specific particles in the *N*-body simulation. In Cooper et al. (2010, hereafter C10), we described a technique in which a semi-analytic model (GALFORM, in our case) is used to predict SFHs, and the tagging operation is carried out for every star-forming halo at every snapshot (so-called 'live' tagging; the common alternative is to tag DM only once per satellite halo, at the time when it crosses the virial radius of the 'main' halo). In each halo, *N*-body particles are ranked in order of binding energy and the stellar mass to be assigned is distributed equally among a fixed fraction ($f_{\rm mb}$) of the most bound (in Cooper et al. 2010, $f_{\rm mb} = 1$ per cent). This technique was applied to MW-like galaxies by C10 and later extended to more massive galaxies by Cooper et al. (2013, hereafter C13) and Cooper et al. (2015a), who give additional details of the method. To distinguish the scheme set out in those three papers from other particle tagging schemes in the literature, we refer to it hereafter as STINGS (*Stellar Tags In N-body Galaxy Simulations*).

The advantage of particle tagging over hydrodynamical methods is that the evolving phase-space distribution of the stellar component can be followed at much higher resolution and at much lower computational cost. This in turn allows phenomena that arise from the dynamics of hierarchical assembly, including stellar haloes and the scaling relations of elliptical galaxies, to be explored with a much wider range of galaxy formation models. In that sense, the particle tagging approximation can be thought of as an extension of the semi-analytic approach to modelling galaxy formation (a comprehensive overview of which is given by Lacey et al. 2016).

It is very clear that particle tagging is only an approximate technique, because the contribution of baryons to the gravitational potential is not treated self-consistently in the dynamical part of the calculation. In that calculation, each *N*-body particle includes a baryonic mass equal to the universal fraction, regardless of how much stellar mass the tagging procedure associates with it and regardless of the inflow and outflow of gas assumed in the semi-analytic component of the model. The aim of this paper is to use hydrodynamical simulations as a benchmark to test how this approximation affects some of the most basic predictions that have been made with particle tagging methods. We use STINGS as the basis for our comparison, but our results are relevant to the particle tagging technique more generally. We present our results in the context of the MW and its stellar halo because this is the regime in which particle tagging has been applied most often. In that context, the most important astrophysical processes that typical *N*-body particle tagging schemes neglect are as follows:

(i) The *internal structure* of the DM haloes of satellite galaxies can be altered by the inflow and outflow of baryons. The rapid dissipative condensation of gas within haloes can increase the central density of the DM ('cusp formation'), and the rapid gas expulsion of gas by supernova feedback may have the opposite effect ('core formation'; e.g. Navarro, Eke & Frenk 1996; Pontzen & Governato 2012; Nipoti & Binney 2015 and references therein). Both contraction and expansion of the potential can affect the kinematics of stars and the resulting rate of mass loss through tidal forces (Peñarrubia et al. 2010; Errani, Peñarrubia & Tormen 2015; Read, Agertz & Collins 2016). This implies that the method should be restricted to satellite galaxies with high mass-to-light ratios, although tagging has also been used to make predictions for the structure of stellar haloes in more massive galaxies (e.g. Cooper et al. 2013).

(ii) The *orbital evolution* of a particular satellite may be different in simulations from the same initial conditions with and without hydrodynamics, because the host halo potential can also change shape and concentration in response to the motion of baryons (e.g. Abadi et al. 2010; Binney & Piffl 2015). The growth of a massive stellar disc could make the initially triaxial inner regions of the host halo more oblate or spherical. Differences in the rate of mass loss due to these changes or other effects (such as ram pressure stripping of gas; e.g. Arraki et al. 2014) could exacerbate initially small divergences in satellite orbits.

(iii) Changes in the host potential will also affect *strong gravitational interactions* involving satellites and associated dynamical heating and disruptive effects, particularly for satellites with orbits that pass through the centre of the host ($r < \sim$20 kpc). Disc shocking (Spitzer & Chevalier 1973) is an example of this kind of interaction (e.g. D'Onghia et al. 2010). The consequences will depend on the strength and extent of the perturbation represented by the disc, its evolution with time, and the number of halo-progenitor satellites that pass through the region concerned.

(iv) In hydrodynamic simulations, stars can form on phase-space trajectories that are not well sampled in an equivalent collisionless *N*-body simulation, most notably those in centrifugally supported discs.

Likewise, subject to the hydrodynamic scheme and subgrid star formation prescription, gas particles stripped from infalling satellites may spawn stars directly on halo-like orbits (Cooper et al. 2015b). Consequently, applications of the method have mostly been restricted to the accreted component of galactic stellar haloes, as opposed to their possible components.

Earlier 'comparative' tests along similar lines to ours have been carried out by Libeskind et al. (2011) and Bailin et al. (2014). A more detailed discussion of these earlier studies is given in Section 6. In summary, the interpretation of this previous work is complicated by (i) the introduction of more complex tagging schemes (Libeskind et al. 2011) or simplified schemes which omit important features of those commonly used in the literature (Bailin et al. 2014); (ii) the use of hydrodynamic galaxy formation models that do not match well to observations in the regime under study; (iii) the use of complex statistics, such as the clustering of the projected stellar density (Bailin et al. 2014) as a basis for comparison; and (iv) the lack of a sufficiently clear distinction between uncertainties that are directly associated with the dynamical approximations involved in particle tagging and uncertainties associated with other differences in the models being compared. The comparison we present here addresses all these points. Points (iii) and (iv) are particularly important because the extent to which particle tagging 'works' as a proxy for a given hydrodynamical model depends on whether the approximations of the tagging scheme are justified for the conditions under which stars form in that particular model. When interpreting tests of particle tagging based on such comparisons, it should be borne in mind that different smooth particle hydrodynamic (SPH) simulations and semi-analytic models currently make very different predictions for when, where and in what quantity stars form, particularly in dwarf galaxies. In addition to the discussion below, we refer the reader to Le Bret et al. (2015), for another perspective on this issue and its implications for the dynamical evolution of sets of tagged particles in hydrodynamical and collisionless simulations.

Finally, we note that Dooley et al. (2016), in their study of stellar haloes in self-interacting DM models, also compare the predictions of a tagging prescription similar to ours against an SPH simulation from the same initial conditions. The stellar masses in their tagging model were obtained from an abundance matching argument, rather than a forward model of star formation; the only control for differences in stellar mass and SFH between their collisionless and SPH realization was a post-hoc renormalization of total stellar mass by an order of magnitude. Nevertheless, Dooley et al. (2016) found close agreement in stellar mass density between 50 and 200 kpc.

We proceed as follows. We concentrate on the spherically averaged density profile of the stellar halo, a particularly straightforward and relevant prediction which features prominently in previous work using particle tagging (Bullock & Johnston 2005; De Lucia & Helmi 2008; Cooper et al. 2010; Tumlinson 2010; Libeskind et al. 2011). We examine high-resolution cosmological SPH simulations of galaxies similar to the MW (Section 2). In Section 3, we first use the DM distribution in our SPH simulations together with the SFH of each halo in the *same* simulation to produce a particle tagging approximation for the distribution of stellar mass, which we compare to the original SPH star particles (Section 3.1). We then repeat this exercise with a separate *N*-body simulation that starts from initial conditions identical to those of our SPH simulation (Section 3.2). These are our main results. Section 4 examines in detail the origin of the (small) discrepancies we find between our SPH simulation and STINGS applied to the *N*-body version. We examine the choice of $f_{\rm mb}$, the single free parameter in the C10 implementation of STINGS (Section 5). In Section 6, we discuss our findings in the context of previous work on particle tagging. We summarize our results in Section 7. In Appendices A and B, we present examples that illustrate why (in the 'comparative' approach used here and in previous work on this topic), it is important to distinguish systematic and stochastic discrepancies that arise from modelling the collisional dynamics of baryons explicitly from less relevant effects that arise from the use of different models for star formation. Appendix C revisits how Cooper et al. (2010) used size distribution of satellite galaxies as a constraint on particle tagging models in light of the results here, and Appendix D discusses numerical convergence.

This paper is about comparing particle tagging and SPH simulations, rather than comparing either of these methods to observational data on stellar haloes in detail. Readers who are more interested in the 'bottom line' performance of particle tagging schemes in the context of the MW's stellar halo than in the technicalities of the method might, therefore, prefer to examine the first two figures and related text in Section 3 and then skip ahead to the comparison with earlier work in Section 6 and the summary of our findings in Section 7.

2 SIMULATIONS

Our hydrodynamic simulation,[2] which we refer to as AqC-SPH, is described by Parry et al. (2012) and Cooper et al. (2015b). It uses initial conditions from the Aquarius project (Springel et al. 2008), specifically those of halo Aq-C at resolution level 4, as the basis for re-simulation with an upgraded version of the SPH scheme described by Okamoto et al. (2010). Particle masses are $2.6 \times 10^5\,{\rm M}_\odot$ for DM and $5.8 \times 10^4\,{\rm M}_\odot$ for gas (assuming the Hubble parameter $h = 0.73$). The Plummer-equivalent softening length is $\epsilon_{\rm phys} = 257$ pc. 128 snapshots of the simulation were stored, spaced evenly by 155 Myr at redshifts $z < 2.58$ and by shorter intervals at higher redshift. The virial mass of the MW analogue is $1.8 \times 10^{12}\,{\rm M}_\odot$, towards the upper end of constraints on the most likely MW halo mass from recent measurements (see for example the compilation of results in fig. 1 of Wang et al. 2015). A stable baryonic disc forms at $z \sim 2.5$ and persists to $z = 0$ (see Scannapieco et al. 2012). Excluding self-bound satellites, the total stellar mass bound to the main halo at $z = 0$ (comprising the disc and spheroid of the MW analogue) is $4.1 \times 10^{10}\,{\rm M}_\odot$. We also make use of a DM-only version of this simulation (halo Aq-C-4 of Springel et al. 2008) with the same initial density perturbation phases and comparable resolution, which we refer to as AqC-DM. Zhu et al. (2016) have recently examined the properties of satellites simulated from these same initial conditions with a moving-mesh hydrodynamical scheme.

3 PARTICLE TAGGING

3.1 Tagging in an SPH simulation

The first question we ask is how well DM particle tagging works *within* our SPH simulation. This is obviously not how particle tagging is applied in practice but it provides a benchmark for interpreting the differences that arise when we tag particles in collisionless simulations.

The mass, formation time and phase-space trajectory of each star particle in AqC-SPH are known precisely. We extract SFHs by building a merger tree and assigning each of these star particles to the halo or subhalo to which it is bound at the first snapshot following its formation. We then use these self-consistent SFHs as the 'input' for tagging of DM particles in the same AqC-SPH haloes, following the STINGS scheme outlined in C10. In this experiment, tagged DM particles and the 'original' star particles experience the same orbital evolution and tidal field, as in Bailin et al. (2014) and Le Bret et al. (2015). This simple experiment allows us to study directly the effects of the differences in the initial phase-space distribution of star particles and tagged DM particles that result from the approximations inherent in the tagging procedure.

Fig. 1 compares the stellar mass density profile of the MW analogue[3] in AqC-SPH at $z = 0$ with the analogous result obtained by tagging DM particles according to the 'fixed-fraction' STINGS scheme, using $f_{\rm mb} = 5$ per cent[4] and the 'self-consistent' SPH SFHs. These curves are the same as those in the lower right panel of fig. 2 in Le Bret et al. (2015). Note that here both the star particle and tagged particle profiles *include* the stellar component (the figures in C10 did not show this component). The agreement is reasonably close, with discrepancies of no more than an order of magnitude at any radius over a density range covering ten orders

[2] Although we only discuss tests based on AqC-SPH in detail here, we refer the reader to Le Bret et al. (2015), who, in the context of particle tagging, compare this simulation to others based on a different code, GASOLINE, with alternative subgrid physical recipes.

[3] All star particles bound to the main halo, comprising the disc, spheroid and halo of the main galaxy and excluding satellites.

[4] Since the point of this exercise is to compare the two simulation techniques rather than to interpret observations, the appropriate value of $f_{\rm mb}$ is that which best approximates the behaviour of the subgrid star formation model in AqC-SPH with regard to the distribution of stellar binding energies after dissipative collapse.

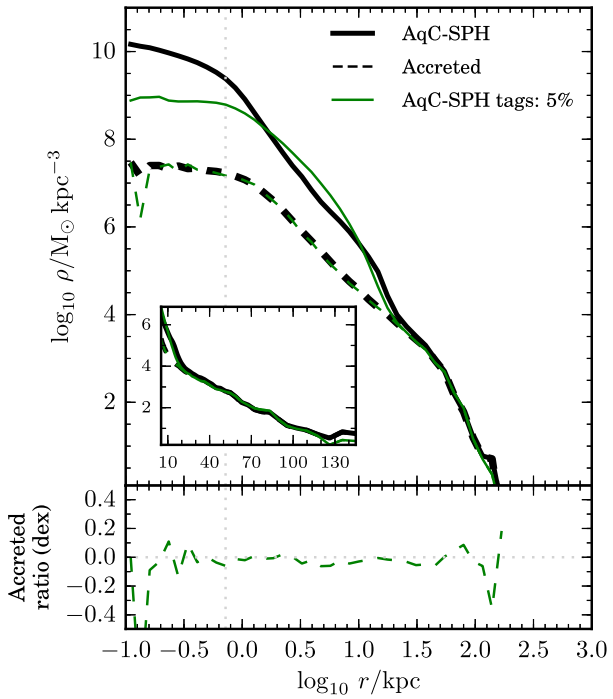

Figure 1. Top: spherically averaged mass density profile of all star particles bound to the main halo in AqC-SPH (black) compared with that obtained by tagging DM particles in AqC-SPH based on the SPH SFH (green), using $f_{\rm mb} = 5$ per cent. Dashed lines of the same colours show accreted stars only. The inset shows the same curves on a linear radial scale from 5 to 150 kpc. Bottom: logarithm of the ratio between the SPH and tagged-particle density profiles for accreted stars. This figure demonstrates that the distribution of star particles and tagged particles in the same SPH simulation agree well, particularly for the accreted component, even for a tagging scheme based only on rank order of particle binding energy.

of magnitude and little or no discernible systematic offset. Physical features in the profile, such as the *in situ* to accreted transition ($r \gtrsim 10$ kpc) and 'breaks' in the density of accreted stars, are much more significant than these discrepancies. The most obvious differences between star particles and tagged DM particles are seen in the inner 10 kpc. These are the result of differences in the density of stars, since accreted stars contribute very little mass in these regions.

If we only consider the accreted stellar component (as in C10 and most other applications of particle tagging), the agreement between star particles and tags is much closer at all radii (dashed lines), well below 0.1 dex for $r < 100$. This implies that the dynamical differences between the two techniques will not dominate the uncertainty in typical comparisons to real data, for example on the shape and amplitude of accreted stellar halo density profiles or their moments, such as total mass and half-mass radius. The observational errors on these quantities are of a similar order (∼0.5 dex; e.g. the density of the MW stellar halo in the Solar neighbourhood; McKee, Parravano & Hollenbach 2015) and the system-to-system scatter likely greater (for example the density of the stellar halo of MW-like galaxies at 30 kpc has a scatter of >1 dex; Cooper et al. 2013).

In Section 4.1, we will show that the mismatch between the spherically averaged density profiles of star particles and their corresponding tags in Fig. 1 arises because a single, universal value of $f_{\rm mb}$ cannot adequately represent the complex energy distribution of star-forming gas particles in the inner regions of our MW analogue. An alternative explanation for this mismatch might be the three-dimensional (3D) shape of the component, which is highly oblate in our SPH simulation. Most star particles belong to a rotationally supported disc, which obviously cannot be reproduced by DM particles selected on the basis of energy alone (even in this case, where the DM particles also feel the potential generated by the stellar disc). However, Section 4.1 demonstrates that the difference in 3D shape is *not* responsible for the majority of the discrepancy seen in Fig. 1.

3.2 Tagging in a collisionless simulation

The SPH experiment above was designed to take the effects of baryons on the galactic potential out of the comparison, to demonstrate that the phase-space evolution of stellar populations in our simulation is then well approximated by the DM particles we select as tags. In practice (e.g. in C10), particle tagging is used to model the phase-space distribution of stars in collisionless simulations, which do not include baryonic effects on gravitational dynamics. We now proceed to a more general comparison between AqC-SPH and STINGS applied to AqC-DM, a 'DM-only' simulation from the same initial conditions. The mass of particles in AqC-DM includes the universal fraction of baryonic mass and is therefore larger than that of the DM particles in AqC-SPH by a factor $\Omega_0/(\Omega_0 - \Omega_{\rm baryon})$, where $\Omega_{\rm baryon} = 0.045$.

3.2.1 'Transplanting' of star formation histories from Aq-SPH

For the experiment in this section, we need to assign SFHs to haloes in AqC-DM. This would be straightforward if the SFH of every halo could be 'transplanted' on a one-to-one basis from AqC-SPH. However, the dynamical histories of corresponding haloes in the two simulations sometimes diverge, which means that any transplanting is unavoidably approximate, particularly in the highly non-linear regime of tidally disrupting satellite galaxies. Some of this divergence is the direct result of baryonic physics (e.g. Sawala et al. 2015); some is stochastic or an indirect consequence of other physical changes; and some is simply the result of ambiguities in the numerical methods we use to identify haloes and subhaloes and link them between snapshots. Transplanting SFHs therefore requires a degree of care that makes it impractical to take this approach for every halo in the simulation.

For these reasons, we carry out a careful individual analysis for only the 10 progenitor haloes that make the most significant contributions of mass to the stellar halo in AqC-SPH. We find that nine of these progenitors each contribute more than 1 per cent of the total mass of stars accreted by our MW analogue; the most massive contributes 30 per cent. For simplicity, we control for the fact that some of these satellites accrete a (small) fraction of their stars from their own hierarchical progenitors by considering only star formation in the 'main branch' of the merger tree of each satellite when transplanting their SFHs.

Fig. 2 compares the density profile of tagged particles in AqC-DM obtained with this transplanting operation (dashed purple line) against the profiles of star particles in AqC-SPH (the figure also shows profiles of tags in AqC-DM based on GALFORM SFHs, which are described in the following subsection). We only compare the accreted stellar component because it makes little sense to tag stars formed with a procedure like this (more details on this point are given below). The 'transplanted' profile agrees well with the SPH result (over five orders of magnitude in density and two in radius above the gravitational softening scale) except for a ∼30 per cent discrepancy in the region $10 < r < 30$ kpc. In this

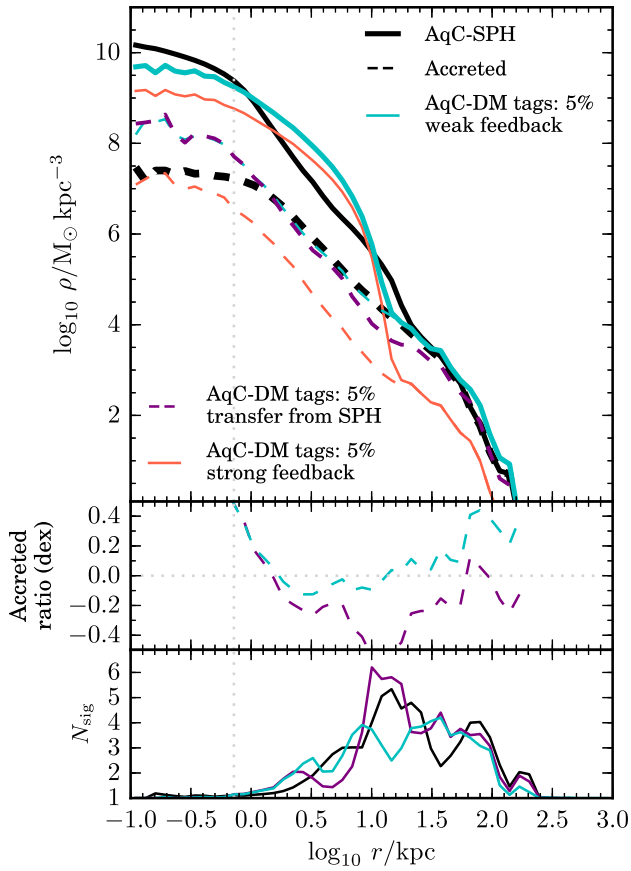

Figure 2. Top: spherically averaged mass density profile of all star particles bound to the main halo in AqC-SPH (black) compared with those obtained by tagging DM particles in a collisionless simulation with the same initial conditions, AqC-DM, based on semi-analytic SFHs predicted by the Bower et al. (2006) GALFORM model (which has 'strong' feedback; orange) and a variant of this model with 'weak' feedback (cyan). Also shown (purple) is the profile resulting from tagging based on a direct transfer of the SFHs from the SPH simulation to the collisionless simulation for the 10 most massive halo progenitors only. Dashed lines show only the accreted stellar component. All tagging results use $f_{\mathrm{mb}} = 5$ per cent. Middle: ratio of accreted density profiles as in Fig. 1. Bottom: the 'diversity' of the stellar halo; lines show $N_{\mathrm{sig}}(r) = [\Sigma_i m_i(r)]^2/\Sigma_i[m_i(r)^2]$, an estimate of the number of progenitors contributing a significant fraction of accreted debris at each radius. The good agreement seen in Fig. 1 holds for this application to a collisionless simulation, which reflects how particle tagging is most commonly applied in practice. The additional discrepancies are dominated by differences in star formation modelling rather than particle tagging (see the text).

region, the contributions to the stellar halo from different progenitors are most equal. The statistic $N_{\mathrm{sig}}(r) = [\Sigma_i m_i(r)]^2/\Sigma_i[m_i(r)^2]$, where $m_i(r)$ is the stellar mass contributed by the ith progenitor at radius r, implies that approximately five progenitors each contribute 1/5 of the stellar mass in this region, in both AqC-SPH and the transplant-tagged version of AqC-DM. Relatively small changes to the balance between the different contributions therefore have a particularly notable effect in this region.

We find no single progenitor dominates the differences we see in Fig. 2. Instead, we find that satellites in AqC-SPH tend to be disrupted (i.e. no longer identified as self-bound systems by SUBFIND) several Gyr earlier than their counterparts in AqC-DM. However, this has a surprisingly small effect on the spherically averaged distribution of their debris in most cases. The characteristic apocentres of the debris are significantly smaller in AqC-SPH in only two case. A third case, the most massive of all the halo contributors, exhibits the same effect but to a lesser extent. We find that the stellar mass density profiles of the individual progenitors at the time of their infall to the main halo are very similar in the two simulations. The differences in their debris distributions must therefore arise from how and when the individual satellites are disrupted. These effects are most notable in the inner region of the halo, where we would expect baryons in the central galaxy to create the most significant differences in the potential. The above results concerning the details of how and why individual satellites differ are presented in Appendix A.

We conclude that neither the self-consistent treatment of baryonic physics in AqC-SPH nor stochastic differences in the orbits of subhaloes between AqC-SPH and AqC-DM give rise to discrepancies with particle tagging that would significantly change our interpretation of predictions for the spherically averaged density profile of the accreted stellar halo.

3.2.2 *Semi-analytic star formation histories*

The final, most general step in our comparison is to use the GALFORM semi-analytic model to predict SFHs for haloes in AqC-DM. This is how particle tagging is usually applied in practice.

In this stage in the comparison, as we stressed in the Introduction, it is important to separate differences that arise from particle tagging from those that arise simply because GALFORM predicts a different SFH to AqC-SPH. Clearly, having reasonably well-matched models of star formation is a pre-requisite for comparing the amplitudes of density profiles and the balance between *in situ* and accreted stars. The parameters of GALFORM can be constrained by comparing predictions derived from cosmological volume simulations to large observational data sets covering a very wide range of galaxy scales. Arguably the most important constraints are $z = 0$ luminosity functions (e.g. Cole et al. 2000). The subgrid star formation model of AqC-SPH, however, was not calibrated in this way. Since differences between the subgrid models used in AqC-SPH and GALFORM are not relevant to our tests of particle tagging, the best GALFORM parameters to use in our comparison are those that most closely reproduce the predictions of AqC-SPH, not necessarily those which satisfy the usual observational constraints. In practice, we find that the GALFORM model used by C10 (essentially that of Bower et al. with refinements to the modelling of dwarf galaxies) reproduces the stellar mass of the MW analogue in AqC-SPH reasonably well, but underpredicts the mass of accreted halo stars. This indicates star formation in low-mass haloes is more strongly suppressed by feedback in the C10 model than in AqC-SPH. We therefore introduce a simple variation on C10 in which we reduce the value of one parameter, V_{hot}, from 450 to 250 km s^{-1} (see Cole et al. 2000 for the definition of V_{hot}). This model, which we refer to as the 'weak feedback' variant, reproduces the mass of the accreted stellar halo in AqC-SPH much more closely.[5] In contrast with this variant, we refer to the default parameters used by C10 as corresponding to relatively 'strong feedback'. The correspondence between individual satellite SFHs in these GALFORM variants and those in AqC-SPH is

[5] Massive haloes, of which we only have one in AqC-SPH by construction, have very different star formation efficiencies to the haloes that host dwarf galaxies. Simply varying the global strength of feedback as we have done here does not take into account this scale dependence and is hence a crude way of 'matching' our SPH and semi-analytic models, but is sufficient for our purposes.

close on average but of course not exact. Further details are given in Appendix B.

Fig. 2 shows the surface brightness profiles arising from tagging AqC-DM, based on SFHs predicted by the two variants of the C10 GALFORM model, and compares these with the original SPH star particle profile. Given the many approximations involved, the weak feedback model reproduces the SPH results remarkably well, particularly in the case of the accreted component. The density of the stellar halo in the weak feedback variant is slightly higher than that of AqC-SPH, which alleviates the discrepancy in the region $10 < r < 30$ kpc identified for the 'transplant-tagging' comparison in the previous subsection. The stellar halo is also less diverse in this region (lower $N_{\rm sig}$) in the semi-analytic realizations, which implies that the contributions of individual progenitors has changed. The SFHs of satellites are at least as important as the dynamics of their host subhaloes in explaining the density profile of the stellar halo at this level of detail.

Finally, it is not surprising that we find more difference in the density of stars formed in the MW analogue than in the accreted component. As noted by C10, the physical assumptions used to justify particle tagging are not expected to hold for this component. It is therefore interesting that the profile the component is reproduced as well as it is,[6] with similar extent and half-mass radius. The differences between particle tagging and SPH results for the component are well within the range of variation between predictions for this specific set of initial conditions using different hydrodynamical schemes (Scannapieco et al. 2012). Overall, the semi-analytic profile for the component in AqC-DM shows a similar discrepancy with the SPH star particles to that seen in Fig. 1, where we tagged DM particles in AqC-SPH simulation itself. This suggests that most of the discrepancy is due to an intrinsic limitation of the tagging scheme, rather than differences in the gravitational potential or the SFH of the main galaxy between AqC-SPH and AqC-DM. If that were true, a different tagging procedure might improve the agreement even further. In the following section, we will explore this idea in order to better understand why particle tagging performs so well in this comparison.

4 LIMITATIONS OF TAGGING FIXED FRACTIONS OF DM BY ENERGY RANK

4.1 Binding energy distributions

One of the fundamental assumptions common to all particle tagging implementations is that it is possible to find, for each newly formed stellar population, a set of DM particles that have similar phase-space trajectories over the time-scale of interest (i.e. a Hubble time). Generally speaking, different implementations assume different forms for the energy distribution of the stellar population and assign weights to particles in the same DM halo in order to reproduce those distributions. For example, the fixed-fraction STINGS scheme assumes the stars uniformly sample the binding energy distribution of the most tightly bound region of the potential at the instant of their formation (i.e. the softened NFW distribution function truncated at a particular relative binding energy). In all these schemes, however, the initial energy distribution of the tags is likely to be a relatively crude approximation to that of an analogous stellar population in a hydrodynamical simulation. Moreover, the phase-space trajectories of star particles are not functions of energy alone. Stars are likely to be formed on circular orbits, at least in the cold, quiescent discs of MW-like haloes. By ignoring the angular momentum of DM particles associated with the stellar tags, only the phase-space excursions of the stars can be approximated, rather than their actual trajectories.[7]

Using our SPH simulation, we now explicitly test the assumption that a suitable set of DM particles *can* be found in a scheme based only on binding energy (putting aside the question of how to find it in practice). We do this by searching for sets of DM particles that match the initial binding energy distributions of each SPH stellar population as closely as possible. For every individual SPH star particle, we identify a 'nearest neighbour' DM particle by sorting all particles in the same host halo in order of their binding energy at the snapshot following its formation. We select the first DM particle with higher rank (lower binding energy) than the star particle as its 'neighbour'. In cases where the star particle is more tightly bound than all the DM particles in its halo, we select the most bound DM particle. A DM particle can be selected as the neighbour of more than one star particle, in which case the associated stellar mass is increased accordingly. There are no free parameters in this selection procedure.[8]

Fig. 3 shows the stellar halo density profile resulting from this idealized scheme, analogous to the results in Figs 1 and 2. The correspondence between the tag and star particle representations of the accreted stellar halo is very close, as in Fig. 2, although here the tagged particle distribution is somewhat nosier because the number of tags is much smaller (roughly one DM tag for each star particle). More remarkable is the spherically averaged distribution of the tags associated with star formation, which in this case agree equally well with the SPH result. The discrepancies at $\lesssim$1 kpc scales are simply due to the fact that the DM particle softening length is larger than that of the SPH star particles. Since the tagged DM particles in this experiment move in a potential that includes the contribution from the gaseous and stellar discs, their distribution is mildly oblate. Nevertheless, it is surprising that such good agreement is obtained in the range $1 < r < 10$ kpc where most star particles are in a thin disc. The agreement here is much better than it is for the identical potential in Fig. 1, the only difference being the use of an idealized tagging scheme rather than the standard fixed-fraction approach.

We might also expect poor agreement at very large radii, because, in the AqC-SPH simulation, an stellar halo is built up by stars that form in streams of weakly bound tidally stripped gas (Cooper

[6] The results of C13 demonstrate that this is not a coincidence for the star formation and assembly history of this particular galaxy. With STINGS, the half-mass radius of the component is explicitly related to that of the host DM halo by construction, at least for plausible Λ cold dark matter SFHs.

[7] As C10 note, it would be possible to include a high angular momentum at a given energy as another criterion in the selection of DM particles. However, that would greatly limit the number of suitable particles available, because relatively few DM particles are on circular orbits.

[7] Since the search for neighbours is limited to the time resolution of the simulation snapshots, the precision of this choice is more limited than it has to be – for increased precision the search could be done 'on the fly', while the simulation was running. Moreover, the phase-space trajectories of tags will not correspond perfectly to those of their associated star particles even in this idealized scheme, because the star particle distribution function need not be a function of energy alone, and because the initial trajectories of star particles may not be well sampled by DM particles (see point (iv) in the Introduction, and footnote 7). As well as being on more circular obits, gas particles may also be more tightly bound than the most bound DM particle when they are converted to stars, in which case the use of the most bound DM particle is somewhat arbitrary and its accuracy dependent on resolution.

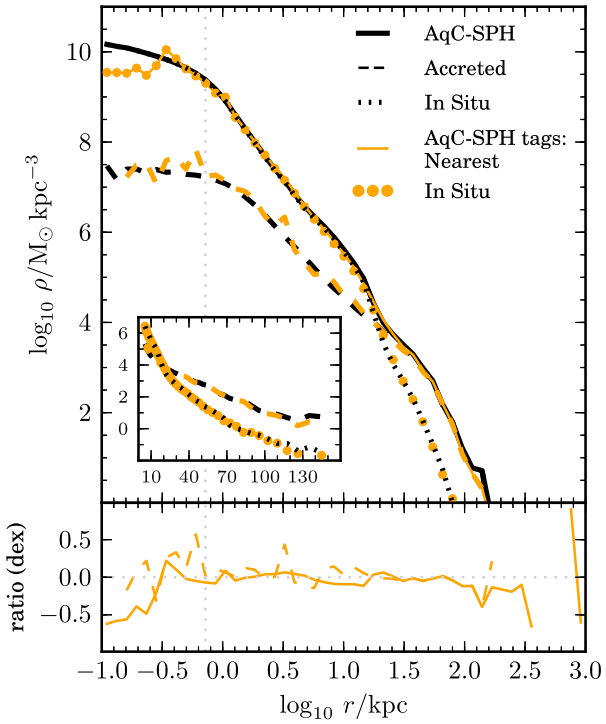

Figure 3. Density profiles for all main halo stars in AqC-SPH (solid), subdivided into accreted (dashed) and (dotted) components. Colours correspond to star particles (black) and stellar mass carried by tags (orange), assigned according to the idealized nearest neighbour scheme. The inset shows the same profiles on a linear scale, with finer binning. The lower panel shows the logarithmic ratio of the SPH and tagged star profiles for all stars (solid) and accreted halo stars only (dashed). A vertical dotted line marks the force softening scale.

et al. 2015b), rather than by the outward scattering of stars formed deep in the potential well. Since star formation in these stripped gas streams is triggered by local fluctuations in gas density, it is unlikely to correlate with the total binding energy of the stripped gas particles. In such cases, the DM particles that are selected as 'nearest neighbours' are much more likely to be smoothly accreted DM than to be associated with the parent stream of the corresponding gas particle. In practice, in our simulation, this effect is not notable – the tagged stellar profile agrees very well with the SPH simulation even at ∼30 kpc.

4.2 Examples of individual stellar populations

To understand why the idealized 'nearest energy neighbour' tagging scheme produces a distribution of 'tagged' DM particles that agrees so well with the distribution of SPH star particles in the same simulation, we now examine the dynamical evolution of individual stellar populations, as in Le Bret et al. (2015).

In Fig. 4, we choose two examples of stellar populations (sets of star particles that form in their host halo between two consecutive snapshots) with very different dynamical histories, labelled 'A' and 'B', respectively. Population A ($M_\star = 1.3 \times 10^7\,\mathrm{M}_\odot$) forms in a low-mass DM halo at $z \sim 5.5$ (a lookback time of 12.5 Gyr). This halo is subsequently accreted by and disrupted within the MW analogue halo. As a result, population A is phase-mixed into the stellar halo of the MW analogue by $z = 0$. Population B ($M_\star = 1.0 \times 10^8\,\mathrm{M}_\odot$) forms in the MW analogue halo itself at much lower redshift ($z \sim 0.4$; lookback 4.2 Gyr) and can be considered a 'MW disc' population.

The figure shows the projected distribution of star particles in each population at the simulation snapshot immediately after their formation (this is the same for all particles in the population, by definition). It is clear that population B forms within the thin baryonic disc of the MW analogue (shown approximately face-on and edge-on), whereas population A has a more amorphous distribution within its initial host. Note the prominent stellar bar in the inner ∼2 kpc of population B, and also the order of magnitude difference in spatial scale between the two populations. Clearly, the definition of a single population in the context of particle tagging does not correspond to the usual concept of a single star-forming region in the case of population B, where stars are forming across the entire disc. The colour of each point corresponds to the fraction of DM more bound than a given star particle (for example 5 per cent of the DM particles are more bound than the star particles shown with blue colours). The clear radial gradient in colour reflects a tight correlation between binding energy and depth in the potential, even for the centrifugally supported disc of population B.

In Fig. 5, we quantify this relationship in more detail using a distribution directly relevant to particle tagging: the fraction of newly formed star particles that are more bound than given fraction of the DM particles in the same halo, when the latter are sorted in rank order of binding energy. We show this distribution for SPH star particles (black) and for the DM particles to which we tag their stellar mass in the nearest energy neighbour scheme (orange). By construction in this scheme, the distributions of stars and tags are almost identical at the time of tagging.

Of more interest in this idealized case is the correspondence between star particles and tagged DM particles at the *final* simulation output time ($z = 0$), shown by the dashed lines in Fig. 5. These distributions show the evolution in a 'relative' sense (including only the DM particles that were part of the set available for tagging at t_i) rather than an absolute sense (which would include *all* the DM in the halo at $z = 0$). Relative evolution occurs because tagged DM particles subsequently diffuse to higher energies, and DM particles that were not tagged diffuse to lower energies. The $z = 0$ distributions for tagged DM particles and stars are very similar in population A, despite the large time interval and the complete disruption of the original host halo which 'scrambles' the initial relationship between the binding energy of stars and DM. The correspondence is even closer for population B; there is almost no evolution in the 'relative' sense even after 4 Gyr, implying that the DM halo and the galactic disc are dynamically stable over this period.

Fig. 6 shows the spherically averaged density profile of the two populations. Black circles show the initial profile of star particles, and black crosses their profile at $z = 0$. These can be compared with the solid and dashed orange lines, respectively, which show the profile of the corresponding tagged DM particles. For scale, a solid grey line shows the density profile of *all* star particles in the host halo (the main MW halo in both these examples) at $z = 0$. Although the evolution in the density of tags and stars does not correspond as closely as their evolution in binding energy, the differences are still relatively small ($\lesssim$0.5 dex). As noted above, population A forms in a dwarf galaxy that is incorporated into the accreted halo of the 'MW' before $z = 0$, hence the initial and final profiles are measured with respect to the centre of the formation halo and the $z = 0$ MW halo, respectively. For population B, the star particle profile has three 'components' (broadly, the inner bar/bulge, an exponential disc and an *in situ* halo). These components are reproduced by tagged particles at both the initial and final times. Remarkably,

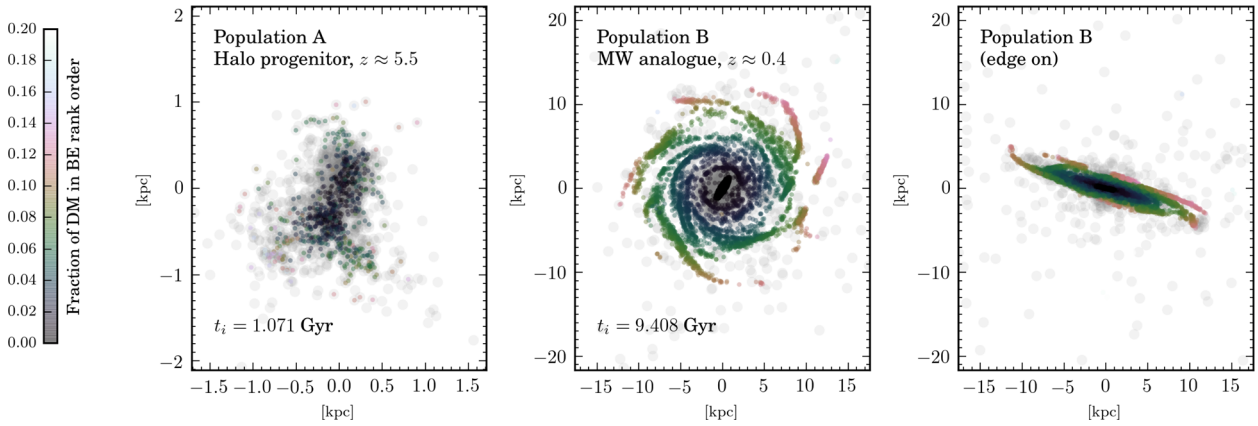

Figure 4. Two examples of the 'initial conditions' of single stellar populations (SSPs) in AqC-SPH at their formation time (t_i, measured from the big bang to the present day at $t = 13.582$ Gyr). *Left:* a 'starburst' population (A) formed at $z \approx 5.5$ in one of the larger progenitors of the accreted stellar halo. The dwarf galaxy host of these stars is fully disrupted in the main MW analogue halo before $z = 0$. *Centre:* a different population (B) formed in the disc of the MW analogue galaxy itself at $z \approx 0.4$. Note the central bulge/bar and associated 'ring' of star formation. *Right:* an orthogonal projection of population B, approximately edge-on. Colour indicates the fraction of DM in the host halo with higher binding energy rank than the particle at t_i (see the text). Grey points are pre-existing stars in the host at t_i (only 1/1000 of these are shown); by definition these do not belong to the SSPs A or B and are shown only for scale.

tagged DM reproduces the spherically averaged scalelength of the SPH disc and bulge and preserve this correspondence over many Gyr, despite the complexities of the separate star-forming regions and their very different distribution in configuration space.

We conclude that, given an SPH simulation, it is possible to select sets of DM particles that trace the evolution of the spherically average density distribution of star particles in the same simulation to an accuracy better than a factor of two. This is the case at least for the subgrid star formation model implemented by our SPH simulation and holds even for stars forming in a thin disc. When particle tagging is applied in practice, an initial energy distribution has to be determined a priori, necessarily with some approximation. We argue that success or failure in reproducing the distribution star particles in SPH simulations with tagged DM particles in the same simulations is almost entirely determined by the accuracy of this approximation with respect to the true initial energy distribution of star particles.

5 WHAT CHOICE OF $f_{\rm mb}$ IS APPROPRIATE FOR FIXED-FRACTION TAGGING?

In practice, particle tagging schemes are applied to simulations that do not already include a separate dynamical component representing stars, and therefore have to use simple approximations for the initial energy distribution of stellar populations. For example, the STINGS scheme assumes these distributions can be approximated by those of DM particles selected in rank order of binding energy from the most bound down to a specified fraction. The free parameter of the method, $f_{\rm mb}$, sets the 'bias' between the energy distribution of newly formed stars and the DM of their host halo. This bias is assumed to be universal, hence 'fixed fraction'. This approximation is simplistic and it is no surprise that it breaks down in detail for complex star formation regions dominated by the baryonic potential and having significant angular momentum, like the MW disc (as illustrated by population A in Fig. 4).

Fig. 7 shows how variations of $f_{\rm mb}$ affect the results of the SPH-based tagging shown in Fig. 1. The good agreement for the accreted halo distribution is largely insensitive to the exact choice of $f_{\rm mb}$. Discrepancies between these four profiles exceed ∼10 per cent only within $r < 1$ kpc and beyond $r > 100$ kpc. Hence, although it is sensible to calibrate $f_{\rm mb}$ with respect to the scale radii of surviving dominated galaxies, results for accreted halo stars are *not* particularly sensitive to this choice. In most cases, the diffusion of stars in phase space associated with the tidal disruption process dominates over small differences in the structure of the progenitor. For reasons discussed by C13 and examined in detail by Le Bret et al. (2015), the differences in the 'initial conditions' of their populations are lost by $z = 0$. This is not the case for the 'disc' populations forming in the very stable central region of the main halo at low redshift, which consequently show large variations in their shape and moderate variations in their half-mass radius as $f_{\rm mb}$ varies from 1 to 10 per cent. To a lesser extent, the same is true for the scale radii of surviving satellites, which C10 compared to observations to support a value of $f_{\rm mb} \sim 1$ per cent (further details are given in Appendix C).

Fig. 8 illustrates this directly by repeating the analysis of the individual populations A and B from Fig. 4 using fixed-fraction tagging with $f_{\rm mb} = 1$ and 10 per cent. In the case of population A (the 'halo' population), it can be seen that the large initial differences between the two sets of tags (and between each set and the corresponding AqC-SPH star particles) are erased by the time the stars and tags have been mixed into the stellar halo of the MW analogue. In the case of population B (the 'disc' population), neither set of tags evolves significantly, except for the diffusion of particles in the low binding energy tail above the initially sharp cut-off energy. The quality of the agreement between tags and SPH star particles at $z = 0$ is therefore dominated by the initial conditions imposed at the time of tagging. The initial energy distribution is multimodal, as shown in the previous section, and this clearly cannot be captured by a single value of $f_{\rm mb}$. The $f_{\rm mb}$ scheme corresponds to a linear form for the 'energy rank distribution', whereas this distribution for actual star particles is at best only approximately linear for the most tightly bound stars in population A. In population B, comparing the blue and red lines in Fig. 8 shows that $f_{\rm mb} = 10$ per cent describes the bulk of the exponential disc reasonably well, while $f_{\rm mb} \lesssim 1$ per cent more closely reproduces the distribution for star particles formed in the compact nuclear region of the galaxy.

The most appropriate value of $f_{\rm mb}$ will clearly differ from galaxy to galaxy, and from snapshot to snapshot. Putting aside the issue of multiple stellar populations forming simultaneously in a galaxy, a good empirical approximation to the optimal value of $f_{\rm mb}$ for

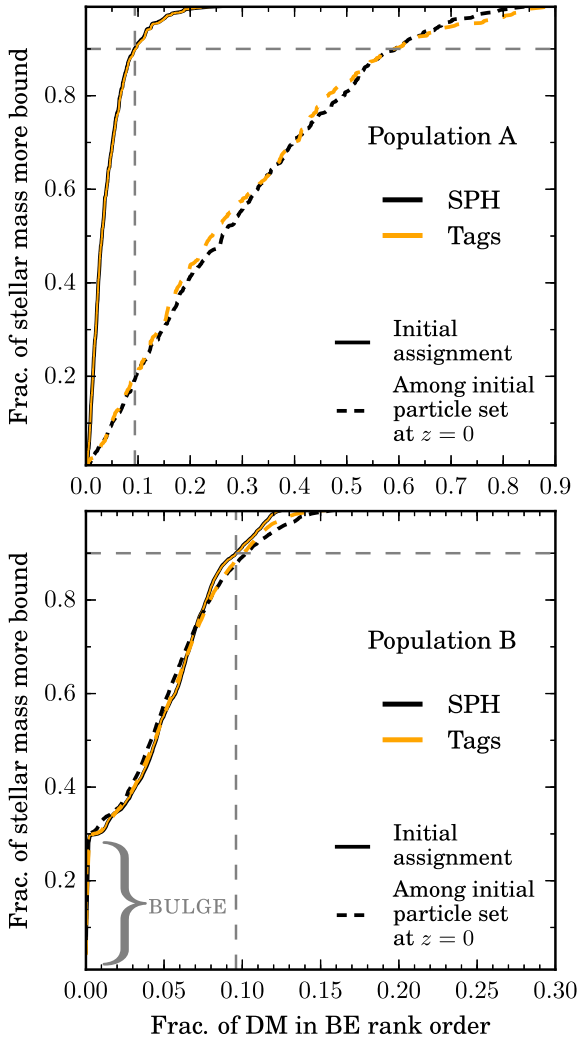

Figure 5. Cumulative fraction of stellar mass in the newly formed populations A and B (Fig. 4) that is more bound than a given fraction of the DM in their corresponding host haloes. The horizontal axis corresponds exactly to the colour scale of the images in Fig. 4. The nearest energy neighbour tagging scheme (orange) applied to DM in AqC-SPH accurately reproduces the distributions of the actual star particles (black) at the formation time (solid lines; essentially by construction in this scheme). Vertical dashed lines indicate f_{90}, the fraction of DM enclosing 90 per cent of the stellar mass at t_i, an empirical equivalent to $f_{\rm mb}$ (see the text). The distribution of the tags still traces that of the star particles at $z = 0$ (dashed lines). We mark the region corresponding to the nuclear 'bar/bulge' of population B in Figs 4 and 6.

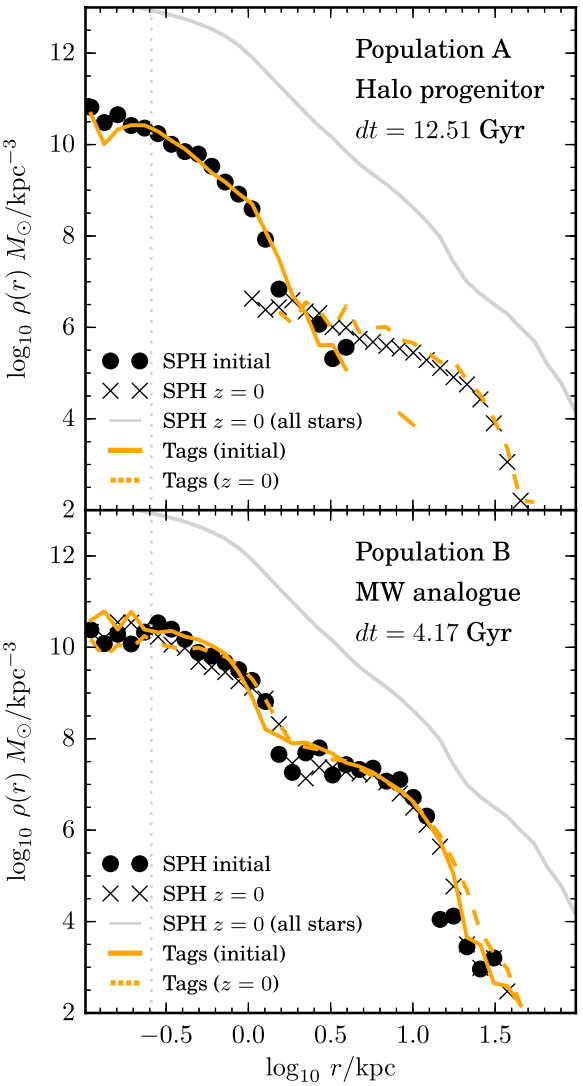

Figure 6. Density profiles of star particles (black) and tagged particles (orange) in the populations shown in Fig. 4. Dots/solid lines correspond to profiles at the formation time, t_i. Crosses/dashed lines correspond to the distribution of the same particles at $z = 0$ (the elapsed time, dt, is indicated). In the case of population A (top), the initial profile is centred on the halo in which it forms, and the final profile is centred on the MW analogue halo, into which the population is accreted by $z = 0$. For scale, the grey line shows the profile of all stars at $z = 0$. Note the transition in profile B around ∼1 kpc, corresponding to the extent of the nuclear 'bar/bulge' region in Figs 4 and 5.

each 'aggregate' coeval population in our SPH simulation can be defined as $f_{\rm mb} \approx f_{90}$, where f_{90} is the fraction of DM in rank order of binding energy enclosing 90 per cent of the newly formed stars (this definition is illustrated for our two example populations by the vertical lines in Figs 5 and 8).

Fig. 9 plots f_{90} for all populations in AqC-SPH against their stellar mass. A clear sequence of points corresponding to the stable disc is apparent at $f_{90} \sim 0.1$, highlighted in the inset panel. For low-mass populations, there is huge scatter, reflecting the complex nature of star formation in low-mass DM haloes (likely in addition to numerical noise). A good understanding of the shape of this distribution under different star-forming conditions would greatly improve the correspondence between particle tagging and SPH simulations, although it is not clear that the complexity of a variable-faction tagging scheme would justified. Given the other approximations inherent in the method, our results suggest that the simple fixed $f_{\rm mb}$ scheme is probably adequate in most cases where particle tagging is significantly more efficient than SPH simulations, namely very high-resolution models of dwarf satellite accretion and computing the statistical properties of large numbers of galaxies in lower resolution cosmological simulations. For those applications, only the approximate scale of the component is important, provided energy diffusion is taken into account as discussed in Le Bret et al. (2015) either by 'live' tagging (as in STINGS) or by explicitly imposing an appropriate distribution function (as in Bullock & Johnston 2005 and Libeskind et al. 2011). For studies of MW-like stellar haloes, calibration $f_{\rm mb}$ with reference to the mass–size relation of *in situ*-dominated galaxies, as in Bullock & Johnston (2005) and C10, is

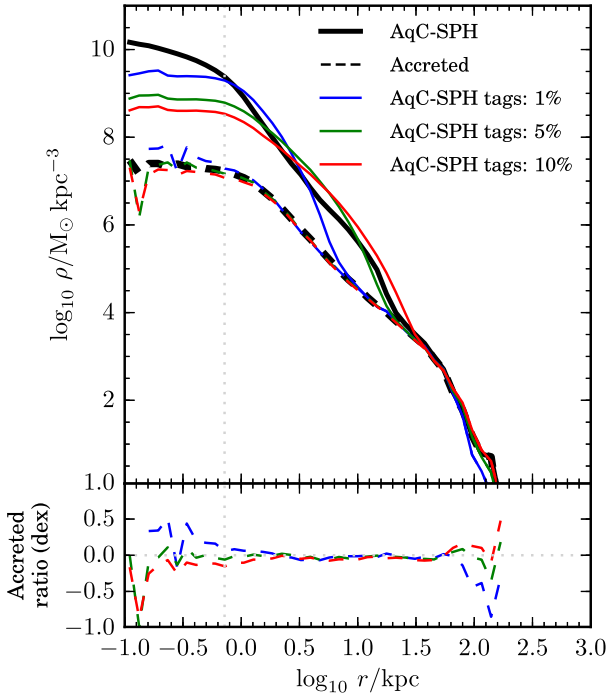

Figure 7. Stellar density profile of the main (MW analogue) halo at $z = 0$ for fixed-fraction tagging based on the SPH SFH with three different choices of $f_{\rm mb}$ (see the legend). These are compared to the result for SPH star particles (black) for all stars (solid) and accreted stars only (dashed). Lower panel shows the ratio of accreted stellar mass density in tagged particles to that in SPH star particles, for each $f_{\rm mb}$.

adequate. In Appendix C, we discuss this calibration further in light of the results above.

6 DISCUSSION

We can now finally return to the question posed by Figs 1 and 2. Given two simulations from the same initial conditions, one hydrodynamical and the other using STINGS (semi-analytic galaxy formation in combination with fixed-fraction particle tagging), are the theoretical limitations of the particle tagging approach responsible for the differences we see in the spherically averaged density profile? In the case of the simulation we analyse, this does not seem to be the case. The discrepancies between SPH and STINGS predictions in Fig. 2 are less than an order of magnitude and can be explained as the result of differences in the modelling of star formation, the simplistic form of energy distribution assumed by the STINGS fixed-fraction tagging scheme and the use of a universal fraction parameter in this scheme. None of these are fundamental to the particle tagging approach and can easily be improved upon. The only clear discrepancy that can be seen as a clear limitation of particle tagging is the more rapid rate of disruption of massive subhaloes in our SPH simulation. We conclude that, of the possible sources of discrepancy listed in the Introduction, the most important limitation of particle tagging is its failure to reproduce this systematic effect on the orbital evolution of subhaloes in collisionless simulations. We have not identified the origin of this effect. It is possible (but not yet proven) that it is due to a modification of the innermost regions of the gravitational potential induced by the baryons associated with the central galaxy. Changes in the internal structure of satellites may also be important, although for the most part these are not evident at the time of infall. Disc-shocking effects, although important for the overall subhalo population, seem unlikely to affect the most significant contributors of the stellar halo. Remarkably, in the context of the stellar halo, the approximate and incomplete sampling of the phase space of hydrodynamic star particles by collisionless tags appears to be less important than any of these factors.

Our findings give a new perspective on previous comparisons between SPH and particle tagging models, which we briefly revisit in this section.

6.1 Energy diffusion

As discussed in detail by Le Bret et al. (2015), the apparently minor simplification of tagging DM particles at the time at which each satellite progenitor falls into main halo, as adopted, for example by De Lucia & Helmi (2008) and Bailin et al. (2014), can change the results of comparisons like that shown in Fig. 1 and make the agreement between star particles and tagged particles substantially worse in some cases. This is because the tag-at-infall approach does not allow for the prior diffusion of tagged particles in energy space after the associated stars form, whereas SPH star particles naturally undergo such diffusion. For that reason, tag-at-infall schemes are not always good proxies for work using live tagging schemes. Live schemes such as STINGS take diffusion into account naturally, although it is also possible to introduce parameters into the tagging scheme to make a posteriori corrections for its effects when tagging at infall (Bullock & Johnston 2005; Libeskind et al. 2011).

The difference between the tag-at-infall simplification and a live scheme also depends on the SFHs of satellites in the simulation, for three reasons. First, if satellites form the bulk of their stars immediately before infall then the approximation is obviously more reasonable than if they form long before. Second, the basic assumptions required for tagging no longer hold if star formation in satellites is efficient enough to significantly alter the density profile in ways that are not captured by a collisionless simulation. Third, satellites themselves are the product of hierarchical mergers and may acquire their own diffuse haloes before infall, but any distinction between and accreted stars *within satellites* is not captured by single-epoch tagging. Another, more minor issue is that the amount of energy diffusion is much reduced if stars are less deeply embedded in their host potential at the time of star formation (for example in the STINGS scheme, if $f_{\rm mb} \sim 10$ per cent rather than 1 per cent). Low numerical resolution has essentially the same effect (as seen perhaps in De Lucia & Helmi 2008) because the central regions of halo potentials, where diffusion effects are strongest, are not well resolved to begin with. In a high-resolution simulation in which star formation in satellites is inefficient and peaks several Gyr before accretion on to the proto-MW, this diffusion effect is critically important. These are the conditions for halo star formation favoured by recent cosmological simulations.

6.2 Comparison with previous tests of particle tagging

6.2.1 Libeskind et al. (2011)

Libeskind et al. (2011) examine some of the issues above using an SPH simulation of a Local Group analogue with comparable resolution to ours, alongside a matched collisionless simulation. They claim that fixed-fraction tagging of satellites at the time of infall does not reproduce the density profile of the accreted stellar haloes in their SPH simulation adequately. They advocate an alternative time-of-infall method, in which the 'absolute' potential, ϕ, of particles to be tagged must satisfy $\phi \geq \kappa\,\phi_{\rm subhalo}$ where

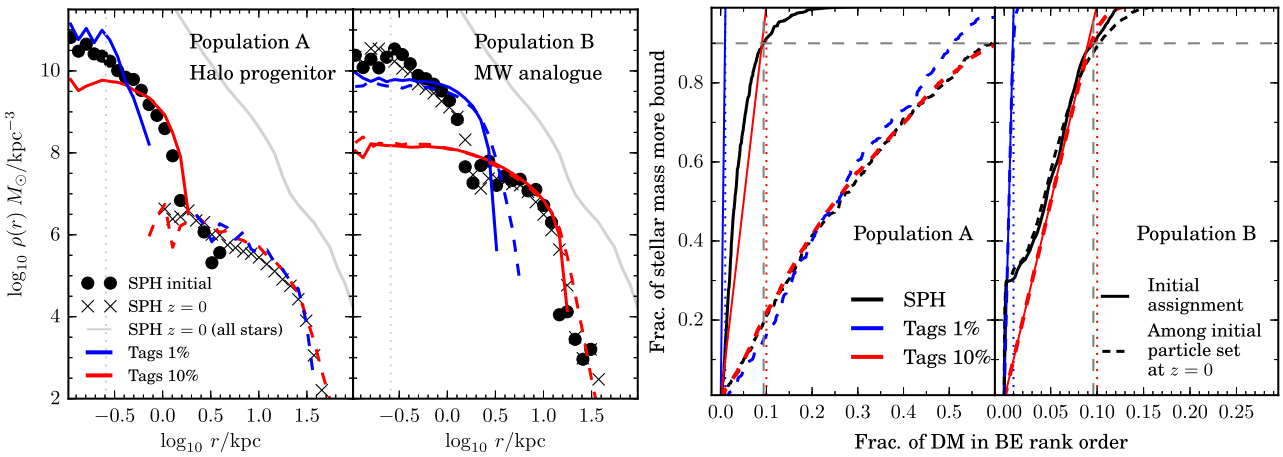

Figure 8. Initial- and final-time density profiles (left) and 'energy rank distributions' (right) of single population examples shown in Figs 4 and 5, here for fixed-fraction tagging schemes with $f_{\rm mb} = 1$ per cent (blue) and 10 per cent (red), compared with SPH results (black). Dashed vertical lines mark f_{90} and dotted vertical lines mark $f_{\rm mb}$ (see the text). Note that fixed-fraction schemes correspond to linear distribution functions in the right-hand panels (solid red and blue lines).

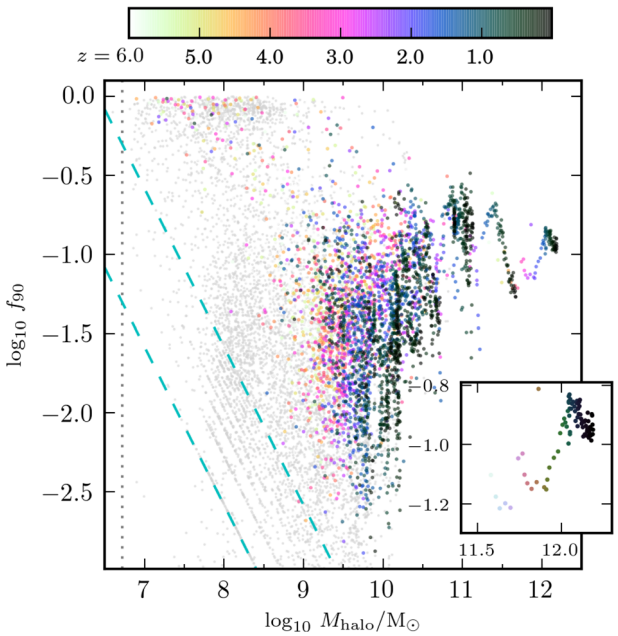

Figure 9. Scatter plot of f_{90} (see the text) for SSPs in AqC-SPH as a function of instantaneous halo mass measured by SUBFIND (main haloes only). Colour coding indicates formation redshift: stellar mass is generally a monotonically increasing function of time for main haloes. Only star formation events forming five or more star particles in haloes with more than 10 DM particles and baryon fractions of $\lesssim$20 per cent (hence excluding spurious clumps of baryons) are coloured. Remaining star formation events are shown in grey and mostly correspond to small baryon-dominated clumps identified as independent haloes. Dashed blue lines indicate the loci for which f_{90} corresponds to 1 particle (lower line) and 10 particles (upper line). Since the DM particle mass is fixed and baryon particle masses vary within a narrow range, the smallest non-zero values of f_{90} are discretized along lines parallel to these loci (there are multiple lines corresponding to small haloes with different mixes of the three particle species in AqC-SPH). Inset shows evolution for the MW analogue halo only.

$\phi_{\rm subhalo} = -GM_{\rm virial}/r_{\rm virial}$ is defined at the 'edge' of the subhalo immediately before infall. Their optimal value is $\kappa \sim 16$, chosen to best match the density profiles of tagged particles and stars in their SPH simulation.

Since the method advocated by Libeskind et al. (2011) is applied at the time of infall, the freedom in choosing κ implicitly compensates for diffusion in energy between the time of star formation and the time of infall, as discussed by Le Bret et al. (2015). If the baryonic physics in their simulation significantly alters the concentration of their potentials, or causes them to depart from the NFW form, this may explain why they find that fixed-fraction tagging at infall performs poorly. If not, their claims in this regard are hard to understand, because although the Libeskind et al. (2011) method requires the explicit calculation of potential energies, in practice it is essentially the same as our STINGS fixed-fraction scheme.[9] They report that this criterion selects about ∼1–3 per cent of the DM particles accreted from subhaloes and bound to their three most massive host haloes at $z = 0$. Their criterion therefore appears roughly equivalent to $f_{\rm mb} \sim 1$ per cent. This is not easy to interpret, however, because, for a simple NFW profile, the minimum of the potential has $\phi_{\rm cen}/\phi_{\rm subhalo} \leq 16$ for concentrations $c_{\rm NFW} \lesssim 46$ (e.g. Cole & Lacey 1996), which implies that, for a typical mass–concentration relation, no mass should be as tightly bound as they require in the majority of haloes. Even for a (rather extreme) halo with $c_{\rm NFW} = 50$, $\kappa = 16$ corresponds to only $f_{\rm mb} \sim 0.2$ per cent. Libeskind et al. do not recommend a way to apply their technique in cases where no particles in a subhalo are more bound than their threshold. Overall,

[9] We do not agree with the statement in section 4 of Libeskind et al. (2011) that tagging a fixed fraction of DM particles by binding energy rank (which they call 'relative') is distinct from (and hence less accurate than) the 'absolute' approach they propose. For a self-bound, virialized collection of equal mass particles, selecting a fixed fraction of mass in order of binding energy rank is equivalent to selecting particles more bound than a fixed multiple of $-GM_{\rm virial}/r_{\rm virial}$, because ϕ is a monotonic function of $M(<r)$. At least part of the discrepancy they discuss is likely to be due to the fact that their method implicitly corrects for the shortcomings of applying a fixed-fraction scheme at the time of infall, as discussed by Le Bret et al. (2015).

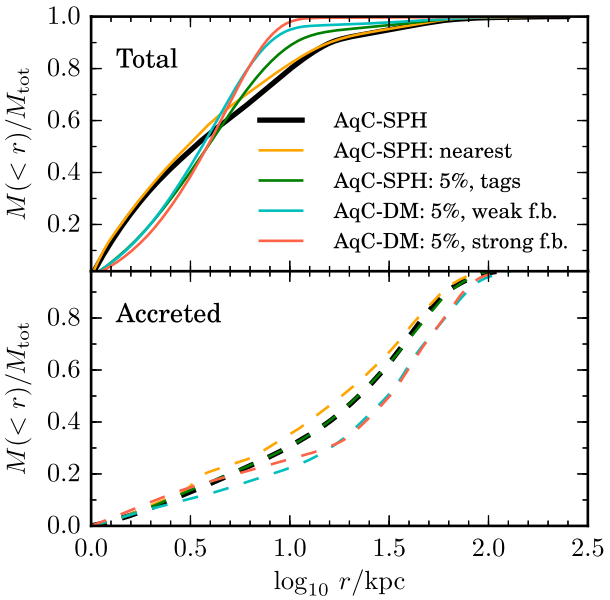

Figure 10. Cumulative mass fraction between 1 and 250 kpc for star particles in AqC-SPH (black), and particle tagging results described above for AqC-SPH (green and orange) and AqC-DM (blue and red). Upper panel: all stars, lower panel: accreted stars only. This figure can be contrasted with fig. 4 of Bailin et al. (2014).

however, the Libeskind et al. (2011) study seems to be in broad agreement with our conclusions and those of Le Bret et al. (2015).

6.2.2 *Bailin et al. (2014)*

Bailin et al. (2014) present a critique of particle tagging, also in the context of the MW stellar halo and based on a comparison between hydrodynamical and collisionless simulations from the same initial conditions. The discussion and results they present underscore several well-known potential pitfalls of particle tagging methods, which were noted (and avoided) in the implementations of Cooper et al. (2010) and Libeskind et al. (2011). Their work emphasizes discrepancies in the 3D shape and smoothness of the stellar halo (the latter quantified by the variance of fluctuations in the density of halo stars in broad 'zones' defined in spherical coordinates). These measures are relevant to the interpretation of observational data (e.g. Cooper et al. 2011; Helmi et al. 2011; Monachesi et al. 2016) and could be sensitive to differences in how satellites are disrupted in hydrodynamical and *N*-body models. However, it is not easy to distinguish the effects of particle tagging on these statistics from other sources of divergence between SPH and collisionless simulations. We therefore believe the most direct point of comparison between our work and Bailin et al. (2014) is their claim that particle tagging artificially reduces the predicted concentration of accreted stellar haloes because it does not account for the baryonic contribution to the potential.

The lower panel of Fig. 10 shows the cumulative stellar mass profiles of accreted stars only, in the region between 1 and 250 kpc, for several of our model variants. These curves can be compared directly to those in the lower panel of fig. 4 in Bailin et al. (2014). We also include the corresponding cumulative profile of the total stellar mass (top panel); this was not shown by Bailin et al. (2014), but is useful for reference here. We show these enclosed mass fraction curves only for comparison with Bailin et al. because they are not straightforward to interpret. Compared to the density profiles shown in Fig. 2, they are more sensitive to differences near the centre of the potential ($\lesssim$1 kpc), which may not be significant in the context of the stellar halo overall. These curves do not provide information about the absolute density of each variant at a given radius, only about the relative concentration of their density profiles. For example, the profiles of the two particle tagging variants with different feedback strengths (red and cyan lines) appear very similar in Fig. 10, but very different in Fig. 2.

Bailin et al. describe a model, SPH-EXACT, in which each stellar halo progenitor in their SPH simulation is tagged only once, at the time of its maximum mass. The sum of the mass of SPH star particles (*in situ* or accreted) accumulated up to that point is distributed evenly among the 1 per cent most bound DM particles.[10] This can be compared with the two tagging schemes we apply to AqC-SPH ($f_{\rm mb} = 5$ per cent, Section 3.1; and 'nearest energy neighbour', Section 4.1). In our case, both of these schemes predict distributions for the accreted component that are nearly identical to that of the original star particles (the nearest neighbour results diverge slightly at $\sim$3 kpc because of a small spike in the density at this radius which can be seen in Fig. 3). Bailin et al., in contrast, find their SPH-EXACT halo does not resemble their star-particle reference model (SPH-STARS). The disagreement they find is $\sim$30 per cent at 10 kpc. This difference is even greater than that shown in the top panel of Fig. 10, which, however, includes *all* the stellar mass, the majority of which formed . We conclude that either the assumptions of particle tagging are violated much more strongly in the simulation of Bailin et al. than in our simulation, or else their SPH-EXACT model diverges from their SPH-STARS model for reasons other than the limitations of particle tagging alone.

Bailin et al. also use a simple linear relation between stellar mass and halo mass obtained from their SPH-STARS model to carry out tagging in their collisionless simulation (their DM-PAINTED model). They find the results of that experiment do not resemble their SPH-STARS or SPH-EXACT models, and do not agree well even with the results of the same scaling relation applied to the SPH simulation (their SPH-PAINTED model). In contrast, we find that tagging AqC-DM using SFHs from GALFORM (which roughly match those predicted by AqC-SPH) produces results very similar to those from tagging DM particles in AqC-SPH according to the SPH SFH. We argued above that the residual discrepancy is largely due to the (relatively minor) differences in the SFHs used as input.

It therefore appears that, in our study, the use of DM particles as proxies for star particles does not in itself create less concentrated haloes. Simple fixed-fraction particle tagging schemes like the C10 implementation of STINGS result in a less concentrated stellar distribution overall for the MW analogue galaxy because they do not account for the presence of separate star-forming regions seen in our SPH simulation, particularly in the nucleus of the galaxy. The apparent concentration difference for the accreted stars in AqC-DM seen in Fig. 10 is hardly notable in Fig. 2, where it is dominated by differences on scales of $\lesssim$2 kpc.

There are many possible explanations for the differences between our findings and those of Bailin et al. We consider the most important to be: that they used an SPH simulation that greatly overpredicts the efficiency of star formation and the resultant baryonic impact on the potential of the central galaxy and its satellites (Stinson et al. 2010; Keller, Wadsley & Couchman 2015); that they used a tag-at-infall scheme, which does not allow for diffusion in the energy of tagged

[10] Bailin et al. do not state why the total stellar mass of their SPH-EXACT realization is almost twice that of the original SPH simulation.

particles between formation and infall (this is particularly important for SPH simulations with strong feedback, as noted by Le Bret et al. 2015); and that they did not quantify the effects of differences between the SFHs used as input to each model in their comparison. Bailin et al. conclude that their findings motivate the development of more elaborate tagging schemes to overcome the shortcomings they identify. Although it would be worthwhile to explore well-constrained extensions of the simple fixed-fraction approach, our results suggest that even straightforward implementations like STINGS are adequate for many applications of particle tagging to the study of stellar haloes. Moreover, the results in Appendix A demonstrate that the most important systematic differences between SPH and particle tagging do not concern how stars are distributed within satellites before they are disrupted, but rather the dynamics of subhalo disruption, which more sophisticated tagging schemes would not be able to address.

7 CONCLUSIONS

We have used an SPH simulation to test the assumptions inherent in the semi-analytic particle tagging scheme of C10 (STINGS) an efficient but dynamically approximate method for modelling the phase-space evolution of galactic stellar haloes. In the case of the simulation we consider, these approximations appear to be reasonable. We were able to recreate the spherically averaged density profile of star particles representing the accreted halo in the SPH version of our simulation by applying GALFORM and STINGS to collisionless version of the same simulation. We also found that the spatial distribution of stellar populations can be reproduced reasonably well by these schemes, under certain more restrictive conditions. Our findings support the conclusions of Le Bret et al. (2015), who explored the role of diffusion in energy space in the comparison of particle tagging schemes to hydrodynamical simulations. We summarize our results as follows:

(i) Given a set of recently formed star particles in an SPH simulation it is possible to select subsets of the DM particle distribution in the *same* simulation that trace the subsequent phase-space evolution of those particles almost exactly, using only their relative binding energies. The 'best' outcome possible under a scheme like this (approximated here by our 'nearest energy neighbour' experiment) is a near-exact correspondence between the star particle and tagged particle realizations of both the accreted and stellar density distributions.

(ii) More approximate fixed-fraction particle tagging schemes, including the STINGS scheme, reproduce SPH star particle results for the accreted stellar halo well provided that the particles used to trace a particular stellar population are selected *at the time when that population actually forms* (this is especially true for heavily phase-mixed populations). Our results on this point reinforce the conclusions of Le Bret et al. (2015).

(iii) A fixed-fraction tagging scheme applied to a DM only simulation (most relevant from a practical point of view) can also yield a close match to SPH results for the accreted stellar component, *in so far as the orbital evolution of subhaloes agrees between the two simulations*. Differences in debris distributions are smallest for heavily phase-mixed populations and streams at large distances from the galaxy, and largest for coherent streams produced by the interaction of heavily stripped satellites with the inner regions of the potential. The spatial extent of the component (e.g. its half-mass radius) is also recovered by tagging, although its detailed 3D distribution may depend on additional factors that are not taken into account even approximately by fixed-fraction schemes (especially for complex galaxies like the MW; see below).

(iv) Star formation modelling is the most important 'nuisance' factor to control for in comparisons between particle tagging and hydrodynamic simulations. In such comparisons, the differences that can be attributed to the dynamical limitations of particle tagging only become significant when much larger discrepancies arising from the use of different star formation prescriptions are eliminated. Controlling for those discrepancies requires either a good understanding of how the star formation prescriptions in different models correspond to one another, or else a robust procedure for 'transplanting' SFHs from a simulation with baryons to its collisionless equivalent. Here, we find that reducing the strength of feedback in our semi-analytic model (which is otherwise identical to that used in C10) matches the SFHs in our AqC-SPH simulation well and substantially improves the correspondence between tagged and star particles, simply by changing when, where and in what quantity stars are predicted to form.

(v) In our simulation, the conditions necessary for good correspondence between SPH and particle tagging representations are met for most of the significant progenitors of the stellar halo. These include: very little long-lasting modification of satellite haloes by contraction or expansion due to the motion of baryons; the formation of the majority of accreted stars before infall into to the MW-like halo; a lack of strong interactions between significant progenitor satellites and the baryon-dominated regions of the central potential; and the slow growth of that central baryonic contribution, also with limited overall contraction or expansion. Particle tagging will naturally provide a worse approximation to hydrodynamical simulations in which some or all these conditions are not met.

Our conclusions are limited by the fact that we have only examined one SPH simulation, as did the similar 'comparative' studies of Libeskind et al. (2011) and Bailin et al. (2014). Moreover, those earlier studies drew strong conclusions about the general merits of using particle tagging models to interpret observational data based on the implicit assumption that specific SPH simulations were themselves suitable for that purpose. That assumption is hard to justify without a robust statistical comparison between a cosmologically representative set of real and simulated galaxies. Our AqC-SPH simulation is a modest improvement in this respect (it is constrained by the MW satellite luminosity function and does not suffer rampant 'overcooling', demonstrated by a ratio of stellar mass to halo mass in agreement with abundance matching, a stable stellar disc and low bulge-to-total mass ratio; Okamoto et al. 2010, Parry et al. 2012). More recent simulations have made further improvements with regard to large-scale constraints on galaxy formation (e.g. Sawala et al. 2016).

Hydrodynamical models remain computationally expensive and subject to large, poorly constrained uncertainties in their 'subgrid' recipes, which can easily overwhelm the advantage of dynamical self-consistency. So long as this remains the case, our results suggest further tests of particle tagging using larger samples of haloes from a wide variety of hydrodynamical schemes would be worthwhile, and motivate the investigation of some improvements to the methodology. Specifically:

(i) Controlled numerical experiments would be helpful to determine how important dynamical differences in the disruption of satellites are for applications of particle tagging, in isolation from the many uncertainties involved in cosmological hydrodynamical simulations and star formation physics. For example, it would be useful to quantify the fraction of significant halo progenitors that

interact strongly with regions of the potential dominated by baryons (i.e. with the disc, in the MW case) and how exactly the orbits of these systems differ between simulations with and without self-consistent hydrodynamics.

(ii) Further constraints on the modification of real galactic potentials by star formation and feedback would help to inform judgements about particle tagging. There is an ongoing debate on this point in the theoretical literature and the observational situation is also uncertain (for a recent summary see e.g. Oman et al. 2016). For a given star formation model applied to an MW-like system, it would be useful to quantify how common heavily modified satellite galaxies are among typical sets of halo progenitors. If semi-analytic models could be used to predict the degree of baryonic modification to satellites, a small number of strongly modified satellites could at least be flagged and treated with appropriate caution in a subsequent particle tagging analysis.

(iii) In the context of fixed-fraction tagging schemes, the most appropriate fraction of most-bound particles varies from population to population. This suggests a refinement to the scheme in which the tagged fraction can vary based on other physical parameters such as galaxy size and the mass of newly formed stars. However, this would require more free parameters and may be unnecessarily complex for many applications of particle tagging. Likewise, any single-fraction scheme cannot reproduce SPH results for *in situ* stars as well in cases where multiple populations with different intrinsic binding energy distributions form in the same halo at the same time. Information about the relative star formation rates in different regions of the central potential could be obtained from the underlying semi-analytic model and used in a more complex tagging scheme to construct a more accurate distribution of stellar binding energies.

(iv) Collisionless simulations that account in some way for the baryonic contribution to the host galaxy potential (for example by adding a smoothly growing disc potential) would likely perform even better in comparisons against SPH simulations. This was the spirit of the approach in Bullock & Johnston (2005) and could be greatly improved on with modern numerical techniques (e.g. Lowing et al. 2011). When implementing a scheme like this, it will be important to ensure that the density, radial extent, stability and growth rate of baryonic components of the potential satisfy observational constraints (Aumer & White 2013; Aumer et al. 2013).

(v) In some hydrodynamic simulations (including our AqC-SPH), the halo is formed mostly from gas stripped from massive satellites – the same satellites whose stars contribute the bulk of the accreted stellar halo (Cooper et al. 2015b). In such a scenario, particle tagging might be adapted to approximate the formation of stars in streams of stripped gas, provided those streams are almost 'ballistic' (i.e. they form soon after the gas is stripped, such that the kinematics of the stars are dominated by the orbital motion of their parent satellite rather than hydrodynamic interactions). For example the fraction of recently unbound gas converted to stars after stripping could be estimated and stars tagged to specific recently unbound collisionless particles.

With or without these improvements, particle tagging models are intended as an approximation and consequently have very clear dynamical limitations. The results we have reported suggest to us that no galaxy formation theory is sufficiently well constrained at present to make those limitations more important than differences between subgrid (or semi-analytic) star formation recipes.

ACKNOWLEDGEMENTS

We thank the anonymous referee for their careful reading of our manuscript. APC is supported by a COFUND/Durham Junior Research Fellowship under EU grant 267209 and thanks Andrew Benson for discussions that led to this work. CSF acknowledges ERC Advanced grant 267291 'COSMIWAY'. APC, SC and CSF acknowledge support from STFC (ST/L00075X/1). This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grants ST/H008519/1 and ST/K00087X/1, STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure. We acknowledge use of matplotlib (Hunter 2007) and NASA's Astrophysics Data System Bibliographic Services.

REFERENCES

Abadi M. G., Navarro J. F., Fardal M., Babul A., Steinmetz M., 2010, MNRAS, 407, 435
Arraki K. S., Klypin A., More S., Trujillo-Gomez S., 2014, MNRAS, 438, 1466
Aumer M., White S. D. M., 2013, MNRAS, 428, 1055
Aumer M., White S. D. M., Naab T., Scannapieco C., 2013, MNRAS, 434, 3142
Bailin J., Bell E. F., Valluri M., Stinson G. S., Debattista V. P., Couchman H. M. P., Wadsley J., 2014, ApJ, 783, 95
Binney J., Piffl T., 2015, MNRAS, 454, 3653
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
Bullock J. S., Johnston K. V., 2005, ApJ, 635, 931
Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001a, ApJ, 555, 240
Bullock J. S., Kravtsov A. V., Weinberg D. H., 2001b, ApJ, 548, 33
Cole S., Lacey C., 1996, MNRAS, 281, 716
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Cooper A. P. et al., 2010, MNRAS, 406, 744 (C10)
Cooper A. P., Cole S., Frenk C. S., Helmi A., 2011, MNRAS, 417, 2206
Cooper A. P., D'Souza R., Kauffmann G., Wang J., Boylan-Kolchin M., Guo Q., Frenk C. S., White S. D. M., 2013, MNRAS, 434, 3348 (C13)
Cooper A. P., Gao L., Guo Q., Frenk C. S., Jenkins A., Springel V., White S. D. M., 2015a, MNRAS, 451, 2703
Cooper A. P., Parry O. H., Lowing B., Cole S., Frenk C., 2015b, MNRAS, 454, 3185
D'Onghia E., Springel V., Hernquist L., Keres D., 2010, ApJ, 709, 1138
De Lucia G., Helmi A., 2008, MNRAS, 391, 14
Dooley G. A., Peter A. H. G., Vogelsberger M., Zavala J., Frebel A., 2016, MNRAS, 461, 710
Errani R., Peñarrubia J., Tormen G., 2015, MNRAS, 449, L46
Helmi A., Cooper A. P., White S. D. M., Cole S., Frenk C. S., Navarro J. F., 2011, ApJ, 733, L7
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Keller B. W., Wadsley J., Couchman H. M. P., 2015, MNRAS, 453, 3499
Lacey C. G. et al., 2016, MNRAS, 462, 3854
Laporte C. F. P., White S. D. M., Naab T., Gao L., 2013, MNRAS, 435, 901
Le Bret T., Pontzen A., Cooper A. P., Frenk C., Zolotov A., Brooks A. M., Governato F., Parry O. H., 2015, preprint (astro-ph/1502.06371)
Libeskind N. I., Knebe A., Hoffman Y., Gottlöber S., Yepes G., 2011, MNRAS, 418, 336
Li Y.-S., Helmi A., 2008, MNRAS, 385, 1365
Lowing B., Jenkins A., Eke V., Frenk C., 2011, MNRAS, 416, 2697
McConnachie A. W., 2012, AJ, 144, 4
McKee C. F., Parravano A., Hollenbach D. J., 2015, ApJ, 814, 13

Monachesi A., Gómez F. A., Grand R. J. J., Kauffmann G., Marinacci F., Pakmor R., Springel V., Frenk C. S., 2016, MNRAS, 459, L46
Navarro J. F., Eke V. R., Frenk C. S., 1996, MNRAS, 283, L72
Nipoti C., Binney J., 2015, MNRAS, 446, 1820
Okamoto T., Frenk C. S., Jenkins A., Theuns T., 2010, MNRAS, 406, 208
Oman K. A., Navarro J. F., Sales L. V., Fattahi A., Frenk C. S., Sawala T., Schaller M., White S. D. M., 2016, MNRAS, 460, 3610
Parry O. H., Eke V. R., Frenk C. S., Okamoto T., 2012, MNRAS, 419, 3304
Peñarrubia J., McConnachie A. W., Navarro J. F., 2008, ApJ, 672, 904
Peñarrubia J., Benson A. J., Walker M. G., Gilmore G., McConnachie A. W., Mayer L., 2010, MNRAS, 406, 1290
Pontzen A., Governato F., 2012, MNRAS, 421, 3464
Rashkov V., Madau P., Kuhlen M., Diemand J., 2012, ApJ, 745, 142
Read J. I., Agertz O., Collins M. L. M., 2016, MNRAS, 459, 2573
Sawala T. et al., 2015, MNRAS, 448, 2941
Sawala T. et al., 2016, MNRAS, 457, 1931
Scannapieco C. et al., 2012, MNRAS, 423, 1726
Schaye J. et al., 2015, MNRAS, 446, 521
Spitzer L. Jr., Chevalier R. A., 1973, ApJ, 183, 565
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2008, MNRAS, 391, 1685
Stinson G. S., Bailin J., Couchman H., Wadsley J., Shen S., Nickerson S., Brook C., Quinn T., 2010, MNRAS, 408, 812
Tumlinson J., 2010, ApJ, 708, 1398
Wang W., Han J., Cooper A., Cole S., Frenk C., Cai Y., Lowing B., 2015, MNRAS, 453, 377
Zhu Q., Marinacci F., Maji M., Li Y., Springel V., Hernquist L., 2016, MNRAS, 458, 1559

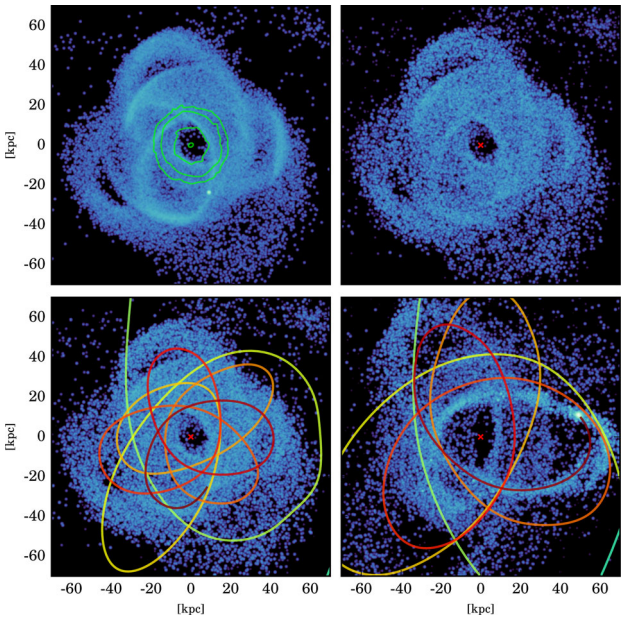

Figure A1. Surface density of debris from the 'trefoil stream' (the most massive halo progenitor) in four realizations. Clockwise from top left: AqC-SPH, AqC-SPH with 1 per cent tags, AqC-DM with 5 per cent tags and AqC-SPH with 5 per cent tags. The centre of the main halo is marked by a red cross. The trajectory of the progenitor is shown in the lower panels, with increasing time running from green to red. Contours of the central disc surface density are shown in grey in the AqC-SPH panel (top left), corresponding to $\log_{10} \Sigma/\mathrm{M}_{\odot}\,\mathrm{kpc}^{-2} = 5, 6, 7$ and 8.

APPENDIX A: USING TRANSPLANTED SPH STAR FORMATION HISTORIES AS THE BASIS FOR PARTICLE TAGGING TESTS

In Section 3.2.1, we stated that transplanting SFHs from an SPH simulation to its DM only equivalent is necessarily approximate. Some divergence is almost unavoidable, which may be physically meaningful in some cases and stochastic in others. Detailed tests of particle tagging that use this approach require careful analysis of such divergence for significant progenitors of the accreted halo in order to isolate effects that are directly attributable to the tagging technique. In Fig. 2, we showed that the distribution of accreted star particles in our SPH simulation can be reproduced reasonably well by applying particle tagging to a collisionless simulation from the same initial conditions, using transplanted SFHs for the 10 most massive progenitors. In this appendix, we analyse each of these progenitors individually to give more insight into the correspondence between our SPH and collisionless simulations. This analysis raises questions beyond the scope of our paper, so we present it mainly to highlight the uncertainties involved and to suggest directions for future work.

A1 The most massive halo progenitor

We begin with an example that illustrates in detail the case of one stellar halo progenitor for which the particle tagging results from AqC-DM do not correspond well to the star particles in AqC-SPH. This progenitor is interesting in several respects, some of which have already been described by Parry et al. (2012, section 6). At $z = 0$, it has been heavily stripped of both stars and DM, and is responsible for the most striking coherent feature in the stellar halo. We refer to this feature as the 'trefoil stream', on account of its morphology in the top left-hand panel of Fig. A1 (these streams lie roughly in the plane of the AqC-SPH disc). Secondly, this halo (the 'trefoil progenitor') has survived as a satellite of the MW analogue for ∼8 Gyr, undergoing many apocentric passages on a decaying orbit. The evolution of satellites on orbits like this, and the streams they produce, should be particularly sensitive to the shape, orientation and depth of the gravitational potential in AqC-SPH, including the baryonic contribution absent in AqC-DM. Finally, at $z = 0$ this object is bound by its remnant *stellar* mass rather than by DM; we do not expect good agreement with AqC-DM in this case where the binding energy of the stars is critical to the survival of the satellite. Fig. A1 also shows the stream as predicted by fixed-fraction tagging of DM particles in AqC-SPH (top right and bottom left) and tagging of AqC-DM based on the transplanted AqC-SPH SFH. The AqC-DM version shows thinner streams, with fewer, wider orbits and a more prominent remnant core at $(x, y) \sim (50, 10)$.

Fig. A2 quantifies some of the features that distinguish the trefoil progenitor from other satellites, comparing its radial position (relative to the main halo centre) and mass evolution between AqC-SPH and AqC-DM. The trajectories diverge around the time of the third apocentre, with the progenitor in AqC-SPH subsequently having a shorter period and a more rapid decay than its AqC-DM counterpart. This divergence seems to be associated with catastrophic mass loss in AqC-SPH between the third apocentre and fourth pericentre. The AqC-DM satellite loses mass more gradually. The mass still bound to the satellite at the present day is similar in AqC-SPH and AqC-DM, as are the relative orbital phases, despite the fact that the SPH version passes through two pericentres more than the DM version. The most likely reason for the divergence, other than stochasticity, is interaction between the SPH satellite and the baryons concentrated at the centre of the potential (the disc and bulge). In the upper panel of Fig. A2, dashed grey lines are drawn at 10 and 30 kpc,

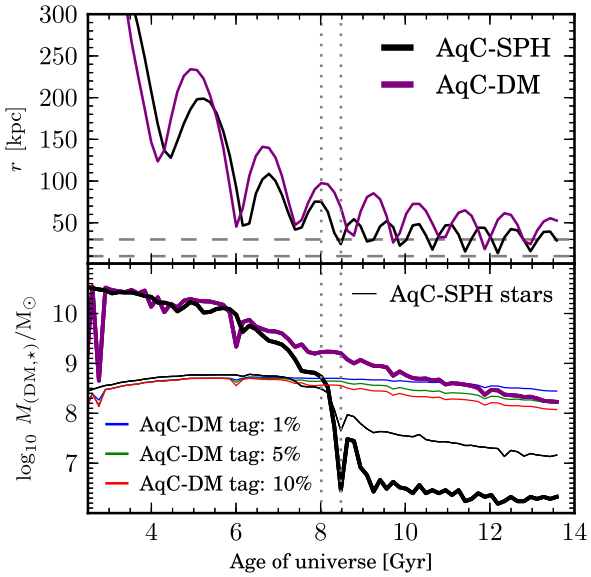

Figure A2. Galactocentric radius (upper panel) and bound mass (lower panel) of the trefoil stream progenitor (see the text and Fig. A1) with age of the universe, starting from the time of infall into the main halo, for AqC-SPH (black) and AqC-DM (purple). In the lower panel, thick lines indicate DM mass and thin lines stellar mass. Thin blue, green and red lines in the lower panel correspond to the bound stellar mass predicted by tagging of AqC-DM with $f_{\rm mb} = 1$, 5 and 10 per cent, respectively.

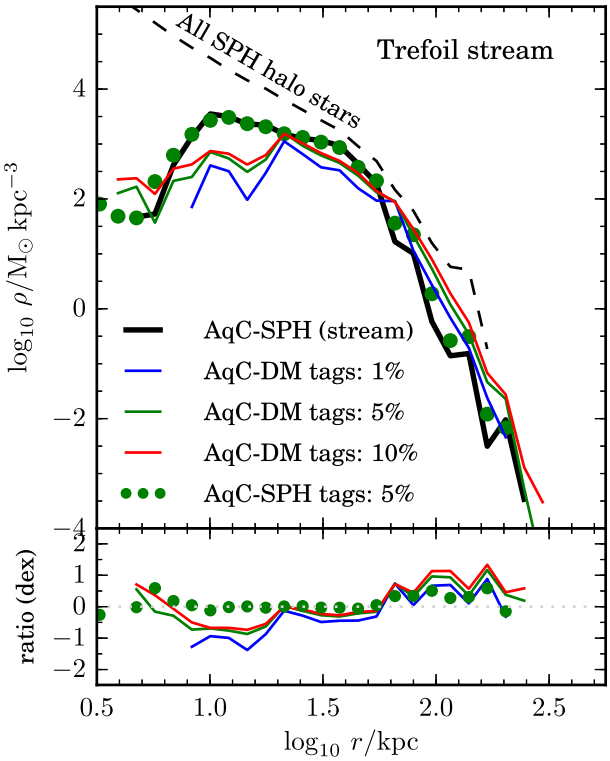

Figure A3. Surface density of stars in the 'trefoil stream' (see Fig. A1) at $z = 0$ in AqC-SPH (black). Green dots show the profile recovered by tagging ($f_{\rm mb} = 5$ per cent) in AqC-SPH, using the SFH of the SPH star particles. Blue, green and red lines show profiles recovered by using this same SFH as the basis for tagging of the matched halo in AqC-DM, with $f_{\rm mb} = 1$, 5 and 10 per cent, respectively. A grey dashed line shows the density of all accreted star particles in AqC-SPH. The lower panel shows the logarithmic ratio of each curve relative to the SPH result. Note the reduced range of radius and shift in density scale relative to previous density profile plots for the main halo.

corresponding to disc surface densities of $\sim$7 and $\lesssim 4 {\rm M}_\odot\,{\rm kpc}^{-2}$, respectively.

Fig. A3 shows the spherically averaged stellar mass density profile of stars from the trefoil stream progenitor at $z = 0$. Tagging with $f_{\rm mb} = 5$ per cent in AqC-SPH itself (shown in the lower left panel of Fig. A1) results in a close match to the star particle profile, consistent with the good overall agreement shown in Fig. 7.

When we use the AqC-SPH SFH for this object as the basis for tagging of the matched satellite in the AqC-DM simulation, the agreement is clearly worse (this result is not sensitive to the exact value of $f_{\rm mb}$). The predicted density is lower by an order of magnitude at galactocentric radii below $\sim$20 kpc, and higher by a similar factor beyond $\sim$100 kpc. This is readily understood by the differences in orbital evolution shown in Fig. A2. The mass of stars tagged to the bound core of the trefoil progenitor in AqC-DM is shown for $f_{\rm mb} = 1$, 5 and 10 per cent. These indicate an increasing mass-to-light ratio approaching $\sim$1 at the present day, with the consequence that the stellar mass in the stream (as opposed to the progenitor) is sensitive to $f_{\rm mb}$ (increasing by roughly a factor of 3 as $f_{\rm mb}$ varies from 1 to 10 per cent).

Parry et al. (2012) show that the SFH of this satellite is dominated by an extreme peak associated with rapid gas dissipation following an early, low mass ratio merger with another halo. This starburst leads to a cusped stellar density profile and 'explosive' feedback which unbinds a large fraction of the remaining gas in a short time. The outcome is a baryon-dominated central cusp and a corresponding low-density DM core. As discussed by Le Bret et al. (2015), particle tagging is a poor approximation in cases where feedback significantly alters the overall density profile. The abnormally low central density of DM after this event in AqC-SPH is not reproduced in AqC-DM, which may then explain why it does not reproduce the rapid loss of DM relative to stars in the fourth pericentric approach seen in AqC-SPH. From this single example, it is hard to divide blame between the abnormal DM density profile and differences in the interaction with the main halo, since each reinforces the effects of the other.

In summary, despite the considerable differences in the evolution of the trefoil progenitor in AqC-SPH and AqC-DM, we find the distribution of its debris to be similar overall. The differences we see illustrate the divergence between results from SPH and DM particle tagging that can be expected in cases where stellar haloes are dominated by individual 'atypical' objects. Satellites are 'atypical' in this case if either feedback significantly alters their phase-space density, or they interact with a strongly modified central galactic potential. Further work with SPH simulations is required to understand the frequency of such cases and their dependence on other aspects of the models. In our particular simulation, divergence between different realizations of the most massive halo progenitor (which is atypical on *both* of the above counts) does not affect the conclusions drawn from the spherically averaged halo density profile as a whole. The 'bias' of particle tagging is not negligible, but neither is it catastrophic. That bias could, however, alter other conclusions, for example regarding the extrema of surface brightness features that might be detected by stream-finding algorithms.

A2 Other massive progenitors

Fig. A4 show more examples of massive halo progenitors matched between AqC-SPH and AqC-DM. These examples highlight some

general systematic differences between the two simulations. As in the preceding section, the results for AqC-DM are obtained with 'transplanted' SFHs from AqC-SPH and a 5 per cent fixed-fraction tagging scheme.

In some cases, the density profiles (leftmost column) of stars accreted into the main halo by $z = 0$ are better matched between AqC-SPH (black solid line) and AqC-DM (green solid line) than in the case of the trefoil progenitor. Only one case, progenitor F,

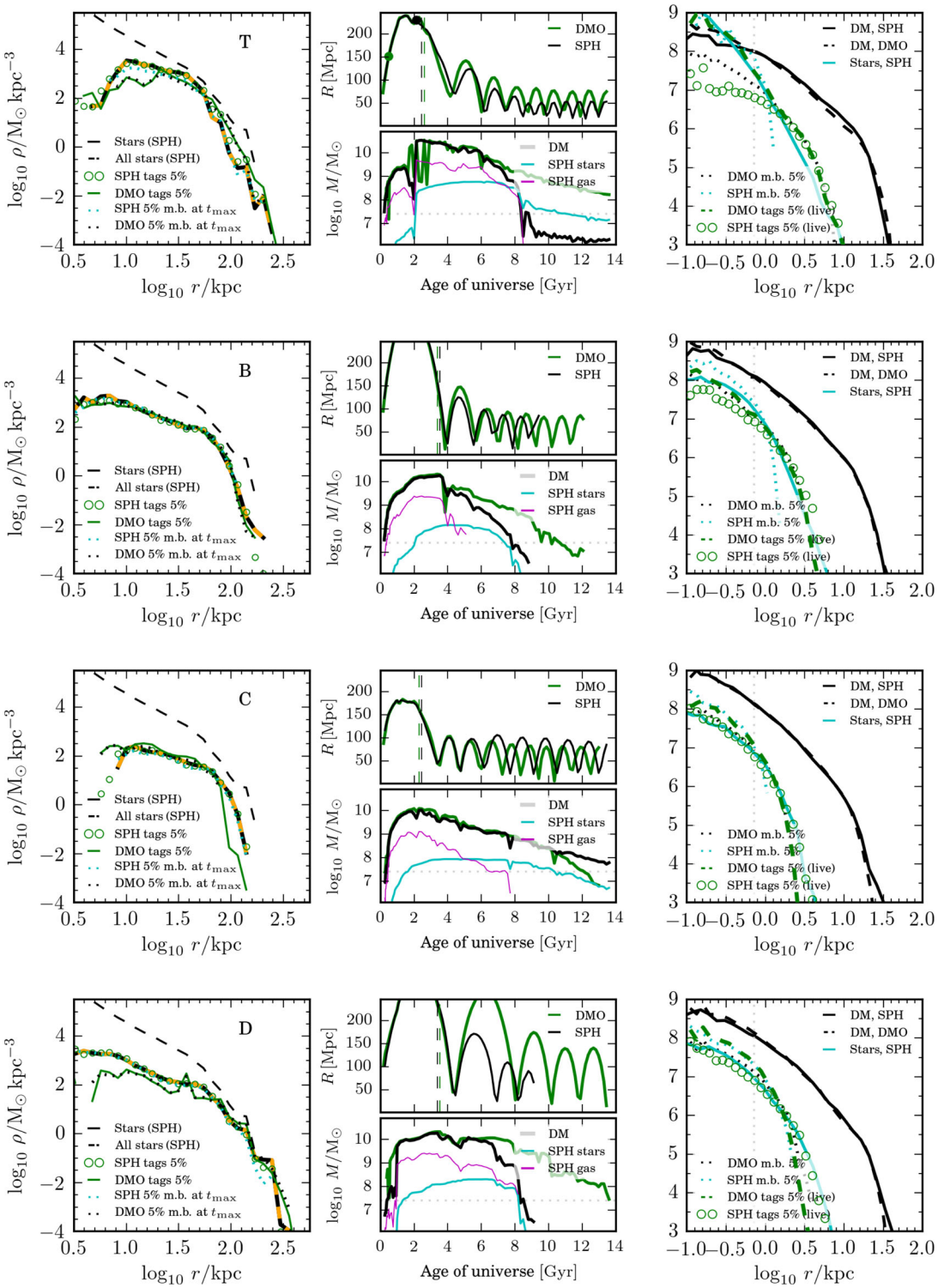

Figure A4. Further examples of individual stellar halo progenitors, comparing AqC-SPH and AqC-DM simulations. Panels show: *left*, density profile of debris at the present day; *centre top*, orbit of the progenitor (dashed line at infall time); *centre bottom*, evolution of its gaseous, stellar and dark mass (dotted line at the mass of 10 particles); *right* density profile of the satellite halo at the time of maximum mass ($t_{max} \approx t_{infall}$). Dotted black and blue lines in left- and right-hand panels correspond to a selection of the most bound 5 per cent of the DM at the time of infall; dashed green line and circles in the right-hand panel correspond to the actual distribution of 'tagged' DM in AqC-DM and AqC-SPH, respectively.

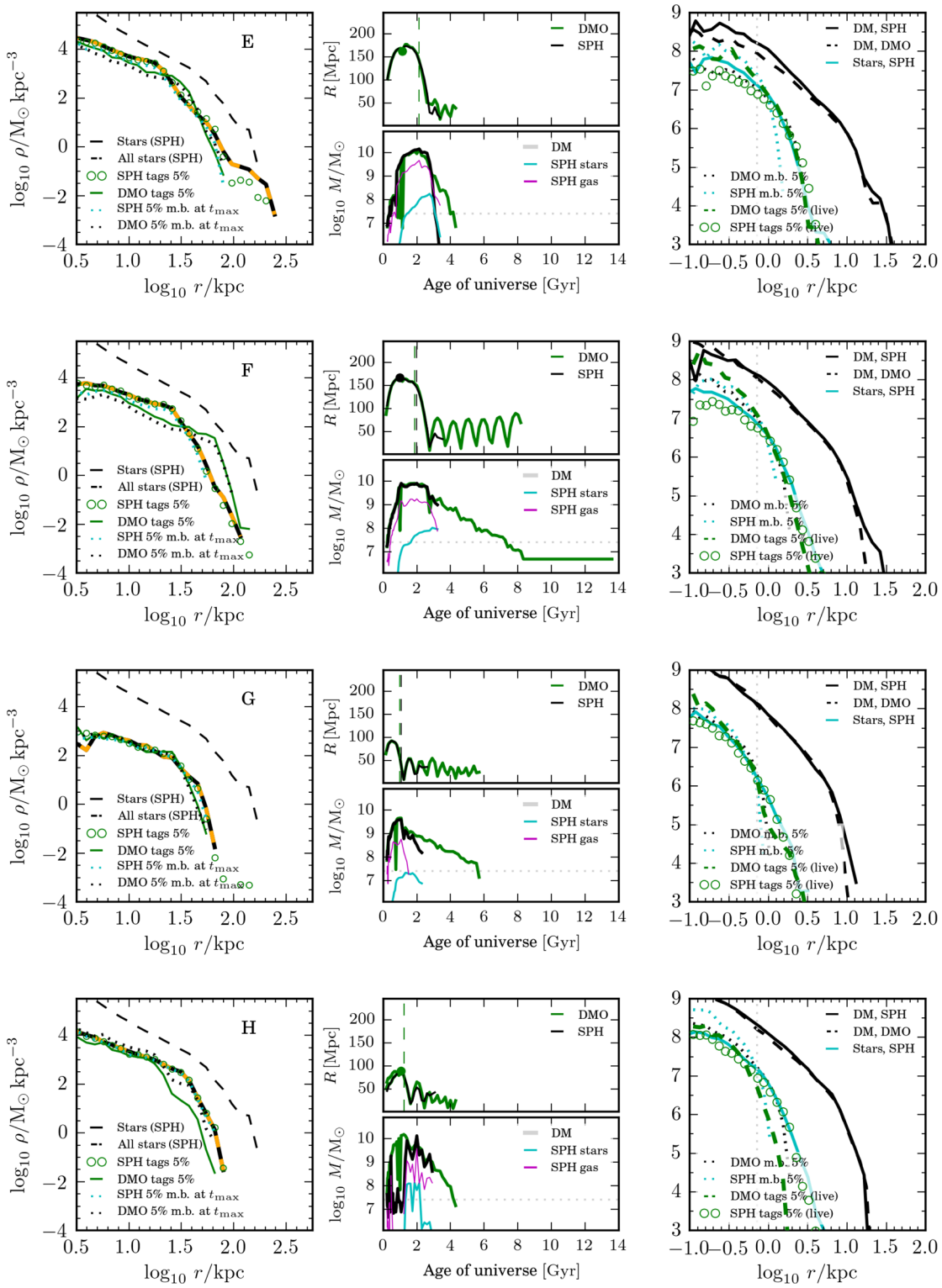

Figure A4 – *Continued.*

clearly shows greater discrepancy. This may be because progenitor F is disrupted less than 1 Gyr after infall in AqC-SPH (after only one pericentre) but survives for ∼6 Gyr (six pericentres) in AqC-DM. Taking all these examples together, the most obvious systematic difference is that most progenitors are disrupted somewhat more quickly in AqC-SPH than they are in AqC-DM (progenitor C being the only exception). Only in the case of progenitor F, however, does this have an obvious effect on the distribution of the resulting halo stars at $z = 0$. Progenitor B is a counterexample, being similar to C in many respects, having greater divergence in mass-loss rate, and yet having little discernible discrepancy in the profile of its debris.

It is beyond the scope of this paper to explore why exactly the orbits of massive satellites diverge in this way between AqC-SPH and AqC-DM. One possibility, noted above, is that it is the result of interaction with a central potential heavily modified (contracted and made more spherical) by baryonic effects. The mass-loss rates shown in Fig. A4 (lower middle panels) suggest another mechanism related to the to the rapid removal of gas from the progenitors (magenta lines) most likely by ram pressure as they pass through the densest regions of the halo (Arraki et al. 2014). It is clear (in particular for B, C and perhaps G) that gas can be removed more rapidly and essentially disappear after few orbits. The total removal of gas seems to correlate with the onset of divergence in orbit and mass-loss rate of DM between AqC-SPH and AqC-DM. Other stochastic effects are possible; for completeness, we note that satellites often arrive in weakly bound groups of DM haloes with and without stars (e.g. Li & Helmi 2008). Small changes in the interactions among members of these groups before and after infall could be another source of stochasticity for satellite orbits in cosmological simulations and are extremely difficult to control for.

The density profiles of debris at $z = 0$ (left) and the intact progenitor at the time of infall (right) also show the predictions of the tag-at-infall approach used by other studies of particle tagging (dotted green and cyan lines; see Section 6.1). Tagging at infall in these examples does not show much difference compared to the results of live tagging because feedback in AqC-SPH is relatively weak overall, as discussed by (Le Bret et al. 2015) (furthermore, these examples neglect any stars accreted by the satellite – including both and accreted stars in a single assignment is another source of inaccuracy in the tag-at-infall approach). The clear counterexample is the trefoil progenitor, which, as discussed above, has a small DM core at infall, the innermost part of the galaxy being dominated by stars. In this case, at infall, the most bound 5 per cent of the DM particles in AqC-SPH are good proxy for the distribution of the innermost star particles, somewhat better even than the live tags in the same simulation, although they truncate at ∼1 kpc whereas star particles (and live tags) extend to ∼10 kpc. Notably, the equivalent set of the 5 per cent most bound particles in AqC-DM are *not* a good proxy for the AqC-SPH star particles. Live tagging of AqC-DM, on the other hand, results in a reasonable match to AqC-SPH. This is the only one of our examples to show such complex behaviour.

APPENDIX B: COMPARING SPH AND GALFORM STAR FORMATION MODELS

In Section 3.2.2, we presented particle tagging results for the AqC-DM simulation based on SFHs predicted by our semi-analytic code, GALFORM. Since there is little reason to expect that GALFORM with our fiducial choice of parameters will predict SFHs similar to those of AqC-SPH, we examined two choices of GALFORM parameters, which correspond to relatively 'weaker' and 'stronger' feedback. The weak feedback choice results in a total stellar mass for the MW analogue and a density profile for its accreted stellar halo more comparable to that of AqC-SPH, whereas the strong feedback case corresponds to the parameter set used by C10.[11] Interestingly, the good agreement in Fig. 2 suggests the SFHs produced by GALFORM for the entire merger tree of AqC-DM are not substantially different from those predicted by the full hydrodynamic calculation.

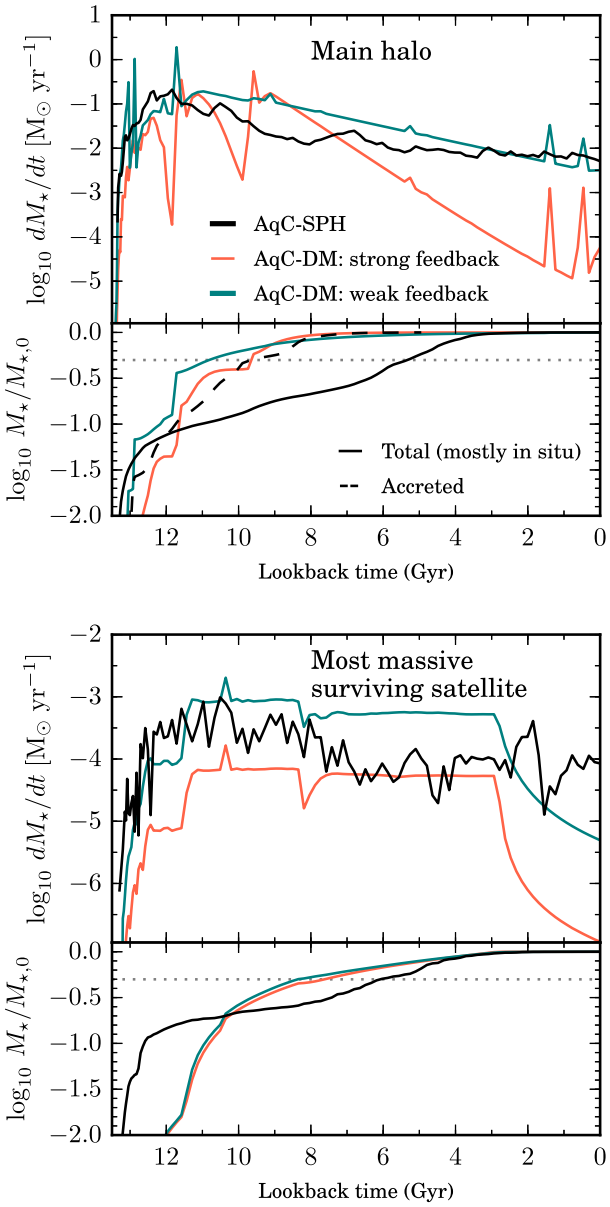

Figure B1. Top: *in situ* SFHs (upper panel) and fraction of $z = 0$ stellar mass in place (lower panels) as a function of lookback time for the MW analogue in our AqC-SPH simulation (black). These are compared with SFHs from two semi-analytic GALFORM models, with relatively 'strong' (red) and 'weak' (blue) feedback (see the text), applied to the AqC-SPH and AqC-DM simulations. The black dashed line in the lower panel is the collective SFH of star particles in the accreted stellar halo at $z = 0$ in AqC-SPH. Bottom: the same, for the most massive satellite of the MW analogue at $z = 0$.

Fig. B1 confirms this similarity by comparing the GALFORM SFHs for the main halo in our simulation and its most massive *surviving* satellite against those from AqC-SPH. For both the weak and strong feedback variants, GALFORM predicts more bursts of star formation in the main halo at redshifts $z > 2$, which leads to more rapid growth of the MW analogue relative to the AqC-SPH calculation. Conversely, the amplitude of the GALFORM SFR at later times is relatively low,

[11] The set of parameters we call 'strong feedback', essentially the model of Bower et al. (2006), was used by C10 because it can match a number of observational constraints from the wider galaxy population when applied to a representative cosmological volume; the weak feedback variant, all other parameters being held fixed, would substantially overpredict the number of galaxies with $L \lesssim L_\star$.

but approximately constant, as in AqC-SPH. The lower panels of Fig. B1 show that, in the main halo, the rapid rate of mass growth at high redshift in the GALFORM models more closely resembles that of the stars in AqC-SPH that are eventually accreted (black dashed line) than the stars formed in AqC-SPH.

The agreement of SFR predictions for the most massive satellite appears slightly better, although the total stellar mass is underpredicted in our strong feedback variant. This overall increase in stellar mass is responsible for the somewhat better agreement between the weak feedback variant and AqC-SPH with regard to the amplitude of the stellar halo density profile, shown in Fig. 2. In AqC-SPH, this satellite halo forms the bulk of its stellar mass at very high redshift, but still considerably earlier than GALFORM predicts. These differences in the rate of growth of the stellar mass are important, because they determine the characteristic scale of the DM halo at the time of tagging and hence the initial scale radius of the *in situ* population. All else being equal, earlier peak star formation will lead to the stellar tags being assigned to more tightly bound DM particles and hence more compact density profiles at $z = 0$. For these two cases, where we see that the time-scale of *in situ* star formation is significantly longer in AqC-SPH, that effect most likely contributes to the more concentrated components seen for AqC-DM in Fig. 2.

APPENDIX C: SATELLITE SIZES AND SURFACE BRIGHTNESS PROFILES

The main text focuses on the MW analogue halo in our AqC-SPH and AqC-DM simulations. Fig. C1 shows comparisons between stars and tagged particles for the most massive surviving satellite subhalo, analogous to Figs 2 and 7. This is relevant because the size–mass relation of surviving satellites provides an important constraint on the choice of $f_{\rm mb}$. Although the precise choice of $f_{\rm mb}$ does not have a strong effect on the distribution of stripped stars (Fig. 7), it directly determines the scalelength of the component in satellites that are not strongly perturbed by tidal forces (as noted by C10 and elaborated on by C13).

The connection between $f_{\rm mb}$ and the sizes of surviving satellites is shown in the left-hand panel of Fig. C1. Clearly, for this particular satellite, $f_{\rm mb} \sim 5$ per cent is close to the optimal choice. In the right-hand panel of Fig. C1, we contrast this result with tagging of AqC-DM based on GALFORM ($f_{\rm mb} \sim 5$). The impact of the different SFHs in our two GALFORM variants is clear, changing the amplitude of the profile (i.e. the total mass of stars in the satellite) by an order of magnitude. Neither variant reproduces the AqC-SPH profile very closely, suggesting that baryonic effects on the potential and/or the orbital evolution of this satellite may differ significantly between AqC-SPH and AqC-DM.

C10 and C13 calibrated $f_{\rm mb}$ according to the median relation between half-mass radius, R_{50}, and stellar mass, $M_\star$. Fig. C2 shows this relationship for galaxies in AqC-SPH. Half-mass radii are measured from the centre of the potential of each subhalo as reported by SUBFIND (Springel et al. 2001). This figure also shows data from galaxies in the Local Group to demonstrate that the distribution of sizes for well-resolved galaxies in AqC-SPH are consistent with those of real dwarf galaxies of similar mass. In very low-mass haloes harbouring the smallest galaxies in AqC-SPH, the potential is artificially cored by the gravitational softening, which artificially inflates the sizes.

C10 found that $f_{\rm mb} = 1$ per cent resulted in a size–mass relation in reasonable agreement with the Local Group observations shown in Fig. C2, although this also corresponded to the lower

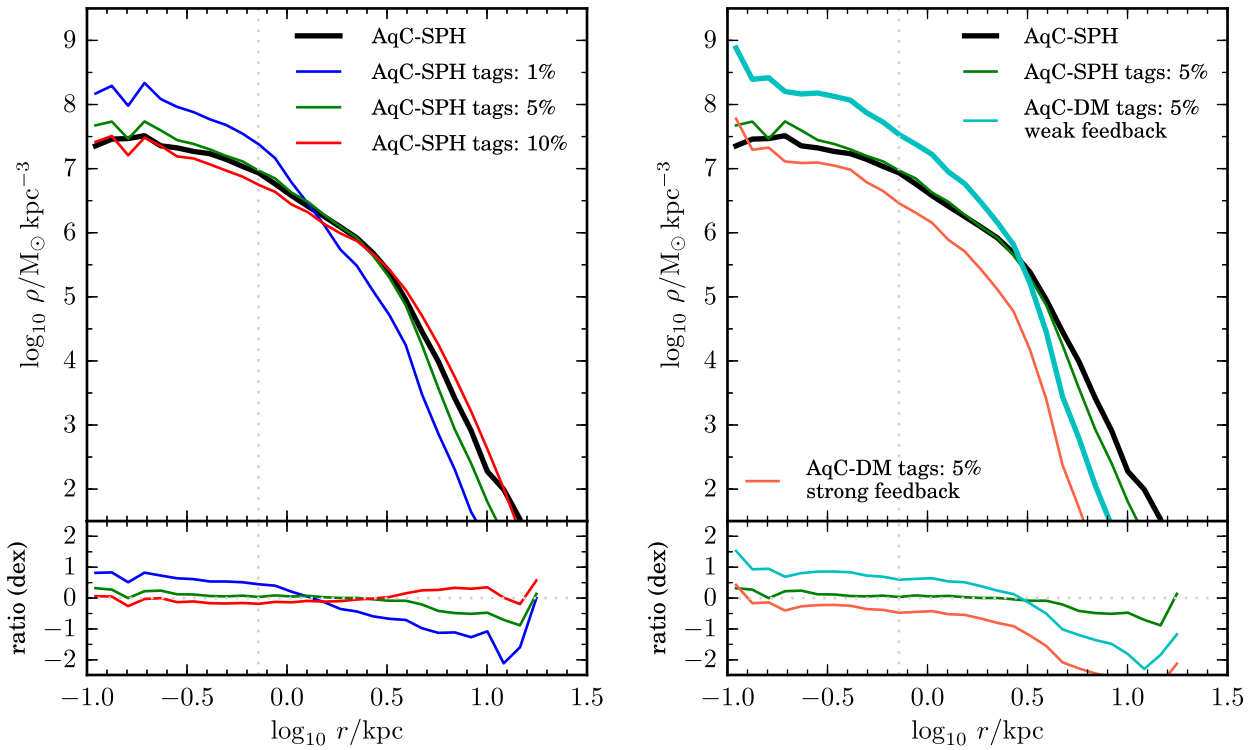

Figure C1. Left: stellar mass density profiles as Fig. 7 for the most massive satellite halo of our MW analogue at $z = 0$. The lower panel shows the ratio of stellar mass density in tagged particles to that in SPH star particles. Right: a similar comparison with the results of tagging ($f_{\rm mb} = 5$ per cent) based on a GALFORM model in AqC-DM (cyan), as in Fig. 2.

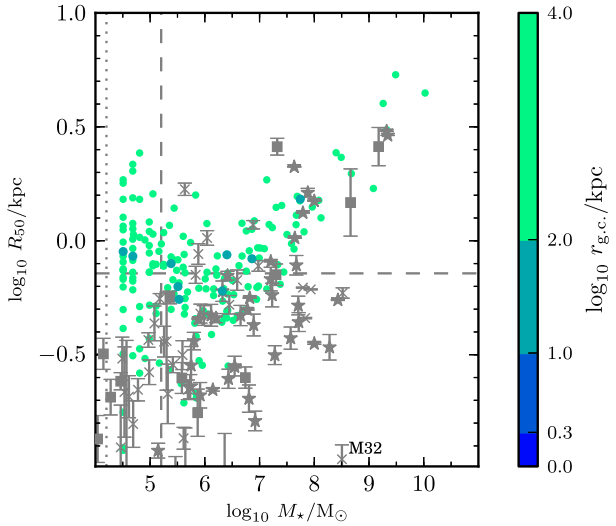

Figure C2. Half-mass radii of individual satellites in our SPH simulation as a function of their stellar mass (points), colour-coded by their distance from the central galaxy. Grey symbols with error bars show corresponding data for the dwarf galaxies around the MW (squares), around M31 (crosses) and in the Local Group (stars) for which both mass and size measurements are available in the compilation of McConnachie (2012). In cases where the error in half-mass radius is unknown (17 of 95 objects), an error of ±50 per cent is assumed. The vertical and horizontal dashed lines indicate the limits in mass and size below which the finite resolution of the simulation renders these results unreliable. The horizontal dashed line shows the force softening scale of our simulation. At low mass, the apparent discretization is due to the approximately quantized mass of individual star particles; vertical lines correspond to the mass of 1 (dotted) and 10 (dashed) star particles.

limit of convergence with numerical resolution for their simulations. For a lower resolution simulation of a much larger volume and using a different semi-analytic model, C13 found values in the range 2–5 per cent best matched the field galaxy size–mass relation for late-type galaxies. Given the agreement between AqC-SPH and these observations, it is not surprising that a similar value of $f_{\rm mb} \sim 5$ per cent best reproduces the results of the subgrid star formation model in AqC-SPH.

The size–mass relations predicted by the tagging model variants discussed in this paper are shown in Fig. C3. The left-hand panel shows the relations that result from tagging AqC-SPH with $f_{\rm mb} = 1$, 5 and 10 per cent. The main halo and the most massive surviving satellite discussed in the previous subsection are highlighted by star symbols.

The right-hand panel of Fig. C3 shows similar results for the idealized 'nearest neighbour' tagging scheme discussed in Section 4.1. Grey lines link the two representations of each satellite in this figure. For the majority of well-resolved satellites (those in the upper right quadrant marked by dashed grey lines), the tagged particle representation has systematically smaller half-mass radius compared to its AqC-SPH star particle counterpart, although the difference is small ($\lesssim$0.1 dex). Conversely, tagged particles bound to the main halo (rightmost point, highlighted) have a larger half-mass radius than the corresponding star particles. The two points representing the most massive satellite are highlighted with pink stars in Fig. C3; both figures show that the distribution of tagged particles is slightly more compact than that of the corresponding star particles.

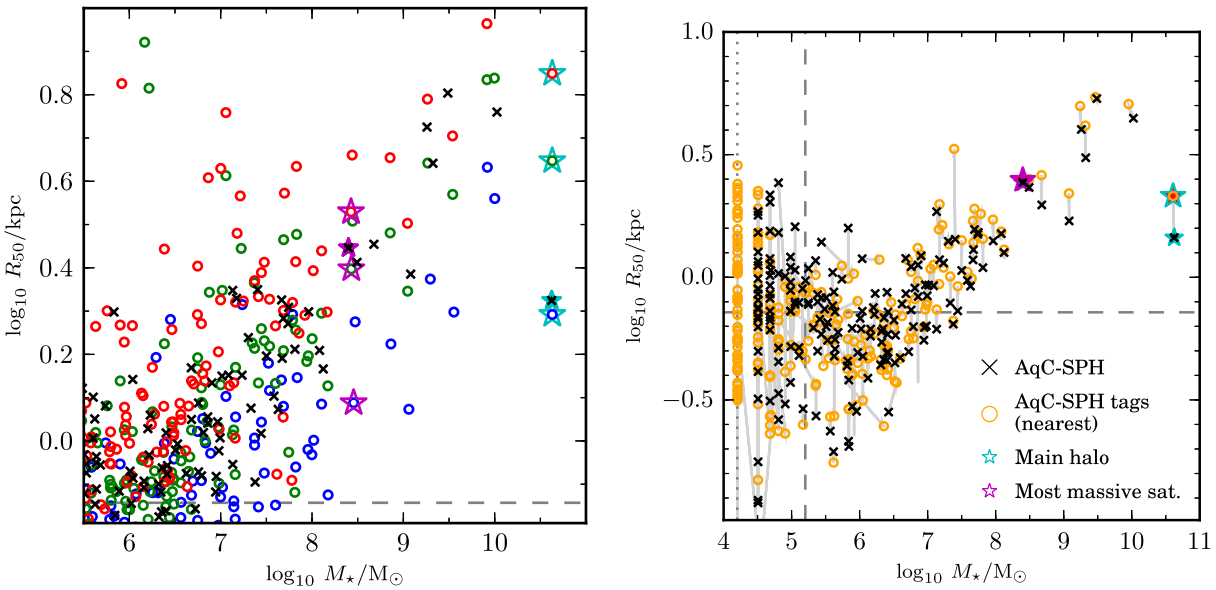

Figure C3. Satellite size–mass relations. Black crosses show the SPH results. Left: comparison against STINGS fixed-fraction tagging with different $f_{\rm mb}$: red 10 per cent, green 5 per cent and blue 1 per cent. The set of four points (three tagging and one SPH) corresponding to the main halo are marked with cyan stars; likewise, points for the most massive satellite are marked with magenta stars. Right: comparison with 'nearest energy neighbour' tagging scheme (orange). The central galaxy and its largest satellite are indicated as in the left-hand panel. We have excluded subhaloes with low-resolution particles. Solid grey lines link the AqC-SPH point for each galaxy to its corresponding tagged particle realization. Broken grey lines mark the softening scale (horizontal dashed) and the mass of 1 and 10 gas particles (vertical dotted and dashed, respectively).

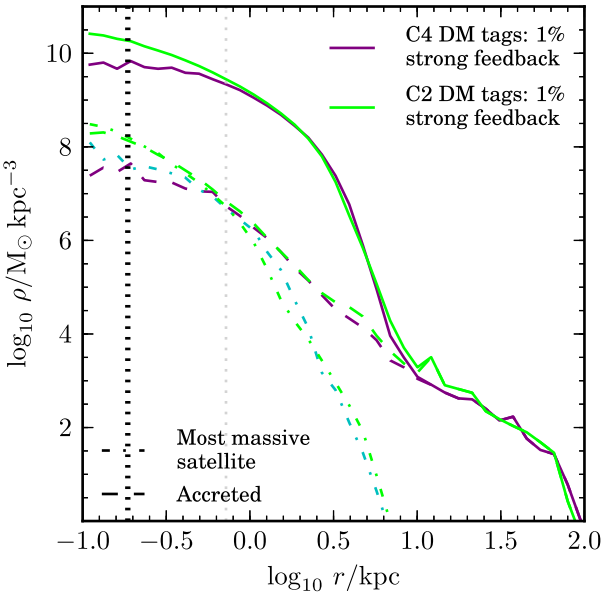

Figure D1. Convergence of the density profiles shown in Fig. 2. The green line results from application of the 'strong feedback' model to a collisionless simulation with identical initial conditions to AqC-DM and particle mass reduced by a factor of 20. The dot–dashed line shows the density profile of the most massive satellite.

APPENDIX D: CONVERGENCE

Fig. 2 shows the density profiles of the main halo and its most massive satellite that result from application of the GALFORM models discussed in this paper to a higher resolution version of our collisionless simulation (AqC-DM), with a particle mass $\sim$20 $\times$ lower. This is the resolution level used by C10. Fig. 2 demonstrates that the particle tagging results we are concerned with here have converged at the resolution limit of AqC-DM.

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