Where did it all come from? Inter-Lagrangian baryon transfer in the SIMBA simulations

Josh Borrow[1], Daniel Anglés-Alcázar[2] & Romeel Davé[2,3]

[1] Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham, DH1 3LE, UK
[2] Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA
[3] Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK

12 September 2018

ABSTRACT

In this work a new framework for analysing the movement of baryonic matter in cosmological simulations is presented. This analysis uses only the initial conditions and final state of the simulation to look at how matter has been transferred between bound structures. This is performed using two independent metrics; the distance between two closest neighbours in the initial conditions in the final simulation state, and the fractions of mass in halos at redshift $z = 0$ that originated in a given lagrangian region as defined by the dark matter. Only 60% of the baryonic matter in a given halo at $z = 0$, roughly independent of halo mass, originates from the lagrangian region defined by the resident dark matter in that halo. The fraction of baryonic mass in a given halo, as a function of radius, from inside the lagrangian region, from outside, and from other lagrangian regions is a very well constrained function providing insights onto the assembly of the Circumgalactic Medium (CGM). This presents interesting problems for semi-analytic models of galaxy formation, as well as zoom-in simulations, as at least 10% of the baryonic mass in a halo originated in the region defined by the dark matter of another halo. We present a number of significant avenues for future research.

Key words: galaxies: formation, galaxies: evolution, methods: N-body simulations

1 INTRODUCTION

Cosmological simulations have been used for decades to study the evolution of the universe. Particles or cells, depending on the choice of numerical method, are tracked over cosmic time until their final resting place (usually at redshift $z = 0$), where the distribution of matter can be compared to observations. In classic galaxy formation theory, dark matter collapses early to form virialised halos, into which gas is also pulled and can cool to form stars at the center (Mo et al. 2010). Early cosmological simulations were only powerful enough to include gravitational forces, and the choice was made to only include the dominant gravitational fluid, dark matter. These dark matter only simulations (see e.g. Frenk et al. 1988; Springel et al. 2005) then had a semi-analytic galaxy formation model applied on top to study the expected properties of bound objects (see e.g. Porter et al. 2014; Henriques et al. 2015; Lacey et al. 2016). Even these galaxy formation models, to accurately predict properties of galaxies, require the consideration of baryonic effects. Feedback from stars and black holes is critical to explain the observed properties of galaxies. Large-scale winds eject gas from galaxies, which can re-accrete back, remain in the Circumgalactic Medium (CGM), or reach the Intergalactic Medium (IGM) outside of halos. This cycling of baryons is an integral part of modern galaxy formation theory.

It is now possible to run full hydrodynamical models of the universe that explicitly include the baryonic component. Using codes such as TreeSPH, RAMSES, and GADGET (Hernquist & Katz 1989; Teyssier 2002; Springel 2005), along with full galaxy formation models, such as GEAR, Illustris, and EAGLE (Revaz & Jablonka 2011; Vogelsberger et al. 2014; Schaye et al. 2015), it is now possible to reproduce a large number of observed properties of galaxies. These codes can include stellar and AGN feedback, star formation, magnetic fields, and many more physical processes that are believed to be important for galaxy formation. Recent cosmological 'zoom-in' simulations from the FIRE project (Hopkins et al. 2014) have shown that gas ejected in winds from satellite galaxies can accrete onto the central galaxy, and this intergalactic transfer of material can be a primary contributor to galaxy growth. Galaxies providing intergalactic transfer material often end up merging with the central galaxy, but the extent to which galactic winds can push gas to larger scales and connect individual central halos at $z = 0$ cannot be addressed in 'zoom-in' simulations of individual galaxies (Anglés-Alcázar et al. 2016).

In this work, we extend the intergalactic transfer analysis of Anglés-Alcázar et al. (2016) to a large cosmological volume using the SIMBA simulations (Dave et al. 2018). More generally, we present a framework for analysing the relative motion of dark matter and baryons on large scales owing to hydrodynamic and feedback processes. We connect the distribution of dark matter and baryonic lagrangian resolution elements at $z = 0$ with their original distributions at the initial conditions, identifying the 'Lagrangian region' of $z = 0$ halos as the region in the initial conditions that will collapse into each dark matter halo. We quantify for the first time the large scale gas flows between Lagrangian regions and the surrounding IGM and the importance of 'inter-Lagrangian transfer' in galaxy evolution. In §2, the Simba suite is described. In §3 a simple analysis based on the distances between particles is considered, with the concept of lagrangian regions being introduced in §4. In §5, the lagrangian region analysis is ensured to be robust, and in §6 the conclusions are presented, with avenues for future research explored in §7.

2 THE SIMBA SIMULATION SUITE

This work uses the Simba simulation suite (Dave et al. 2018), which inherits a large amount of physics from Mufasa (Davé et al. 2016). Simba uses a variant of the GIZMO code (Hopkins 2015), with the Meshless-Finite-Mass (MFM) hydrodynamics solver. This solver uses a Wentland C2 kernel with 64 neighbours. In the $50h^{-1}$Mpc, 512^3, box used here, the mass resolution for the hydrodynamically active particles is $1\times10^7 h^{-1}$ M$_\odot$. The gravitational forces are solved using the Tree-PM method as described in Springel (2005) for Gadget-2, of which GIZMO is a descendent. There are 512^3 dark matter particles in the box, with a dark matter mass resolution of $7\times10^7 h^{-1}$ M$_\odot$. The cosmology used in Simba comes from Planck Collaboration (2016), with $\Omega_\Lambda = 0.7$, $\Omega_{\rm m} = 0.3$, $\Omega_{\rm b} = 0.048$, $H_0 = 68$ km s^{-1}, $\sigma_8 = 0.82$, and $n_s = 0.97$.

On top of this base code, the Simba sub-grid model is implemented. This model is fully described in Dave et al. (2018), but it is summarised here. Radiative cooling and photoionisation are included from Grackle-3.1 (Smith et al. 2016). Stellar feedback is modelled using decoupled two-phase winds that have 30% of their ejected particles set at a temperature given by the supernova energy minus the kinetic energy of the wind. In Simba, the mass loading factor of these winds scales with stellar mass using scalings from Muratov et al. (2015) that were calibrated using the FIRE zoom-in simulations (Hopkins et al. 2014).

Black holes are now fully modelled in Simba, using the torque-limited accretion model from Anglés-Alcázar et al. (2017) and Bondi (1952) accretion for the hot gas. The model for the black holes includes both kinetic (winds) and energetic (X-ray) feedback. At high Eddington ratios, the radiative-mode winds are ejected at ISM temperature at low velocity $\sim 10^3$ km s^{-1}. At low Eddington ratios, the jet-mode winds are ejected at a much higher ($\sim 10^4$ km s^{-1}) velocity; these high-velocity winds are however only allowed for black hole masses $M_{\rm BH} > 10^8$ M$_\odot$. We refer the interested reader to the full description of this feedback model in Dave et al. (2018).

3 HALO-INDEPDENT MEASURES OF BARYONIC FLOWS

3.1 Using inter-particle distances to describe transfer

Usually, analysing how baryonic and dark matter move differently requires the use of a halo finder, to identify structures between which gas can flow. However, if the gas and dark matter become decoupled, it should be possible to see this effect without having to consider bound structures at all.

To find how separated the dark matter becomes from the gas, two distances *in the final snapshot at* $z = 0$ are considered. First, the distance from each dark matter particle to the corresponding closest dark matter neighbour in the *initial conditions*, i.e. the distance

$$r_{ij,\, z=0} = \sqrt{\left|\mathbf{x}_{i,\, z=0} - \mathbf{x}_{j \ni \min(r_{ij,\, z=z_{ini}}),\, z=0}\right|^2} \qquad (1)$$

where $\mathbf{x}_i$ is the position of particle i, and $\mathbf{x}_i - \mathbf{x}_j$ is wrapped within the periodic box. The corresponding distance between the dark matter particle and the closest gas neighbour in the initial conditions is also considered, $r_{\rm gas}$. A diagrammatic representation of this is given in Figure 1. Star particles at $z = 0$ are ID matched with their gas progenitors, where at all possible.

3.2 Distribution of Distances

Now that the distances at $z = 0$ have been identified it is possible to compare how mixed the dark matter has become, compared to the gas, beginning with the distribution of final particle distances from their initial neighbour. In Figure 2 the similar dynamical distributions of the stellar and dark matter components are shown, with the gaseous component having a significantly longer tail. By the end of the simulation, gas particles can end up around $15h^{-1}$Mpc away from their original nearest neighbour; this can only occur due to the strong wind velocities that are powered by the AGN in the simulation.

Such a large separation is certainly possible over the course of the simulation in the Simba model. AGN winds are powered at around 10^4 km s^{-1}, meaning that over the whole run-time of the simulation the maximal distance that a wind can travel is nearly 150 Mpc; enough to wrap the whole box three times. This is, however, clearly an extreme upper limit, with the AGN winds being slowed immediately by interaction with the potential well of the halo.

The similarity between the dark matter and stellar distribution is clear here. Both follow extremely similar power-laws, something that must be investigated further in the future. That the stars must form out of particles that were initially gas in the simulation is non-trivial. The similar distribution could simply owe to the dynamics of the gas that eventually forms stars being dominated by gravitational forces, or the mixing timescale being short enough that the majority of the stars become well mixed with the dark matter by the end of the simulation. This could also possibly be a signature of the gravitational softening, or be a signature of tidal stripping of satellite halos, now thought to be a significant effect, as was shown by van den Bosch et al. (2018).

3.3 Particle-by-particle comparison to dark matter

Comparing solely the distributions of each particle type prevents the use of the fine-grained particle data that is available. It is possible to compare, for a given dark matter particle in the initial conditions, how much the nearest baryonic particle has moved, compared to the nearest dark matter particle. For each particle, the distance that the nearest dark matter neighbour has travelled, compared to the nearest baryonic neighbour, is plotted in Figure 3. The stellar distribution is highly symmetric and peaks around $r_{\rm star}/r_{\rm DM} = 1$, implying that the stellar and dark matter components have a very similar dynamical distribution (see also Figure 2) but that this is not a *local* effect. The gas and dark matter do not become separated from the gas, as implied by the original histogram, causing this effect; the stellar and dark matter are both mixed in a similar way by the gravitational dynamics, completely randomly. If there was a

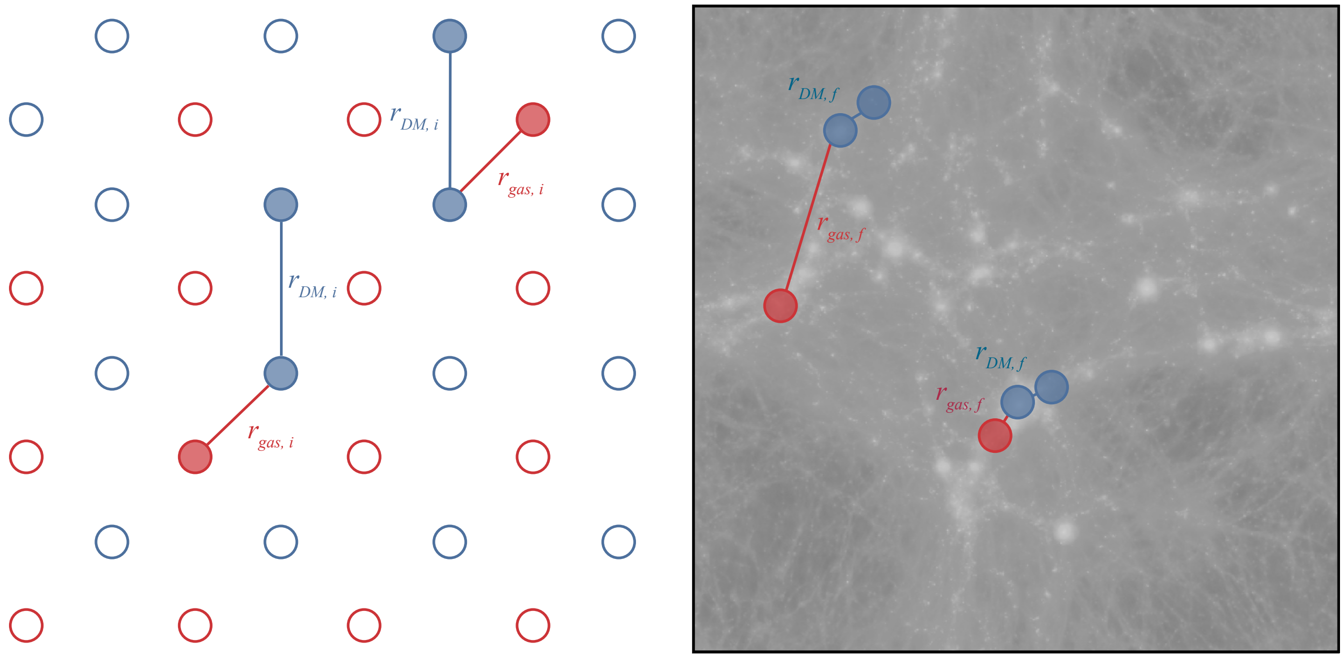

Figure 1. A diagrammatic representation of the distance measure. On the left, the initial conditions are shown. The blue dark matter particles each find their closest dark matter and gas (red) neighbour. These particles are then tracked to the final state of the simulation at z=0 (right) and the distance between them calculated again to assess the relative motion of dark matter and baryons.

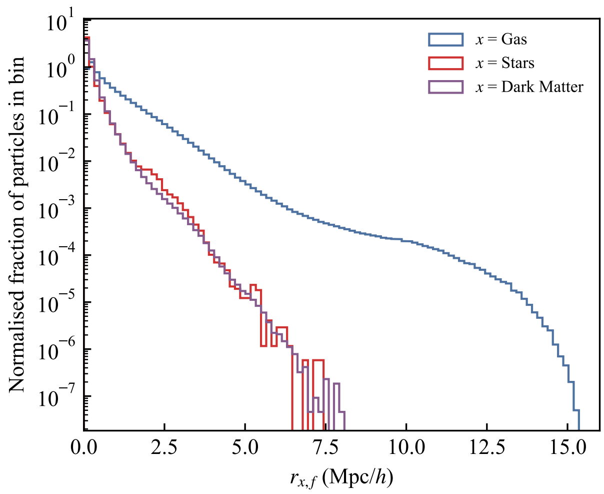

Figure 2. The distribution of all distances to particles at $z = 0$. Note the similarity between the dark matter distribution and the stellar distribution, and how different they are to the gas with the associated long tail.

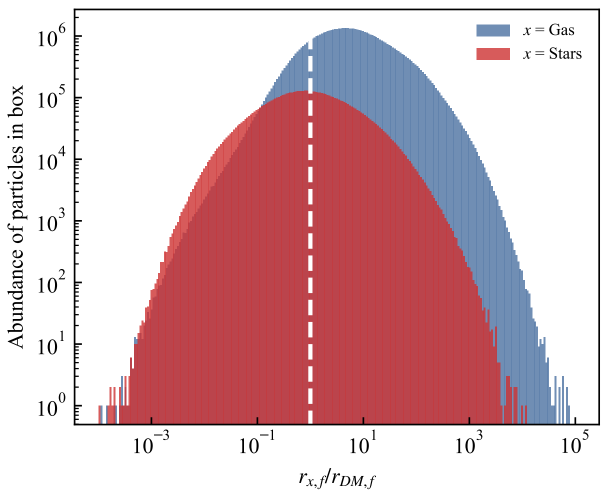

Figure 3. The distribution of relative final state distances for each dark matter particle to the appropriate gas and stellar (these correspond to i in the bottom label) particles at $z = 0$ normalised by the distance to the dark matter particle that was closest in the initial conditions. Note the significantly different distributions for gas and stars.

local effect here, we would see that the ratio of $r_{\rm star}/r_{\rm DM}$ would be much more tightly constrained.

The same distribution for the gas, however, has a different signature. $r_{\rm gas}/r_{\rm DM}$ peaks around 10, not 1, showing that gas ends up preferentially further away than the neighbouring dark matter particle. That gas behaves differently to dark matter is unsurprising; gas particles feel repulsive forces from hydrodynamics, can be heated, and even get blown out of galaxies. The strength of this effect, to blow gas out to distances of around $\sim 15h^{-1}$Mpc, corresponds approximately to the size of superbubbles in the IGM created by AGN, suggesting that this is a signature of feedback (Dave et al. 2018). The effect of the gravitational dynamics can not be decoupled from this result; the 9 orders of magnitude wide distribution is also affected by mixing in the dark matter. The specific details of the dynamics that causes this spread is still not understood, and must be investigated in further work.

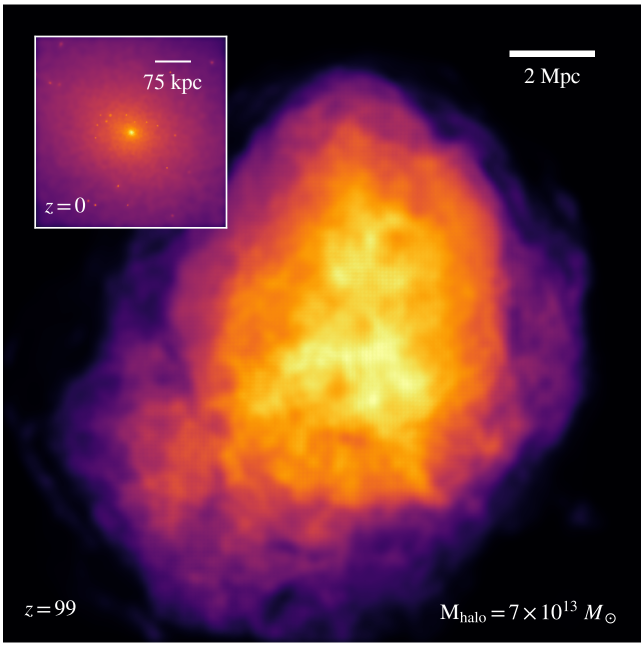

Figure 4. The lagrangian region associated with a 7×10^{13} $M_\odot$ halo at $z = 0$, shown in the inset plot. The local particle density in this region is approximately constant; the structure shown here is due to the particle selection, rather than any real structure in the overall distribution. Colour encodes projected density.

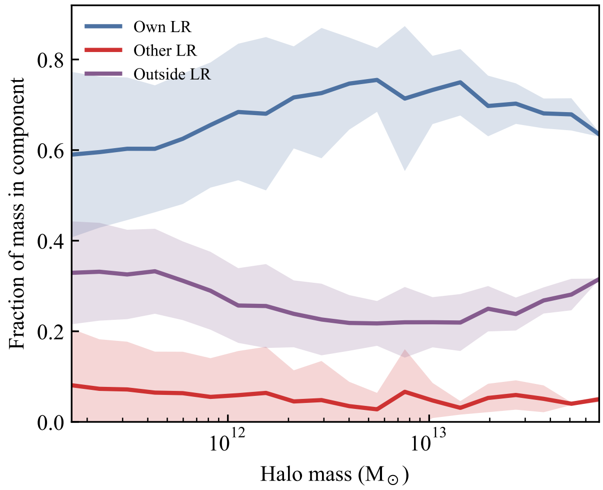

Figure 5. The fraction of total baryonic mass in a given halo at $z = 0$ as a function of halo mass, originating from the lagrangian region defined by the halo (blue), from outside any lagrangian region (purple), and from the lagrangian region defined by another halo (red). These values are computed using the lagrangian region IDs that were assigned to the gas and star particles in the initial conditions. See Figure 6 for a breakdown into stellar and gaseous components. The shaded regions show a single standard deviation of variance in the mass fraction for that bin, and do not include errors from halo sampling bias or cosmic variance.

4 LAGRANGIAN REGIONS

4.1 Definition of a lagrangian region

A lagrangian region is defined as the region in the initial conditions where the dark matter from a given collapsed object at lower redshift resides. The following discussion describes comparison with redshift $z = 0$ compact objects but this definition is easily extensible to higher redshift.

To extract the collapsed objects at $z = 0$ the AMIGA halo finder (AHF, Gill et al. 2004; Knollmann & Knebe 2009) is used. This spherical overdensity finder determines the halo centers by using a nested grid, and then fits parameters based on the Navarro-Frenk-White (NFW, Navarro et al. 1995) profile.

4.2 Matching lagrangian IDs

Each dark matter particle in the simulation is assigned, based on the known unique particle ID, a lagrangian region identifier that corresponds to the halo ID of the associated $z = 0$ collapsed object. Particles which end up outside of any halo at $z = 0$ are assigned a lagrangian region identifier of -1. To extend the definition of the lagrangian region from dark matter to the baryonic particles, a nearest neighbour search for each gas particle at $z = z_{\rm ini}$ is performed and each particle assigned a lagrangian region identifier that corresponds to that of the nearest dark matter particle. This ensures that the very fine-grained detail present in the lagrangian region is preserved (see Figure 4).

Once all particles in the initial conditions have been assigned a lagrangian region, they must be ID matched with particles in the final, $z = 0$, snapshot of the simulation. This is performed by looping through all of the (sorted) particles and assigning a lagrangian region ID for the final-state particles that is equal to that of their initial state progenitor. To assign a lagrangian region to the star particles at $z = 0$, their gas progenitor (which can be tracked using the unique particle ID) is used. Particles which either become black holes, or are consumed by a black hole, are ignored for this analysis. This re-matching also takes place for the dark matter particles, to ensure that the ID matching is working correctly.

4.3 Quantifying inter-lagrangian transfer

Once this analysis has been performed it is possible to calculate the fraction of baryonic mass at $z = 0$ that originates from the lagrangian region of a given halo (see Figure 5). There is a significant difference in the contributions from the gaseous and stellar components to this mass fraction; see Figure 6. This data is for every halo in the box, and hence does not include any cuts based on the particular neighbourhood of these halos; a significant fraction of the scatter here is likely to come from isolated halos, versus those in clusters and other noisier environments. This analysis shows that a significant portion (up to 20%) of the stellar mass of a Milky-Way mass halo may come from the lagrangian region as defined by a *different halo*.

It is important to note that this spread in mass fractions as a function of halo mass is still to be quantified. It could be that those halos which end up having less mass transfer are those which are more isolated; in future analysis we hope to include the isolation criteria used in, for example, the APOSTLE project (Fattahi et al. 2016) to select halos and compare their mass fractions. This inter-lagrangian transfer could have a significant effect on these high-resolution zoom-in galaxies that is not correctly captured using the current isolation criteria.

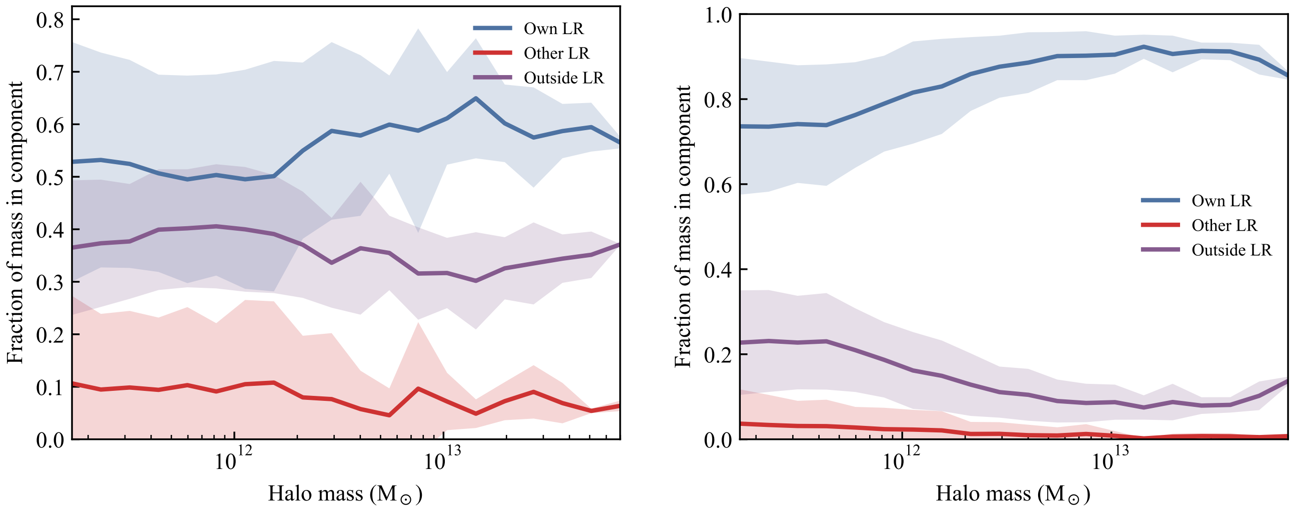

Figure 6. Left: fraction of gaseous mass at $z = 0$ in each halo from each component; right: fraction of stellar mass at $z = 0$ from each component. Note that there is significantly more transfer shown in the gaseous component. Gas that is transferred between lagrangian regions must be given time to cool before being able to form stars. As the events that enable transfer are typically very energetic (AGN, stellar feedback, accretion), it is unlikely that the cooling time will be short enough to form stars by the end of the simulation for most transfer.

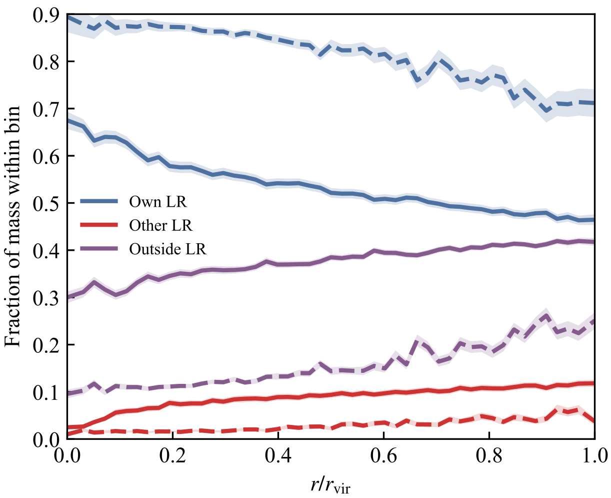

Figure 7. The dependence of mass fraction split by component as a function of radius, normalised by the virial radius for each halo. Every halo with $10^{12} \leq \mathrm{M_{halo}} < 10^{13}$ M$_\odot$ is stacked in this plot, with the shaded regions giving standard errors around the mean. Solid lines show the trends for gas, with the dashed lines showing the same but for the stellar component of the halo. 50 bins were spaced linearly in radius.

4.4 Radial trends

The mass fractions contributed from various components to each halo appear to be relatively independent of halo mass (see Figure 5). In Figure 7 the mass fraction contributed to the halo by each component is shown as a function of radius. As expected, more gas in the center of the halo comes from the corresponding lagrangian region, but interestingly this only approaches 70%, suggesting significant transfer *into halos* still takes place for gas that ends up at the bottom of the potential well at $z = 0$.

Also note how the mass fraction of stars originating from the lagrangian region defined by each halo drops as a function of radius. Note that around 10% of the stars at the center of these halos (within $0.1r_{\rm vir}$) are formed from gas that was not present in the initial lagrangian region of this halo. This is surprising, as this transfer must have taken place relatively early to ensure that the gas was able to cool and form stars.

These radial trends tell us something about the assembly of the Circumgalactic Medium (CGM) (Tumlinson et al. 2017). These trends are very tight and show strong convergence when multiple halos are stacked. Crain et al. (2013) found that the CGM assembles from the 'inside-out', with feedback within the galaxy establishing a strong negative metallicity gradient. However, these results seem to suggest there is a significant effect from inter-lagrangian (and hence inter-galactic) transfer from gas that has been blown out of *other* galaxies falling in to add to this metallicity gradient. Future work will focus on separating these two variables using this analysis.

5 ENSURING THE METHOD IS ROBUST

The above results are quite striking; only around 50% of the gaseous mass in a given halo (at Milky-Way halo mass). To ensure that the methodology discussed above is robust, several extra checks have been put in place, that are discussed below.

5.1 Expanding the halo boundary

A possible criticism of the above definition of lagrangian regions is that they are too sensitive to halo boundary effects. In the above analysis, dark matter particles are identified as belonging to the lagrangian region of a given halo if they lie within the virial radius of that halo at $z = 0$. In this section, changing the radius at which particles are selected to be part of the eventual lagrangian region is explored.

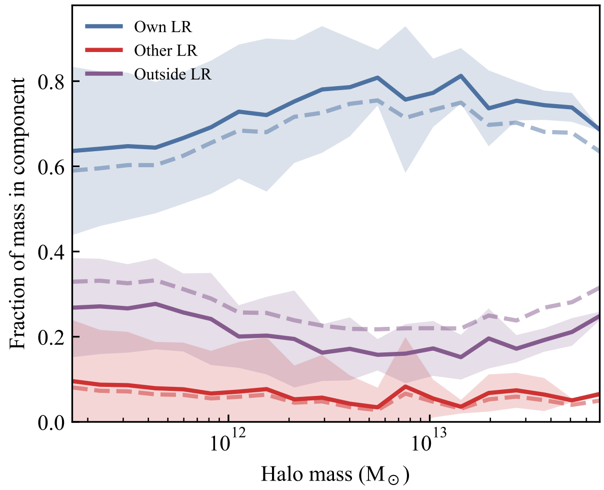

Figure 8. The analogue of Figure 5 but after re-running the analysis with lagrangian regions re-defined such that $r_{\rm vir,new} = 1.2 r_{\rm vir,old}$, see the text for details. The dashed lines represent the exact curves as shown in Figure 5 for direct comparison by the reader. Similar trends follow if the lagrangian region is extended to include particles that lie within $1.5 r_{\rm vir}$.

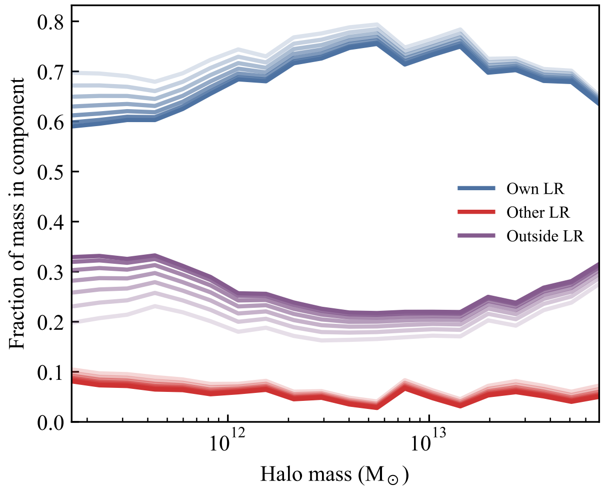

Figure 9. The analogue of Figure 5, but with lagrangian regions filled out with a local nearest neighbour search of the nearest n =1 (i.e. only the particle itself), 2, 4, 8, 16, 32, and 64 particles, with the lines getting lighter as more particles are included in the smoothing. Note the general trend of the mass fraction from the lagrangian region of a halo increasing as they are filled out by the smoothing, with the fraction from other lagrangian regions remaining relatively unchanged.

The procedure for extending the lagrangian region is as follows:

(i) For every halo in the box, search for the centre of that halo (by looking for the extreme particles in each direction and finding their centre point, as well as taking into account the periodic boundaries), and the corresponding radius. This is used instead of the halo centers from AHF to ensure the code remains generic, and provides the same data.
(ii) Multiply this radius (which by definition, from the halo finder, is the virial radius $r_{\rm vir}$) by a factor, such as 1.2, or 1.5
(iii) For each halo in the box, use a periodic KDTree to search for the neighbours of the centre point within that radius, going from highest mass (in dark matter) to lowest mass to ensure that lower-mass halos 'steal' from the higher mass ones, should they be embedded or nearby. From this we would expect to see slightly increased transfer from other regions.
(iv) Label these particles as belonging to the lagrangian region of that halo, but not to the halo themselves if they lie outside of $r_{\rm vir}$.
(v) Re-run the original analysis with this definition of the particles that that lagrangian region.

In Figure 8, the baryonic mass fraction from each lagrangian component is shown as a function of halo mass, where the lagrangian region has been extended to include particles within $1.2 r_{\rm vir}$ of the halo centre. Note that no extra particles are included in the halo, i.e. the halo definition still ends at $r_{\rm vir}$. This ensures that the same edge effects that we are trying to remove are not simply present at the increased radius. This also means that the mass of the particles in the lagrangian region is no longer the same as the mass of the particles in the final, $z = 0$, halo; there will necessarily be a higher fraction of mass originating from the lagrangian region that is defined by the larger volume.

Figure 8 shows that around 5% of the mass of the halo has been re-characterised as originating in the same lagrangian region as the one that the halo defines instead of originating outside any lagrangian region. There has been little change in the mass originating from the lagrangian regions defined by other halos, which shows that there has been very little 'stealing' of material when the re-definition of the lagrangian region took place.

The same overall trends with halo mass are observed even with this extended definition for the lagrangian region, with the expected change observed; there will be some gas in the CGM that is strongly mixed on the halo boundary.

5.2 The definition of lagrangian region

In the above analysis, we considered a very diffuse notion of a lagrangian region, defined particle-by-particle. The definition of what constitutes a lagrangian region is somewhat open to interpretation; there may be particles enclosed in the convex hull of such a region that do not end up in the collapsed object, especially if an incorrect choice is made for the gravitational softening. Whilst increasing the virial radius will go some way to filling the 'holes' in these regions, due to the large transfer of dark matter that still occurs (§3), a different methodology is required.

Some of the holes that are present in lagrangian regions are vitally important; these holes will collapse down to independent halos. An effort must be made to ensure that those holes remain, whilst others are erased, with lower-mass halos taking priority over their higher-mass cousins. With this hole-filling exercise, it is important to note that even dark matter particles may now have a different halo ID to lagrangian ID. The methodology that is proposed here is as follows:

(i) Initially define lagrangian regions in the same way as before for the dark matter.
(ii) Find the first n neighbours of every dark matter particle which has a lagrangian region ID of -1, i.e. it is outside of any region.

(iii) Overwrite the lagrangian ID of these particles with the lowest (i.e. corresponding to the lowest mass halo) in the group.

(iv) Extend the lagrangian region definition to the gas particles in the same way as previously, by finding the closest gas neighbour particle to every dark matter particle.

This aims to both fill out the lagrangian regions, increasing their volume-filling fraction, ensure that no particles are 'stolen' from higher-mass halos, and reduce surface effects leading to spurious lagrangian transfer from outside any region. See Figure 9 for the results for $n = 1$ to $n = 64$ neighbours.

The effect of this smoothing, by design, is to move more particles from outside any lagrangian region to within one. Hence, we see significant movement of mass being labelled as originating outside any lagrangian region to originating inside the lagrangian region of the host halo. This is expected as the surface of the lagrangian regions grow in the initial conditions, as well as particles that lie within the convex hull of the initial lagrangian region being considered as part of that region.

The more surprising result from this is that the fraction of baryonic mass in a halo originating from another lagrangian region is almost constant throughout this process. This implies, as we fill the regions from low-mass to high-mass, that this transfer is extremely robust; we would expect to see huge gains if it was that halos with only a few particles were being extended to include many more particles than they should.

6 CONCLUSIONS

In this work, a new method of analysis for cosmological simulations has been presented, that looks primarily at where gas *originated* in the initial conditions, instead of looking at the final state of matter for comparisons to observations. This gives us a handle on key insights into galaxy formation that previously were only speculated about.

This work is highly preliminary and mainly focuses on the methods behind identifying inter-lagrangian transfer. We refer the interested reader to §7 for details of planned future work.

The key results and concepts of this paper are summarised below:

- By considering the inter-particle distances of nearest neighbours in the initial conditions, it is possible to see that stellar populations form out of gas that, generally, is tightly coupled to dark matter. This is explained by gas requiring enough time to cool to become dense enough to form stars, and this is only possible in the deep potential wells at the centre of halos. Gas, on the other hand, can be spread much more, relative to the dynamical motions of the background dark matter.
- The baryonic matter that resides in any given halo does not necessarily follow the same path as the dark matter to get there. It is possible to quantify this 'inter-lagrangian transfer' by looking at the particles that resided in the lagrangian region as defined by the dark matter in the initial conditions and comparing this to the final conditions. On average, independent of halo mass, only around 60% of the baryons in a given dark matter halo at $z = 0$ originated in that lagrangian region. Around 10% of that mass originated in regions defined by the dark matter in other halos, with the remaining 30% originating outside any lagrangian region. The former contribution can, possibly, be explained by stellar winds and AGN feedback blowing gas out of halos, allowing it to accrete onto neighbours. The latter contribution from outside any lagrangian regions can be explained by two things: the non-complete filling fraction of the lagrangian region definition chosen in this work, and the strong flow of matter that is provided by the nonlinear dependence of cooling on density. This allows gas to cool and fall into the halos quicker than dark matter, which due to a lack of cooling cannot lose angular momentum.
- Despite requiring cooling to take place, some small halos can still receive up to a 20% fraction of their stellar mass from either in-situ or ex-situ star formation provided by material from other regions.
- As expected, due to the above observations and explanations, the fraction of baryonic mass from the lagrangian region of a halo decreases as a function of radius. As a function of radius, the contribution from outside and from other lagrangian regions grows, showing that the externally contributed material plays a significant role in the formation and dynamics of the CGM.
- Several ways to ensure that the analysis is robust, including increasing the radius over which material is included in the lagrangian region, and smoothing the regions themselves, were presented. These had little effect on the overall qualitative results from this work.

This area of simulation analysis promises to be a fertile ground for the future. The authors hope that other simulators interested in such an analysis will download `LTCaesar`, and either independently, or with us, run the same analysis. `LTCaesar` is open-source and in open development, with details on how to find and run the code available in Appendix 9.1. Feedback models underpin a large amount of the dynamics involved in this transfer, and hence it is clearly important to compare various feedback models and modes in future analysis.

7 FUTURE WORK

As noted in §6, this work is preliminary and mainly focuses on the methods behind this new approach. Below is a list of possible opportunities, improvements, and science that we wish to accomplish using the newly developed `LTCaesar` code.

7.1 Improvements to the analysis

7.1.1 *Improvements to particle ID handling*

Due to a bug in the simulation code when this particular box was simulated, gas particles which have formed a star, and star particles that formed out of a gas particle which had already created a star particle, must be excluded. This is unfortunate, as these particles are the ones that would probably have extended out to the largest distances due to their interactions with stellar and AGN feedback. In future analysis this problem will be fixed; however at this time we do not expect it to have a significant effect as this is a very small contribution (less than 0.1% of gas particles).

7.1.2 *Filling-out work*

Currently we fill out the lagrangian regions by using a nearest-neighbour search in the initial conditions. This works fine, but it would be helpful to include other definitions, such as the convex hull, for comparison to other works and to zoom-in simulations.

7.1.3 *Distance distribution improvements*

The interesting thing about these distance metrics that is still not well understood by the authors is the significant distances that dark matter particles can end up with between them despite starting the simulation next to each other. These particles may be on either side of a void, for instance, but it is unclear where the origin of this 7.5 Mpc maximal distance is (this appears to be independent of resolution and box size based on preliminary testing).

It should be possible to consider various distance distribution metrics; at the moment only the nearest neighbour for each dark matter particle is considered. However, there is much more exploratory work to be done here. Consider, for instance, the median distance among the nearest n neighbours, the variance in distance among neighbours, and this metric as a function of position in the initial conditions.

7.2 Science work

7.2.1 *Comparison of models*

In a future paper, once this methods paper is finalised, we hope to re-run this analysis on *at least* the MUFASA and EAGLE simulations. We also hope to re-run the analysis on the Illustris and Illustris-TNG models, however this may require a little more code development work due to the tracer particle model available there for tracking the flows of gas. Ideally, this should be completely transparent, but this is unclear at the moment. This work really highlights the inherrent advantages of particle-based methods, and we hope that it becomes a rationale behind people using these (over the often easier to conceptualise eulerian grid methods) in the future.

7.2.2 *The structure of the CGM*

This work shows that the baryonic matter in a given halo, as a function of radius, has a well-defined origin profile. This should lead to a well-defined metallicity profile. These can be investigated using the SIMBA simulations and matched to observations.

This external contribution may also lead to a different density profile than that is expected from analytical and semi-analytical work. This differing density profile can then be broken into components based on their origin, thanks to this work. This could affect many cosmological probes, including weak-lensing measurements that are used to constrain modified gravity models.

7.2.3 *Zoom-in simulations*

This work should have some implications for zoom-in simulations. Current zoom-ins neglect sub-grid physics for particles outside of the high-resolution zone, but based on this analysis it appears that this may no longer be sufficient to capture the full physics; particles can travel up to 15 Mpc from their initial neighbour, and significant transfer is seen between galaxies. This is especially important in work that aims to capture the physics of the CGM in those galaxies.

8 ACKNOWLEDGEMENTS

This work was initiated as a project for the Kavli Summer Program in Astrophysics held at the Center for Computational Astrophysics of the Flatiron Institute in 2018. The program was co-funded by the Kavli Foundation and the Simons Foundation. We thank them for their generous support.

JB is supported by STFC studentship ST/R504725/1. This work used the ARCHER UK National Supercomputing Service. This work also used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1, and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.

REFERENCES

Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2016,] 10.1093/mnras/stx1517, 4719, 4698
Anglés-Alcázar D., Davé R., Faucher-Giguère C.-A., Özel F., Hopkins P. F., 2017, Monthly Notices of the Royal Astronomical Society, 464, 2840
Benitez-Llambay A., 2015, py-sphviewer: Py-SPHViewer v1.0.0, doi:10.5281/zenodo.21703
Bondi H., 1952, Monthly Notices of the Royal Astronomical Society, 112, 195
Crain R. A., McCarthy I. G., Schaye J., Theuns T., Frenk C. S., 2013, Monthly Notices of the Royal Astronomical Society, 432, 3005
Davé R., Thompson R., Hopkins P. F., 2016, Monthly Notices of the Royal Astronomical Society, 462, 3265
Dave R., Angles-Alcazar D., Rafieferantsoa M. H., Appleby S., Narayanan D., Li Q., 2018, 10, 1
Fattahi A., et al., 2016, Monthly Notices of the Royal Astronomical Society, 457, 844
Frenk C. S., White S. D. M., Davis M., Efstathiou G., 1988, The Astrophysical Journal, 327, 507
Gill S. P., Kneb A., Gibson B. K., 2004, Monthly Notices of the Royal Astronomical Society, 351, 399
Henriques B. M., White S. D., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, Monthly Notices of the Royal Astronomical Society, 451, 2663
Hernquist L., Katz N., 1989, The Astrophysical Journal Supplement Series
Hopkins P. F., 2015, Monthly Notices of the Royal Astronomical Society, 450, 53
Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C. A., Quataert E., Murray N., Bullock J. S., 2014, Monthly Notices of the Royal Astronomical Society, 445, 581
Jones E., Oliphant T., Peterson P., Others 2001, SciPy.org, `https://www.scipy.org/`
Knollmann S. R., Knebe A., 2009, Astrophysical Journal, Supplement Series, 182, 608
Lacey C. G., et al., 2016, Monthly Notices of the Royal Astronomical Society, 462, 3854
Mo H., van den Bosch F., White S. D. M., 2010, Galaxy Formation and Evolution.. (`arXiv:1104.0012`), doi:10.1017/CBO9780511807244, `papers://e1f660dc-e86d-4910-82ad-4b2b33c2f9ed/Paper/p10838`
Muratov A. L., Kereš D., Faucher-Giguère C. A., Hopkins P. F., Quataert E., Murray N., 2015, Monthly Notices of the Royal Astronomical Society, 454, 2691
Navarro J. F., Frenk C. S., White S. D. M., 1995, The Astrophysical Journal, 462, 563
NumPy 2018, NumPy, `http://www.numpy.org/`
Pavlyk O., Nagorny D., Guzman-Ballen A., Malakhov A., Liu H., Totoni E., Anderson T. A., Maidanov S., 2017, 15th Python in Science Conf. (SciPy 2017), pp 1–8
Planck Collaboration 2016, Astronomy & Astrophysics, 596, A108
Porter L. A., Somerville R. S., Primack J. R., Johansson P. H., 2014, Monthly Notices of the Royal Astronomical Society, 444, 942
Revaz Y., Jablonka P., 2011,] 10.1051/0004-6361/201117402
Schaye J., et al., 2015, Monthly Notices of the Royal Astronomical Society, 446, 521

Smith B. D., et al., 2016,] 10.1093/mnras/stw3291, 2234, 2217
Springel V., 2005, The cosmological simulation code GADGET-2 (arXiv:0505010), doi:10.1111/j.1365-2966.2005.09655.x
Springel V., et al., 2005, Nature, 435, 629
Teyssier R., 2002, Astronomy & Astrophysics, 385, 337
Thompson R., 2018, Caesar, https://caesar.readthedocs.io/en/latest/index.html
Tumlinson J., Peeples M. S., Werk J. K., 2017,] 10.1146/annurev-astro-091916-055240, pp 389–433
Vogelsberger M., et al., 2014, Nature, 509, 177
van den Bosch F. C., Ogiya G., Hahn O., Burkert A., 2018, Monthly Notices of the Royal Astronomical Society, 474, 3043

9 APPENDIX

9.1 Code

The code used in this analysis, `LTCaesar` is made available to the community for use with their own simulations[1]. This code can also be installed with the '`pip`' python packaging manager. In the repository, there is also a series of scripts that can be used to convert between halo catalogues, and a robust suite of unit tests that ensure that `LTCaesar` performs the analysis that we claim it does. This code is highly extensible and can use any halo finder, simulation code (so long as it includes at least tracer particles), and can run on a 512^3 particle box in under an hour (including the production of all of the figures in this report) on a single supercomputing node. The authors acknowledge that this section, unfortunately, is under-developed and significantly under-referenced, based on the possible claims presented here.

`LTCaesar` uses the numerical routines from `numpy`, the KDTree from `scipy` for nearest neighbour searching, and the halo finder wrapper from `caesar` (NumPy 2018; Jones et al. 2001; Thompson 2018). Full documentation, and more information, can be found on the code webpages. The work in this paper in particular made use of the Intel Distribution for Python that provides optimized routines for `numpy` and `scipy` (Pavlyk et al. 2017).

The visualisations in this work made use of `py-sphviewer` (Benitez-Llambay 2015).

[1] https://www.github.com/jborrow/lagrangian-transfer