

# Modelling the tightest relation between galaxy properties and dark matter halo properties from hydrodynamical simulations of galaxy formation

Jian-hua He <sup>1,2</sup>★

<sup>1</sup>*School of Astronomy and Space Science, Nanjing University, Nanjing 210093, China*

<sup>2</sup>*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

Accepted 2020 February 26. Received 2020 February 26; in original form 2019 May 9

## ABSTRACT

We investigate how a property of a galaxy correlates most tightly with a property of its host dark matter halo, using state-of-the-art hydrodynamical simulations of galaxy formation: EAGLE, Illustris, and IllustrisTNG. Unlike most of the previous work, our analyses focus on all types of galaxies, including both central and satellite galaxies. We find that the stellar mass of a galaxy at the epoch of the peak circular velocity with an evolution correction gives the tightest such correlation to the peak circular velocity  $V_{\text{peak}}$  of the galaxy's underlying dark matter halo. The evolution of galaxy stellar mass reduces rather than increases scatter in such a relation. We also find that one major source of scatter comes from star stripping due to the strong interactions between galaxies. Even though, we show that the size of scatter predicted by hydrodynamical simulations has a negligible impact on the clustering of dense  $V_{\text{peak}}$ -selected subhalo from simulations, which suggests that even the simplest subhalo abundance matching (SHAM), without scatter and any additional free parameter, can provide a robust prediction of galaxy clustering that can agree impressively well with the observations from the Sloan Digital Sky Survey (SDSS) main galaxy survey.

**Key words:** galaxy clustering – hydrodynamical simulations – galaxy-halo relation.

## 1 INTRODUCTION

In modern analysis of the large-scale structure of galaxies, galaxies are usually treated as discrete points. The overdensity field of galaxies is a Poisson sample of the underlying dark matter field (Peebles 1980) and the statistics of galaxies does not depend on properties of galaxies. However, this assumption is indeed oversimplified (Casas-Miranda et al. 2002). Galaxy formation is far more complicated than just a discrete point process. Galaxies are not formed in an isolated environment but rather interact frequently with one another. Not only can such interactions change the probability distribution of galaxies but also the properties of galaxy themselves. As a result, galaxies are biased tracers of the dark matter field. The way they trace the dark matter field depends both on their properties and their assembly histories. For instance, it is well known that red galaxies are more clustered than the blue ones; more luminous galaxies are more clustered than the fainter ones as well (Zehavi et al. 2011). Even for galaxies with similar properties, such as those selected in the Baryon Oscillation Spectroscopic Survey (BOSS) constant-mass (CMASS) sample (Dawson et al. 2012), where the galaxies have a fairly uniform mass distribution that peaks at around  $\log(M/M_{\odot}) \sim 11.3$ , galaxy clustering is still dependent on the stellar mass assembly history (Montero-Dorta et al. 2017).

Since in the realistic case galaxies are not good tracers of the dark matter field, the key question then becomes how to accurately quantify the relation between them. The first approach for this is to do modelling. In  $\Lambda$  cold dark matter ( $\Lambda$ CDM), there has been significant progresses in modelling such biases over the past decade (Scoccimarro 2004; Taruya, Nishimichi & Saito 2010; Bianchi et al. 2012; de la Torre & Guzzo 2012; Bianchi, Chiesa & Guzzo 2015; de la Torre et al. 2017). The galaxy bias model now can yield reasonable accuracy in the quasi-linear regime. However, despite the progress, this approach is still unsatisfactory in several aspects. First, the accuracy and utility of these models have to be tested against mock galaxy surveys, as the galaxy bias models are essentially phenomenological and empirical. However, the mock catalogues are usually built for a particular survey or specific to a particular selection of galaxies. It is unclear how accurate these bias models can be applied to samples with a wide range of properties and complex biases. Moreover, it also remains difficult to assess the extent to which the underlining simplified assumptions and intrinsic limitations in the mock catalogues affect the calibration of these models in the first place. In addition, aside from the accuracy and utility issue, another concern is that current bias models are tested against mock galaxy catalogues based on simulations only in  $\Lambda$ CDM. It is unknown whether or not these models can still work in a modified gravity (MG) model. This is an important test because, unlike in  $\Lambda$ CDM, the relationship between galaxies and the dark matter field is much more complicated in MG models. Galaxies in

\* E-mail: [hejianhua@nju.edu.cn](mailto:hejianhua@nju.edu.cn)

MG models are not even directly related to the dark matter field, as the formation, clustering, and motion of galaxies are dictated by a different potential  $\phi_+ = (\psi + \phi)/2$  rather than the lensing potential  $\phi_- = (\psi - \phi)/2$  that is directly related to the true dark matter field. The relationship between  $\phi_+ = (\psi + \phi)/2$  and  $\phi_- = (\psi - \phi)/2$  can be very complex (He, Li & Hawken 2015) in a MG model.

The second approach makes use of  $N$ -body simulations, coupled to some phenomenological frameworks, such as the halo occupation distribution (HOD; Jing, Mo & Borner 1998; Peacock & Smith 2000; Berlind & Weinberg 2002; Zheng et al. 2005; Zheng, Coil & Zehavi 2007; Zehavi et al. 2011; Leauthaud et al. 2012; Guo et al. 2015, 2016) or the conditional luminosity function (CLF; Yang, Mo & van den Bosch 2003; Yang, Mo & Bosch 2009), to link galaxies to dark matter haloes. The basic idea of this approach is that galaxies reside in dark matter haloes, namely, the densest regions of the underlying dark matter field. And the probability of the distribution of galaxies is only dependent on the masses of dark matter haloes. Although these assumptions are obviously oversimplified that neglect some important effects such as the assembly bias (Gao, Springel & White 2005; Gao & White 2007), the HOD/CLF modelling turns out to be very successful in reproducing galaxy clustering even at very small scales for galaxy samples with a wide variety of different properties and complex biases (Zheng et al. 2007; Zehavi et al. 2011; Guo et al. 2015, 2016). However, despite this, when applied to testing the underlining cosmological models, these frameworks have obvious limitations: they are too flexible and, in general, lack of strong physical motivations. In a MG model, for example, even naively using the same HOD parametrizations as in  $\Lambda$ CDM, the framework can still be tuned to reproduce desired galaxy clustering (Hernández-Aguayo et al. 2018), which is clearly unfeasible given that the dark matter halo properties and the processes of galaxy formation in a MG model ought to be very different from those in  $\Lambda$ CDM.

In this work, rather than adhering to the doctrine that galaxies are tracers of the dark matter field, we regard galaxies as tracers of the dark matter haloes: galaxies illuminate the dark matter haloes and their properties are tracers of the properties of the dark matter haloes. This idea is in line with the philosophy of subhalo abundance matching (SHAM; Kravtsov et al. 2004; Vale & Ostriker 2004; Conroy, Wechsler & Kravtsov 2006; Guo et al. 2010; Moster et al. 2010; Behroozi, Wechsler & Conroy 2013; Reddick et al. 2014). In this paper, we will further explore how a property of galaxy physically correlates to a property of its host dark matter halo. In order to achieve this, we use hydrodynamical simulations of galaxy formation. In contrast to the semi-analytical galaxy formation models that are based on the dark matter only (DMO) ones (Cole et al. 2000; Guo et al. 2011), hydrodynamical simulations can trace the complicated interactions between baryons and dark matter in a self-consistent way based on the first principles of gravity and hydrodynamics. This is of paramount importance as the motion and clustering of galaxies, along with their host dark matter haloes, are primarily dictated by gravity and hydrodynamics. In order to further strength our results and prevent potential biases due to the choice of a particular simulation, in this work we adopt three different simulations: EAGLE (Crain et al. 2015; Schaye et al. 2015; McAlpine et al. 2016), Illustris (Vogelsberger et al. 2014; Nelson et al. 2015), and IllustrisTNG (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018a,b; Pillepich et al. 2018; Springel et al. 2018). These simulations are different in many aspects such as the numerical methods used, the model of subgrid baryonic physics, and, most importantly, the properties of simulated galaxies.

In addition, in contrast to most of the previous analysis that only focuses on central galaxies (e.g. Matthee et al. 2017; Matthee & Schaye 2019), our analysis includes all types of galaxies, especially including a significant fraction of satellite galaxies. Unlike the central galaxies, satellite subhaloes cannot be easily matched from the DMO simulations to the full baryonic physics hydrodynamical ones by identifying their dark matter particles. It can be as high as 30 per cent low-mass satellites in the DMO simulations that cannot find their counterparts in the hydrodynamical ones (Chaves-Montero et al. 2016). We, therefore, present a new approach to analyse the simulation data, which is different from the method used in Chaves-Montero et al. (2016).

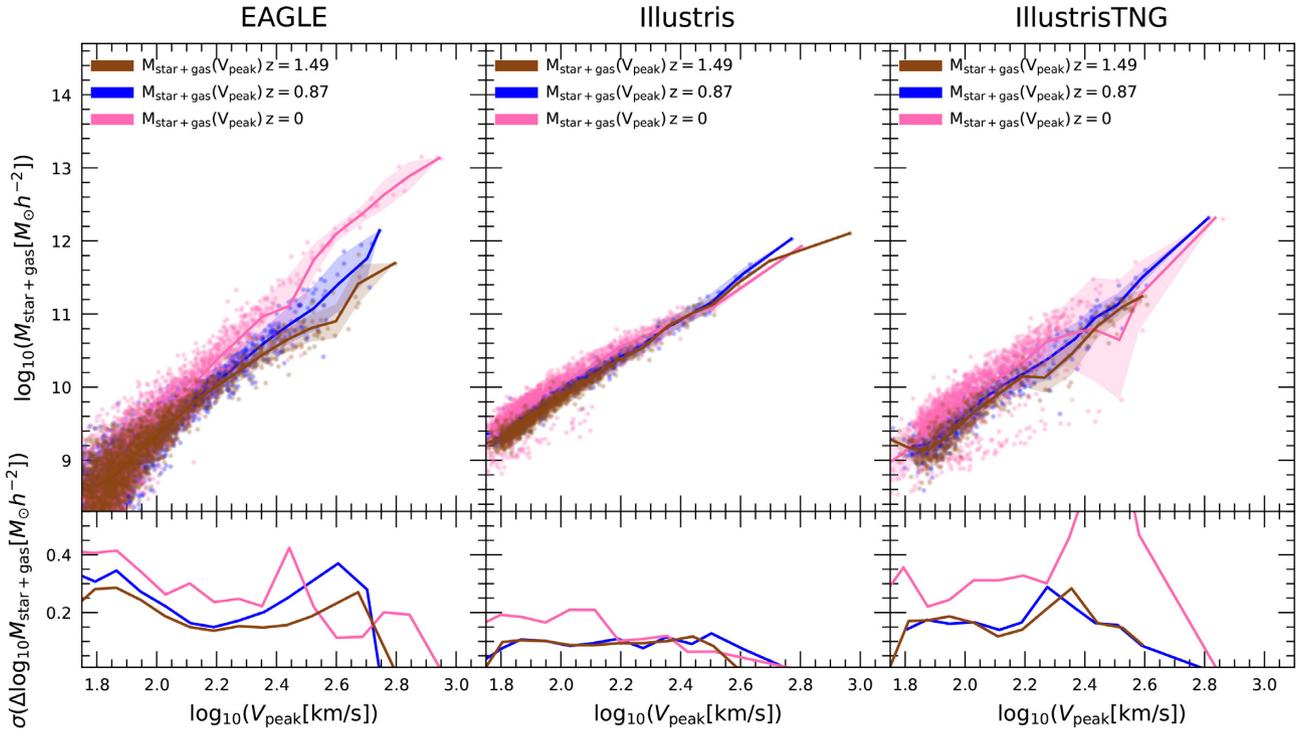
This paper is organized as follows. In Section 2, we introduce the hydrodynamical simulations and the galaxy catalogue used. In Section 3, we discuss the physical relationship between a property of galaxy and a property of its host dark matter halo before major disruptions. In Section 4, we discuss the impact of disruptions on the galaxy property–halo property relation. In Section 5, we discuss how to efficiently model galaxy clustering using subhaloes from simulations. In Section 6, we summarize and conclude this work.

## 2 HYDRODYNAMICAL SIMULATIONS

The first simulation we adopt is the EAGLE simulation (Crain et al. 2015; Schaye et al. 2015; McAlpine et al. 2016). Among the EAGLE suite, we use the largest simulation with a box size of  $67.77 \text{ Mpc } h^{-1}$  along one side. The mass resolution of gas particle in this simulation is  $1.81 \times 10^6 M_\odot$  and a dark matter particle is  $9.70 \times 10^6 M_\odot$ . We only use well-resolved galaxies with  $M_{\text{star}} > 8 \times 10^7 M_\odot$ . For the Illustris (Vogelsberger et al. 2014; Nelson et al. 2015) simulation, we use the highest resolution run with a box size of  $75 \text{ Mpc } h^{-1}$  along one side. A gas particle in this simulation is  $1.3 \times 10^6 M_\odot$  and a dark matter particle is  $6.3 \times 10^6 M_\odot$ . We focus on galaxies with  $M_{\text{star}} > 8 \times 10^7 M_\odot$ . In the IllustrisTNG suite (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018a,b; Pillepich et al. 2018; Springel et al. 2018), we use the simulation with a box size of  $75 \text{ Mpc } h^{-1}$  along one side, the same as Illustris. A gas particle in this simulation is  $1.4 \times 10^6 M_\odot$  and a dark matter particle is  $7.5 \times 10^6 M_\odot$ . Again, we focus only on galaxies with  $M_{\text{star}} > 8 \times 10^7 M_\odot$ . Further, the stellar (or gas) mass used in this work is the total mass of stars (or gases) that are bounded to a galaxy.

## 3 PRE-DISRUPTION

The first galaxy property we should look at is the total baryonic mass  $M_{\text{star+gas}}$  of a galaxy, namely, the total masses of stars and cold gas in the galaxy. This is motivated by the recent discovery of the mass discrepancy relation (McGaugh, Lelli & Schombert 2016), which finds that there is a tight empirical relation between the radial dependence of the enclosed baryonic-to-dynamical mass ratio and the baryonic acceleration. Such a relation also indicates that the total baryonic mass of a galaxy should be tightly correlated with the circular velocity of its host dark matter halo. In this work, we investigate this issue using hydrodynamic simulations. In order to simplify our analysis, we adopt the following strategy: before discussing a galaxy’s property at the current time, we first examine the galaxy’s property at the epoch of the peak value of the maximum circular velocities during its merger history  $V_{\text{peak}}$ . This is to use a galaxy’s property before major disruption, which can avoid any drastic changes in the properties of galaxies. We then discuss the impact of disruption on a galaxy’s property later on.



**Figure 1.** The total baryonic mass of a galaxy (the value at the epoch of  $V_{\text{peak}}$ ) as a function of  $V_{\text{peak}}$  at three representative redshifts  $z = 1.49, 0.87,$  and  $0$ . The samples used here include both centrals and satellites. The upper panels show the total baryonic mass– $V_{\text{peak}}$  relations for three simulations EAGLE (left), Illustris (middle), and IllustrisTNG (right). The lower panels show the scatter in these relations. The EAGLE and IllustrisTNG simulations show relatively big scatter that also depends on redshift. The Illustris simulation, however, shows a smaller scatter.

The upper panels of Fig. 1 show the total baryonic mass of a galaxy as a function of  $V_{\text{peak}}$  in the EAGLE (left), Illustris (middle), and IllustrisTNG (right) simulations at three representative redshifts  $z = 1.49, 0.87,$  and  $0$ . Galaxies in Fig. 1 are selected when their maximum circular velocities peak at the three specified redshifts during their merger histories. Hence, the value of the total baryonic mass for each galaxy in Fig. 1 is at  $V_{\text{peak}}$ . Furthermore, note that the samples used here include both centrals and satellites. From the lower panels of Fig. 1, EAGLE and IllustrisTNG simulations show relatively big scatter in the total baryonic mass– $V_{\text{peak}}$  relation, which is around  $\sigma(\Delta \log_{10} M_{\text{star+gas}}) \sim 0.3$ . The scatter also depends on redshift. However, Illustris simulation shows a smaller scatter, which is around  $\sigma(\Delta \log_{10} M_{\text{star+gas}}) \sim 0.2$ . That hydrodynamical simulations do not give a consistent result on the total baryonic mass– $V_{\text{peak}}$  relation indicates that the behaviour of the gas component in a galaxy is dependent on the details of baryonic physics implemented in the simulations.

Next, we turn to the stellar component of a galaxy. The upper panels of Fig. 2 show the stellar mass of a galaxy (the value at the epoch of  $V_{\text{peak}}$ ) as a function of  $V_{\text{peak}}$  for the EAGLE (left), Illustris (middle), and IllustrisTNG (right) simulations at three different redshifts  $z = 1.49, 0.87,$  and  $0$ , the same as in the previous case. In contrast to the total baryonic mass, all three simulations show a very similar tight relation between stellar mass and  $V_{\text{peak}}$  at a fixed redshift. From the lower panels of Fig. 2, different simulations even at different redshifts also show a very similar size of scatter. However, the zero-points in the scaling relations are different. This is due to the fact that a galaxy’s stellar mass evolves with time. A longer time of star formation can lead to more stellar mass formed in a galaxy and the zero-points of the scaling relation are, therefore, higher at low redshift and lower at high redshift. Clearly,

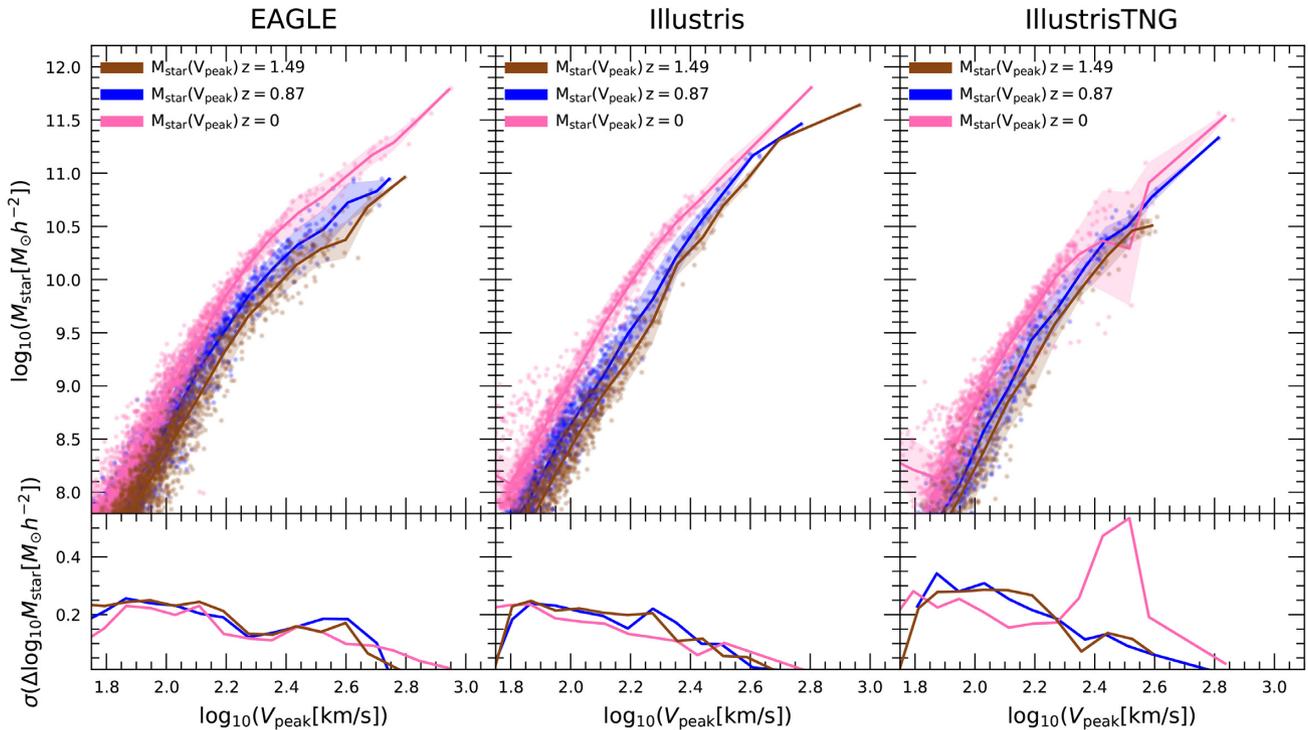
the evolution of the stellar mass plays an important role here. If we neglect such evolution effect and simply take the stellar mass at the epoch of  $V_{\text{peak}}$ , the scatter in the stellar mass– $V_{\text{peak}}$  relation will be very large. Therefore, the evolution of galaxy stellar mass should be accounted for. However, before we proceed to present how to correct for this effect, we first look at the specific star-forming rate (sSFR), namely, the star-forming rate per unit stellar mass of a galaxy.

Fig. 3 shows the sSFR of a galaxy as a function of  $V_{\text{peak}}$  at three representative redshifts in the EAGLE simulation. Around the mean values (represented by solid lines and usually called the main-sequence galaxies), the scatter of the sSFR is very big. However, the average value of the sSFR is nearly a constant at a fixed redshift, which is almost independent of  $V_{\text{peak}}$ , except the most massive ones at redshift zero. Thus, one can model such evolution of stellar mass as

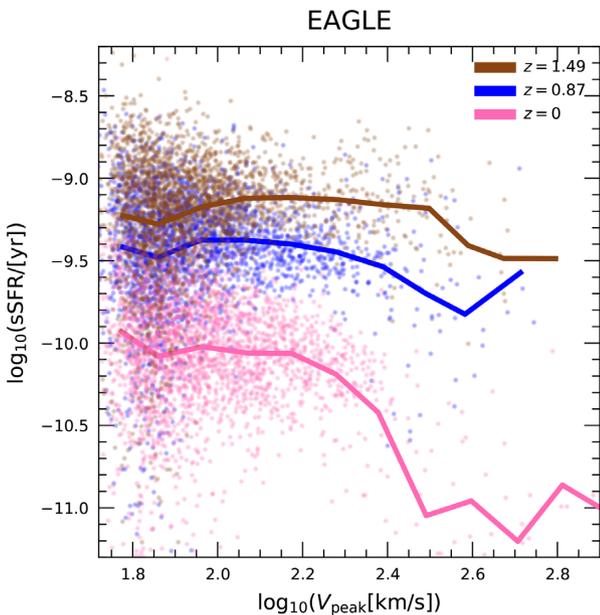
$$\log_{10}(M_{\text{star}}[M_{\odot} h^{-2}])|_{z=0} = \log_{10}(M_{\text{star}}[M_{\odot} h^{-2}])(z) + \alpha \times z, \quad (1)$$

where  $\alpha$  is a constant, which is different among the EAGLE, Illustris, and IllustrisTNG simulations ( $\alpha_{\text{EAGLE}} = 0.324$ ,  $\alpha_{\text{Illustris}} = 0.243$ ,  $\alpha_{\text{IllustrisTNG}} = 0.194$ ). This is because the stellar mass functions and the star-forming rates in these simulations are different. The stellar mass– $V_{\text{peak}}$  relation at different redshifts then can be corrected to redshift zero using the above equation.

Fig. 4 shows the evolution corrected stellar mass– $V_{\text{peak}}$  relation at redshift zero. We only use the EAGLE simulation here for illustrative purposes. In Fig. 4, we also show the results at three representative redshifts as control. After taking into account the effect of stellar mass evolution, the evolution corrected stellar mass of galaxies is tightly correlated with  $V_{\text{peak}}$  (right-hand panel in Fig. 4). The intrinsic scatter of this relation is very small. Note



**Figure 2.** The stellar mass of a galaxy (the value at the epoch of  $V_{\text{peak}}$ ) as a function of  $V_{\text{peak}}$  at three representative redshifts  $z = 1.49, 0.87,$  and  $0$ , the same as in Fig 1. The samples used here include both centrals and satellites. The upper panels show the stellar mass ( $V_{\text{peak}}$ )– $V_{\text{peak}}$  relations and the lower panels show the scatter in these relations. At a fixed redshift, the stellar mass correlates tightly with  $V_{\text{peak}}$  in all the three simulations and the scatter almost does not depend on redshift. The zero-point of the scaling relation varies with redshift. This is due to the evolution of stellar mass. Since a longer time of star formation can lead to more stellar mass formed in a galaxy, the zero-points of the scaling relation are higher at low redshift than those at high redshift.



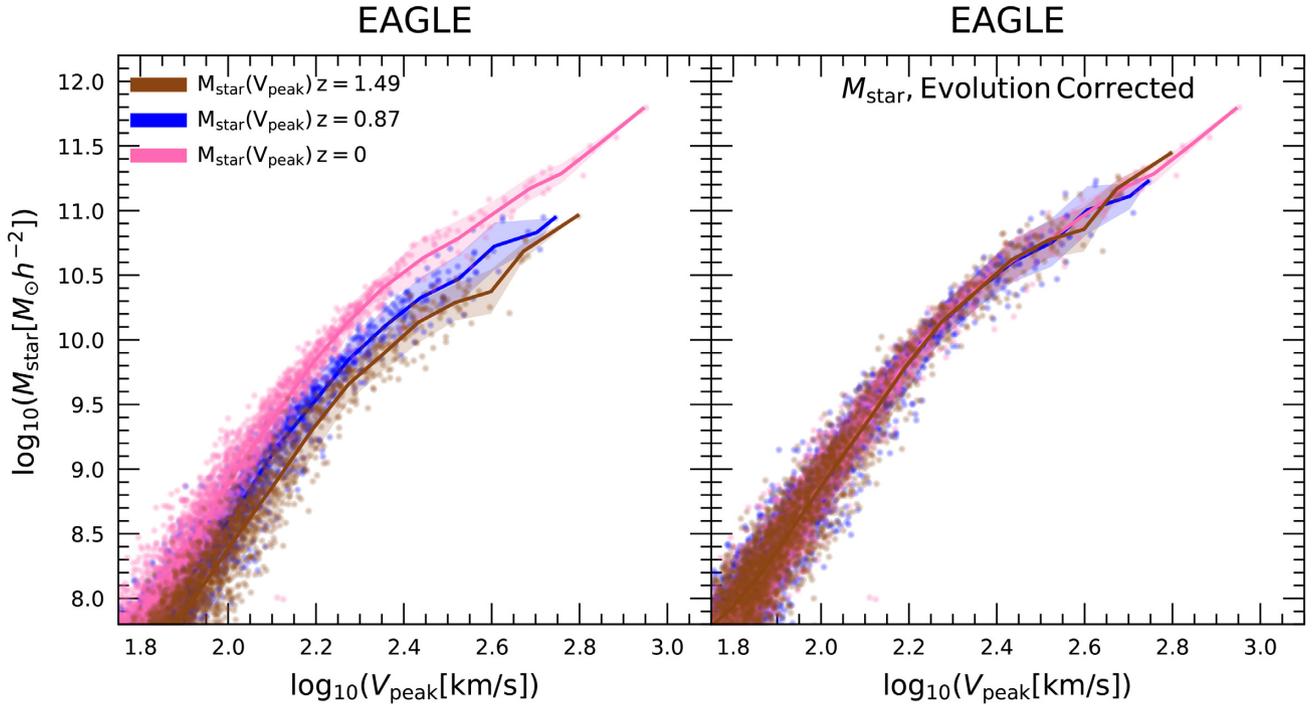
**Figure 3.** The specific star-forming rate (sSFR) as a function of  $V_{\text{peak}}$  at three representative redshifts in the EAGLE simulation. Despite the big scatter around the main-sequence galaxies, the average sSFR is nearly a constant, which is almost independent of  $V_{\text{peak}}$ , except the most massive ones at redshift zero. A very similar result can be found in the Illustris and IllustrisTNG simulations as well. However, we do not present here for simplicity.

that here we have used the galaxy’s stellar mass that would be at redshift zero if there were no disruption rather than the true current stellar mass of the galaxy.

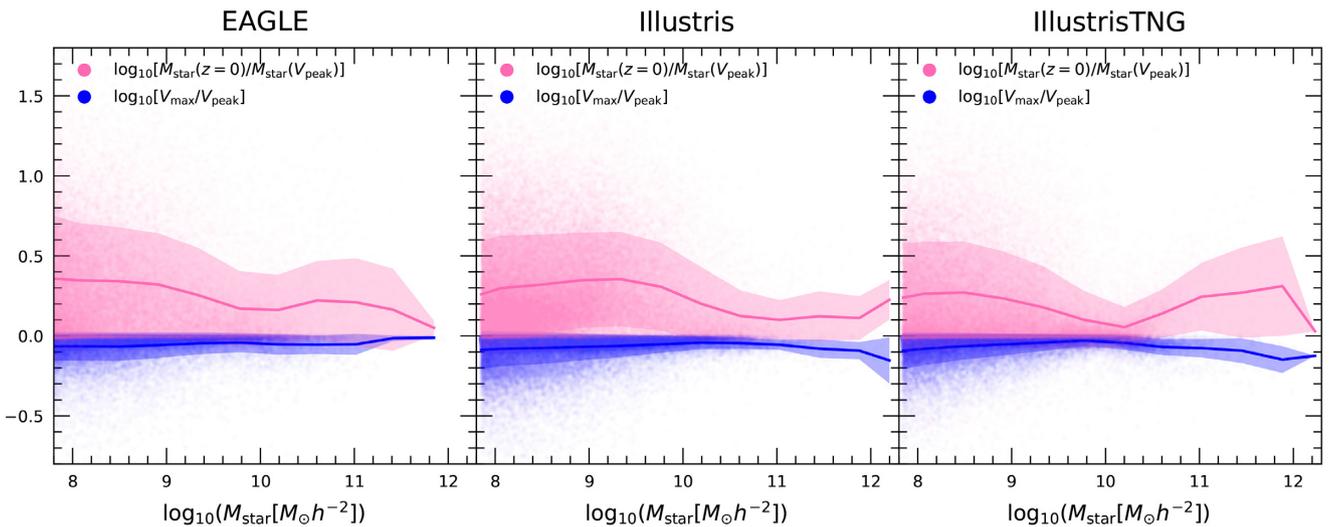
In addition, it is worth noting that although the scatter in the sSFR– $V_{\text{peak}}$  relation is very large (see Fig. 3), the stellar mass and  $V_{\text{peak}}$  end up with a very tight correlation. This is indeed not a surprise because the sSFR has big fluctuations during the accretion history of stellar mass (Matthee & Schaye 2019), which leads to the big scatter in the sSFR– $V_{\text{peak}}$  relation. However, in the stellar mass– $V_{\text{peak}}$  relation, this effect can be largely smoothed out as the stellar mass is the temporal cumulation of the sSFR.

#### 4 POST-DISRUPTION

Unlike the pre-disruption stellar mass– $V_{\text{peak}}$  relation, after  $V_{\text{peak}}$  great complexity comes in due to the interactions between galaxies. In order to illustrate this, in Fig. 5, we show the ratio of the stellar mass (pink points) and the maximum circular velocity  $V_{\text{max}}$  (blue points) of a galaxy at the present day to those values at the epoch of  $V_{\text{peak}}$  as a function of the stellar mass for the EAGLE (left), Illustris (middle), and IllustrisTNG (right) simulations. As shown by the blue points, the values of  $\log_{10}(V_{\text{max}}/V_{\text{peak}})$  for a significant fraction of galaxies are much less than zero, which means that the disruption of the dark matter component of the galaxies is prevalent (in most cases  $V_{\text{max}}$  at the present day is much smaller than the value at the epoch of  $V_{\text{peak}}$ ). In contrast, the stellar component of most galaxies (pink points), on the other hand, grows rather than decreases significantly from the epoch of  $V_{\text{peak}}$  to the present day. This happens even for the cases that the dark matter and gas components undergo a dramatic loss. This indicates that even when



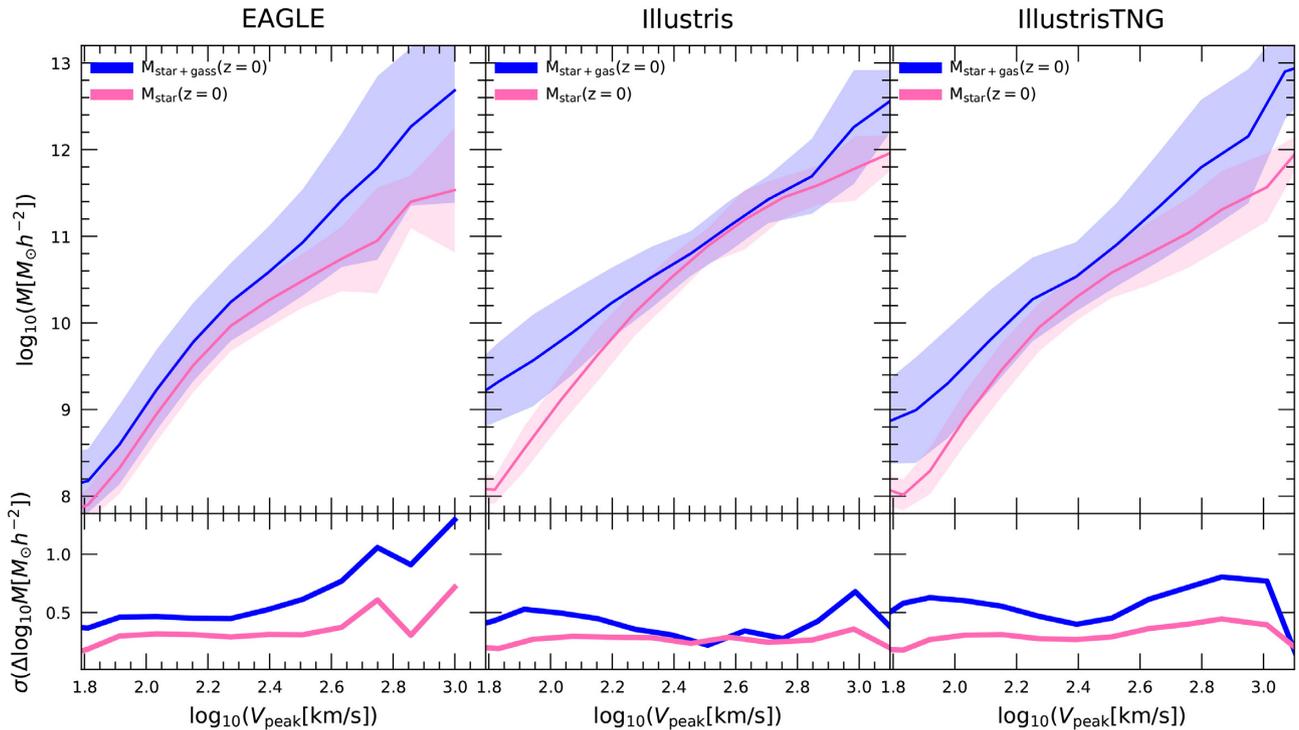
**Figure 4.** Right: the evolution corrected stellar mass– $V_{\text{peak}}$  relation at redshift zero. Left: for comparison, the stellar mass– $V_{\text{peak}}$  relation at three different redshifts  $z = 1.49, 0.87,$  and  $0$ , which are the same as the left-hand panel of Fig. 2. We only show the result from the EAGLE simulation for illustrative purposes.



**Figure 5.** The ratio of the stellar mass (pink points) and the maximum circular velocity  $V_{\text{max}}$  (blue points) of a galaxy at the present day to those values at the epoch of  $V_{\text{peak}}$  as a function of stellar mass for the EAGLE (left), Illustris (middle), and IllustrisTNG (right) simulations. As a significant fraction of galaxies have  $\log_{10}(V_{\text{max}}/V_{\text{peak}})$  much less than zero (blue points), which indicates that the disruption of the dark matter component of galaxies is prevalent. In contrast, for most galaxies, the stellar component at the present day is much greater than the value at the epoch of  $V_{\text{peak}}$  ( $\log_{10}[M_{\text{star}}(z=0)/M_{\text{star}}(V_{\text{peak}})] > 0$ ), which means that even the dark matter and gas components undergo disruption, star forming in most galaxies can still last for a quite long time. However, it is also evident that in some extreme cases, the stellar components suffer dramatic losses along with their dark matter components (pink points with  $\log_{10}[M_{\text{star}}(z=0)/M_{\text{star}}(V_{\text{peak}})] \ll 0$ ).

a galaxy becomes a satellite merging into a large system, galaxies in most cases can still continue forming stars for a long time. Only in very extreme cases, the stellar components can suffer dramatic losses along with their dark matter components due to disruptions (pink points with  $\log_{10}[M_{\text{star}}(z=0)/M_{\text{star}}(V_{\text{peak}})] \ll 0$ ).

Fig. 6 shows the post-disruption total baryonic mass– $V_{\text{peak}}$  relation, compared with the post-disruption stellar mass– $V_{\text{peak}}$  relation for different hydrodynamical simulations. The left-hand panels are for the EAGLE simulation, the middle ones are for the Illustris simulation, and the right ones are for the IllustrisTNG simulations.



**Figure 6.** The upper panels show the post-disruption total baryonic mass– $V_{\text{peak}}$  relation, compared with the post-disruption stellar mass– $V_{\text{peak}}$  relation for different hydrodynamical simulations. In contrast to Figs 1 and 2, the values of the total baryonic mass and stellar mass of galaxies in this plot are taken at redshift zero. The left-hand panels are for the EAGLE simulation, the middle ones are for the Illustris simulation, and the right ones are for the IllustrisTNG simulations. The lower panels show scatters for the above relations. All hydrodynamical simulations consistently show that in the post-disruption case, the stellar mass– $V_{\text{peak}}$  relations are tighter than those in the total baryonic mass– $V_{\text{peak}}$  relation.

The lower panels of Fig. 6 show scatters for the above relations. After the epoch of  $V_{\text{peak}}$ , both stars and gases suffer disruption. However, since there is only gravitational interaction between stars and there are significant non-gravitational interactions between gases (e.g. energetic feedbacks), gas is more prone to disruption than that of stars. The stellar mass– $V_{\text{peak}}$  relation, therefore, is much tighter than that of the total baryonic mass– $V_{\text{peak}}$  relation. Further, all three hydrodynamical simulations yield a consistent result.

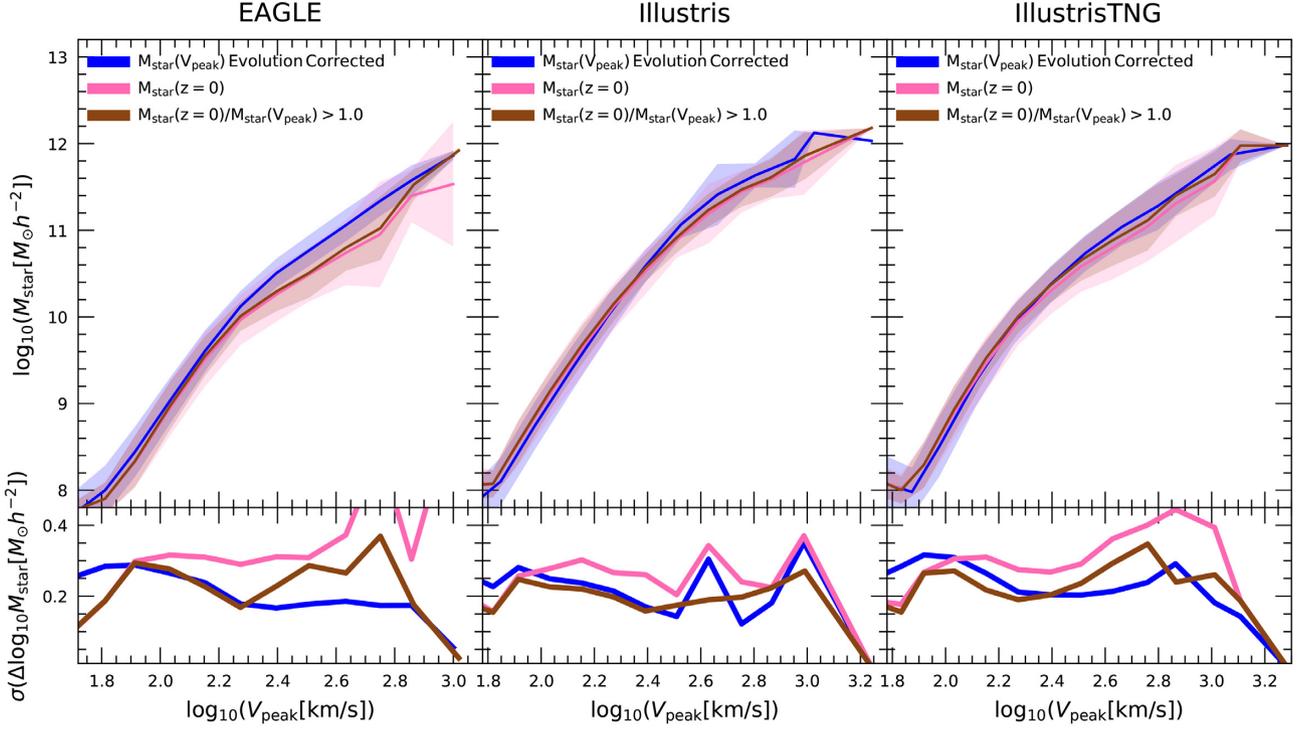
Fig. 7 shows various stellar mass– $V_{\text{peak}}$  relations for the EAGLE (left), Illustris (middle), and IllustrisTNG simulations. We show the evolution corrected relation (blue), in which stellar mass is taken at the epoch of  $V_{\text{peak}}$  but evolution corrected to redshift zero using equation (1). In the same plot, we also present the current stellar mass (post-disruption)– $V_{\text{peak}}$  relation (pink) and the current stellar mass– $V_{\text{peak}}$  relation but without disrupted galaxies  $\log_{10}[M_{\text{star}}(z=0)/M_{\text{star}}(V_{\text{peak}})] > 0$  (brown). The lower panels show scatters in the various relations. The typical scatter on stellar mass in the post-disruption stellar mass– $V_{\text{peak}}$  relation is around  $\sigma(\Delta \log_{10} M_{\text{star}}) \sim 0.3$  in the EAGLE and IllustrisTNG simulations but slightly smaller  $\sigma(\Delta \log_{10} M_{\text{star}}) \sim 0.25$  in the Illustris simulation. As in our analysis we include both centrals and satellites, the scatters here are slightly bigger than the values reported in the previous work that only use centrals (e.g. Matthee et al. 2017; Wechsler & Tinker 2018). However, if we do not take into account the disrupted galaxies and only consider the galaxies that gain stellar mass after  $V_{\text{peak}}$ , namely,  $\log_{10}[M_{\text{star}}(z=0)/M_{\text{star}}(V_{\text{peak}})] > 0$ , the scatter can be significantly reduced below  $\sigma(\Delta \log_{10} M_{\text{star}}) < 0.2$  for galaxies with  $V_{\text{peak}} > 100$  [km s $^{-1}$ ] (brown). In this case, the scatter is close to the ideal case of the evolution corrected value using stellar mass at the epoch of  $V_{\text{peak}}$  (comparing the blue and brown lines in the lower

panels of Fig. 7). This indicates that after  $V_{\text{peak}}$  when a galaxy is accreted by a larger system, stellar mass stripping is a major source of scatter to the relation between the current stellar mass and  $V_{\text{peak}}$ . Therefore, the tightest relation between a galaxy property and a halo property can only be achieved in terms of quantities before this disruption.

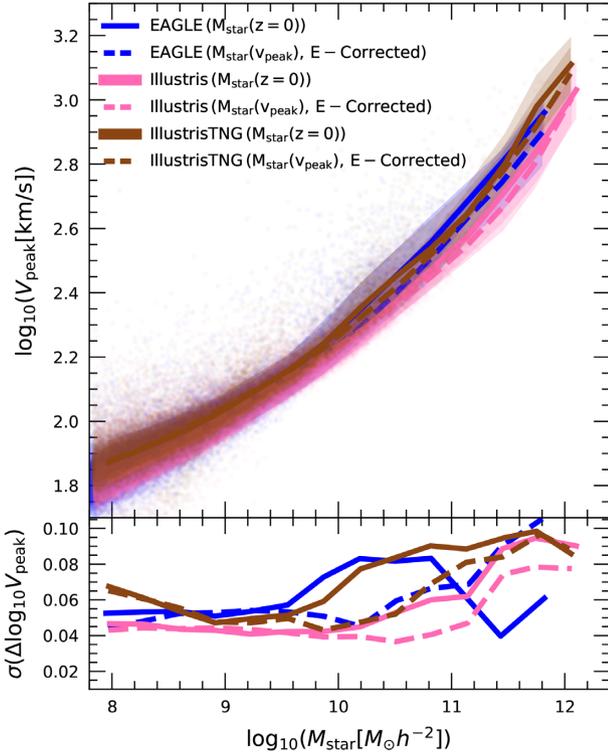
Nevertheless, it is worth pointing out that even in the case of post-disruption, the current stellar mass and  $V_{\text{peak}}$  still exhibit a very tight correlation. In Fig. 8, we compare such a relation (solid lines) with the evolution corrected value (dashed lines). Unlike Fig. 7, in Fig. 8 we show the scatter on  $\log_{10} V_{\text{peak}}$ . This has more practical meaning as in observations we can only measure a galaxy’s current stellar mass (post-disruption). Given the observed galaxy catalogue, we usually add scatter in the  $V_{\text{peak}}$ -ranked subhalo catalogues in  $N$ -body simulations and then compare them with observations. From Fig. 8, even considering the post-disruption galaxies, the scatter on  $\log_{10} V_{\text{peak}}$  is still very small  $\sigma(\Delta \log_{10} V_{\text{peak}}) < 0.08$ , only except the most massive cases, which, however, account for only a small fraction of the total galaxies.

## 5 MODELLING GALAXY CLUSTERING

The existence of such a tight correlation between galaxy stellar mass and  $V_{\text{peak}}$  has an important implication on how to efficiently model galaxy clustering from  $N$ -body simulations. As demonstrated in the previous sections, even for hydrodynamical simulations with very different subgrid baryonic physics and numerical methods used, they still produce similar predictions on such relation (see Fig. 8). Therefore, the predicted galaxy stellar mass– $V_{\text{peak}}$  relation should be robust unless there are serious limitations in our current



**Figure 7.** The upper panels show the evolution corrected stellar mass– $V_{\text{peak}}$  relation, in which the stellar mass is taken at the epoch of  $V_{\text{peak}}$  but evolution corrected to redshift zero (blue), the current stellar mass (post-disruption)– $V_{\text{peak}}$  relation (pink) and the current stellar mass– $V_{\text{peak}}$  relation but without disrupted galaxies  $\log_{10}[M_{\text{star}}(z=0)/M_{\text{star}}(V_{\text{peak}})] > 0$  (brown). The left-hand panels are for the EAGLE simulation, the middle ones are for the Illustris simulation, and the right ones are for the IllustrisTNG simulations. The lower panels show scatters for the various relations.



**Figure 8.** Upper panels: comparisons of the  $V_{\text{peak}}$ –evolution corrected stellar mass relation (solid lines) and  $V_{\text{peak}}$ –current stellar mass relation (dashed lines) derived from different hydrodynamical simulations. Lower panels: scatter on  $\log_{10} V_{\text{peak}}$ .

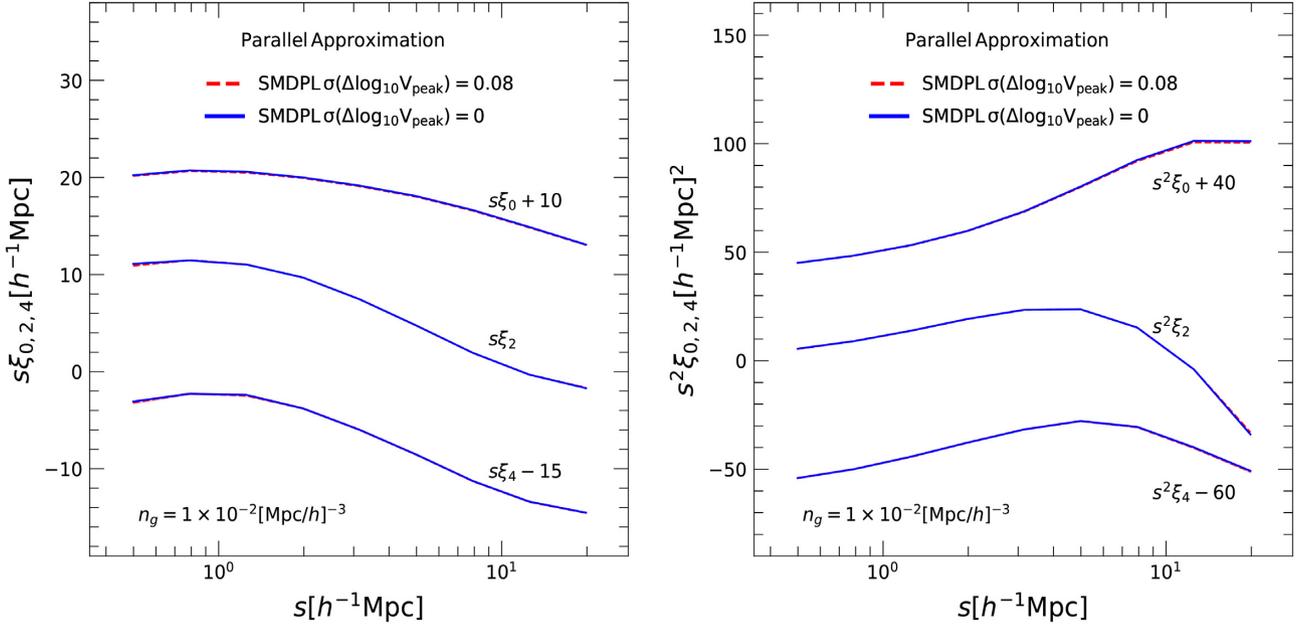
understanding of galaxy formation. Given such a tight correlation, galaxy clustering should be able to be modelled by using  $V_{\text{peak}}$ -selected subhaloes, which corresponds to stellar mass selected galaxies in the real observations. In this section, we will test this point with observations.

Before comparing simulations with observations, we first test the impact of scatter in the stellar mass– $V_{\text{peak}}$  relation on the clustering of subhaloes. In our analyses, we use the Small MultiDark Planck (SMDPL) simulation (Klypin et al. 2016), which uses  $3840^3$  dark matter particles in a box of  $400 \text{ Mpc } h^{-1}$  along one side. We focus on the redshift space since the redshift-space clustering can be directly measured in the real observations. We add scatter in the halo catalogue as follows. We first take subhaloes from the original catalogue and then replace their values of  $V_{\text{peak}}$  by drawing a random number around the logarithm of  $V_{\text{peak}}$  ( $\log_{10} V_{\text{peak}}$ ) within  $1\sigma$  ( $\Delta \log_{10} V_{\text{peak}} = 0.08$  scatter). Then we re-rank all the subhaloes in the catalogue and generate 200 realizations for such catalogues. The choice of the value of scatter is motivated by Fig. 8, from which  $\sigma(\Delta \log_{10} V_{\text{peak}}) = 0.08$  is large enough for most galaxies only except the most massive ones. The most massive galaxies, however, only account for a small fraction of the total number of galaxies.

In order to measure the redshift-space two-point correlation function  $\xi(r_{\sigma}, r_{\pi})$ , we use the Landy and Szalay estimator (Landy & Szalay 1993),

$$\xi(r_{\sigma}, r_{\pi}) = \frac{DD(r_{\sigma}, r_{\pi}) - 2DR(r_{\sigma}, r_{\pi}) + RR(r_{\sigma}, r_{\pi})}{RR(r_{\sigma}, r_{\pi})}, \quad (2)$$

where  $DD$ ,  $DR$ , and  $RR$  are the data–data, data–random, and random–random pair counts and  $r_{\sigma}$ ,  $r_{\pi}$  are the separations of galaxy pairs perpendicular and parallel to the line-of-sight direction, respectively.  $\xi(r_{\sigma}, r_{\pi})$  can be expanded in terms of Legendre



**Figure 9.** Left: the impact of scatter on the predicted RSD multipoles. The solid lines are for the results without scatter. The dashed lines are for the results with scatter. The shaded regions on the dashed lines are the  $1\sigma$  uncertainty around the expectation value, derived from 200 realizations. The shaded regions are so small that they cannot be clearly seen in the plot. The RSD multipoles  $\xi_{0,2,4}$  are multiplied by the redshift-space separation  $s$ . This rearrangement makes the RSD multipoles more sensitive to changes at small scales. Right: similar to the left-hand panel, but the RSD multipoles  $\xi_{0,2,4}$  are multiplied by  $s^2$ . The larger index of  $s$  can make the RSDs multipoles more sensitive to changes at large scales.

polynomials,

$$\xi_l(s) = \frac{2l+1}{2} \int_{-1}^1 d\mu \xi(s, \mu) P_l(\mu), \quad (3)$$

where  $P_l(\mu)$  is the Legendre polynomial of order  $l$ ,  $s = \sqrt{r_\sigma^2 + r_\pi^2}$ , and  $\mu = r_\pi/s$ . We use logarithmic bins in  $s$  ( $d \log_{10} s = 0.2$ ) and measure  $s$  up to  $s_{\max} = 25 \text{ Mpc } h^{-1}$ . We numerically work out the integration in equation (3) with  $d\mu = 0.05$ .

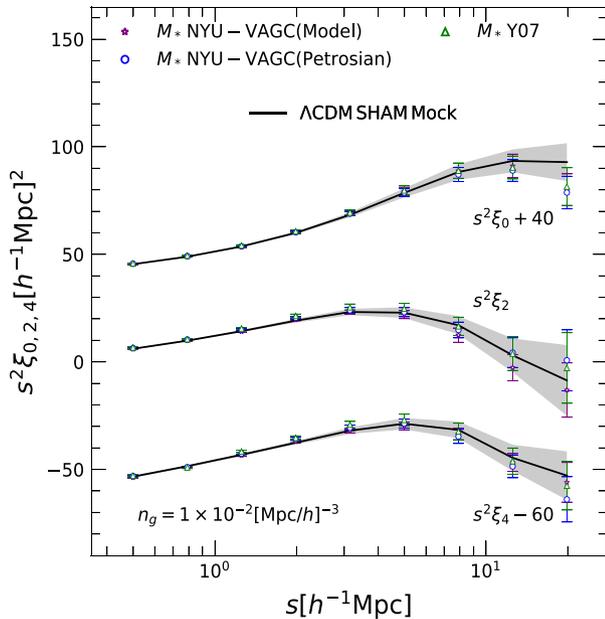
Fig. 9 shows the impact of scatter on the predicted multipoles of redshift-space distortions (RSD) of subhaloes with a number density of  $n = 0.01 [\text{Mpc } h^{-1}]^{-3}$ . The solid lines are for the results without scatter. The dashed lines are for the results with scatter. The left-hand panel of Fig. 9 shows the RSD multipoles  $\xi_{0,2,4}$  multiplied by the redshift-space separation  $s$ , which makes the RSD multipoles more sensitive to changes at small scales. The right-hand panel shows similar results but for RSD multipoles  $\xi_{0,2,4}$  multiplied by  $s^2$ . The larger index of  $s$  in this case can make the RSD multipoles more sensitive to changes at large scales. Overall, in both cases, the scatter has a very limited impact on the RSD multipoles of subhaloes. This result indeed can be expected. Given the high number density of our samples, the subhaloes that would be in and out of the catalogue due to scatter only account for a small fraction of the total samples. The clustering of subhaloes, therefore, is stable against such scatter. Furthermore, note that in the above analysis, we have adopted the parallel approximation to add the RSD effects for subhaloes in simulations.

Fig. 10 shows the predicted multipoles (monopole  $\xi_0$ , quadrupole  $\xi_2$ , and hexadecapole  $\xi_4$ ) of RSDs (black solid lines) compared to the measurements from the Sloan Digital Sky Survey (SDSS) data (symbols with error bars), which is volume limited and complete in galaxy stellar mass with a number density of  $n = 0.01 [\text{Mpc } h^{-1}]^{-3}$  (see He et al. 2018 for details). The predictions from subhaloes agree impressively well with observations even at relatively large scales

$r \sim 20 \text{ Mpc } h^{-1}$ . Note that here we use a realistic mock catalogue by collating eight replicas of the box and place the observer at the centre. This mock catalogue has the same survey mask as the real data. The subhaloes are simply selected by ranking  $V_{\text{peak}}$  from the SMDPL simulation. So there is no free parameter in our mock catalogue. In order to demonstrate the robustness of our RSD measurements, we test three different estimators of stellar masses: a template-fit method originally adopted in the New York University (NYU) catalogue with the SDSS model magnitudes (stars; Blanton et al. 2005), the same template-fit method but using SDSS Petrosian magnitudes (circles), and a single-colour method (triangles; Yang et al. 2007).

## 6 CONCLUSIONS

In this work, we have investigated the galaxy property–halo property relation using hydrodynamical simulations of galaxy formation. Unlike the conventional phenomenological frameworks such as HOD or CLF, the advantage of hydrodynamical simulations is that it can provide a clear physical picture about how a property of a galaxy relates to a property of its host dark matter halo. Based on state-of-the-art hydrodynamical simulations, such as EAGLE, Illustris, and IllustrisTNG, we study under what circumstance a galaxy’s property could correlate most tightly with a property of its host dark matter halo. In addition, in contrast to most of the previous work, our analyses include all types of galaxies, especially containing a significant fraction of satellites. Unlike the central galaxies, satellite subhaloes cannot be easily matched from the DMO simulations to the full baryonic physics hydrodynamical ones. Therefore, we adopt a novel approach that is different from the one used in Chaves-Montero et al. (2016) to analyse the simulation data. The main findings of our work are summarized as follows.



**Figure 10.** Redshift-space multipoles (monopole  $\xi_0$ , quadrupole  $\xi_2$ , and hexadecapole  $\xi_4$ ) for subhaloes (solid black lines) and the measurements from the SDSS main galaxy samples (symbols with error bars). The galaxy samples are constructed as volume limited and complete in stellar mass. The number density is chosen as  $n = 0.01 [\text{Mpc } h^{-1}]^{-3}$ . The predictions from subhaloes agree impressively well with the observations. Note that the subhaloes are simply selected by ranking  $V_{\text{peak}}$ . So there is no free parameter in our mock galaxy catalogue. In addition, we also show the observational measurements obtained using three different stellar mass estimators: a template-fit method as adopted in the NYU catalogue with the SDSS model magnitudes (stars), the same template-fit method but with SDSS Petrosian magnitudes (circles), and a single-colour method (triangles). The black shaded regions represent the  $1\sigma$  uncertainty in the theoretical predictions and the error bars are derived from 133 realizations using the jackknife resampling technique.

(i) Despite the presence of the observed mass discrepancy relation, hydrodynamical simulations do not show a consensus on the tight correlation between the total baryonic mass of a galaxy and the circular velocity of its host dark matter halo if considering all types of galaxies at the present time (post-disruption; see Fig. 6). Within  $\Lambda\text{CDM}$ , this can be expected as the gas component of a galaxy is strongly affected by various feedbacks in the processes of galaxy formation in the first place. The gas component can also have non-gravitational interactions and, therefore, is more prone to disruptions than that of the dark matter and stellar components since the latter two only have gravitational interactions among them. Therefore, such a tight correlation cannot be expected to exist in satellite galaxies.

(ii) The stellar mass of a galaxy at the epoch of  $V_{\text{peak}}$  with an evolution correction correlates most tightly to  $V_{\text{peak}}$  of the galaxy (see Fig. 4).

(iii) After accretion, star formation in most galaxies can last for a quite long time. Most galaxies, therefore, can still gain significant amounts of stellar mass (see Fig. 5). The stellar mass evolution reduces rather than increases the scatter in the current stellar mass– $V_{\text{peak}}$  relation.

(iv) Aside from the intrinsic scatter, star stripping is the main cause for the scatter in the current stellar mass (post-disruption)– $V_{\text{peak}}$  relation (see Fig. 7).

(v) Even for the current stellar mass (post-disruption)– $V_{\text{peak}}$  relation, hydrodynamical simulations still predict a very small scatter on  $\log_{10} V_{\text{peak}}$  (see Fig. 8), which means that the clustering of stellar mass selected galaxies in observations can be well modelled by  $V_{\text{peak}}$ -selected subhaloes in simulations.

(vi) Since the scatter predicted by hydrodynamical simulations has a very limited impact on the clustering of dense  $V_{\text{peak}}$ -selected subhaloes (see Fig. 9), even simple SHAM (without scatter and free parameter) can yield a robust prediction of galaxy clustering. We show that when compared with the SDSS main galaxy samples, the predictions of SHAM mock catalogue agree impressively well with the observations (see Fig. 10).

It is worth noting that, in the above conclusions, we have assumed that baryons have a limited impact on the positions and motions of subhaloes. This assumption is based on the findings reported in Chaves-Montero et al. (2016), Hellwing et al. (2016), and McAlpine et al. (2016). The authors there have compared the clustering of subhaloes that can be well matched from DMO simulations to the hydrodynamical ones with the clustering of the simulated galaxies. They find that the effect of baryons is very small on scales greater than  $> 1 \text{ Mpc } h^{-1}$ . However, an important issue here is that Chaves-Montero et al. (2016) and Hellwing et al. (2016) only focus on the matched subhaloes. But in fact a significant fraction of subhaloes in DMO runs cannot find their counterparts in the hydrodynamical ones in the first place as dark matter substructures are prone to disruptions due to numerical errors. It is unclear yet how to gauge the effect of baryons for those unmatched subhaloes. Moreover, as recently pointed out by van den Bosch & Ogiya (2018) and van den Bosch (2017), even modern state-of-the-art DMO simulations might not reliably resolve the dark matter substructures due to the insufficient force resolution in simulations. So to exactly pin down the physical effect of baryons, rather than numerical artefacts, on the clustering and motions of dark matter substructures is indeed highly non-trivial. A detailed analysis is needed, which, however, is beyond the scope of this work and will be explored in our future work.

## ACKNOWLEDGEMENTS

I acknowledge the Virgo Consortium for making their simulation data available. The EAGLE simulations were performed using the DiRAC-2 facility at Durham, managed by the ICC, and the PRACE facility Curie based in France at TGCC, CEA, Bruyères-le-Châtel. J-hH acknowledges support of Nanjing University and this work used the DiRAC at Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility ([www.dirac.ac.uk](http://www.dirac.ac.uk)). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University, and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

## REFERENCES

- Behroozi P. S., Wechsler R. H., Conroy C., 2013, *ApJ*, 770, 87  
 Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587  
 Bianchi D., Guzzo L., Branchini E., Majerotto E., de la Torre S., Marulli F., Moscardini L., Angulo R. E., 2012, *MNRAS*, 427, 2420  
 Bianchi D., Chiesa M., Guzzo L., 2015, *MNRAS*, 446, 75  
 Blanton M. R. et al., 2005, *AJ*, 129, 2562  
 Casas-Miranda R., Mo H. J., Sheth R. K., Boerner G., 2002, *MNRAS*, 333, 730

- Chaves-Montero J., Angulo R. E., Schaye J., Schaller M., Crain R. A., Furlong M., Theuns T., 2016, *MNRAS*, 460, 3100
- Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168
- Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
- Crain R. A. et al., 2015, *MNRAS*, 450, 1937
- Dawson K. S. et al., 2012, *AJ*, 145, 10
- de la Torre S., Guzzo L., 2012, *MNRAS*, 427, 327
- de la Torre S. et al., 2017, *A&A*, 608, A44
- Gao L., White S. D. M., 2007, *MNRAS*, 377, L5
- Gao L., Springel V., White S. D. M., 2005, *MNRAS*, 363, L66
- Guo Q., White S., Li C., Boylan-Kolchin M., 2010, *MNRAS*, 404, 1111
- Guo Q. et al., 2011, *MNRAS*, 413, 101
- Guo H. et al., 2015, *MNRAS*, 453, 4368
- Guo H. et al., 2016, *MNRAS*, 459, 3040
- He J.-h., Li B., Hawken A. J., 2015, *Phys. Rev. D*, 92, 103508
- He J.-h., Guzzo L., Li B., Baugh C. M., 2018, *Nat. Astron.*, 2, 967
- Hellwing W. A., Schaller M., Frenk C. S., Theuns T., Schaye J., Bower R. G., Crain R. A., 2016, *MNRAS*, 461, L11
- Hernández-Aguayo C., Hou J., Li B., Baugh C. M., Sánchez A. G., 2018, 485, 2194
- Jing Y. P., Mo H. J., Borner G., 1998, *ApJ*, 494, 1
- Klypin A., Yepes G., Gottlober S., Prada F., Hess S., 2016, *MNRAS*, 457, 4340
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlober S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Leauthaud A. et al., 2012, *ApJ*, 744, 159
- McAlpine S. et al., 2016, *Astron. Comput.*, 15, 72
- McGaugh S. S., Lelli F., Schombert J. M., 2016, *Phys. Rev. Lett.*, 117, 201101
- Marinacci F. et al., 2018, *MNRAS*, 480, 5113
- Matthee J., Schaye J., 2019, *MNRAS*, 484, 915
- Matthee J., Schaye J., Crain R. A., Schaller M., Bower R., Theuns T., 2017, *MNRAS*, 465, 2381
- Montero-Dorta A. D. et al., 2017, *ApJ*, 848, L2
- Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, *ApJ*, 710, 903
- Naiman J. P. et al., 2018, *MNRAS*, 477, 1206
- Nelson D. et al., 2015, *Astron. Comput.*, 13, 12
- Nelson D. et al., 2018a, Preprint ([arXiv:1812.05609](https://arxiv.org/abs/1812.05609))
- Nelson D. et al., 2018b, *MNRAS*, 475, 624
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*, Princeton University Press
- Pillepich A. et al., 2018, *MNRAS*, 475, 648
- Reddick R. M., Tinker J. L., Wechsler R. H., Lu Y., 2014, *ApJ*, 783, 118
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Scoccimarro R., 2004, *Phys. Rev. D*, 70, 083007
- Springel V. et al., 2018, *MNRAS*, 475, 676
- Taruya A., Nishimichi T., Saito S., 2010, *Phys. Rev. D*, 82, 063522
- Vale A., Ostriker J. P., 2004, *MNRAS*, 353, 189
- van den Bosch F. C., 2017, *MNRAS*, 468, 885
- van den Bosch F. C., Ogiya G., 2018, *MNRAS*, 475, 4066
- Vogelsberger M. et al., 2014, *MNRAS*, 444, 1518
- Wechsler R. H., Tinker J. L., 2018, *ARA&A*, 56, 435
- Yang X.-h., Mo H. J., van den Bosch F. C., 2003, *MNRAS*, 339, 1057
- Yang X., Mo H. J., Bosch F. C. v. d., Pasquali A., Li C., Barden M., 2007, *ApJ*, 671, 153
- Yang X., Mo H. J., Bosch F. C. v. d., 2009, *ApJ*, 695, 900
- Zehavi I. et al., 2011, *ApJ*, 736, 59
- Zheng Z. et al., 2005, *ApJ*, 633, 791
- Zheng Z., Coil A. L., Zehavi I., 2007, *ApJ*, 667, 760

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.