Galaxy assembly bias of central galaxies in the Illustris simulation

Xiaoju Xu and Zheng Zheng

Department of Physics and Astronomy, University of Utah, 115 South 1400 East, Salt Lake City, UT 84112, USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Galaxy assembly bias, the correlation between galaxy properties and halo properties at fixed halo mass, could be an important ingredient in halo-based modelling of galaxy clustering. We investigate the central galaxy assembly bias by studying the relation between various galaxy and halo properties in the Illustris hydrodynamic galaxy formation simulation. Galaxy stellar mass $M^*$ is found to have a tighter correlation with peak maximum halo circular velocity $V_{\text{peak}}$ than with halo mass $M_h$. Once the correlation with $V_{\text{peak}}$ is accounted for, $M^*$ has nearly no dependence on any other halo assembly variables. The correlations between galaxy properties related to star formation history and halo assembly properties also show a cleaner form as a function of $V_{\text{peak}}$ than as a function of $M_h$, with the main correlation being with halo formation time and to a less extent halo concentration. Based on the galaxy-halo relation, we present a simple model to relate the bias factors of a central galaxy sample and the corresponding halo sample, both selected based on assembly-related properties. It is found that they are connected by the correlation coefficient of the galaxy and halo properties used to define the two samples, which provides a reasonable description for the samples in the simulation and suggests a simple prescription to incorporate galaxy assembly bias into the halo model. By applying the model to the local galaxy clustering measurements in Lin et al. (2016), we infer that the correlation between star formation history or specific star formation rate and halo formation time is consistent with being weak.

Key words: galaxies: haloes – galaxies: statistics – cosmology: theory – large-scale structure of Universe

1 INTRODUCTION

It has been well established that galaxies form in dark matter haloes (White & Rees 1978). As the first step to study galaxy formation and clustering, halo formation and clustering, which is dominated by gravity, have been extensively studied with analytic models (e.g., Press & Schechter 1974; Bardden et al. 1986; Mo & White 1996; Sheth & Tormen 1999) and cosmological N-body simulations (e.g., Springel 2005; Prada et al. 2012). It has been found that halo clustering depends not only on halo mass but also on halo assembly history or environment (e.g., Gao et al. 2005; Gao & White 2007; Paranjape et al. 2018; Xu & Zheng 2018; Han et al. 2019). This is called halo assembly bias, whose nature is still under investigation (e.g., Dalal et al. 2008; Castorina & Sheth 2013).

If galaxy properties are affected by halo formation and assembly history, halo assembly bias would translate to galaxy assembly bias. Operationally, galaxy assembly bias can be defined as that at fixed halo mass, the statistical galaxy content shows dependence on other halo variables or galaxy properties show correlations with halo assembly history. The widely adopted halo model (e.g., Cooray & Sheth 2002) of interpreting galaxy clustering, such as the halo occupation distribution (e.g., Berlind & Weinberg 2002; Zheng et al. 2005) or conditional luminosity function (e.g., Yang et al. 2003), makes the implicit assumption of no galaxy assembly bias. Such methods have been successfully applied to galaxy clustering (e.g., Zehavi et al. 2005; Zheng et al. 2007; Xu et al. 2018). However, if assembly bias is significant, neglecting it in the model would lead to incorrect inference of galaxy-halo connections and introduce possible systematics in cosmological constraints (e.g., Zentner et al. 2014, 2016; but see also McEwen & Weinberg 2016; McCarthy, Zheng & Guo 2018). Conversely, observa-
tionally inferred galaxy assembly bias would help understand galaxy formation.

The existence and strength of galaxy assembly bias are still a matter far from settled, either in theory or in observation. Galaxy assembly bias has been investigated in hydrodynamic or semi-analytic galaxy formation models (e.g., Berlind et al. 2003; Croton et al. 2007; Mehta 2014; Chaves-Montero et al. 2016; Zehavi et al. 2018; Contreras et al. 2018), focusing on the effect on galaxy occupation function and galaxy clustering. The results seem to depend on the implementation details of star formation and feedback. Studying galaxy assembly bias from observation has the difficulty of determining halo mass, and the results are not conclusive (e.g., Yang et al. 2006; Berlind et al. 2006; Lin et al. 2016; Zu et al. 2017; Guo et al. 2017). Given the potential importance of galaxy assembly bias in modelling galaxy clustering, in this paper we study the correlation between various central galaxy properties and halo properties in the Illustris hydrodynamic simulation (Vogelsberger et al. 2014a) at the halo level, aiming at providing useful insights in describing galaxy assembly bias.

The structure of the paper is as follows. In section 2, we introduce the simulation and the galaxy and halo catalogues. Then in section 3, we investigate the relation between galaxy and halo properties, with primary galaxy-halo properties in section 3.1 and general galaxy-halo properties in section 3.2. In section 3.3, we present a simple model to connect galaxy assembly bias with halo assembly bias. Finally, we summarise and discuss the results in section 4.

2 SIMULATION AND GALAXY-HALO CATALOGUE

In this work, we use galaxies and haloes from the state-of-the-art hydrodynamic galaxy formation simulation Illustris\(^\dagger\) (Vogelsberger et al. 2014a; Nelson et al. 2015) to study galaxy assembly bias, which is able to produce different type of galaxies seen in observation (Vogelsberger et al. 2014b; Genel et al. 2014). In particular, we use the Illustris-2 simulation, which has a box size of \(75h^{-1}\)Mpc on a side, and contains 910\(^3\) dark matter particles of mass \(5 \times 10^7 M_{\odot}\) and the same number of baryon particles of mass \(1 \times 10^7 M_{\odot}\). The mass resolution is sufficient for our purpose of studying (central) galaxies in haloes of more massive than a few times \(10^{10} h^{-1} M_{\odot}\). The simulation adopts a spatially-flat cosmology with the following parameters: \(\Omega_m = 0.27, \Omega_b = 0.0456, h = 0.70, \sigma_8 = 0.963, \) and \(\sigma_8 = 0.809\).

The haloes in the Illustris database are identified with the friends-of-friends (FoF) algorithm. As this algorithm is not conclusive (e.g., Yang et al. 2006; Berlind et al. 2006; Lin et al. 2016; Zu et al. 2017; Guo et al. 2017). Given the potential importance of galaxy assembly bias in modelling galaxy clustering, in this paper we study the correlation between various central galaxy properties and halo properties in the Illustris hydrodynamic simulation (Vogelsberger et al. 2014a) at the halo level, aiming at providing useful insights in describing galaxy assembly bias.

The central galaxy properties we consider include:

1. \(M_g\), halo mass enclosed in a volume with mean density of 200 times the background density of the universe;
2. \(V_{\text{peak}}\), peak maximum circular velocity of the halo over its accretion history;
3. \(c\), halo concentration parameter, defined as the ratio of halo virial radius to scale radius;
4. \(a_{\text{M}/2}\), cosmic scale factor when the halo obtains half of its current \((z = 0)\) total mass;
5. \(M_h\), halo mass accretion rate near \(z = 0\) (averaged between \(z=0\) and \(z=0.197\), about 2.4 Gyr, one dynamical time), in units of \(h^{-1} M_{\odot} \) yr\(^{-1}\);
6. \(M_h/ M_{\odot}\), specific halo accretion rate, in units of Gyr\(^{-1}\).

The central galaxy properties we consider include:

1. \(M_*\), stellar mass (sum of masses of star particles within twice the stellar half mass radius);
2. SFR, star formation rate within twice the stellar half mass radius;
3. sSFR, specific star formation rate, the ratio of SFR to \(M_*\);
4. \(g-r\), galaxy colour defined by the \(g\)-band and \(r\)-band luminosity.

3 RESULTS

We aim at presenting the relation between central galaxies and haloes to learn about the correlation between halo formation and assembly and galaxy properties. In section 3.1 We study how galaxy stellar mass depends on the primary halo properties \((M_h\) and \(V_{\text{peak}}\)). Then we investigate how various halo and galaxy properties are correlated in section 3.2. Finally, in section 3.3 we use a simplified model to describe the connection between galaxy and halo assembly bias factor.

3.1 Relationship between stellar mass and halo properties

Galaxy stellar mass is a primary property inferred from observation. The relation between this primary galaxy property and certain primary halo property (e.g., \(M_h\) and \(V_{\text{peak}}\)) can be established based on subhalo abundance matching or modelling the stellar mass dependent clustering, which encodes information about galaxy formation. Here we show

---

\(^\dagger\) http://www.illustris-project.org
the relation predicted by the Illustris simulation and study how tight stellar mass correlates with $M_h$ and $V_{\text{peak}}$.

The top-left panel of Fig. 1 shows $M_*$ as function of $M_h$, colour-coded with values of $V_{\text{peak}}$. Galaxy stellar mass $M_*$ increases steeply with $M_h$ at $\log(M_*/(h^{-1}M_\odot)) < 12$ and then slowly at $\log(M_*/(h^{-1}M_\odot)) > 12$, a trend similar to that inferred from observation (e.g. Behroozi et al. 2010; Leauthaud et al. 2012; Zu & Mandelbaum 2015). The scatter in $M_*$ at fixed halo mass in the $M_* - M_h$ relation decreases with increasing $M_h$ (solid curve in Fig. 2), varying from about 0.3 dex at the low-mass end to about 0.17 dex at the high-mass end. The scatter at the high-mass end is consistent with the value $\sim 0.16$ dex inferred from galaxy clustering modelling (e.g. Tinker et al. 2017). One source of the scatter can be the halo formation history (Tinker 2017), which may affect the growth history of stellar mass (either from star formation or galaxy merging; e.g. Gu et al. 2016).

The colour code in $V_{\text{peak}}$ in the top-left panel enables us to see how the scatter in the $M_* - M_h$ relation may be connected to halo assembly. On average, $V_{\text{peak}}$ and $M_h$ are correlated, and the mean relation is found to be well described by

$$V_{\text{peak}} = 170 \left(\frac{M_h}{10^{12}h^{-1}M_\odot}\right)^{1/3} \text{ km s}^{-1}$$

in the Illustris simulation. However, there is scatter on top of the mean relation, and at fixed $M_h$, the distribution of $V_{\text{peak}}$ reflects that in the assembly history. As can be seen in the top-left panel, the assembly of haloes encoded in $V_{\text{peak}}$ does contribute to the scatter in the $M_* - M_h$ relation – at fixed $M_h$, galaxies residing in haloes of higher $V_{\text{peak}}$ tend to have higher stellar mass, especially at the low mass end.

To see how well the scatter in $M_*$ can be attributed to the scatter in $V_{\text{peak}}$, in the top-right panel of Fig. 1, we plot the $M_* - V_{\text{peak}}$ relation. It follows a similar trend seen in the $M_* - M_h$ relation, steeper (shallower) dependence of $M_*$ on $V_{\text{peak}}$ at the low (high) $V_{\text{peak}}$ end, which is expected given the correlation between $V_{\text{peak}}$ and $M_h$. The $M_* - V_{\text{peak}}$ relation appears to be tighter than the $M_* - M_h$ relation, in the sense that at fixed $V_{\text{peak}}$ the scatter in $M_*$ is lower than that at $V_{\text{peak}}$. 

---

**Figure 1.** Top-left: $M_*$ as function of $M_h$ for central galaxies. The galaxies are colour-coded according to $\log(V_{\text{peak}}/(\text{km s}^{-1}))$. For galaxies in each bin of $\log(V_{\text{peak}})$, the contours correspond to the 68.3 and 95.4 per cent distribution, respectively. Top-right: $M_*$ as function of $V_{\text{peak}}$ for central galaxies, colour coded according to $\log(M_h/(h^{-1}M_\odot))$. Bottom-left: $M_*$ as function of $M_h$ for central galaxies, with the mean relation colour-coded according to the values of $\alpha_{M/2}$. For clarity, the scatter in the mean relation is only shown for the bin with the highest $\alpha_{M/2}$ (latest forming haloes). Bottom-right: $M_*$ as function of $V_{\text{peak}}$ for central galaxies, colour-coded according to $\alpha_{M/2}$, with the shaded region illustrating the scatter for the bin with the highest $\alpha_{M/2}$. Note the remarkable result that $M_*$ does not depend on $\alpha_{M/2}$ at fixed $V_{\text{peak}}$ (compared to the $M_h$ case in the bottom-left panel).
the corresponding $M_h$ (see Matthee et al. 2017 for a similar result in terms of $z = 0$ maximum halo circular velocity with the EAGLE simulation). The scatter varies from $\sim 0.28$ dex at low $V_{\text{peak}}$ to $\sim 0.13$ dex at high $V_{\text{peak}}$ (dashed curve in Fig. 2). In the $M_\ast - V_{\text{peak}}$ plot (top-right panel), the contours are colour-coded by $M_h$. Unlike the $M_h$ case in the top-left panel, we find that $M_\ast$ does not show a clear dependence on $M_h$ – at fixed $V_{\text{peak}}$ (i.e. by taking a vertical cut in the plot), $M_\ast$ in haloes of different $M_h$ appears to follow a similar trend on the left 7 columns of contour panels) show the correlation between each pair of the halo properties, indicating how strong the correlation is. It is calculated as

$$\rho = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\sigma_x \sigma_y}, \quad \text{(2)}$$

where $x$ and $y$ denote the two properties, $\langle \rangle$ means average, and $\sigma_x$ and $\sigma_y$ are the standard deviations of $x$ and $y$. The panel at the top of each column shows the marginalised distribution of the property labelled at the $x$-axis of the column.

The panels with red contours (i.e. the top 6 rows and left 6 columns of contour panels) display the correlation between halo properties. Within the small but finite halo mass bin, all but one halo property shows almost no correlation with $M_h$ (correlation coefficient close to zero). The exception is $V_{\text{peak}}$, and the correlation is simply driven by the $V_{\text{peak}} \propto M_h^{1/3}$ mean relation. At fixed $M_h$, any pair of halo properties show a substantial correlation (with $|\rho|$ above 0.2). The nearly perfect correlation ($\rho = 0.983$) between $M_h$ and $M_\ast/M_h$ is a consequence of fixed $M_h$. Overall the correlation trend is that haloes of higher $V_{\text{peak}}$ are more concentrated, form earlier, spin more slowly, and have lower accretion rate, which have been seen in previous work (e.g. Jeeson-Daniel et al. 2011; Han et al. 2019; Xu & Zheng 2018).

The panels with black contours (i.e. the bottom 4 rows and left 7 columns of contour panels) show the correlation between halo and galaxy properties. The correlation between $M_\ast$ and $M_h$ shows up because of the finite size of the halo mass bin. At fixed $M_h$, the central galaxy stellar mass $M_\ast$ correlates with all other halo properties – haloes of higher $V_{\text{peak}}$: higher concentration, earlier formation, lower

![Figure 2. Standard deviation in log $M_\ast$ as a function of $M_h$ (solid) and $V_{\text{peak}}$ (dashed). The correspondence between $M_h$ and $V_{\text{peak}}$ is from the mean relation $V_{\text{peak}} \propto M_h^{1/3}$ in equation (1).](image)
Table 1. Summary of the correlation coefficients

| Variable | Description | Correlation Coefficient |
|----------|-------------|-------------------------|
| $M_{\text{peak}}$ | Peak halo mass | 0.464 |
| $c$ | Concentration parameter | -0.029 |
| $a$ | Angular momentum | 0.656 |
| $M_{\text{halo}}$ | Halo mass | -0.090 |
| $\lambda$ | Angular momentum | -0.351 |
| $M_{\text{star}}$ | Stellar mass | -0.033 |
| $M_{\text{sSFR}}$ | Specific star formation rate | 0.378 |
| $g-r$ | Rest-frame color | 0.107 |

Figure 3. Relation between each pair of galaxy and/or halo properties at $\log(M_{\text{halo}}/(h^{-1}M_\odot)) = 12$. In each contour panel, the two contours show the central 68.3 and 95.4 per cent of the distribution of the pair of properties. The panels with red contours (i.e. the top 6 rows and left 6 columns of contour panels) display the correlations between halo properties. Those with black contours (i.e. the bottom 4 rows and the left 7 columns of contour panels) show the correlations between galaxy and halo properties, and those with blue contours (i.e. the right 3 columns of contour panels) are for the correlations between pairs of galaxy properties. The number in each contour panel is the Pearson correlation coefficient for the pair of properties. The histogram at the top panel of each column is the probability distribution function of the variable of that column.
Figure 4. Same as Fig. 3, but at fixed log($V_{\text{peak}}/(\text{km s}^{-1})] = 2.23. Note particularly the lack of correlation of $M_*$ with other assembly variables (including $c$, $\alpha_{M/2}$, $\lambda$, $M_h$, $M_h/M_\text{uni}$), in contrast with the case in Fig. 3.

spin, and lower accretion rate tend to host more massive central galaxies. On the contrary, the SFR shows no strong correlation with any halo properties. The most significant one is with halo formation time ($\rho \sim 0.16$), with on average higher SFR in haloes of later formation. Given the substantial correlation between $M_*$ and halo properties and the weak or lack of correlation between SFR and halo properties, the sSFR ($\equiv \text{SFR}/M_*$) is expected to correlate well with halo properties, but in an trend opposite to and weaker than that with $M_*$. This is indeed the case. The most significant correlation is with $V_{\text{peak}}$ or $\alpha_{M/2}$ (both with $|\rho| \sim 0.4$). The correlation between sSFR and the average halo accretion rate over the past dynamic time is there but not strong ($\rho \sim 0.16$). The correlation between galaxy colour $g-r$ and halo properties...
Figure 5. Pearson correlation coefficient $\rho$ of each pair of galaxy and/or halo properties as a function of $M_h$ (solid) and $V_{\text{peak}}$ (dashed). The galaxy and halo properties are marked to the far left of each row and at the bottom of each column. For clarity, we only label the values of $V_{\text{peak}}$ on the horizontal axis, and the values of $M_h$ can be inferred from $M_h \propto V_{\text{peak}}^3$ from equation (1). As with Fig. 3, panels with red, black, and blue curves are for correlations between halo-halo, galaxy-halo, and galaxy-galaxy properties, respectively. In each panel, the dotted horizontal line indicates no correlation. Note particularly the lack of correlation of $M_*$ with other assembly variables (including $c$, $a_{M/2}$, $\lambda$, $M_h$, $M_h/M_h$) for the $V_{\text{peak}}$ dependence case, in contrast with the $M_h$ dependent case. Also the correlations of SFR, sSFR, and colour with halo assembly variables show more consistent behaviours in the $V_{\text{peak}}$ dependence case.
essentially follows the case of sSFR (with a sign change in $\rho$; redder galaxies having lower sSFR).

The panels with blue contours (i.e. the right 3 columns of contour panels), the correlations between pairs of galaxy properties at fixed halo mass are shown. The SFR positively correlates with $M_*$ ($\rho \sim 0.54$), and the mean relation has a slope close to unity, $SFR \propto M_*$. It resembles the star-forming main sequence (e.g. Brinchmann et al. 2004; Speagle et al. 2014; Santini et al. 2017). That is, even if we only consider central galaxies in haloes of fixed mass, the star-forming main sequence emerges. Note that this SFR–$M_*$ correlation is not driven by the correlation of SFR and $M_*$ with a common halo variable we consider here. In fact, from Fig. 3, it can be seen that their correlation with a common halo variable may lead to the opposite effect. For example, haloes of earlier formation tend to host central galaxies of higher $M_*$ and lower SFR, and naively this would imply an anti-correlation between SFR and $M_*$, opposite to what is found here. While it is possible that the halo-level star-forming main sequence is related to a halo variable not considered here, it is more likely that the sequence is driven by baryonic physics, which may have complicated dependence on or decouple from halo formation history. Unlike the SFR, the sSFR shows little dependence on $M_*$. However, the sSFR is tightly correlated with the SFR ($\rho \sim 0.88$). Given that sSFR= SFR/$M_*$, the pattern in the mutual correlations among $M_*$, SFR, and sSFR can be achieved if the SFR–$M_*$ correlation coefficient is close to the ratio of the scatters in $\log M_*$ and $\log SFR^2$, which appears to be the case. Galaxy $g$–$r$ colour strongly correlates with sSFR and follows the same trends as sSFR in its correlations with $M_*$ and SFR.

Overall, at fixed halo mass, galaxy properties other than SFR show significant correlations with one or more halo properties, manifesting galaxy assembly bias at the level of haloes. The correlations among galaxy properties, however, may largely result from baryonic physics, given that the trend cannot be simply explained by their correlation with halo properties. In section 3.1, it is found that switching from $M_*$ to $V_{\text{peak}}$ can remove the dependence of $M_*$ on other halo properties. We now extend the investigation to other galaxy properties.

3.2.2 At fixed $V_{\text{peak}}$

Fig. 4 is similar to Fig. 3, but the correlations are presented for haloes at a fixed $V_{\text{peak}}$ bin, $\log[V_{\text{peak}}/(\text{km s}^{-1})] = 2.23 \pm 0.03$. The correlations among halo properties (in panels with red contours) are similar to those in Fig. 3, and there are additional correlations between $M_*$ and other halo properties.

The galaxy-halo correlations are shown in panels with black contours. The finite bin size in $V_{\text{peak}}$ makes the $M_*$–$V_{\text{peak}}$ correlation show up. Other than this (and the one with $M_*$), $M_*$ does not correlate with any other halo properties at fixed $V_{\text{peak}}$, reinforcing the result in section 3.1. It indicates that the correlations of $M_*$ with halo properties seen at fixed $M_*$ (Fig. 3) can be attributed to the $M_*$–$V_{\text{peak}}$ correlation and the correlation of $V_{\text{peak}}$ with other halo properties. For the SFR, at fixed $V_{\text{peak}}$, it correlates significantly with $M_*$ and $a_{M/2}$ higher SFR in haloes of higher mass and later formation. The correlations between SFR and other halo properties are weak. The correlations between sSFR (or colour) and halo properties closely follow the SFR case.

In terms of the galaxy-halo correlations, the trends are similar to those seen at fixed halo mass, except that the sSFR and colour now show clear correlations with $M_*$.

As a whole, using $V_{\text{peak}}$ as the halo variable largely removes the correlations between $M_*$ and other halo assembly properties, and the dependences of SFR, sSFR, and colour on halo assembly variables follow each other.

3.2.3 Dependence on $M_h$ and $V_{\text{peak}}$

The correlations shown in Fig. 3 and Fig. 4 are for haloes of $\log[M_h/(h^{-1}M_\odot)] \sim 12.0$ and $\log[V_{\text{peak}}/(\text{km s}^{-1})] \sim 2.23$. To obtain a full picture, in Fig. 5 we present the $M_h$ and $V_{\text{peak}}$ dependent Pearson correlation coefficients for the various pairs of galaxy and halo properties, by performing the calculation in different $M_h$ and $V_{\text{peak}}$ bins, respectively. The panels correspond to those in Fig. 3 and Fig. 4, and the correlation shown in a panel of a given row and column is between the property as labelled at the far left of the row and that at the bottom of the column. In each panel, the solid (dashed) curve is the dependence of $\rho$ on $M_h$ ($V_{\text{peak}}$), with zero correlation marked by the black dotted curve. Note that only $V_{\text{peak}}$ is shown on the $x$-axis and the corresponding $M_h$ can be obtained according to $M_h \propto V_3$ from equation (1).

The panels with red curves show the correlations among halo properties. If we limit to halo properties other than $M_h$ and $V_{\text{peak}}$, we find that the correlation of any pair of the assembly variables only weakly depends on $M_h$ or $V_{\text{peak}}$ if any (manifested by the nearly flat curves) and that the correlation strength does not depend on whether we use $M_h$ or $V_{\text{peak}}$ bins (manifested by the highly overlapped solid and dashed curves).

For the galaxy-halo correlations (in panels with black curves), in terms of $M_h$ dependence, the strongest correlation between galaxy and halo property is found between $M_*$ and $V_{\text{peak}}/c/a_{M/2}$ in low mass halos, with $|\rho| \sim 0.5$–0.6. It holds true in the full range of haloes considered here that using $V_{\text{peak}}$ largely removes the correlation between $M_*$ and any other halo assembly variable (dashed curves around zero). The only exception is that $M_*$ appears to be slightly anticorrelated with $c$ in low-$V_{\text{peak}}$ haloes. With $V_{\text{peak}}$ the dependence of correlation on $V_{\text{peak}}$ for SFR, sSFR, and colour closely track each other, which is not the case for those on $M_h$. With $V_{\text{peak}}$ as the primary halo variable, star formation related properties (SFR, sSFR, and colour) mainly show dependence on halo formation time and then halo concentration.

For galaxy properties (in panels with blue curves), the correlation between SFR and $M_*$ reaches a maximum in haloes of $\sim 10^{12}h^{-1}M_\odot$. It weakens in haloes of higher $M_h$ or $V_{\text{peak}}$, probably because galaxies move away from the star-forming main sequence and passive evolution starts to dominate. However, the tight sSFR–SFR correlation persists over the full $M_h$ or $V_{\text{peak}}$ range. For colour, the correlation with

---

2 To see this, let $x = \log M_*$, $y = \log \text{SFR}$, and $z = \log \text{sSFR}$, then $\rho_{xz}/\rho_{yz} = (\rho_{xy} - \rho_{xz}\rho_{yz})/(1 - \rho_{xy}\rho_{xz})$. For $\rho_{xz}$ to be near zero, we have $\rho_{xy} \sim -\rho_{yz}$.
SFR and sSFR is weak in low-$M_h$ or low-$V_{peak}$ haloes and becomes stronger in haloes of higher $M_h$ or $V_{peak}$.

The results indicate that in the Illustris simulation galaxy formation ties to $V_{peak}$ more closely than $M_h$. In comparison with the $M_h$-based results, we find that in terms of $V_{peak}$, galaxy properties show a cleaner trend in the correlation with other halo assembly variable, such as the lack of correlation for $M_c$ and the similar correlation pattern for SFR/sSFR/colour. It suggests that $V_{peak}$-based halo model would be a good choice for capturing galaxy assembly bias effect (e.g. with $M_c$-based galaxy samples) and for studying galaxy assembly bias (e.g. with SFR/sSFR/colour-based samples). With $V_{peak}$ as the primary halo variable in the model, halo formation time and concentration would be the main options for the secondary variable to describe the relation between haloes and star formation related quantities, with the former preferable.

### 3.3 Assembly bias of central galaxies

With the set of galaxy and halo properties investigated in section 3.2, we do not find a galaxy property that 100 per cent correlates with a halo assembly property. For the $M_h$ dependence, the strongest correlation has $|p| \approx 0.5-0.6$, between $M_c$ and $V_{peak}/c/\sigma_{M/2}$. It means that halo assembly bias cannot be fully inherited by galaxies and be fully translated to galaxy assembly bias. Galaxy assembly bias should be different from halo assembly bias. For example, for haloes of the same mass, we can split them into two halo samples of low and high concentrations and then split central galaxies into two galaxy samples with low and high $M_c$. There would be a difference in the clustering of the two halo samples, as well as in that of the two galaxy samples. Given that $M_c$ is not perfectly correlated with $c$, we expect that the difference in the galaxy samples is smaller than that in the halo samples. As connecting galaxy assembly bias to halo assembly bias at the halo level can be an important ingredient in incorporating assembly bias effect into clustering model, we develop a simplified model below to understand the connection between galaxy and halo assembly bias.

Let us consider haloes at fixed $M_h$ (or $V_{peak}$) and focus on one halo assembly variable $x$ (e.g. $a_{IM}/c$ or concentration) and one galaxy property $y$ (e.g. $M_h$ or SFR). Without losing generality, $y$ is assumed to be positively correlated with $x$.

The joint distribution of $x$ and $y$ is illustrated by an ellipse in the top panel of Fig. 6. We can form a halo sample by selecting the fraction $f$ of haloes with the highest $x$ (to the right of the vertical dashed line, the region inside the dashed curve) and a galaxy sample by selecting the same fraction of central galaxies with the highest $y$ (above the red-purple dividing line, the region in red). Then, what is the relation between the bias factors of the galaxy and halo sample?

To proceed, we make the following assumptions – (1) The joint distribution of galaxy and halo properties follows a 2-dimensional (2D) Gaussian $p(x, y)$, characterised by the centre $(x_c, y_c)$, standard deviations $\sigma_x$ and $\sigma_y$, and the correlation $\rho$ between $x$ and $y$. We can take $(x_c, y_c) = (0, 0)$ by shifting $x$ and $y$. A 2D Gaussian function is a reasonable approximation for the distributions seen in Fig. 3 and Fig. 4, which also follows the Taylor expansion of the logarithm of the distribution function to the second order. (2) Galaxy property $y$ has a strong dependence on the halo assembly property $x$, and only weakly on other halo assembly variables. (3) At the fixed $M_h$ (or $V_{peak}$), halo bias factor is linear with respect to halo property $x$, $b(x) = kx + b_c$, the first order approximation from Taylor expansion. The value $b_c$ is the bias factor at $x = 0$, which is also the average halo bias factor for haloes at mass $M_h$ (or $V_{peak}$). The slope $k$ is the first derivative of $b$ with respect to $x$, $k = \partial b/\partial x$.

For the top $f$ fraction of haloes with the highest $x$ and...
The bias factors for the halo and galaxy samples are then defined as

$$\delta^b_h = (b_h - b_c)/b_c$$  \hspace{1cm} (6)$$

and

$$\delta^b_g = (b_g - b_c)/b_c.$$  \hspace{1cm} (7)

We can characterise the assembly bias effect by the fractional difference between the bias factor of the selected halo/galaxy sample and the average halo bias factor at $M_h$ (or $V_{\text{peak}}$), $\delta^b_h = (b_h - b_c)/b_c$ and $\delta^b_g = (b_g - b_c)/b_c$. Based on equations (3)–(7), we have

$$\delta^b_g = \rho \delta^b_h.$$  \hspace{1cm} (8)

That is, the assembly bias effect of the galaxy sample is weaker than that of the halo sample, by a factor equal to the correlation coefficient of the galaxy and halo property. Only in the case that galaxy and halo properties are tightly correlated (with zero scatter; $|\rho| = 1$) does halo assembly bias effect completely translate to galaxy assembly bias effect. The connection in equation (8) is also valid for samples defined by the property range bounded by two percentiles (i.e. bin samples rather than threshold samples considered here).

To test how well the simple model works, we choose a pair of halo and galaxy properties to construct the halo and galaxy samples. We then measure the two-point correlation functions (2PCFs) of the halo and galaxy samples in each $M_h$ and $V_{\text{peak}}$ bin. To reduce the uncertainty, the large scale bias factor of a given halo sample is derived from the ratio of the halo-matter two-point cross-correlation function and the matter auto-correlation function (e.g. Xu & Zheng 2018), $b_h = (\xi_{\text{hm}}(r)/\xi_{\text{mm}}(r)$, averaged over scales of 5–18 h−1 Mpc. The bias factor for the galaxy sample is similarly derived. We consider samples based on halo formation time $t_{\text{form}}$ and galaxy colour/SFR/sSFR, which show $M_h (V_{\text{peak}})$ dependent correlation coefficient (Fig. 5). The results are shown in Fig. 7. In the top-left panel, the thick dotted red (blue) curves show the assembly bias quantity $\delta^b_h$ for the 50 per cent oldest (youngest) haloes as a function of $M_h$. The solid red (blue) curves are $\delta^b_g$ for the 50 per cent reddest (bluest)
central galaxies. Both quantities decrease with increasing halo mass, i.e. the assembly bias effect becomes weaker for more massive haloes. The thin dashed curves are the same as the dotted curves but modulated by the correlation coefficient between colour and $a_{M/2}$, i.e. $\rho b^h_\delta$, the prediction for $\delta^h_\delta$ from the simple model. The model works well in reproducing the halo mass dependent galaxy assembly bias effect based on halo assembly bias and the galaxy-halo correlation. The top-middle and top-right panels are for galaxies selected according to SFR and sSFR. The bottom panels show the assembly effect as a function of $V_{\text{peak}}$. In all the cases, $\delta^h_\delta$ can be well described by $\rho b^h_\delta$, which supports the effectiveness of the simple model.

The success of the model suggests that an easy recipe could be developed to incorporate galaxy assembly bias into the halo model. The contribution of central galaxies to the galaxy bias factor, in its full form in the simple model, is

$$b_g = b_c + \frac{\partial b}{\partial x}\left| x_c + \sigma_x \frac{\rho}{\sqrt{2\pi}} \exp(-x^2/2)/f \right|,$$

where $x = t(f)$ is from equation (5). As an example, let us use the halo mass $M_h$ as the primary variable in the halo model and consider a galaxy property (colour) that correlates with halo formation time ($a_{M/2}$ or $\log a_{M/2}$). In the halo model, besides the average halo bias $b_c$, we also need to know how the halo bias changes with $a_{M/2}$ ($\partial b/\partial x$), the mean value of $a_{M/2}$ ($x_c$), and the scatter in $a_{M/2}$ ($\sigma_x$), all as a function of $M_h$. As usual, we can construct fitting formulae for those four quantities based on $N$-body simulations. The quantities $f$ and $\rho$ belong to the description of the galaxy-halo relation, which can be parameterised. For $f$, it is simply the occupation fraction for haloes at $M_h$. For $\rho$, the results in Fig. 5 suggest that a quadratic form with $M_h$ would suffice. To compute the galaxy bias factor for a galaxy sample, we only need to perform a 1D integral over halo mass. Certainly it is straightforward to include assembly bias effect by populating dark matter haloes in $N$-body simulations. However, the above proposal has its virtue for analytic calculations in theoretical investigations.

The simple model can also be applied to observation to infer the correlation between galaxy and halo properties. Lin et al. (2016) construct samples of central galaxies from the Sloan Digital Sky Survey data and study the assembly bias effect from the 2PCF measurements. Early and late galaxy samples are defined according to either star formation history (SFH) or sSFR. Weak lensing measurements are used to verify that the host haloes of the early and late galaxies are of similar halo mass (around $10^{12} h^{-1} M_\odot$). Lin et al. (2016) compare the difference in the early and late galaxy clustering to that in the early and late formed haloes, and do not find evidence for galaxy assembly bias. For SFH-based galaxy samples, they find the ratio of the early to late galaxy bias factor to be $1.06 \pm 0.12$ (see their fig.5). If we take the mean of the bias factors of the two galaxy samples as the average halo bias and the uncertainty comes from two similar error bars added in quadrature, the measurement gives $\delta^h_\delta = 0$ with an uncertainty 0.085. For haloes around $10^{12} h^{-1} M_\odot$, $\delta^h_\delta \sim 0.25$, with early- and late-forming haloes (Fig. 5). The coefficient of the correlation between SFH and halo formation time is then inferred to be $\rho = \delta^h_\delta/\delta^h_\delta = 0.00 \pm 0.34$. For the sSFR-based samples, the ratio of the early to late galaxy bias factor is $1.07 \pm 0.14$ (their fig.5). We infer $\delta^h_\delta = 0.034 \pm 0.099$, and with $\delta^h_\delta \sim 0.25$ the coefficient of the correlation between sSFR and halo formation time is constrained to be $\rho = 0.14 \pm 0.40$. For both cases, the correlation between galaxy and halo property is consistent with being small. It implies that galaxy SFH and sSFR at most only loosely track halo formation. Similar measurements with large samples can reduce the uncertainty in the inferred correlation, which would help test galaxy formation models (e.g. by comparing to those in Fig. 5).

4 SUMMARY AND DISCUSSION

Properties in galaxies residing in haloes of the same mass may have a dependence on certain aspects of halo assembly or formation history, which is termed as galaxy assembly bias. Studying galaxy assembly bias and its relation to halo properties can help improve the halo model of galaxy clustering and yield insights in galaxy formation and evolution. Using the Illustris cosmological hydrodynamic galaxy formation simulation, we investigate the central galaxy assembly bias effect through studying the relation among a set of galaxy and halo properties.

The main results can be summarised as follows.

(1) Central galaxy stellar mass $M_*$ has a tighter relation with $V_{\text{peak}}$ than with $M_h$, manifested by the smaller scatter in $M_*$ at fixed $V_{\text{peak}}$ than at that at fixed $M_h$. Once the assembly effect of $M_*$ on $V_{\text{peak}}$ is included, $M_h$ shows nearly no correlation with any other halo assembly properties.

(2) The correlations between halo assembly properties and other galaxy properties also appear cleaner if studied at fixed $V_{\text{peak}}$, which reveal that galaxy SFR, sSFR, and colour mainly correlate with halo formation time (and to a less extent with halo concentration).

(3) A simple model is presented to show the relation between galaxy and halo assembly bias, which is linked by the correlation coefficient of the galaxy and halo property in consideration.

The Illustris simulation produces a relation between central galaxy stellar mass and halo mass ($M_* - M_h$) similar to that inferred from observation. We find that the scatter in the relation is closely related to halo assembly properties. For example, at fixed $M_h$, haloes of higher $V_{\text{peak}}$ or earlier formation tend to host galaxies of higher $M_*$. If we choose $V_{\text{peak}}$ to be the primary halo variable, the scatter in $M_*$ at fixed $V_{\text{peak}}$ is reduced compared to that at fixed $M_h$. Remarkably, once switched to $V_{\text{peak}}$, $M_h$ appears to have nearly no dependence on other halo assembly properties, at least for those considered in our study (including halo concentration, formation time, spin, accretion rate, and specific accretion rate). The property $V_{\text{peak}}$, which is an indication of the maximum potential depth over the assembly history of haloes, is able to capture almost all the assembly effect in galaxy stellar mass. The results are in broad agreement with the study using the EAGLE simulation in terms of the $z = 0$ maximum halo circular velocity (Matthee et al. 2017). The reason for $V_{\text{peak}}$ to be the fundamental property in determining $M_*$ is likely related to the accretion and response of baryons in the gravitational potential. As for the scatter, the simulation noise in Illustris does not contribute much
(Genel et al. 2018). It could be related to chaotic or stochastic processes in star formation and feedback (Matthee et al. 2017; Genel et al. 2018). Further study is needed to understand the cause of the correlation between $M_\star$ and $V_{\text{peak}}$ and the origin of the scatter.

We present the correlation between each pair of galaxy and halo property in terms of the Pearson correlation coefficient. Besides $M_\star$, the other galaxy properties (SFR, sSFR, and colour) show a more consistent and clear trend with $V_{\text{peak}}$ than with $M_\star$. At fixed $V_{\text{peak}}$, those other galaxy properties related to star formation are found to mainly correlate with halo formation time and concentration, with stronger correlation with the former. The relatively nice behaviour in the correlations with galaxy properties in the $V_{\text{peak}}$-based investigation suggests that it would be advantageous to use $V_{\text{peak}}$ as the primary variable in halo model of galaxy clustering, in particular in modelling stellar-mass-based samples.

To further model SFR-, sSFR-, or colour-selected samples of galaxies, our investigation suggests to introduce halo formation time as the secondary halo variable (and to a less extent, halo concentration). This is in line with the age-matching model of Hearin & Watson (2013), who assumes monotonic mapping between galaxy colour and some variant of halo formation time at fixed galaxy luminosity (or stellar mass). That is, there exists a perfect correlation between the galaxy and halo property. Our investigation shows, however, that the correlation coefficient between galaxy SFR/sSFR and halo property in terms of the Pearson correlation coefficient of the galaxy and halo properties used to define the two samples. It gives a reasonable description for the samples constructed with the simulation. It suggests a simple prescription to incorporate galaxy assembly bias into the halo model. By applying the simple model to the galaxy clustering measurements in Lin et al. (2016), we infer that the correlation between SFR/sSFR and halo formation time is consistent with being weak ($\rho \sim 0.14$). The simple model can be further tested with other hydrodynamic simulations, like EAGLE (Schaye et al. 2015) and IllustrisTNG (Nelson et al. 2018), and semi-analytic models, which can also provide further insights on parameterising the correlations between galaxy and halo properties. While our study in this paper focuses on central galaxies, we plan to carry out similar investigations for satellite galaxies to complete the picture of galaxy assembly bias at the halo level to help improve the halo model.

ACKNOWLEDGEMENTS

The support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged.

REFERENCES

Bardell M. J., Bond J. R., Kaiser N., & Szalay A. S. 1986, ApJ, 304, 15
Behroozi P. S., Conroy C., & Wechsler R. H. 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., & Wu H.-Y. 2013, ApJ, 762, 109
Behroozi P. S., Wechsler R. H., Wu H.-Y., et al. 2013, ApJ, 763, 18
Berlind A. A., & Weinberg D. H. 2002, ApJ, 575, 587
Berlind, A. A., Weinberg, D. H., Benson, A. J., et al. 2003, ApJ, 593, 1
Berlind A. A., Kazin, E., Blanton, M. R., et al. 2006, arXiv:astro/ ph/0610524
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Castorina E., Sheth R. K. 2013, MNRAS, 433, 1529
Chaves-Montero J., Angulo R. E., Schaye J., et al. 2016, MNRAS, 460, 3100
Contreras S., Zehavi I., Padilla N., et al. 2018, arXiv:1808.02986
Cooray A., & Sheth R. 2002, Phys. Rept., 372, 1
Croton, D. J., Gao, L., & White, S. D. M. 2007, MNRAS, 374, 1303
Dalal N., White, M., Bond J. R., & Shirokob 2008, ApJ, 687, 12
Gao L., Springel V., & White S. D. M. 2005, MNRAS, 363, L66
Gao L., & White S. D. M. 2007, MNRAS, 377, L5
Genel S., et al., 2014, MNRAS, 445, 175
Genel S., et al., 2018, arXiv e-prints, arXiv:1807.07084
Gu M., Conroy C., & Behroozi P. 2016, ApJ, 833, 2
Guo H., et al., 2016, MNRAS, 459, 3040
Guo H., Li C., Zheng Z., et al. 2017, ApJ, 846, 61
Han J., Li Y., Jing Y. P., Nishimichi T., Wang W., & Jiang C. 2019, MNRAS, 482, 1900
Hearin A. P., Watson D. F., 2013, MNRAS, 435, 1313
Jeeon-Daniel, A., Dalla Vecchia, C., Haas, M. R., et al. 2011, MNRAS, 415, L69.
Leauthaud A., Tinker J., Bundy K., et al. 2012, ApJ, 744, 159
Lin Y.-T., Mandelbaum R., Huang Y.-H., et al. 2016, ApJ, 819, 119
Matthee J., Schaye J., Crain R. A., Schaller M., Bower R., Theuns T., 2017, MNRAS, 465, 2381
McCarthy K. S., Zheng Z., Guo H., 2018, arXiv e-prints, arXiv:1810.05183
McEwen J. E., Weinberg D. H., 2016, arXiv e-prints, arXiv:1601.02693
Mehta, K. T. 2014, Ph.D. Thesis,
Mo H. J., & White S. D. M. 1996, MNRAS, 282, 347
Nelson D., et al., 2015, A& C, 13, 12
Nelson D., et al., 2018, arXiv e-prints, arXiv:1812.05609
Paranjape A., Hahn O., & Sheth R. K. 2018, MNRAS, 476, 3631
Press W. H., & Schechter P. 1974, ApJ, 187, 425
Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., & Primack J. 2012, MNRAS, 423, 3018

MNRAS 000, 1–13 (2018)
Galaxy assembly bias of central galaxies in the Illustris simulation

Reddick R. M., Wechsler R. H., Tinker J. L., & Behroozi P. S. 2013, ApJ, 771, 30
Santini P., et al., 2017, ApJ, 847, 76
Schaye J., Crain R. A., Bower R. G., et al. 2015, MNRAS, 446, 521
Sheth R. K., & Tormen G. 1999, MNRAS, 308, 119
Sparre M., et al., 2015, MNRAS, 447, 3548
Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15
Springel V., White S. D. M., Jenkins A., et al. 2005, NATURE, 435, 629
Tinker J. L., et al., 2017, ApJ, 839, 121
Tinker, J. L. 2017, MNRAS, 467, 3533.
Vogelsberger M., Genel S., Springel V., et al. 2014, MNRAS, 444, 1518
Vogelsberger M., Genel S., Springel V., et al. 2014, NATURE, 509, 177
White S. D. M., Rees M. J. 1978, MNRAS, 183, 341
Xu H., Zheng Z., Guo H., Zu Y., Zehavi I., & Weinberg D. H. 2018, MNRAS, 481, 5470
Xu X., & Zheng Z. 2018, MNRAS, 479, 1579
Yang X., Mo H. J., & van den Bosch F. C. 2003, MNRAS, 339, 1057
Yang X., Mo H. J., & van den Bosch F. C. 2006, ApJ, 638, L55
Zehavi I., Zheng Z., Weinberg D. H., et al. 2005, ApJ, 630, 1
Zehavi I., Contreras S., Padilla N., et al. 2018, ApJ, 853, 847
Zentner A. R., Hearin A. P., & van den Bosch F. C. 2014, MNRAS, 443, 3044
Zentner, A. R., Hearin, A., van den Bosch, F. C., Lange, J. U., & Villarreal, A. 2016, arXiv:1606.07817
Zheng Z., Berlind A. A., Weinberg D. H., et al. 2005, ApJ, 633, 791
Zheng Z., Coil A. L., & Zehavi I. 2007, ApJ, 667, 760
Zu Y., & Mandelbaum R. 2015, MNRAS, 454, 1161
Zu Y., Mandelbaum R., Simet M., Rozo E., & Rykoff E. S. 2017, MNRAS, 470, 551

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