University of Hertfordshire UH

Citation for the published version:

Martin, G., Kaviraj, S., Devriendt, J. E. G., Dubois, Y., Pichon, C., & Laigle, C. (2018). Identifying the progenitors of present-day early-type galaxies in observational surveys: correcting `progenitor bias' using the Horizon-AGN simulation. Monthly Notices of the Royal Astronomical Society. DOI: 10.1093/mnras/stx3057

Document Version: Accepted Version

Link to the final published version available at the publisher:

https://doi.org/10.1093/mnras/stx3057

This article has been accepted for publication in Monthly Notices of the Royal Astronomical Society. © 2018 The Author(s). Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

General rights

Copyright© and Moral Rights for the publications made accessible on this site are retained by the individual authors and/or other copyright owners.

Please check the manuscript for details of any other licences that may have been applied and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (http://uhra.herts.ac.uk/) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Take down policy

If you believe that this document breaches copyright please contact us providing details, any such items will be temporarily removed from the repository pending investigation.

Enquiries

Please contact University of Hertfordshire Research & Scholarly Communications for any enquiries at rsc@herts.ac.uk

Identifying the progenitors of present-day early-type galaxies in observational surveys: correcting 'progenitor bias' using the Horizon-AGN simulation

G. Martin,[1⋆] S. Kaviraj,[1] J. E. G. Devriendt,[2] Y. Dubois,[3] C. Pichon,[3,4] and C. Laigle[2]
[1]Centre for Astrophysics Research, School of Physics, Astronomy and Mathematics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK
[2]Dept of Physics, University of Oxford, Keble Road, Oxford OX1 3RH UK
[3]Institut d'Astrophysique de Paris, Sorbonne Universités, UMPC Univ Paris 06 et CNRS, UMP 7095, 98 bis bd Arago, 75014 Paris, France
[4]Korea Institute of Advanced Studies (KIAS) 85 Hoegiro, Dongdaemun-gu, Seoul, 02455, Republic of Korea

arXiv:1711.06694v2 [astro-ph.GA] 6 Dec 2017

7 December 2017

ABSTRACT

As endpoints of the hierarchical mass-assembly process, the stellar populations of local early-type galaxies encode the assembly history of galaxies over cosmic time. We use Horizon-AGN, a cosmological hydrodynamical simulation, to study the merger histories of local early-type galaxies and track how the morphological mix of their progenitors evolves over time. We provide a framework for alleviating 'progenitor bias' – the bias that occurs if one uses only early-type galaxies to study the progenitor population. Early-types attain their final morphology at relatively early epochs – by $z \sim 1$, around 60 per cent of today's early-types have had their last significant merger. At all redshifts, the majority of mergers have one late-type progenitor, with late-late mergers dominating at $z > 1.5$ and early-early mergers becoming significant only at $z < 0.5$. Progenitor bias is severe at all but the lowest redshifts – e.g. at $z \sim 0.6$, less than 50 per cent of the stellar mass in today's early-types is actually in progenitors with early-type morphology, while, at $z \sim 2$, studying only early-types misses almost all (80 per cent) of the stellar mass that eventually ends up in local early-type systems. At high redshift, almost all massive late-type galaxies, regardless of their local environment or star-formation rate, are progenitors of local early-type galaxies, as are lower-mass ($M_\star < 10^{10.5}$ $M_\odot$) late types as long as they reside in high density environments. In this new era of large observational surveys (e.g. LSST, JWST), this study provides a framework for studying how today's early-type galaxies have been built up over cosmic time.

Key words: methods: numerical – galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: statistics

1 INTRODUCTION

In the standard ΛCDM paradigm, galaxy formation proceeds hierarchically. Dark matter halos, which arise as a result of primordial fluctuations in the initial matter density field (Starobinsky 1982; Guth & Pi 1982; Hawking 1982), merge to form progressively more massive haloes over cosmic time (e.g. Blumenthal et al. 1984; Kauffmann et al. 1993; Somerville & Primack 1999). Cold gas condenses into these halos where it forms rotationally-supported discs (Franx & van Dokkum 1996). The rate of star-formation is determined by the local density of this cold gas (Kennicutt 1998), with feedback from supernovae (Scannapieco et al. 2008) and active galactic nuclei (AGN) (e.g. Silk & Rees 1998; Kaviraj et al. 2017) regulating the process of stellar mass growth. A consequence of this paradigm is that the stellar mass of an individual galaxy is assembled through a combination of in-situ star-formation i.e. by gas turning into stars within a galaxy's own halo, and ex-situ star-formation i.e. stars formed in another halo which have become members of the halo in question as a result of merging (e.g. Kauffmann et al. 1993). As 'end-points' of this hierarchical assembly process, local 'early-type' (i.e. elliptical and S0) galaxies are a particularly significant class of objects (e.g. Kaviraj et al. 2007). These galaxies dominate the stellar mass density in today's Universe (e.g. Bernardi et al. 2003) and thus encode, in their stellar populations, the signatures of galaxy mass assembly over cosmic time (e.g. Worthey 1994; van Dokkum & Franx 2001). Studying these galaxies offers unique insights into the build up of the observable Universe (e.g. Barrientos & Lilly 2003; Longhetti et al. 2005; McDermid et al. 2015) and significant effort in the literature has, therefore, been dedicated to understanding these systems.

Galaxies in the early universe form with disc-like (late-type) morphologies and, through interactions and secular pro-

⋆ E-mail: g.martin4@herts.ac.uk

© 2017 The Authors

cesses, acquire more spheroidal (early-type) morphologies over time (e.g. Franx & van Dokkum 1996; van Dokkum & Franx 2001; Buitrago et al. 2014; Conselice 2014; Kaviraj 2014b). Consequently, at progressively higher redshift, the progenitors of today's early-types become increasingly dominated by late-type systems. Understanding the formation and evolution of today's early-types therefore requires us to consider their entire progenitor population, especially their late-type progenitors at earlier epochs. If, as is often assumed, early-type galaxies do not revert back to late-types (e.g. Hau et al. 2008; Fang et al. 2012)[1], techniques that identify *late-type* galaxies that are progenitors of today's early-types in observational surveys are essential. This becomes particularly important at high redshift, e.g. $z > 2$, the redshift regime at which the early-type population is rapidly assembled (e.g. Conselice 2014), and which can be routinely accessed by forthcoming facilities such as JWST (Gardner et al. 2006), EUCLID (Laureijs et al. 2011), etc.

Past observational studies (e.g. Gladders et al. 1998; Stanford et al. 1998) have attempted to trace the assembly of present-day early-types by focussing only on the population of early-type galaxies at high-redshift. Since the stellar populations of present-day early-types are largely in place in the early Universe (e.g. Trager et al. 2000; Thomas et al. 2005; Kaviraj et al. 2011; Kaviraj 2014a), this is a reasonable approximation at low and intermediate redshift ($z \sim 1$), assuming that galaxies, once they achieve early-type morphology, cannot revert back to being late-type systems. However, as noted above, at earlier epochs, an increasing proportion of the mass in present-day early-types is contained in late-type progenitors. Thus, considering only early-type galaxies introduces a bias (the so-called 'progenitor bias') in any study of their evolution, which becomes progressively more severe with increasing redshift.

In a similar vein, other observational work (e.g. Bell et al. 2004) has often used the optical red sequence (Faber et al. 2007; Barro et al. 2013) as a proxy for the population of progenitors of early-type galaxies. However, since the stellar populations in late-type galaxies tend to be younger (and therefore bluer), such a colour cut misses the majority of the late-type galaxies that are progenitors of local early-types. In addition, since a wide variety of star-formation histories are observed in early-type galaxies themselves, particularly at high redshift (Kaviraj et al. 2013; Fitzpatrick & Graves 2015; Lofthouse et al. 2017a), a large fraction of blue galaxies that already have early-type morphology will also be missed (Shankar et al. 2015), if such a colour cut is employed. While the red sequence traces the progenitors of early-type galaxies well at the highest end of the luminosity function (Kaviraj et al. 2009), it becomes less reliable at low masses (where galaxies are bluer) and fails to identify early and late-type progenitors (of all stellar masses) that lie blueward of the red sequence. In a general sense, therefore, a simple colour cut is not a reliable approximation for the progenitor population of today's early-type galaxies. And, in a similar vein to using an early-type selection to identify the progenitor population, the red-sequence approximation becomes progressively less effective with increasing redshift.

Progenitor bias is difficult to overcome observationally, since individual galaxies cannot be observed as they evolve and, therefore, in any given survey, it is difficult to directly identify objects that will end up in early-type galaxies at $z \sim 0$. Nevertheless, observational methods have been applied to reduce or mitigate this issue. For example, a widely-used technique, proposed by van Dokkum et al. (2010) assumes that cumulative co-moving number density is conserved as the galaxy mass function evolves with time (i.e. as a galaxy evolves, the number of galaxies more massive than it remains constant), so that a given galaxy maintains the same rank. However, while this method is able to account for the mass evolution of the galaxy population, the assumptions made may be too simplistic. For example, it ignores the fact that a galaxy's rank may change over time, resulting in the evolution of its co-moving number density. This may occur if, for instance, galaxies are removed from the population as a result of mergers (Ownsworth et al. 2014), or as a result of the spread in specific star-formation rate (sSFR) and its dependence on mass (Leja et al. 2013; Shankar et al. 2015).

Improvements on the method of van Dokkum et al. (2010) which include prescriptions for the number density evolution of galaxies have been proposed. For example, Behroozi et al. (2013) uses abundance matching in order to match observed galaxies to corresponding dark matter halos in ΛCDM simulations, allowing the median and dispersion of the cumulative co-moving number density tracks to be quantified for the progenitors of galaxies of a given mass. Work by Torrey et al. (2015, 2017) introduces an analytic framework, which accounts for the effects of the merger rate ('coagulation') and stochasticity of galaxy growth rates ('scatter rate'), and describes the median (Torrey et al. 2015) and intrinsic scatter (Torrey et al. 2017) of the evolution of the galaxy population in co-moving density space. Wellons & Torrey (2017) present a probabilistic method based on this framework, which they show is able to more effectively predict progenitor properties than the methods of van Dokkum et al. (2010) and Behroozi et al. (2013).

While this group of methods is widely applicable to any given galaxy property (e.g. Torrey et al. 2015; Clauwens et al. 2016), they may lose predictive power in cases where there is no expectation that the rank of the property of interest will be conserved (e.g. morphology). In such cases, leveraging the constraining power of additional galaxy properties becomes essential. Additionally, all of these methods retain a weakness of the van Dokkum et al. (2010) method, in that they still assume that early-type progenitors follow the same distribution of properties as the general population at a given redshift.

An appealing alternative is to employ a simulation that reproduces the properties of galaxies over cosmic time. Since the identities of the progenitors of local early-type galaxies are precisely known in the model, they can be used to calculate the probability that a galaxy is a progenitor, as a function of its observable properties (e.g. redshift, stellar mass, star-formation rate (SFR), local environment). Kaviraj et al. (2009) have previously addressed the problem of progenitor bias using the GalICS semi-analytical model (Hatton et al. 2003). However, while the semi-analytical approach has successfully reproduced the phenomenology of many aspects of the galaxy formation process (e.g. Somerville & Primack 1999; Cole et al. 2000; Benson et al. 2003; Bower et al. 2006; Croton et al. 2006), the recent advent of hydrodynamical simulations in cosmological volumes provides a more accurate route to addressing the problem. Unlike their semi-analytical counterparts, hydrodynamical models resolve the gas and baryonic content of galaxies, typically on kpc scales, (e.g. Devriendt et al. 2010; Dubois et al. 2014; Vogelsberger et al. 2014; Schaye et al. 2015; Khandai et al. 2015; Taylor & Kobayashi 2016; Kaviraj et al. 2017). This enables them to more accurately model a greater range of physical processes, without the need for semi-analytical recipes (Schaye et al. 2015), although some processes,

[1] Dubois et al. (2016) have shown that transformation from early-type to late-type morphology is prevented by AGN feedback.

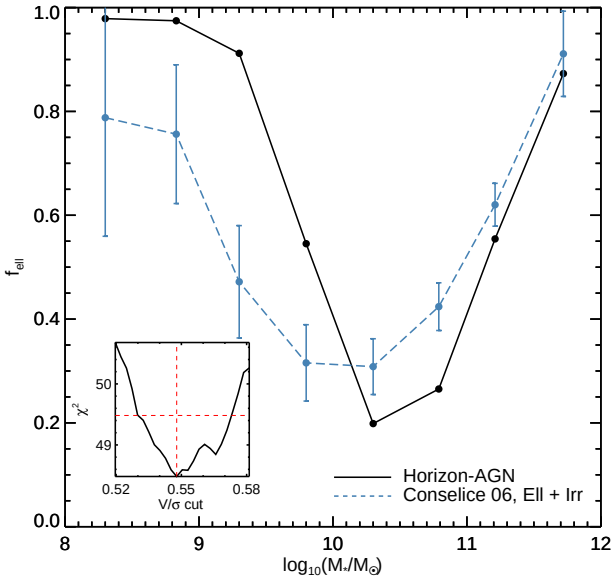

Figure 1. The predicted early-type fraction in Horizon-AGN, for the $^{V}/_{\sigma}$ value ($0.55\pm^{0.03}_{0.02}$) that best reproduces the observed early-type fractions in the local Universe (Conselice 2006). The inset shows the value of χ^2 between the predicted and observed early-type fractions for different $^{V}/_{\sigma}$ thresholds (the red dashed lines show the minimum χ^2 and $\Delta\chi^2 = 1$).

such as AGN feedback, that cannot be resolved, must still be described using sub-grid models (e.g. Katz 1992; Booth & Schaye 2009; Kimm et al. 2015). Such hydrodynamical simulations typically rely on a smaller number of free parameters and offer a more realistic treatment of the physical processes involved in the formation and evolution of galaxies, yielding better agreement with the observed Universe without the need for tuning.

In this study, we use the Horizon-AGN[2] cosmological simulation (Dubois et al. 2014; Kaviraj et al. 2017) to (1) quantify the evolution of the progenitor population of today's early-type galaxies and (2) provide a route for identifying late-type galaxies that are progenitors of present-day early-types in observational surveys, by estimating the probability of a given late-type to be the progenitor of a local early-type system, as a function of measurable observables like redshift, stellar mass, star-formation rate and local density.

The structure of this paper is as follows. In Section 2, we describe the main characteristics of the simulation, describe our simulated galaxy sample and define how observables are measured. In Section 3, we probe the redshift evolution of the progenitors of present-day early-type galaxies. In Section 4, we present probabilistic prescriptions to identify late-type galaxies that are progenitors of local early-types as a function of redshift, stellar mass, environment and star-formation rate. We summarize our findings in Section 5.

[2] http://www.horizon-simulation.org

2 THE SIMULATION

We begin by briefly describing the Horizon-AGN simulation, the extraction of galaxies and merger trees and the prediction of observable quantities.

2.1 Horizon-AGN

Horizon-AGN is a cosmological hydrodynamical simulation (Dubois et al. 2014) that employs the adaptive mesh refinement Eulerian hydrodynamics code, RAMSES (Teyssier 2002). The size of the simulation box is 100 h^{-1} coMpc, which contains 1024^3 dark matter particles and uses initial conditions from a *WMAP7* ΛCDM cosmology (Komatsu et al. 2011). The simulation has a dark matter mass resolution of 8×10^7 $M_\odot$, a stellar-mass resolution of 2×10^6 $M_\odot$ and a spatial resolution of ∼1 kpc. We direct readers to Kaviraj et al. (2017) for details of the recipes (e.g. star-formation and stellar and AGN feedback) employed to model the baryonic evolution of galaxies. Briefly, star-formation follows a standard Schmidt-Kennicutt law (Kennicutt 1998), with the model implementing continuous stellar feedback, that includes momentum, mechanical energy and metals from stellar winds, Type II SNe and Type Ia SNe. Black-hole (BH) feedback on ambient gas operates via two separate channels, depending on the gas accretion rate. For Eddington ratios > 0.01 (high accretion rates), 1.5 per cent of the accretion energy is injected as thermal energy (a quasar-like feedback mode), whilst for Eddington ratios < 0.01 (low accretion rates), bipolar jets are employed with a 10 per cent efficiency. The parameters are chosen to produce agreement with the local cosmic black-hole mass density, and the $M_{BH} - M_\star$ and $M_{BH} - \sigma_\star$ relations (Dubois et al. 2012). Apart from choosing the BH-feedback parameters to match the $M_{BH} - M_\star$ and $M_{BH} - \sigma_\star$ relations at $z = 0$, Horizon-AGN is not otherwise tuned to reproduce the bulk properties of galaxies at $z \sim 0$. As described in Kaviraj et al. (2017), the simulation reproduces key quantities that trace the aggregate stellar mass growth of galaxies: stellar mass and luminosity functions, rest-frame UV-optical-near infrared colours, the star-formation main sequence and the cosmic star-formation history.

2.2 Identifying galaxies and building merger trees

In order to track their progenitors, we build merger histories for each early-type galaxy in the final snapshot of the simulation ($z = 0.06$). In the sections below, we describe the process of galaxy identification, followed by the process of building merger trees.

2.2.1 *Identifying the galaxy sample*

For each snapshot, we produce a catalogue of galaxies, using the ADAPTAHOP structure finder (Aubert et al. 2004; Tweed et al. 2009), operating directly on the star particles. Galactic structures are selected using a local threshold of 178 times the average matter density, where the local density of individual particles is calculated using the 20 nearest neighbours. Only galactic structures with more than 100 star particles are considered, corresponding to a minimum galaxy stellar mass of $M_\star \sim 2\times10^8$ $M_\odot$, which is a consequence of a minimum star particle mass of $m_\star \sim 2 \times 10^6$ $M_\odot$. We identify an average of ∼150,000 galaxies above the 100 particle threshold in each snapshot. We restrict our study to galaxies with stellar mass $M_\star > 10^{9.5}$ $M_\odot$, for reasons outlined in Section 2.2.2.

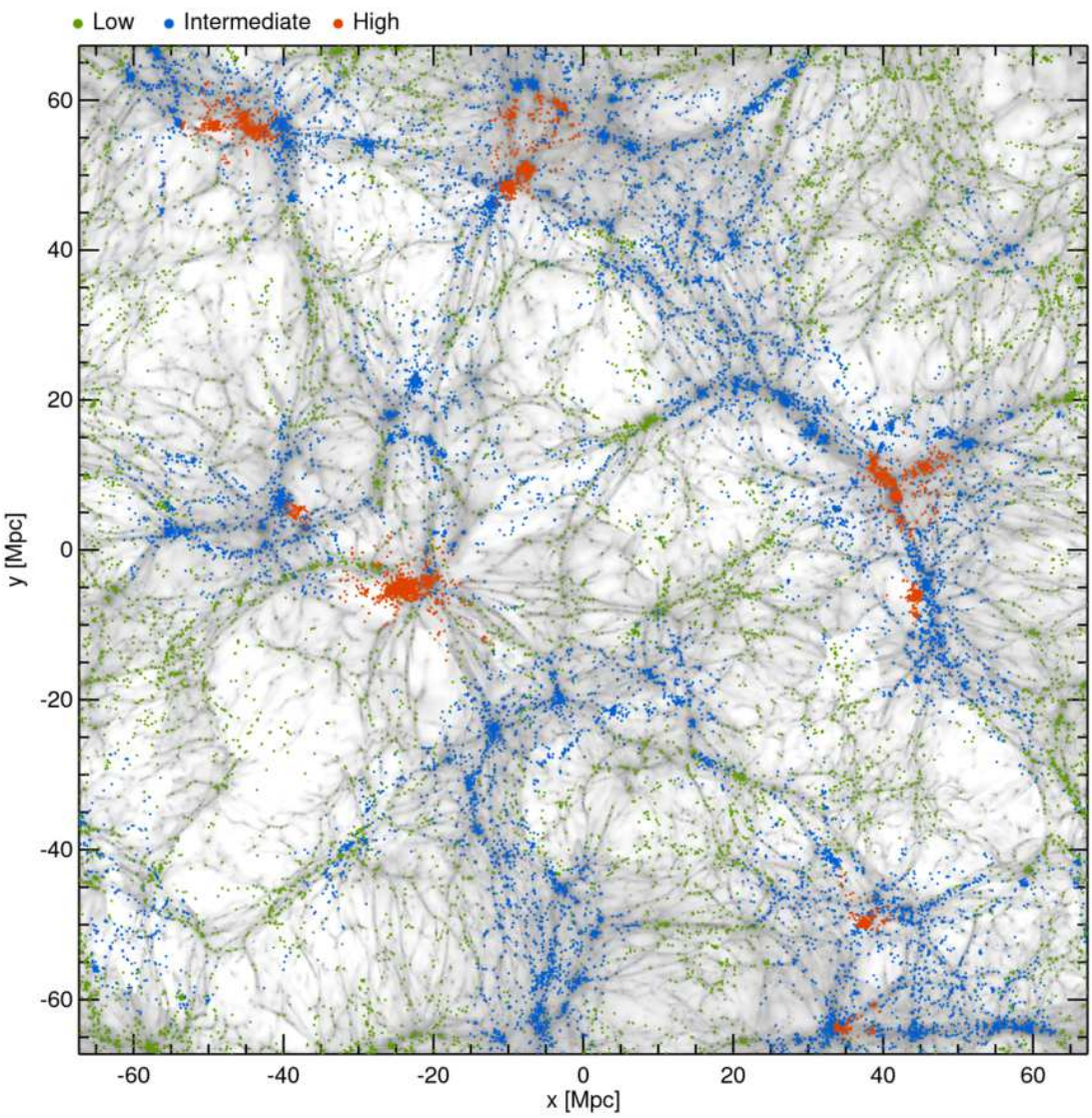

Figure 2. Greyscale map showing the projected gas density in a ~ 20 Mpc slice through the simulation volume at the final time output ($z = 0.06$). Green, blue and red points correspond to the positions of galaxies ($M_\star > 10^{9.5}$ $M_\odot$) in low, intermediate and high density environments respectively (see text in Section 2.3.2 for more details). Axes are in units of proper Mpc.

2.2.2 *Producing merger histories*

Using the catalogue of galaxies identified by ADAPTAHOP, we extract merger histories for each early-type galaxy at the final snapshot. We produce merger trees using 91 snapshots in the range $z \in [0.06, 7.04]$, with an average time-step of $\sim$130 Myr. Merger trees are produced for each early-type, by identifying their progenitors at each snapshot, using the method described in Tweed et al. (2009). Since our threshold for identifying structures is 100 star particles, only mergers where the satellite galaxy has $M_\star \gtrsim 2 \times 10^8$ $M_\odot$ are considered, regardless of mass ratio. Given that our sample excludes galaxies less massive than $M_\star \sim 10^{9.5}$ $M_\odot$, and the minimum galaxy mass identified is $M_\star \sim 2 \times 10^8$ $M_\odot$, our sample is complete for mergers down to a mass ratio of at least 1:15.

2.3 Prediction of observables

We produce observables that can be used in conjunction with contemporary and future observational datasets. These are stellar mass (derived using the total mass of the star particles in a galaxy), star-formation rate, local number density and stellar kinematics (which we use as a proxy for morphology). The following sections describe how we derive each of these measures.

2.3.1 *Morphology*

The morphology of each model galaxy in our analysis is inferred using its stellar kinematics. Morphology is defined using $^V/_\sigma$, which is the ratio of the mean rotational velocity (V) to the mean velocity dispersion (σ), both measured using the entire star particle distribution. Higher values of this ratio correspond to more late-type (disc-like) morphologies. $^V/_\sigma$ is calculated by first rotating the coordinate system so that the z-axis is oriented along the angular momentum vector of the galaxy. Rotational velocity is defined as the mean tangential velocity component in cylindrical co-ordinates, V_θ, and velocity dispersion is computed using the standard deviations of the radial, tangential and vertical star particle velocities, σ_r, σ_θ and σ_z, summed in quadrature. $^V/_\sigma$ is then given by

$$^V/_\sigma = \frac{\sqrt{3}\bar{V}_\theta}{\sqrt{\sigma_r^2+\sigma_\theta^2+\sigma_z^2}}. \quad (1)$$

To separate galaxies morphologically into early and late types using $^V/_\sigma$, we consider a range of values for $^V/_\sigma$ and compare how the resulting predicted early-type fractions compare to their observed counterparts at low redshift. The value of $^V/_\sigma$ which produces the best agreement with the observational data is then selected as the threshold value that separates early-types and late-types in the model. Figure 1 shows the predicted early-type fractions in Horizon-AGN for this best-fitting $^V/_\sigma$ value, $0.55\pm^{0.03}_{0.02}$, compared with early-type fractions derived from observations (Conselice 2006). The largest discrepancy between the observations and the simulation occurs at the low mass end, and is likely a result of insufficient mass resolution (Dubois et al. 2016). Nevertheless, over the mass range considered in this study ($M_\star > 10^{9.5}$ $M_\odot$), the early-type fractions predicted by Horizon-AGN, for our $^V/_\sigma$ threshold of 0.55, are in reasonable agreement with the observations. We note that the progenitor fractions presented in Section 4 are resistant even to relatively large changes in our $^V/_\sigma$ threshold. Varying the $^V/_\sigma$ threshold by as much as 50 per cent introduces only a ± 0.05 variation in our calculated progenitor fractions.

We note that the minimum refinement of the AMR grid is increased at z =4 ,1.5 and 0.25, in order to keep the minimum physical cell size approximately constant (Dubois et al. 2014; Peirani et al. 2016). While it is possible that this refinement may result in the production of sudden instabilities in previously stable discs that increase galaxy velocity dispersions, the smooth nature of the $^V/_\sigma$ evolution of the galaxy population (Dubois et al. 2014), indicates that this is not a significant effect.

2.3.2 *Local environment*

We compute the local environment of each galaxy using an estimate of the 3-D local number density. For each galaxy, we estimate the surrounding number density of galaxies above our mass cut ($10^{9.5}$ $M_\odot$). This is achieved using the adaptive kernel density estimation method of Breiman et al. (1977) which utilises a finite-support Epanechnikov kernel (instead of the more typically-used infinite-support Gaussian kernel) and using the density estimator itself as a pilot to steer the local kernel size (Wilkinson & Meijer 1995; Ferdosi et al. 2011). Galaxies at our final snapshot ($z = 0.06$) are classified into 'high', 'intermediate' and 'low' density environments, by comparing the density percentile they occupy to the corresponding percentiles that observed galaxies inhabiting cluster, group and field environments occupy in the low-redshift universe.

Observations indicate that around 10 per cent of galaxies occupy rich clusters, while around 50 per cent of galaxies occupy groups and poor clusters (e.g Bahcall 1996; Dekel & Ostriker 1999; Tempel et al. 2012), with the remaining 40 per cent of galaxies occupying the field. We use these values as a guide to make density cuts, in order to separate our simulated galaxies into our three density bins (high, intermediate and low), that are roughly analogous to cluster, group and field environments. To allocate our simulated galaxies into these density bins, we first rank the objects by density. Then, the 10 per cent of galaxies with the highest local density (i.e. the $90^{\rm th} - 100^{\rm th}$ percentile range) are classified as occupying high density environments, galaxies in the $40^{\rm th} - 90^{\rm th}$ range are classified as occupying intermediate density and the remaining galaxies are classified as occupying low-density environments. Figure 2 shows the environment classifications of galaxies with $M_\star > 10^{9.5}$ $M_\odot$ superimposed over a map of the gas density. Not unexpectedly, galaxies classified as being in high density environments lie at the nodes of the cosmic web, while intermediate density galaxies lie largely along filaments, with low-density galaxies typically being found in voids. Regions of high gas density also host a high number density of galaxies which, as we show later in Section 3, leads to more rapid morphological transformation. More massive galaxies (those above our detection threshold of 2×10^8 $M_\odot$) first appear in the simulation around nodes, owing to a greater abundance of gas from which to accrete, and also undergo a higher incidence of interactions or mergers. These galaxies are thus likely to experience earlier and accelerated evolution compared to galaxies in less dense environments.

In all of our analysis below, we always cast the local environment in terms of the density percentile of individual galaxies. This is because observers (and theorists) inevitably use different metrics for measuring local density. However, while the absolute values of density depends on the actual metric being used, the density percentile that a galaxy occupies is likely to be roughly independent of the actual estimation method (as we demonstrate later in Section 4).

2.3.3 *Star-formation rate*

Star-formation rates are calculated by computing the change in stellar mass, $m_\star$, of the galaxy in question between two snapshots, dividing by the time difference between those snapshots, and subtracting the mass of stars formed *ex-situ* that has merged with the galaxy between the two snapshots, $m_{\star,merged}$:

$$\mathrm{SFR} = \frac{m_{\star,t=t_i} - m_{\star,t=t_{i-1}} - m_{\star,merged}}{\Delta t}, \quad (2)$$

where t_i is the time at the current snapshot and t_{i-1} is the time at the previous snapshot. Δt is equal to $t_i - t_{i-1}$ and, as discussed in Section 2.2.2, the average time-step used in the simulation is $\sim$130 Myr. We note that the SFR in not sensitive to the exact value of Δt. For example, if we double or halve the value Δt, in either case, our calculated SFRs change by less than 30 per cent.

3 REDSHIFT EVOLUTION OF THE PROGENITORS OF LOCAL EARLY-TYPE GALAXIES

We begin our analysis by performing a broad exploration of the redshift evolution of the progenitors of local early-type galaxies in the simulation. The left-hand panel of Figure 3 presents density plots

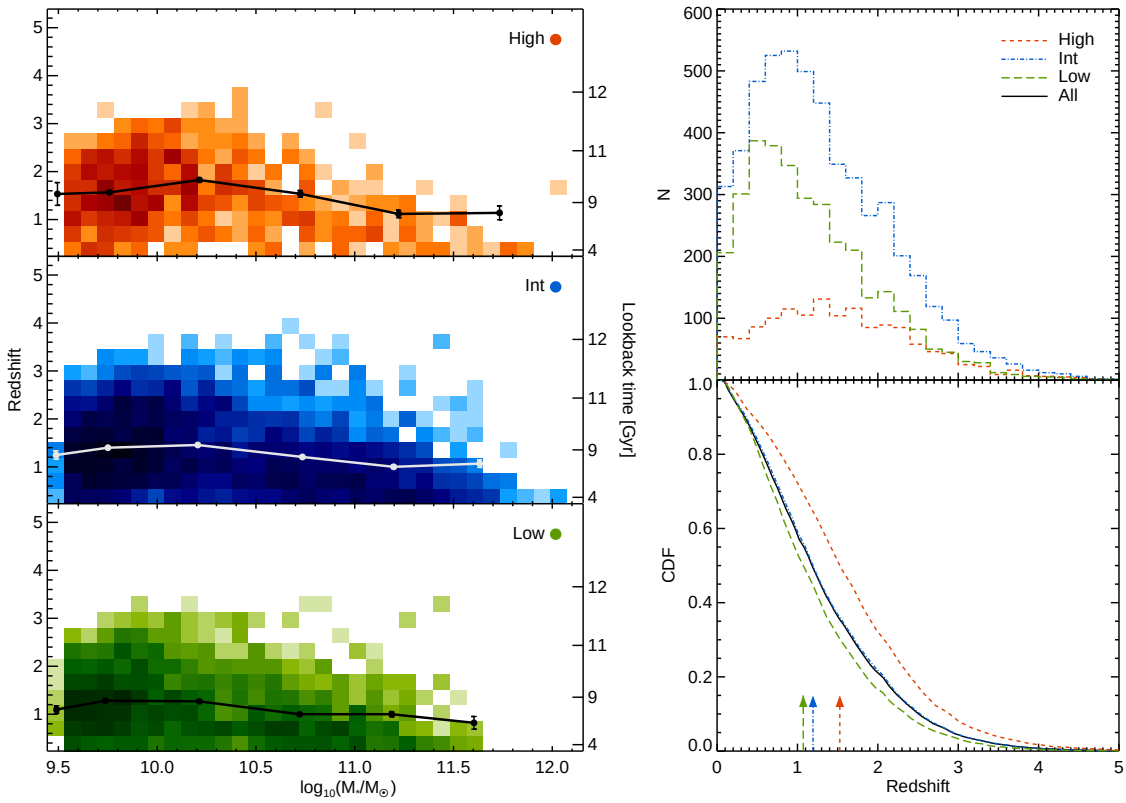

Figure 3. Left: Density plots showing the redshift at which local early-type galaxies in different environments had their last significant merger (i.e. a merger with a mass ratio greater than 1:10). Green, blue and red colours correspond to low, intermediate and high density environments. Over-plotted are lines showing the mean redshift in each bin and its error. Right: Histograms (top) and associated cumulative distribution functions (bottom) of last merger redshifts. The arrows in the bottom panel indicate the redshift where the cumulative distribution functions reach values of 0.5 (i.e the median value) for each environment.

showing the last-merger redshift of each local early-type galaxy. We define this as a galaxy's last significant merger, i.e. a merger which has a mass ratio greater than 1:10. The right-hand panels of Figure 3 show histograms (top-right) and their associated cumulative density functions (bottom-right) as a function of redshift for each environment. The arrows in the bottom-right panel indicate the redshift where the cumulative distribution functions reach values of 0.5 (i.e. the median value) for each environment.

Local early-type galaxies that inhabit denser environments tend to have higher *dynamical ages* i.e. their last significant mergers take place at earlier epochs. While 50 per cent of galaxies in the high density bin have their last significant merger by $z = 1.5$, this is only the case for galaxies in the lowest density bin at $z = 1.1$ (as indicated by the coloured arrows in the cumulative density function plot (bottom-right panel in Figure 3)). Local early-types in higher-density environments also tend to have higher final masses indicating that the bulk of their evolution takes place earlier and is more rapid, although within a specific environment (high, intermediate or low), the most massive galaxies finish assembling at later epochs (see also De Lucia & Blaizot 2007; Dubois et al. 2016).

In Figure 4, we quantify the extent of progenitor bias at various redshifts, as a function of galaxy mass and local environment. We show both the fraction of progenitor galaxies that have already acquired early-type morphology (top panel) and the mass fraction in the progenitor population that is contained in progenitors with early-type morphology (bottom panel). We find that, across all environments, only ~ 50 per cent of the progenitors have acquired early-type morphology by $z \sim 0.6$. This is also true of the mass fraction contained in these early-type progenitors i.e. at $z \sim 0.6$ only half of the stellar mass that eventually ends up in early-type galaxies today is contained in progenitors that have early type morphology. In other words, looking *only* at early-type systems to trace the evolution of today's early-types would miss half of the progenitor population at $z \sim 0.6$.

Since morphological transformation is more rapid in regions of higher density, progenitor bias is less severe. Thus, 50 per cent of the progenitors of local early-types in high-density environments (i.e. today's clusters) have already acquired early-type morphology by $z = 0.7$ (compared to $z = 0.6$ across all environments). Note, however, that the bias remains reasonably high regardless of environment. Very similar trends are seen when quantifying progenitor bias as a function of stellar mass (not shown in Figure 4), with more massive galaxies ($M_\star > 10^{11.5}$ $M_\odot$) following the same trend as the 'high' density environment in Figure 4. This is simply because the most massive galaxies occur overwhelmingly in dense environments.

Finally, we note that the early-type fraction appears to decrease (somewhat counter-intuitively) from $z = 5$ and begins to in-

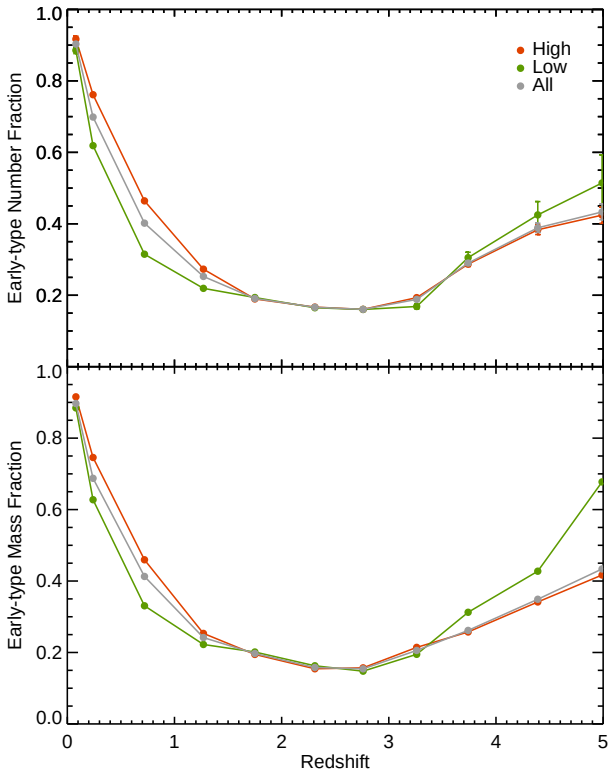

Figure 4. Top: The number fraction of progenitors that already have early-type morphology, split by the local environment of the early-type galaxy that the progenitor ends up in at the present day. Bottom: The mass fraction contained in progenitors that have early-type morphology, split by the environment of the early-type galaxy that the progenitor ends up in at the present day. Poisson error bars are shown (note that most lie within the data points).

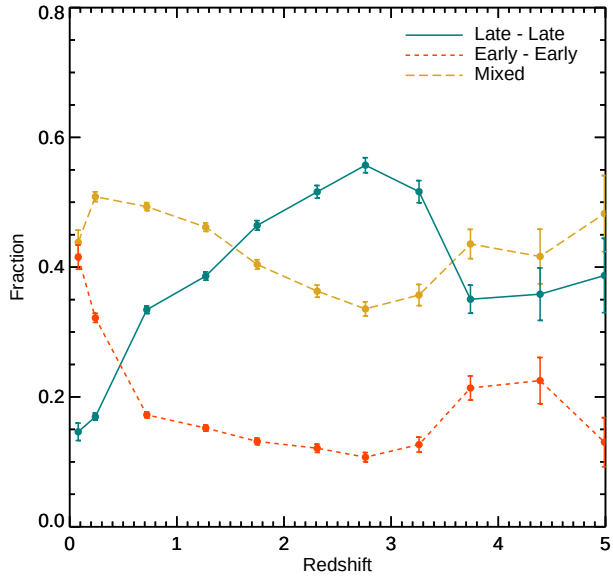

Figure 5. The morphological composition of binary mergers within the progenitor population, that have mass ratios greater than 1:10, as a function of redshift. 'Late-late' indicates mergers between two late-type galaxies, 'mixed' indicates mergers between one early-type and one late-type galaxy and 'early-early' indicates a merger between two early-type galaxies. Poisson errors are shown.

crease again around $z = 3$. This is not an artefact of our 2×10^8 $M_\odot$ detection threshold, which might cause the most rapidly evolving (and more massive) galaxies to be detected first, potentially biasing our result. If we limit our study to narrow mass bins or follow only the evolution of galaxies that are detected by the structure finder at $z \sim 5$, we observe the same non-monotonic evolution. This is partially the result of generally more clumpy star formation and more disturbed morphologies at high redshift (e.g. Ceverino et al. 2010), and is consistent with observational work (e.g. Kassin et al. 2012, 2014), which has shown that star-forming galaxies steadily settle into flat, rotationally supported discs, through the process of 'disc settling' at these epochs. This kinematic settling is driven by the fact that many processes which are able to 'puff up' (i.e. increase the dispersional motion) of gas in the disc without disturbing stellar orbits significantly, become less frequent and/or less intense with time - examples of this include (minor) mergers and gas accretion episodes (e.g. Covington et al. 2010; Bournaud et al. 2011; Lofthouse et al. 2017b; Martin et al. 2017), strong stellar feedback as a result of the high star-formation rates at high redshift (e.g. Silk & Norman 2009) and the high gas fractions at early epochs that lead to increased disc instability. In essence, the more gentle evolution that galaxies undergo at later times is thought to allow the gas in the disc to settle into a more ordered state (Kassin et al. 2012). Star formation then proceeds primarily in a planar disc, gradually reducing the mean $^V/\sigma$ as more stars form. Indeed, for Horizon-AGN galaxies that still host a significant disc at $z = 0.06$, we find that old stars (those formed before $z = 3$) are more likely to be found in orbits outside of the plane of the disc, symptomatic of the fact that, at these early epochs, gas fractions and merger rates were typically higher on average. Disc settling has also been observed in other simulations (e.g. Kassin et al. 2014; Ceverino et al. 2017). In the case of Horizon-AGN, this effect may be compounded slightly by changes to the maximum refinement of the gas grid, which, as mentioned in Section 2.3.1 is increased towards lower redshifts in order to keep the minimum physical cell size approximately constant. This can have the effect of artificially thickening discs of some galaxies, although we still observe non-monotonic evolution regardless of refinement level.

We conclude this section by exploring the morphologies of the progenitors of early-type galaxies that are involved in mergers. We focus only on binary mergers because, although non-binary mergers do occur, they are rare (around two orders of magnitude less frequent than binary mergers). In Figure 5, we show the fraction of mergers at a given redshift that involve two late-type galaxies ('late-late'), one late-type and one early-type galaxy ('mixed') and two early-type galaxies ('early-early'). Mergers between two late-type galaxies dominate in the early Universe i.e. around the epoch of peak cosmic star-formation and beyond ($z > 1.5$). The fraction of mergers involving two early-type galaxies climbs rapidly in the low redshift Universe ($z < 0.5$). However, at all redshifts, the majority of mergers involve at least one late-type galaxy.

4 IDENTIFYING LATE-TYPE GALAXIES THAT ARE PROGENITORS OF PRESENT DAY EARLY-TYPES

We proceed by constructing probabilistic prescriptions for identifying late-type progenitors of local early-type galaxies in observational surveys, as a function of quantities that are measurable in today's datasets: redshift, stellar mass, local environment and star-formation rate. As noted in the introduction, the overall aim is to provide a means for correcting progenitor bias in observational studies, by allowing for the inclusion of late-type progenitors of today's early-type systems. We do this by calculating the fraction of late-type galaxies that are progenitors of local early-types, as a function of the measurable quantities mentioned above. This fraction can then be thought of as a probability that a galaxy with the given properties is the progenitor of an early-type galaxy at present day. Observers who wish to include late-type progenitors of early-type galaxies can then use these probabilities to 'weight' objects in observational surveys, thus enabling them to reduce progenitor bias by including, in a probabilistic sense, the late-type members of the progenitor population. These probabilistic prescriptions are likely to be particularly useful in the new era of deep-wide surveys (e.g. DES (Abbott et al. 2005), EUCLID (Laureijs et al. 2011), LSST (Abell et al. 2009), JWST (Gardner et al. 2006) etc.) which will routinely offer large datasets that probe the early Universe, where progenitor bias becomes most severe, and simplifying assumptions, such as using only early-type galaxies or the red sequence to trace the progenitor population, break down. In what follows, we first explore 1-D progenitor probabilities as a function of stellar mass (split by local environment) and then 2-D probabilities as a function of mass and environment and mass and star-formation rate.

Since we are interested in probing progenitor probabilities as a function of local environment, we calculate, at each redshift of interest, the 3-D local number density using the method described in Section 2.3.2. As in Section 3, we consider galaxies in the $90^{\rm th} - 100^{\rm th}$ percentile range to be inhabiting high density environments, those in the $40^{\rm th} - 90^{\rm th}$ percentiles to be inhabiting intermediate-density environments and those in the $0^{\rm th} - 40^{\rm th}$ percentile range to be inhabiting low density environments. As noted before, the density percentile in which a galaxy lies (which is driven by its rank in density) is likely to be reasonably resistant to the exact method used for the density estimation.

To check this, we compare two different density estimation methods. These are the adaptive kernel density estimator used in Section 2.3.2 and the $k^{\rm th}$ nearest neighbours density estimator, that is commonly used in many observational studies (e.g. Dressler 1980; Baldry et al. 2006; Ferdosi et al. 2011; Shattow et al. 2013). Note that for consistency with the adaptive kernel method, we choose a definition for the $k^{\rm th}$ nearest neighbour algorithm whereby each galaxy is considered to be its own neighbour. Specifically, we choose, $k = 6$, which is almost equivalent[3] to the commonly used case where $k = 5$ and each galaxy is not considered to be its own neighbour (e.g. Baldry et al. 2006). We test the two methods on a 7 Mpc (proper) slice through the simulation snapshot at $z = 0.5$, which corresponds to a difference in velocity due to the Hubble flow of $\Delta V = 500$ km s^{-1}, and implies a requisite precision in redshift of $\Delta z = 0.002$. Such precision will be achievable at intermediate and high redshift using spectroscopic and grism redshifts from future instruments, such as MOONS (Cirasuolo et al. 2011), PFS (Takada et al. 2014), 4MOST (de Jong et al. 2012) and JWST (Gardner et al. 2006). We explore estimates of both the 2-D surface density and the 3-D density. We find that the rank of each galaxy indeed remains approximately constant, regardless of either the exact estimator used, or whether we consider the 2-D or 3-D densities. Typically, the rank of a galaxy does not change by more than 10 per cent and, therefore, changing the density estimator leaves our conclusions unchanged.

Figure 6 shows the fraction of late-type galaxies at a given redshift that are the progenitors of a local early-type, as a function of stellar mass and split by local environment. We show the 2-D progenitor probability as a function of both stellar mass and local density in Figure 7, with the colour bar indicating the progenitor probabilities. At all redshifts, there is a positive trend of progenitor probability with stellar mass i.e. more massive late-type galaxies are more likely to be progenitors of local early-type remnants.

At high redshifts, almost all massive galaxies, regardless of their local environment, are progenitors of present-day early-types. While the progenitor probabilities increase with redshift, for the most massive galaxies the progenitor probability remains close to ~ 1 until $z \sim 0.5$. The principal reason for an increase in the progenitor fraction with redshift is simply the fact that late-type galaxies have more time to merge with other galaxies and undergo morphological transformation before the present day. The rate of morphological transformations is regulated by the merger rate per galaxy, which rises with redshift (e.g. Welker et al. 2015; Rodriguez-Gomez et al. 2015; Kaviraj et al. 2015) and thus controls the rate of change in the progenitor fraction as a function of redshift. Lower-mass ($M_\star < 10^{10.5}$ $M_\odot$) galaxies can also exhibit high progenitor probabilities at high redshift, but *only* if they occupy regions of high density (e.g. the 80-100$^{\rm th}$ density percentile, see Figure 7). Note that the progenitor probabilities decline for all galaxies towards low redshifts, because these systems will not have had time to undergo enough merging to achieve early-type morphology.

We proceed, in Figure 8, by exploring the progenitor probabilities as a function of stellar mass and star-formation rate (the parameter space that is commonly referred to as the 'star-formation main sequence'). The colour bar indicates the progenitor probabilities. Mirroring the trends found earlier, massive late-type galaxies are more likely to be progenitors of early-types. At high redshift, these massive late-type progenitors can be some of the most highly star-forming systems in the Universe. However, this is simply a consequence of star-formation activity being, on average, more vigorous in the early Universe. It is worth noting, however, that at all redshifts, progenitor fractions increase *at fixed stellar mass* as the star-formation rate decreases. In other words, at fixed stellar mass, late-type galaxies with lower specific star-formation rates are more likely to be progenitors. For example, at $z \sim 2$ (top-right hand panel of Figure 8), a late-type galaxy with a stellar mass of $10^{10.7}$ $M_\odot$, which resides at the upper end of the star-formation main sequence, has a progenitor probability of around 70 per cent. A galaxy with a similar mass which sits at the bottom of the star-formation main sequence has a progenitor probability of close to 100 per cent.

We release the progenitor probabilities calculated here, as a function of different observables, as described in the Appendix A. For a series of redshifts, we provide tables of progenitor probabilities as a function of the three principal observables studied here: stellar mass, star-formation rate and local environment. Since not all observational datasets may offer access to all three quantities, we also provide separate tables for projections of this 3-D param-

[3] i.e. both methods calculate the density within the same radius, but the number of objects differs by 1, meaning the $k = 5$ (not-own-neighbour) number density estimate simply differs by a factor of $5/6$ from the $k = 6$ (own-neighbour) estimate.

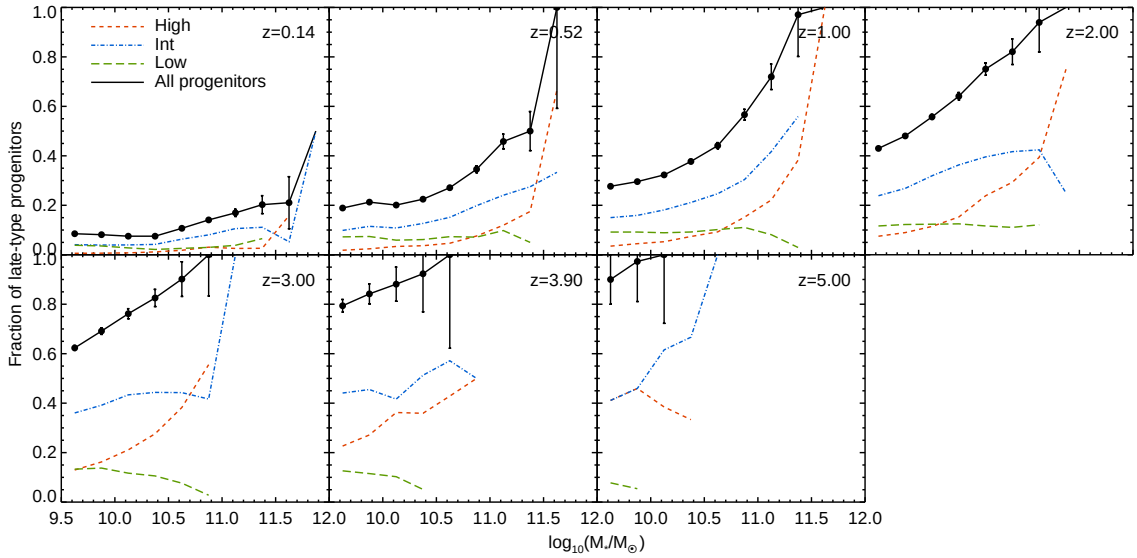

Figure 6. The evolution with redshift of the fraction of late-type galaxies that are progenitors of early-type galaxies at $z \sim 0$. Each panel shows, in black, the fraction of late-type galaxies that are progenitors of local early-types for a given redshift, as a function of the stellar mass of the progenitor. The sample of late-type galaxies is split further into high (red), intermediate (blue) and low (green) density environments. Error bars show Poisson errors. The sample becomes smaller (leading to a corresponding increase in the size of the Poisson error bars) towards higher redshifts, because there are fewer galaxies with stellar masses above $M_\star > 10^{9.5}$ in the simulation. For clarity, errors bars and points are are not shown where the errors are larger than 0.5.

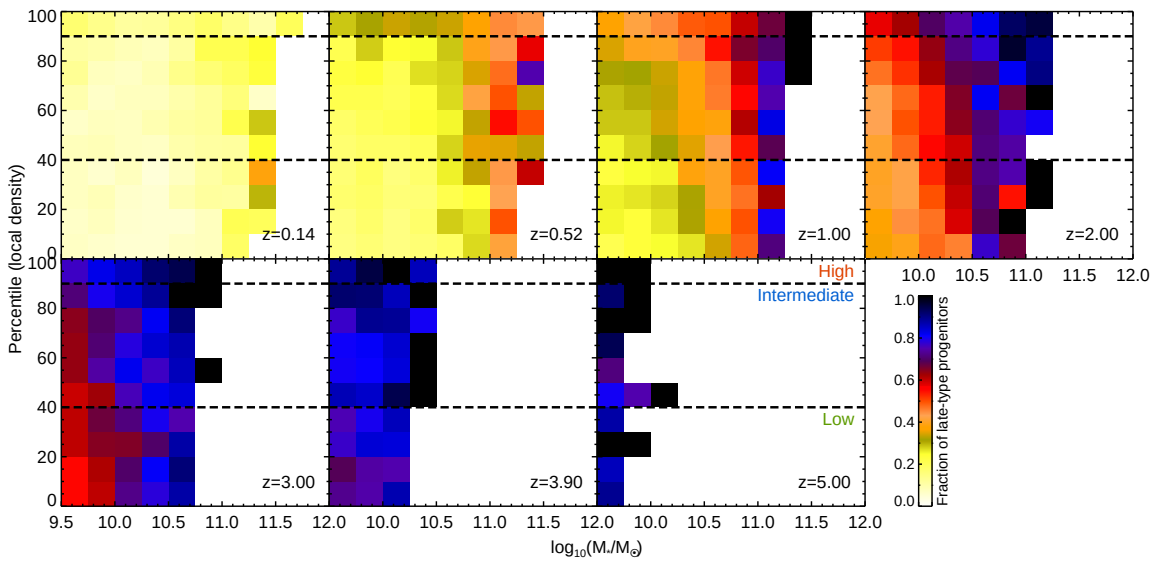

Figure 7. Density plots showing the redshift evolution of the fraction of late-type galaxies that are progenitors of early-types at the present day, as a function of the stellar mass and density percentile of the late-type galaxies in question. The late-type progenitor fraction is represented by the colour bar. We do not plot bins containing 3 or fewer galaxies.

eter space i.e. progenitor probabilities as a function of stellar mass only, stellar mass and local environment and stellar mass and star-formation rate. As noted above, the properties of individual galaxies in current and future observational surveys which provide these observables can be compared to these tables in order to estimate the probability that they are progenitors of an early-type galaxy in the local Universe. The structure of the files containing these tables, and scripts to read them, are provided in the Appendix A.

5 SUMMARY

As end-points of the hierarchical mass assembly process, early-type galaxies host more than half of the stellar mass density in

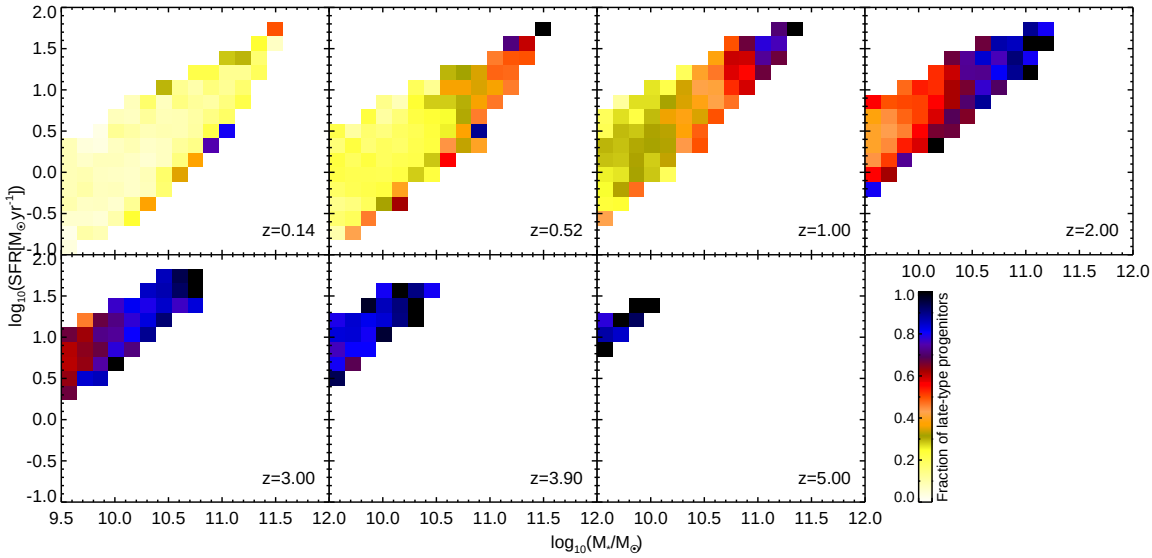

Figure 8. Left: Star-formation main sequence (i.e. star-formation rate plotted against stellar mass) for late-type galaxies, as a function of redshift, colour-coded by the fraction of late-type galaxies that are progenitors of early-types at the present day. We do not plot bins containing 3 or fewer galaxies.

the local Universe, their stellar populations encoding the assembly history of galaxies over cosmic time. Studying these galaxies in the local Universe and probing their progenitors at earlier epochs offers a unique perspective on the evolution of the observable Universe. However, since morphological transformations progressively convert late-type (disc-like) galaxies into early-type systems, the progenitors of today's early-type galaxies become increasingly dominated by late-types at high redshift. Understanding the evolution of early-types over cosmic time therefore requires a reliable method for identifying these late-type progenitors of local early-types. Here, we have used the Horizon-AGN cosmological hydrodynamical simulation, which produces good agreement with the observed properties of galaxies in the redshift range $0 < z < 5$, to study how the progenitors of local early-type galaxies evolve over cosmic time.

We have studied the merger histories of local early-types and the morphologies of galaxies that are involved in these mergers and traced how the morphological mix of galaxies in the progenitor population changes over cosmic time. We have then used the simulation to study the fraction of late-type galaxies that are progenitors of present-day early-types, as a function of redshift, stellar mass, local environment and star-formation rate: observables that can be routinely measured in current and future datasets. As noted earlier, these fractions can be treated as probabilities that can then be used to include late-type progenitors of local early-types in observational surveys by 'weighting' these late-type systems by these probabilities. The benefit of this approach is to alleviate progenitor bias i.e. the bias that occurs if one considers only early-type galaxies (or proxies like the red sequence) to study the progenitor population of today's early-types. Our key conclusions are as follows:

• The merger history of early-type galaxies indicates that these systems finish assembling their stellar mass at relatively early epochs. By $z \sim 1$, around 60 per cent of today's massive early-types, averaged over all environments, have had their last significant merger (i.e. a merger with mass ratio greater than 1:10). For early-type galaxies that inhabit high density environments at the present day (e.g. clusters) this value is 70 per cent, while it is ∼50 per cent in early-types that inhabit low-density environments (e.g. the field). On average, morphological transformation is ∼50 per cent faster in high-density environments compared to low-density regions.

• Progenitor bias is significant at all but the lowest redshifts. Until $z \sim 0.6$ less than half of the progenitors of today's early-types actually have early-type morphology. Similarly, less than half of the stellar mass that ends up in an early-type today is actually hosted by a progenitor that has early-type morphology at this redshift. Around the epoch of peak cosmic star-formation, which is also the epoch at which morphological transformation occurs most rapidly, studying only early-type galaxies misses almost all (at least 80 per cent) of the stellar mass that eventually ends up in early-types at the present day.

• The morphological mix of progenitor galaxies that are involved in mergers evolves over time. At all redshifts, the majority of mergers have at least one late-type progenitor. Mergers between two late-type galaxies dominate at early times i.e. around the epoch of peak cosmic star-formation and beyond ($z > 1.5$) and the fraction of mergers involving two early-type galaxies climbs rapidly at low redshift ($z < 0.5$).

• At all redshifts, late-type galaxies with larger stellar masses are more likely to be progenitors of local early-type remnants. At high redshifts, almost all massive ($M_\star < 10^{11}$ $M_\odot$) late-type galaxies, regardless of their local environment, are progenitors of present-day early-type galaxies. While the progenitor probabilities increase with redshift, for these massive galaxies, the progenitor probability remains close to ~ 1 until $z \sim 0.5$. Lower-mass ($M_\star < 10^{10.5}$ $M_\odot$) galaxies also exhibit high progenitor probabilities at high redshift, as long as they occupy regions of high density (e.g.

the 80-100th percentiles in density).

- At high-redshift, massive late-type galaxies that are progenitors of present-day early-types can be some of the most highly star-forming systems in the Universe, simply because star-formation activity is, on average, more vigorous in the early Universe. However, at fixed stellar mass, progenitor fractions increase as the star-formation rate decreases i.e. late-type galaxies with lower specific star-formation rates are more likely to be progenitors of early-type galaxies.

In the impending era of large observational surveys (e.g. LSST, EUCLID, JWST), this paper provides a framework for studies of how the stellar mass hosted by the local early-type galaxy population is built up over cosmic time.

ACKNOWLEDGEMENTS

We thank the anonymous referee for constructive comments that improved this paper. GM acknowledges support from the Science and Technology Facilities Council [ST/N504105/1]. SK acknowledges a Senior Research Fellowship from Worcester College Oxford. JD acknowledges funding support from Adrian Beecroft, the Oxford Martin School and the STFC. CL is supported by a Beecroft Fellowship. This research has used the DiRAC facility, jointly funded by the STFC and the Large Facilities Capital Fund of BIS, and has been partially supported by grant Spin(e) ANR-13-BS05-0005 of the French ANR. This work was granted access to the HPC resources of CINES under the allocations 2013047012, 2014047012 and 2015047012 made by GENCI. We thank Stephane Rouberol for running and maintaining the Horizon cluster hosted by the Institut d'Astrophysique de Paris. Elias Brinks and Chiaki Kobayashi are thanked for useful comments. This work is part of the Horizon-UK project.

REFERENCES

Abbott T., et al., 2005, preprint, (arXiv:astro-ph/0510346)
Abell P. A., et al., 2009, preprint, (arXiv:0912.0201)
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Bahcall N. A., 1996, ArXiv Astrophysics e-prints,
Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, MNRAS, 373, 469
Barrientos L. F., Lilly S. J., 2003, ApJ, 596, 129
Barro G., et al., 2013, ApJ, 765, 104
Behroozi P. S., Marchesini D., Wechsler R. H., Muzzin A., Papovich C., Stefanon M., 2013, ApJ, 777, L10
Bell E. F., et al., 2004, ApJ, 608, 752
Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38
Bernardi M., et al., 2003, AJ, 125, 1817
Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Bournaud F., Dekel A., Teyssier R., Cacciato M., Daddi E., Juneau S., Shankar F., 2011, ApJ, 741, L33
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Breiman L., Meisel W., Purcell E., 1977, Technometrics, 19, 135
Buitrago F., Conselice C. J., Epinat B., Bedregal A. G., Grützbauch R., Weiner B. J., 2014, MNRAS, 439, 1494
Ceverino D., Dekel A., Bournaud F., 2010, MNRAS, 404, 2151
Ceverino D., Primack J., Dekel A., Kassin S. A., 2017, MNRAS, 467, 2664
Cirasuolo M., Afonso J., Bender R., Bonifacio P., Evans C., Kaper L., Oliva E., Vanzi L., 2011, The Messenger, 145, 11
Clauwens B., Franx M., Schaye J., 2016, MNRAS, 463, L1
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Conselice C. J., 2006, MNRAS, 373, 1389
Conselice C. J., 2014, ARA&A, 52, 291
Covington M. D., et al., 2010, ApJ, 710, 279
Croton D. J., et al., 2006, MNRAS, 365, 11
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Dekel A., Ostriker J. P., eds, 1999, Formation of Structure in the Universe
Devriendt J., et al., 2010, MNRAS, 403, L84
Dressler A., 1980, ApJ, 236, 351
Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012, MNRAS, 420, 2662
Dubois Y., et al., 2014, MNRAS, 444, 1453
Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, MNRAS, 463, 3948
Faber S. M., et al., 2007, ApJ, 665, 265
Fang J. J., Faber S. M., Salim S., Graves G. J., Rich R. M., 2012, ApJ, 761, 23
Ferdosi B. J., Buddelmeijer H., Trager S. C., Wilkinson M. H. F., Roerdink J. B. T. M., 2011, A&A, 531, A114
Fitzpatrick P. J., Graves G. J., 2015, MNRAS, 447, 1383
Franx M., van Dokkum P. G., 1996, in Bender R., Davies R. L., eds, IAU Symposium Vol. 171, New Light on Galaxy Evolution. p. 233 (arXiv:astro-ph/9603029)
Gardner J. P., et al., 2006, Space Sci. Rev., 123, 485
Gladders M. D., López-Cruz O., Yee H. K. C., Kodama T., 1998, ApJ, 501, 571
Guth A. H., Pi S.-Y., 1982, Physical Review Letters, 49, 1110
Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75
Hau G. K. T., Bower R. G., Kilborn V., Forbes D. A., Balogh M. L., Oosterloo T., 2008, MNRAS, 385, 1965
Hawking S. W., 1982, Physics Letters B, 115, 295
Kassin S. A., et al., 2012, ApJ, 758, 106
Kassin S. A., Brooks A., Governato F., Weiner B. J., Gardner J. P., 2014, ApJ, 790, 89
Katz N., 1992, ApJ, 391, 502
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Kaviraj S., 2014a, MNRAS, 437, L41
Kaviraj S., 2014b, MNRAS, 440, 2944
Kaviraj S., et al., 2007, ApJS, 173, 619
Kaviraj S., Devriendt J. E. G., Ferreras I., Yi S. K., Silk J., 2009, A&A, 503, 445
Kaviraj S., Tan K.-M., Ellis R. S., Silk J., 2011, MNRAS, 411, 2148
Kaviraj S., et al., 2013, MNRAS, 428, 925
Kaviraj S., Devriendt J., Dubois Y., Slyz A., Welker C., Pichon C., Peirani S., Le Borgne D., 2015, MNRAS, 452, 2845
Kaviraj S., et al., 2017, MNRAS, 467, 4739
Kennicutt Jr. R. C., 1998, ApJ, 498, 541
Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, MNRAS, 450, 1349
Kimm T., Cen R., Devriendt J., Dubois Y., Slyz A., 2015, MNRAS, 451, 2900
Komatsu E., et al., 2011, ApJS, 192, 18
Laureijs R., et al., 2011, preprint, (arXiv:1110.3193)
Leja J., van Dokkum P., Franx M., 2013, ApJ, 766, 33
Lofthouse E. K., Kaviraj S., Conselice C. J., Mortlock A., Hartley W., 2017a, MNRAS, 465, 2895
Lofthouse E. K., Houghton R. C. W., Kaviraj S., 2017b, MNRAS, 471, 2311
Longhetti M., et al., 2005, MNRAS, 361, 897
Martin G., Kaviraj S., Devriendt J. E. G., Dubois Y., Laigle C., Pichon C., 2017, MNRAS, 472, L50
McDermid R. M., et al., 2015, MNRAS, 448, 3484
Ownsworth J. R., Conselice C. J., Mortlock A., Hartley W. G., Almaini O., Duncan K., Mundy C. J., 2014, MNRAS, 445, 2198
Peirani S., et al., 2016, preprint, (arXiv:1611.09922)

Rodriguez-Gomez V., et al., 2015, MNRAS, 449, 49
Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2008, MNRAS, 389, 1137
Schaye J., et al., 2015, MNRAS, 446, 521
Shankar F., et al., 2015, ApJ, 802, 73
Shattow G. M., Croton D. J., Skibba R. A., Muldrew S. I., Pearce F. R., Abbas U., 2013, MNRAS, 433, 3314
Silk J., Norman C., 2009, ApJ, 700, 262
Silk J., Rees M. J., 1998, A&A, 331, L1
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Stanford S. A., Eisenhardt P. R., Dickinson M., 1998, ApJ, 492, 461
Starobinsky A. A., 1982, Physics Letters B, 117, 175
Takada M., et al., 2014, PASJ, 66, R1
Taylor P., Kobayashi C., 2016, MNRAS, 463, 2465
Tempel E., Tago E., Liivamägi L. J., 2012, A&A, 540, A106
Teyssier R., 2002, A&A, 385, 337
Thomas D., Maraston C., Bender R., Mendes de Oliveira C., 2005, ApJ, 621, 673
Torrey P., et al., 2015, MNRAS, 454, 2770
Torrey P., Wellons S., Ma C.-P., Hopkins P. F., Vogelsberger M., 2017, MNRAS, 467, 4872
Trager S. C., Faber S. M., Worthey G., González J. J., 2000, AJ, 119, 1645
Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, A&A, 506, 647
Vogelsberger M., et al., 2014, MNRAS, 444, 1518
Welker C., Dubois Y., Devriendt J., Pichon C., Kaviraj S., Peirani S., 2015, preprint, (arXiv:1502.05053)
Wellons S., Torrey P., 2017, MNRAS, 467, 3887
Wilkinson M., Meijer B., 1995, Computer methods and programs in biomedicine, 47, 35
Worthey G., 1994, ApJS, 95, 107
de Jong R. S., et al., 2012, in Ground-based and Airborne Instrumentation for Astronomy IV. p. 84460T (arXiv:1206.6885), doi:10.1117/12.926239
van Dokkum P. G., Franx M., 2001, ApJ, 553, 90
van Dokkum P. G., et al., 2010, ApJ, 709, 1018

APPENDIX A: TABLES OF PROGENITOR PROBABILITIES

The joint progenitor probability distribution is tabulated as a function of redshift, stellar mass, local environment and star-formation rate and stored in an $a \times b \times c \times d$ binary file. The python routine `tabulate_progenitor_probability.py` outputs progenitor probabilities from the joint probability distribution and can be run as follows:

```
python tabulate_progenitor_probability.py -z <red-
shift> -m <mass> -p <percentile> -s <SFR>,
```

where mass is in units of $\log_{10}(M_\star/M_\odot)$ and SFR is in units of $M_\odot$ yr^{-1}. At least 1 of the 4 keywords must be supplied, if fewer than 4 keywords are supplied the routine outputs the progenitor probability marginalised over the missing dimension(s).

e.g. calling '`python tabulate_progenitor_probability.py -z 2.0 -m 10.0 -s 3`' returns the joint progenitor probability for $z = 2$, $M_\star = 10^{10}$ $M_\odot$ and $SFR = 3$ $M_\odot$ yr^{-1} with environment marginalised out.

We also provide progenitor probabilities in ASCII format. For each redshift $z \in [0,5]$ progenitor probabilities are tabulated in 4 files:

- `m_rho_SFR.txt` contains 4 columns listing stellar mass [$\log_{10}(M_\star/M_\odot)$], percentile, star-formation rate [$M_\odot$ yr^{-1}] and the joint progenitor probability.

- `m_rho.txt` contains 3 columns listing stellar mass, percentile and the progenitor probability marginalised over SFR.

- `m_SFR.txt` contains 3 columns listing stellar mass, SFR and the progenitor probability marginalised over percentile.

- `m.txt` contains 2 columns listing stellar mass and the progenitor probability marginalised over percentile and SFR.

The routine and files above are available from `http://www.star.herts.ac.uk/~gmartin/bias`.

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