A homogeneous measurement of the delay between the onsets of gas stripping and star formation quenching in satellite galaxies of groups and clusters

Kyle A. Oman[1,2]⋆, Yannick M. Bahé[3], Julia Healy[2,4], Kelley M. Hess[2,5],
Michael J. Hudson[6,7,8], Marc A. W. Verheijen[2]

[1] *Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, United Kingdom*
[2] *Kapteyn Astronomical Institute, University of Groningen, Postbus 800, NL-9700 AV Groningen, The Netherlands*
[3] *Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[4] *Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa*
[5] *ASTRON, the Netherlands Institute for Radio Astronomy, Postbus 2, NL-7990 AA Dwingeloo, the Netherlands*
[6] *Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
[7] *Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
[8] *Perimeter Institute for Theoretical Physics, Waterloo, ON N2L 2Y5, Canada*

3 September 2020

ABSTRACT

We combine orbital information from N-body simulations with an analytic model for star formation quenching and SDSS observations of galaxy positions, redshifts, and colours to infer the differential effect of the group/cluster environment on star formation in satellite galaxies. We repeat a similar exercise with a similar model for gas stripping and supplement the same input catalogue with H I fluxes from the ALFALFA survey. The statistical models are motivated by and tested on the Hydrangea cosmological hydrodynamical simulation suite. Our models recover the typical time when galaxies in the satellite population are stripped and quenched: stripping in massive ($M_{\rm vir} \sim 10^{14.5}\,{\rm M_\odot}$) clusters typically occurs at or just before the first pericentric passage, and up to $\sim 4\,{\rm Gyr}$ later in low mass ($\sim 10^{12.5}\,{\rm M_\odot}$) groups. Quenching occurs later: Balmer emission lines typically fade $\sim 4\,{\rm Gyr}$ ($6.5\,{\rm Gyr}$) after first pericentre in clusters (groups), followed a few hundred Myr later by reddenning in $(g - r)$ colour. These 'delay timescales' are remarkably constant across the entire satellite stellar mass range probed ($\sim 10^{9.5}$–$10^{11}\,{\rm M_\odot}$); this feature is closely tied to our treatment of 'group pre-processing'. While our measurements are qualitatively consistent with the 'delayed-then-rapid' quenching scenario advocated for by several other studies, we find significantly longer delay times. The origin of this apparent discrepancy is difficult to pinpoint. However, our unique use of a homogeneous analysis and input catalogues yields new insight into the sequence of events leading to quenching across wide intervals in host and satellite mass.

Key words: galaxies: clusters: general – galaxies: evolution

1 INTRODUCTION

Galaxies do not exist in isolation. They occupy the various environments embedded within the cosmic web – nodes, filaments, walls and voids, in rough order of decreasing galaxy density – and may be either a locally dominant 'central' object, or a 'satellite' embedded in a group or cluster. Many of the processes which shape the build-up to a present-day galaxy are sensitive to its surroundings. Gas accretion, for instance, is suppressed in higher density environments where galaxies typically move faster relative to a hotter ambient medium. Satellite galaxies may lose material to the ram pressure felt as they plough through the coronal gas of their host systems, or to tides. Mergers become more common with increasing density, but drop off again once the typical relative velocities increase enough

⋆ kyle.a.oman@durham.ac.uk

to make fly-by encounters more common. Some environmental processes couple non-linearly. For instance, 'harassing' fly-by encounters make the galaxies involved more susceptible to stripping mechanisms; ram pressure compresses gas as well as stripping it, which may trigger increased star formation, in turn feeding energy back into the ISM and driving outflows, further accelerating the loss of gas (see also Boselli & Gavazzi 2006, sec. 4 for a more detailed overview of the outline above).

1.1 Quenching and the environment

Despite the numerous physical processes involved and complex interactions between them, some simple broad trends emerge: galaxy populations in denser environments include proportionally more early-types (Dressler 1980), and less star-forming galaxies (Balogh et al. 2004b; Hogg et al. 2004). Not all galaxies are equally sensitive to environment: in a seminal study, Peng et al. (2010) showed that the dependence of star formation (as traced by colour) on density is at least approximately separable from its dependence on stellar mass. Furthermore, star formation in lower mass galaxies is only shut down – 'quenched' – in the highest density environments, while the most massive galaxies ($M_\star/M_\odot \gtrsim 10^{11}$) tend to be non-star-forming ('passive') independent of local density.

Given the large number of potentially important physical processes at play, it is interesting to ask which are dominant in shaping the evolution of galaxies at different stages of their assembly and in environemnts of different densities. One possible approach is to look for relatively sharp transitions in the properties of the galaxy population – for instance as a function of time spent in a high density environment – and measure their timings and durations. These timescales can then be compared to theoretical predictions for the timescales on which the various processes should operate, linking each observed transition to a physical explanation. This general approach has been used by many authors, for a few notable examples see e.g. Balogh et al. (2000); Kauffmann et al. (2004); Peng et al. (2010); Wetzel et al. (2013); Haines et al. (2015); Fossati et al. (2017); see also other specific examples cited below.

In this work we will focus our attention specifically on the regions of highest galaxy density – galaxy groups and clusters. It has been shown that the influence of the group/cluster environment on star formation extends out to several (~ 5) virial[1] radii (Balogh et al. 1999; von der Linden et al. 2010; Lu et al. 2012; Rasmussen et al. 2012). On their first orbit through a cluster, some galaxies may reach apocentric distances of up to $\sim 2.5\,r_{\rm vir}$ (e.g. Mamon et al. 2004; Oman et al. 2013), and such 'backsplash' objects are required to explain the suppression of star formation in galaxies in the immediate vicinities of groups/clusters (Wetzel et al. 2014).

---

[1] Throughout this work, we define 'virial' quantities at an overdensity corresponding to the solution for a spherical top-hat perturbation which has just virialized, see e.g. Bryan & Norman (1998). The virial overdensity at $z = 0$ is $\sim 360\times$ the background density $\Omega_{\rm m}\rho_{\rm c}$, where $\rho_{\rm c}$ is the critical density for closure and $\Omega_{\rm m}$ is the cosmic matter density in units of the critical density. For readers accustomed to a virial overdensity of $200\times$ the critical density, approximate conversions are $M_{200}/M_{\rm vir} \approx 0.81$ and $r_{200}/r_{\rm vir} \approx 0.73$, assuming a typical concentration parameter for cluster-scale systems.

1.2 Group pre-processing

That star formation is reduced relative to more isolated galaxies to much larger radii indicates that star formation is sensitive to more than the direct influence of the cluster. For instance, galaxies may feel ram pressure as they fall through the filaments feeding material onto the cluster (Bahé et al. 2013). The hierarchical clustering of galaxies also implies that groups are proportionally more common around clusters. Star formation in many cluster satellites is thus likely affected by processes particular to the group environment long before they actually reach the cluster itself, especially within filaments (Gouin et al. 2019). Similarly, group satellites fall into their hosts as members of smaller groups. This broad notion has become known as 'group pre-processing', and is recognized as a crucial ingredient in models seeking to explain the environmental dependence of star formation (e.g. Taranu et al. 2014; Hou et al. 2014; Jaffé et al. 2015; Haines et al. 2015; Pallero et al. 2019; Gouin et al. 2019; Rhee et al. 2020).

1.3 Projected phase space

The three observationally accessible 'projected phase space' (PPS) coordinates of a satellite, i.e. its on-sky position and line-of-sight velocity offsets from its host system, correlate with parameters describing its orbit, such as the time since infall, or distance of closest approach (Mahajan et al. 2011; Oman et al. 2013; Pasquali et al. 2019). Note that the sign of the line-of-sight velocity usually carries no information: in the absence of high-precision distance measurements, it is impossible to discriminate between e.g. a foreground satellite receding toward a background host (negative radial velocity with respect to the host), or a background satellite receding away from a foreground host (positive radial velocity with respect to the host). The possible orbits corresponding to a given set of PPS coordinates can be inferred by sampling in simulations the orbits of satellites with similar PPS coordinates, suitably normalized, around broadly similar hosts; the dependence of the resulting orbital distribution on the host mass and the ratio of the satellite and host masses are relatively weak (Oman et al. 2013).

Analyses of the PPS coordinates of galaxies and their correlations with other galaxy properties have led to many inferences on the environmental influence of groups and clusters, for instance: that star formation is nearly completely quenched after a single passage through a rich cluster (Mahajan et al. 2011); that the predicted strength of the ram-pressure force as a function of PPS coordinates is anti-correlated with the PPS distribution of star-forming galaxies (Hernández-Fernández et al. 2014, see also Arthur et al. 2019; Roberts & Parker 2020); that atomic hydrogen-deficient blue galaxies are on average further along their orbits within clusters than atomic hydrogen-rich blue galaxies, suggesting that they have been ram-pressure stripped (Jaffé et al. 2015); that warm dust is also ram-pressure stripped (Noble et al. 2016); that galaxies are morphologically segregated in PPS in the Coma cluster, with late-types preferentially exhibiting tails of stripped H α-emitting gas (Gavazzi et al. 2018); that galaxies exhibiting the most prominent tails in H α are preferentially found at low projected position offset and high projected velocity offset, consistent with being at their first pericentre (Jaffé et al. 2018); that the ages of stellar populations are strongly correlated with PPS coordinates, and that morphological transformation lags the shutdown in star formation (Kelkar

et al. 2019); that the shapes of SED-fitting based galaxy star formation histories correlate with PPS coordinates (Smith et al. 2019).

1.4 Quenching timescales

In this work we focus particularly on the timescales associated with the quenching of satellites in groups and clusters. There is a growing consensus in the literature that quenching of satellites of $M_\star \gtrsim 10^9 \, M_\odot$ in low-redshift groups ($M_{\mathrm{vir}} \gtrsim 10^{13} \, M_\odot$) proceeds in a 'delayed-then-rapid' fashion, with star forming galaxies continuing their activity for several Gyr after infall into a host, followed by an abrupt cessation on a much shorter timescale. This was initially suggested based on semi-analytic models (De Lucia et al. 2012) and measured shortly thereafter (Wetzel et al. 2013, hereafter W13), followed by corroboration by several other authors (e.g. Taranu et al. 2014; Haines et al. 2015; Oman & Hudson 2016, hereafer OH16; Maier et al. 2019; Rhee et al. 2020). Whether different measurements agree in detail, for instance regarding the delay timescale, is more difficult to assess. Each study makes a different set of modelling assumptions, and in most cases chooses a different reference 'infall time', e.g. infall is defined at different radii, or as first infall into any host system rather than infall into the present host system. Given the diversity in satellite orbits, translating between different definitions is not straightforward. In this study we propose (yet) another reference time, tied to the first pericentric passage of the satellite rather than its infall. This is the time where both the tidal and ram-pressure forces acting on the satellite are expected to first peak and so this definition may simplify the interpretation of our measurement, though we note that 'infall', however defined, is likely also relevant in that this is approximately the time when accretion of fresh gas onto the satellite would be expected to cease.

The physical interpretation of the 'delayed-then-rapid' measurement is debated. While it is generally agreed that accretion of fresh gas should cease around the time a satellite enters the intra-group/cluster medium of its host, the extent to which the remaining gas supply is depleted by continued star formation versus removed by ram pressure is unclear. While some authors argue that a starvation[2] model alone can adequately explain the measurements (e.g. W13; Taranu et al. 2014), others contend that ram-pressure stripping (RPS) plays a significant role (Haines et al. 2013; Jaffé et al. 2015; Roberts et al. 2019; Rhee et al. 2020). Curiously, some analyses of hydrodynamical simulations strongly favour a RPS dominated scenario (Bahé & McCarthy 2015; Lotz et al. 2019, see also Sec. 4.1 below), though others argue that starvation alone may be sufficient, particularly for low-mass galaxies (van de Voort et al. 2017). Bahé et al. (2017b, see also Bahé et al. 2016) also caution that, due to their limited resolution in terms of length and temperature, current hydrodynamical simulations of clusters likely overpredict the efficiency of RPS.

Further insight into the physics of quenching can be gleaned by considering the redshift and host mass dependences of the delay timescale, in particular. Mok et al. (2013, 2014) and

Muzzin et al. (2014) find that a 'delayed-then-rapid' scenario also seems to hold for galaxy groups and clusters at $z \sim 1$, but that the delay timescale must be much shorter (but see also Fossati et al. 2017, who infer a somewhat longer timescale). This argues either for an increased importance of RPS (Muzzin et al. 2014; Bahé & McCarthy 2015), or strong, wind-driven outflows (McGee et al. 2014; Balogh et al. 2016, see also Lemaux et al. 2019 who argue against the importance of RPS). While for group and cluster satellites of $M_\star \gtrsim 10^9 \, M_\odot$ the delay timescale decreases with increasing stellar mass, Fillingham et al. (2015, see also Fillingham et al. 2019; Miyoshi & Chiba 2020) point out that this trend must eventually turn over at lower host and/or satellite masses, as the satellites of the Milky Way seem to have very short delays between infall and quenching. They interpret this as evidence of a transition from a starvation-dominated scenario at higher masses to a RPS dominated scenario for the Milky Way satellites. The interpretation of a measurement of a very long ($\sim 10 \, \mathrm{Gyr}$) delay time for dwarfs in group-mass hosts (Wheeler et al. 2014) remains an open question.

1.5 Outline

In this work we aim to measure timescales pertinent to quenching in groups and clusters. We use a homogeneous modelling process and, as far as possible, homogenous input data drawn from the Sloan Digital Sky Survey (SDSS), in order to enable a straightforward comparison across ~ 1.5 decades in satellite stellar mass and ~ 3 decades in host (total) mass. Our model is broadly motivated by the 'delayed-then-rapid' paradigm, but allows for substantial scatter in the timing of the 'rapid' phase for individual galaxies within a population. The form of our model is further inspired by, but not explicitly tied to, an analysis of quenching in the Hydrangea cluster zoom-in cosmological hydrodynamical simulation suite Bahé et al. (2017b, see also Barnes et al. 2017). We adopt a similar methodology to OH16, using a large cosmological N-body simulation to infer the probable orbits of observed satellites based on their PPS coordinates. A crucial difference with that study, however, is that we build our model around the probability distribution for the time of the first pericentric passage, rather than the time of first infall into the final host system.

While many studies explicitly model group pre-processing, in this work we adopt a qualitatively different approach, following OH16. Rather than use a galaxy population well-removed from the group or cluster under consideration as a reference (often termed 'field') sample, we compare group/cluster members to galaxies in the immediate vicinity of the host system. We thus aim to isolate and measure the differential effect of the host on the star formation of galaxies relative to what it would have been had they not fallen into the host but otherwise kept the same evolutionary path up to that point.

We also aim to measure the timescale for the depletion of H I gas (either through conversion into stars, stripping, or a change in phase), using as far as possible the same methodology and a subsample of the same input galaxies where the SDSS overlaps with the Arecibo Legacy Fast ALFA survey (ALFALFA). The combination of constraints on the gas-stripping and star formation-quenching timescales constitutes a powerful probe of the physics regulating star formation in satellite galaxies. For instance, an abrupt decline in H I content synchronized

---

[2] Some authors distinguish between 'starvation', the consumption of gas in the absence of accretion, and 'strangulation', the stripping of hot gas truncating cooling into colder phases. We do not attempt to make this distinction, and use 'starvation' to encompass both.

with a decline in SFR would argue strongly in favour of rapidly-acting RPS, especially if occurring near pericentre, whereas a more gradual decline in H I content accompanied by sustained star formation would be more consistent with a starvation scenario.

This paper is organised as follows. In Sec. 2 we outline the survey catalogues we use as input. The SDSS data used for our quenching analysis are described in Sec. 2.1; the additional ALFALFA data used to supplement the SDSS catalogue for our stripping analysis is described in Sec. 2.2. In Sec. 3 we outline the simulation datasets we use to motivate the form of our model for quenching (Sec. 3.1) and infer the orbital parameters of obseved galaxies given their PPS coordinates (Sec. 3.2). We describe our quenching model in Sec. 4, including its motivation (Sec. 4.1), formal definition and statistical analysis methodology (Sec. 4.2), and tests of our ability to accurately recover model parameter values (Sec. 4.3). We present our measurements of the quenching and stripping timescales as a function of stellar mass and host mass in Sec. 5, discuss our interpretation thereof in Sec. 6, and summarize in Sec. 7.

2 OBSERVED GALAXY SAMPLE

2.1 Sample for quenching analysis

We make use of the SDSS Data Release 7 (Abazajian et al. 2009) catalogue, supplemented with star formation rates (Brinchmann et al. 2004; Salim et al. 2007) and stellar masses (Mendel et al. 2014). We use the spectroscopic sample of galaxies, which introduces incompleteness in the sample at the 10 per cent level globally (likely higher in dense clusters) due to fibre collisions. This bias is not strongly dependent on e.g. colour, such that the effect on our statistical analysis should be minimal. We discard all galaxies with $m_r > 17.5$, which yields a complete (except for fibre collisions) magnitude limited sample (Strauss et al. 2002)[3]. Within a given group, all galaxies are at approximately the same distance, so a magnitude limit translates approximately to a stellar mass limit (the r-band is a reasonable tracer of total stellar mass; Maraston 1999). Since we fit our model (Sec. 4.2) to data in narrow bins in stellar mass, the net effect of the magnitude limit is simply to change the number of groups contributing to any given stellar mass bin, with more distant groups dropping out of the sample for bins at lower stellar masses. Thus, each fit is actually to an approximately *volume*-limited sample of galaxies, with the volume covered varying with stellar mass.

We further prune our sample of galaxies, removing those with $M_\star < 10^{9.5}\,\mathrm{M}_\odot$. This is because these low-mass galaxies have a relative r-band magnitude-dependent bias in their colour distribution, such that there are relatively more faint red galaxies in the catalogue, which could significantly bias our analysis. This issue is discussed further in Sec. 6.1.1.

In order to obtain a sample with a wide range in host halo mass, we select satellite galaxy candidates around groups/clusters from the von der Linden et al. (2007) and Lim et al. (2017) group catalogues: the group virial masses of these two catalogues peak at $\sim 10^{14}$ and $10^{13}\,\mathrm{M}_\odot$, respectively. We discard the small number of groups with redshifts $z < 0.01$,

which typically have very bright members not covered by SDSS spectroscopy. Although the algorithms used to construct the two group catalogues are quite different – von der Linden et al. (2007) search for overdensities of galaxies which share similar colours, while Lim et al. (2017) use a friends-of-friends-based approach – our methodology is minimally sensitive to the any resulting differences as we use only the group centres, redshifts, and velocity dispersions from these catalogues, and not galaxy membership information (the reasons for this are further elaborated below). We estimate host virial masses based on their velocity dispersions as $M_\mathrm{vir} = (\sigma_\mathrm{1D}/9.9 \times 10^{-3}\,\mathrm{km\,s^{-1}})^3$ (an empirical relation obtained by fitting the relevant distribution in an N-body simulation, see Oman 2013); the virial radii follow from the mean enclosed density as $R_\mathrm{vir} = \left(\frac{3}{4\pi}\frac{M_\mathrm{vir}}{\Delta_\mathrm{vir}\rho_\mathrm{crit}}\right)^{\frac{1}{3}}$, where Δ_vir is the virial overdensity and $\rho_\mathrm{crit} = \frac{3H_0^2}{8\pi G}$. The host halo mass, satellite stellar mass and redshift distributions, and the H I masses vs. redshift for ALFALFA-detected satellite candidates, for the two samples are shown in Fig. 1. The host halo mass and redshift distributions are weighted by the number of candidate members used in our analysis (see below): they represent relative numbers of galaxies, not of groups.

We do not use group membership information from the group catalogues because our analysis is designed for a sample which includes the infalling galaxies around each group, as well as foreground and background 'interlopers'. We therefore select satellite candidates from the SDSS catalogue within an aperture of $2.5\,r_\mathrm{vir}$, and $\pm 2\sigma_\mathrm{3D}$ (we assume isotropic velocity distributions such that $\sqrt{3}\sigma_\mathrm{1D} = \sigma_\mathrm{3D}$). These apertures are large enough that essentially all galaxies which enter them and begin orbiting the central group never orbit back out of them (e.g. Oman et al. 2013). For each satellite candidate, we determine normalized position and velocity offsets from the group centre as $R = d_\mathrm{A}\Delta\theta/r_\mathrm{vir}$ and $V = c|z_\mathrm{sat} - z_\mathrm{host}|/((1 + z_\mathrm{host})\sigma_\mathrm{3D})$, where d_A is the angular diameter distance, z_sat and z_host are the satellite and host redshifts, and c is the speed of light. This results in a sample of 6.9×10^4 satellite candidates around 1.8×10^3 groups and clusters.

In order to assign galaxy halo mass estimates to observed satellites, we adopt the abundance matching prescription of Behroozi et al. (2013a). Our analysis, described below, does not rely on precision halo masses since the distribution of possible orbits for a halo with given (R, V) is only a weak function of halo mass (Oman et al. 2013) – estimates within ~ 0.5 dex should suffice. Abundance matching is not directly applicable to satellites, which can be stripped of dark matter and/or stars. However, in the simulations we use the peak halo mass, which is still reasonably well estimated for satellites using abundance matching provided that the stellar component is not substantially stripped. As satellites heavily stripped of stars, but not yet completely destroyed, form only a small fraction of the satellite population (Bahé et al. 2019), we do not attempt to account for these explicitly, and simply accept that their halo masses will be underestimated, introducing a weak bias in our analysis.

Finally, we use two diagnostics of star formation activity, as illustrated in Fig. 2. The first uses the broad-band $(g - r)$ colour: we draw a line 'by eye' just below the red sequence, defined as $(g - r) = 0.05 \log_{10}(M_\star/\mathrm{M}_\odot) + 0.16$, and classify galaxies above (below) this line as 'red' ('blue'). Since our analysis relies only on a binary separation of the two populations, our results are not strongly sensitive to the exact location of this

---

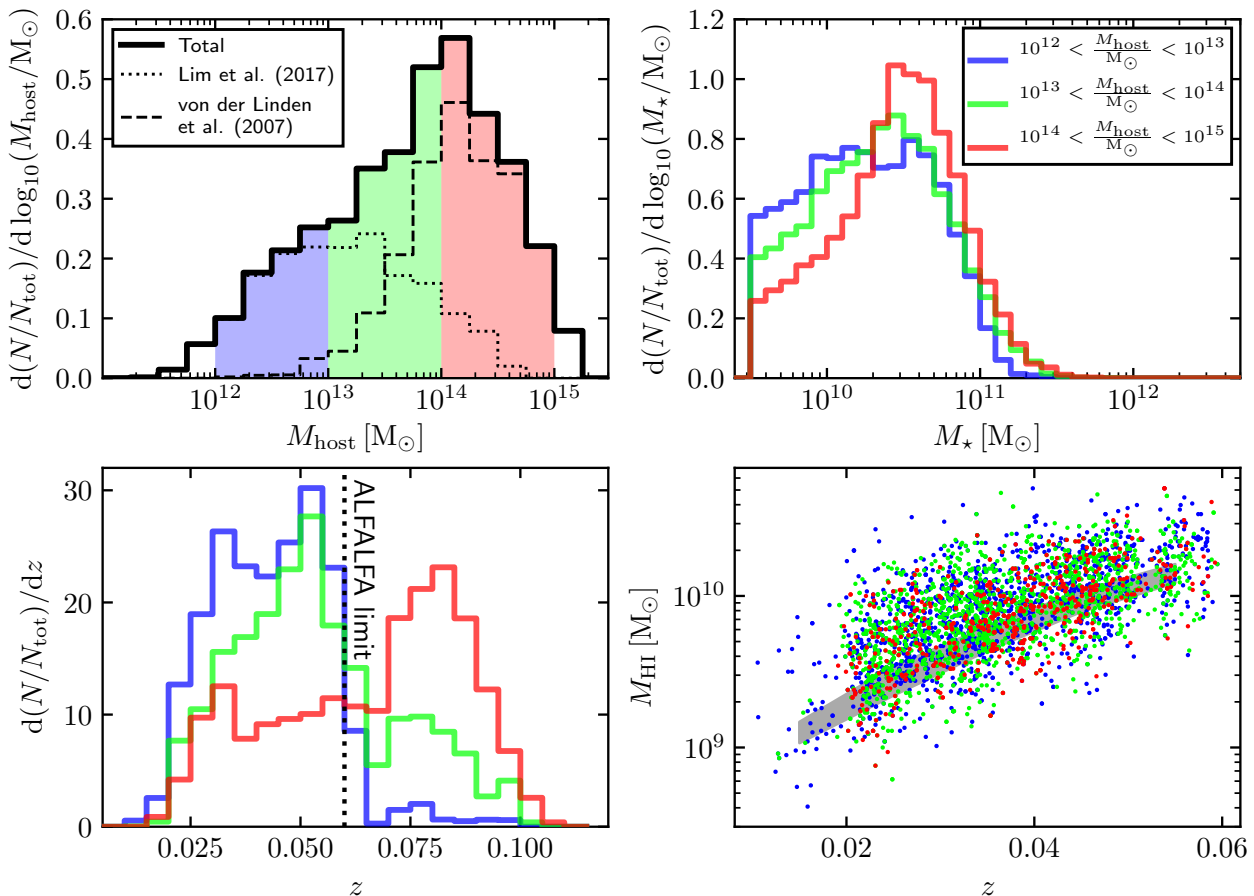[3] See also https://classic.sdss.org/dr7/products/general/target_quality.html.

Figure 1. Overview of the observational samples. *Upper left*: Normalized host virial mass distribution for the combined matches to the von der Linden et al. (2007) and Lim et al. (2017) group catalogues (solid line). The matches to the individual catalogues are shown with broken lines, as labelled. The histogram reflects the number of satellite candidates around hosts of each mass, not the number of host systems. The low-, intermediate- and high-mass host sample ranges are highlighted in blue, green and red, respectively. *Upper right*: Normalized satellite stellar mass distribution in each host mass bin, as labelled. We truncate the sample at $M < 10^{9.5}$ M$_\odot$ (see Sec. 2.1). *Lower left*: Normalized redshift distribution, weighted by satellite candidate count, of hosts for each host mass bin. The redshift limit of $z \sim 0.06$ of the (Haynes et al. 2018) source catalogue is marked with the vertical dashed line; for brevity, we do not show the distributions for the ALFALFA cross-matched galaxy sample. *Lower right*: Redshifts and H I masses of ALFALFA-detected satellite candidates. The gray band shows the interquartile range of upper limit estimates for non-detections (see Sec. 2.2).

cut, provided it reasonably separates the red and blue populations. The second diagnostic classifies galaxies as 'active' or 'passive' based on their specific SFR (sSFR), using the same limit as OH16: $\mathrm{sSFR/yr}^{-1} = -0.4 \log_{10}(M_\star/\mathrm{M}_\odot) - 6.6$. Our model parameters which describe the fractions of star-forming and quiescent galaxies inside/outside groups and clusters have a straightforward dependence on this choice – for instance, moving the colour cut up in Fig. 2 would simply universally increase the fraction of blue galaxies. The more interesting parameters which describe the timing and timescale of the transition from red to blue, or active to passive, are insensitive to the location of the cut, within reason.

We show a representative visualization of the input to our models in the upper and middle panels of Fig. 3, which illustrate the fraction of blue and active galaxies, respectively, as a function of position in the PPS plane. In this example we have selected galaxies with $10 < \log_{10}(M_\star/\mathrm{M}_\odot) < 10.5$ around hosts of $13 < \log_{10}(M_{\mathrm{host}}/\mathrm{M}_\odot) < 14$ (similar figures for more selections in $M_\star$ and M_{host} can be found in Appendix A), and have smoothed the distributions of active and passive galax-

ies with a Gaussian kernel of width 0.25 in each coordinate in order to bring out the overall trend: that there is a relative deficit of blue and active galaxies at low R and V. This smoothing is for visualization only and does not enter into our analysis below.

2.2 Sample for stripping analysis

Our objective is to measure both the timing and timescale for gas stripping and star formation quenching, using a common sample of galaxies, and a common methodology. We could hope to measure, for instance, a delay, or absence thereof, between the removal of atomic hydrogen gas (e.g. by ram pressure) and the shutdown of star formation, which would provide qualitatively new constraints on the environmental quenching process at the galaxy population level.

We begin with the same galaxy sample as in Sec. 2.1 but supplement it with H I gas masses from the ALFALFA survey source catalogue (Haynes et al. 2018). We match the optical counterparts in an extended version of that catalogue (Haynes et al. 2011, Durbala et al. in prep) against the SDSS galaxy sam-
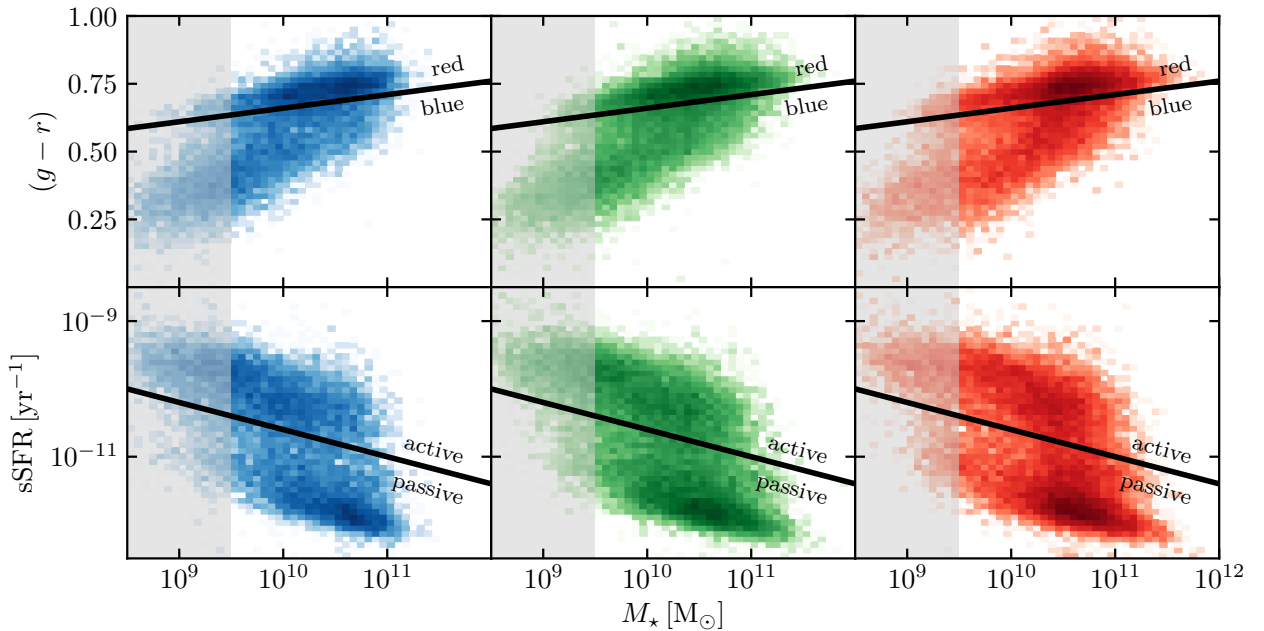
Figure 2. Distribution of galaxies in the low- (left panels), intermediate- (centre panel) and high-mass host samples in the $(g - r)$ colour–stellar mass (upper panels) and sSFR–stellar mass (lower panels) planes. The pixel colour is logarithmically scaled. The solid lines show our adopted divisions between the red and blue populations, $(g - r) = 0.05 \log_{10}(M_\star/M_\odot) + 0.16$, and the active and passive populations, $\mathrm{sSFR}/\mathrm{yr}^{-1} = -0.4 \log_{10}(M_\star/M_\odot) - 6.6$. Galaxies with $M_\star < 10^{9.5}\,\mathrm{M_\odot}$ are excluded from our analysis (see Sec. 2.1) – this region is shaded in gray.

ple described above, requiring matches to be within 2 arcsec on the sky and with a redshift difference of less than 0.0005 ($\sim 150\,\mathrm{km\,s^{-1}}$; note that we are matching the positions of optical sources already associated to ALFALFA H I detections, so these small tolerances are reasonable). The net result is a subset of the sample described in Sec. 2.1, occupying the overlap in sky and redshift coverage of the SDSS and ALFALFA surveys, with either an H I mass measurement, or an H I flux upper limit for non-detections. For non-detections, we derive approximate upper limits on the H I mass assuming an inclination estimated from the r-band axis ratio reported in the SDSS catalogues ($i = \cos^{-1}(b/a)$), an (inclined) width W_{50}^i for the H I line estimated from the g-band Tully-Fisher relation of Ponomareva et al. (2017, table 3), Hubble flow distance estimates derived from the SDSS redshifts, and the ALFALFA 90 per cent completeness limit in H I flux S_{21} as a function of W_{50} (Haynes et al. 2011, eq. 4). The reduced survey volume results in a factor of ~ 2 fewer galaxies, for a total sample of 3.3×10^4 satellite candidates within $R < 2.5$ and $V < 2.0$ of 1.5×10^3 groups and clusters.

Analogous to the colour cut used to separate star-forming and quiescent galaxies (Sec. 2.1), we experiment with a variety of criteria to classify galaxies as 'gas-rich' or 'gas-poor'. However, the large fraction of H I non-detections – 90 per cent of SDSS galaxies in the region and redshift interval where the surveys overlap have no ALFALFA counterpart – makes this challenging. Unlike the distribution of galaxies in colour-magnitude space, in $M_{\rm HI}$–$M_\star$ space there is no obvious separation into two populations, except perhaps the 'detected' and 'undetected' populations. This is intuitive: while a quiescent galaxy reddens but remains relatively easy to detect in an optical survey, a gas-poor galaxy becomes very challenging to detect in 21-cm emis-

sion. We have therefore experimented with divisions in $M_{\rm HI}$–$M_\star$ with various slopes – e.g constant $M_\star/M_{\rm HI}$, constant $M_{\rm HI}$, intermediate slopes – and normalizations. In each case we also consider different treatments of the $M_{\rm HI}$ upper limits, for instance: treating all upper limits as gas-poor; considering only upper limits which are constraining enough to discriminate between gas-rich and gas-poor, given a particular definition; discarding all upper limits from consideration. We reliably find a gradient in the fraction of gas-rich galaxies as a function of position in PPS, with less gas-rich galaxies at low R and V, independent of the definition of 'gas-rich' used, within reason. None of the options explored being obviously superior to the others, we have opted to pursue our analysis using the simplest: we label 'gas-rich' those galaxies which are detected in ALFALFA, and those not 'gas-poor'.

In the lower panel of Fig. 3, we show the resulting gas-rich fraction as a function of position in the PPS plane. A clear gradient is visible, such that there are fewer H I-detected galaxies in groups and clusters, even though the fraction of 'gas-rich' galaxies (i.e., detected in H I) is globally much lower than would be expected for a deeper survey (e.g. Eckert et al. 2015). Because the information pertaining to the timing and timescale for gas stripping along a satellite orbit is encoded in the 'shape' of the transition in the lower panel of Fig. 3, rather than in the absolute normalization of the distribution, we will be able to recover physically meaningful constraints on the relevant model parameters in our analysis below in spite of the highly incomplete nature of the input catalogue. This depends crucially on the probability of a galaxy being detected in H I (given its unknown H I mass) being independent of its position in PPS. This is approximately true; we will return to a more detailed discussion of possible biases in Sec. 6. Of course, an input catalogue
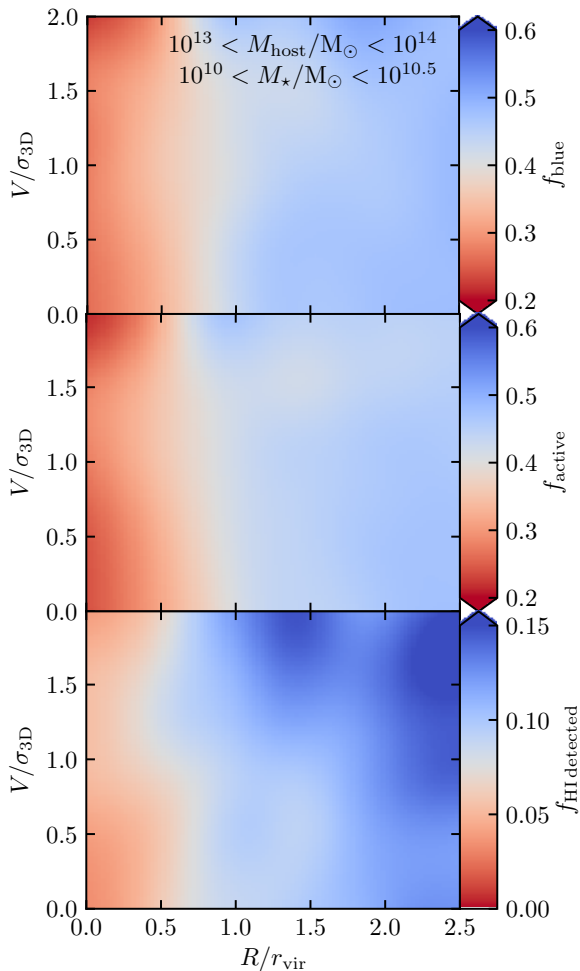
Figure 3. *Upper panel:* Fraction of blue (see Fig. 2) galaxies as a function of normalized projected position $R/r_{\rm vir}$ and velocity $V/\sigma_{\rm 3D}$ offset from the host group centre, for galaxies with $10 < \log_{10}(M_\star/{\rm M}_\odot) < 10.5$ and $13 < \log_{10}(M_{\rm host}/{\rm M}_\odot) < 14$, of which there are $\sim 10^4$. There is a clear relative deficit of blue galaxies at low R and V. We smooth the distributions of blue and red galaxies with a Gaussian kernel of width 0.25 in both coordinates before computing the fraction to better highlight the overall trend. *Middle panel:* As upper panel, but replacing the blue fraction with the 'active fraction' (see Fig. 2). *Lower panel:* As above, but showing the fraction of H I-detected galaxies; only those galaxies in the region where the ALFALFA and SDSS survey volumes overlap are included (see Sec. 2.2), of which there are 6×10^3.

based on a deeper survey would be preferable, however no other current surveys achieve sufficient depth covering a large enough volume for use in our statistical analysis. It seems probable, however, that this will change soon, as SKA precursor facilities come online and begin their surveys (Koribalski et al. 2020, and references therein); we plan to revisit our analysis as new data become available.

3 SIMULATIONS

We make use of two simulation data sets: the Hydrangea cosmological hydrodynamical zoom-in simulations of clusters help to

guide the form of our model for gas stripping and star formation quenching (Sec. 3.1), and a periodic N-body volume run to a scale factor of 2 (redshift $z = -\frac{1}{2}$) to allow us to infer the probability distributions for the orbits of observed galaxies in groups and clusters (Sec. 3.2).

3.1 Hydrangea

The Hydrangea simulations are a suite of 24 cosmological hydrodynamical 'zoom-in' simulations. The zoom regions are selected around rich clusters ($M_{\rm vir} > 10^{14}\,{\rm M}_\odot$) but extend out to $\sim 10 r_{\rm vir}$ and so include many surrounding groups as well as field galaxies. The same galaxy formation model as in the EAGLE project (Schaye et al. 2015; Crain et al. 2015) is used – specifically the 'AGNdT9' model – at the same fiducial resolution level used for the 100 Mpc EAGLE simulation: a baryon particle mass of $1.81 \times 10^6\,{\rm M}_\odot$ and a force softening of 700 pc (physical) at $z < 2.8$. Full details of the simulation setup and key results are described in Bahé et al. (2017b, see also Barnes et al. 2017), and we refer to the papers describing EAGLE, cited above, for details of the algorithms, models and calibration strategy.

The Hydrangea sample broadly reproduces many properties of galaxy clusters. Of particular relevance here is that the scaling with stellar mass (for $M_\star \gtrsim 10^{10}\,{\rm M}_\odot$) of the strength of the differential quenching effect due to the cluster environment is in quantitative agreement with observations by Wetzel et al. (2012), although the absolute quenched fraction is too low both in clusters and in the field (see Bahé et al. 2017b, their fig. 6). Since our model (Sec. 4) is explicitly designed to capture the differential effect of the cluster (or group) environment, these simulations are well-suited to offer guidance on its functional form.

Furthermore, Hydrangea offers a compromise between number of clusters (~ 40 with $M_{\rm vir} > 10^{14}\,{\rm M}_\odot$; several of the 24 zoom regions contain additional clusters besides that centered in the volume) and resolution (we define 'well-resolved' galaxies as those with $M_\star \geq 2 \times 10^9\,{\rm M}_\star$, or about 10^3 star particles) not found in other current cosmological hydrodynamical simulations. We stress, however, that the cold ISM is beyond the resolving power of these simulations: gas at densities $n_{\rm H} > 10^{-5}\,{\rm cm}^{-3}$ is explicitly prevented from cooling below 8000 K and instead falls onto an effective equation of state. Gas in this regime follows empirically calibrated prescriptions for star formation and feedback. This means that dynamically cold, thin gas discs cannot exist in Hydrangea – the gas discs are therefore somewhat too weakly bound and are likely more susceptible to e.g. stripping by ram pressure than they should be. The Hydrangea cluster environment is therefore likely to strip satellites, especially low mass satellites, of gas and quench their star formation somewhat more efficiently than real clusters (see Bahé et al. 2017b, fig. 6).

3.2 N-body

We broadly follow the methodology of OH16 to derive orbit parameter probability distributions from a library of orbits extracted from an N-body simulation, with some improvements.

We extend the 'level 0' N-body simulation from the voids-in-void-in-voids (VVV; Wang et al. 2019) project, using exactly the same configuration as for the original simulation ex-

cept as described below. The simulation has a box size of $500\,h^{-1}\,\mathrm{Mpc}$, mass resolution elements of $10^9\,h^{-1}\,\mathrm{M_\odot}$, and a force softening of $4.6\,h^{-1}\,\mathrm{kpc}$, which offer a reasonable compromise between abundance of group- and cluster-sized structures and smallest resolved satellite galaxies. We run the simulation to a final scale factor of $a = 2$ ($z = -\frac{1}{2}$, $\approx 10\,\mathrm{Gyr}$ into the future). This allows us to tabulate probability distributions for additional orbital parameters, in particular the time of first pericentre, even when it occurs in the future. We use the ROCKSTAR halo finder (Behroozi et al. 2013b) and the related CONSISTENT-TREES utility (Behroozi et al. 2013c) to generate halo merger trees. We then identify satellites of host systems with $\log_{10}(M_{\mathrm{vir}}/\mathrm{M_\odot}) > 12$ as those haloes within $2.5\,r_{\mathrm{vir}}$ at $z = 0$. We trace the primary progenitors/descendants of the satellite sample backward/forward in time and record their orbits relative to the primary progenitor/descendant of their host system at $z = 0$. We compile a table containing properties of the satellites at $z = 0$: the projected offset from the cluster centre $R = \sqrt{(r_{\mathrm{host},x} - r_{\mathrm{sat},x})^2 + (r_{\mathrm{host},y} - r_{\mathrm{sat},y})^2}/r_{\mathrm{vir}}$, the projected velocity offset $V = |(v_{\mathrm{host},z} - v_{\mathrm{sat},z})| + H(r_{\mathrm{host},z} - r_{\mathrm{sat},z})$, and the halo mass of the host, M_{host}. r and v are the coordinate and velocity vectors of the simulated systems, with subscripts (x, y, z) denoting the orthogonal axes of the simulation volume; H is the Hubble parameter. For the satellite mass, M_{sat}, we use the maximum mass at $z \leq 0$, which is better correlated with the stellar mass for moderately stripped satellites (Conroy et al. 2006, see also appendix A in Wetzel et al. 2013). Finally, we also tabulate the lookback time to the first pericentric passage $t - t_{\mathrm{fp}}$ of the satellite within the $z = 0$ host system, with negative times corresponding to future times ($a > 1$ or $z < 0$).

We also compile a similar sample of interlopers around each host system. These are haloes which are within $2.5\,r_{\mathrm{vir}}$ in projection (arbitrarily along the simulations z-axis), but outside $2.5\,r_{\mathrm{vir}}$ in 3D – foreground and background objects. We also require the line-of-sight (z-axis) velocities of interlopers to be within $\pm 2.0\,\sigma_{\mathrm{3D}}$ of the host halo velocity along the same axis. We compile the values of R, V, M_{host} and M_{sat} for all interlopers.

In order to estimate the probability distribution for the pericentric time for an observed satellite (or interloper) galaxy with a given (R, V, M_{host}, M_{sat}), we use the distribution of t_{fp} for all satellites and interlopers within (0.05, 0.04, 0.5 dex, 0.5 dex) of each of these parameters, respectively. Our results are not sensitive to the exact intervals chosen for each parameter; we find that these values offer a good compromise between keeping a narrow range around the properties of the galaxy of interest and selecting a large enough subsample to construct a well-sampled probability distribution for t_{fp}. The interlopers do not have measurements of t_{fp}, but their abundance relative to the selected satellites defines the probability that the observed galaxy is an interloper rather than a satellite. In this way we compute probability distributions for t_{fp} individually tailored to each observed galaxy. We illustrate example t_{fp} probability distributions for satellites with $10^{11} < M_{\mathrm{sat}}/\mathrm{M_\odot} < 10^{12}$ in hosts with $10^{14} < M_{\mathrm{host}}/\mathrm{M_\odot} < 10^{15}$ at various locations in the PPS plane in Fig. 4.

4 ENVIRONMENTAL PROCESSING MODEL

In this section we describe the statistical model which we use in Sec. 5 to infer parameters describing quenching and stripping in groups and clusters. In Sec. 4.1 we present results from the Hydrangea cosmological hydrodynamical simulations (see Sec. 3.1) which motivate the form we adopt for our model. In Sec. 4.2 we provide the formal definition of the model, and describe the method we use to constrain its parameters. Finally, in Sec. 4.3 we describe two tests which demonstrate the limits within which our method can reliably recover the model parameters.

4.1 Motivation

We use results from the Hydrangea simulations to guide the form of our model linking the orbital histories of galaxies to their current star formation (or gas content). Inspired by previous studies (W13; OH16; Lotz et al. 2019), we first parametrized the orbital history by the infall time t_{inf}, here defined as when the satellite first crosses $2.5\,r_{\mathrm{vir}}$. The left panels of Fig. 5 show the evolution of the active fraction (with 'active' defined as sSFR $> 10^{-11}\,\mathrm{yr}^{-1}$) as a function of the time since infall for a sample of Hydrangea satellites which fell into their hosts between 4 and 10 Gyr ago[4] (heavy black line). The upper panels are for ~cluster-mass hosts with $M_{\mathrm{host}} > 10^{14}\,\mathrm{M_\odot}$, while the lower panels are for ~group-mass hosts with $10^{13} < M_{\mathrm{host}}/\mathrm{M_\odot} < 10^{14}$. We see the expected monotonic decline in the active fraction as the population orbits for longer in the cluster. Perhaps surprisingly, the decline begins several Gyr before infall into the cluster-mass hosts. The reason for this becomes apparent when the galaxies are separated according to whether they were centrals (dotted line) or satellites (solid line) at infall: the early decline is predominantly driven by satellites, pointing to 'pre-processing' in groups. The trend is also sensitive to the peak stellar mass of the satellites (coloured lines), with low-mass satellites (darker colour) feeling the influence of the host more strongly than high-mass satellites (paler colour).

In the middle panels of Fig. 5, we repeat the same exercise as in the left panels, except that the orbits are aligned on the time of first pericentre, t_{fp}, rather than the infall time. The same broad trends as in the left panels are seen, but a well-defined, sharp drop in f_{active} appears near $t - t_{\mathrm{fp}} = 0$. This suggests that star formation quenching in Hydrangea is more tightly tied to the pericentric passage than initial infall into the group/cluster environment.

Finally, the right panels of Fig. 5 show the same as the central panel, except that the active fraction f_{active} has been replaced with the gas-rich fraction f_{rich}, where gas-rich[5] galaxies

---

[4] This selection allows us to track each individual galaxy in the sample across the entire $t - t_{\mathrm{inf}}$ range plotted without running into the beginning/end of the simulation. Note that this approach involves 'shifting' the orbits of the satellites in time to align them on their infall or pericentre times; the results in Fig. 5 are therefore not representative of a fixed redshift. Part of the overall declining trend seen in all panels of Fig. 5 comes from the decline in the global fraction of star forming galaxies with time. An example 'snapshot' at a fixed time is shown in Fig. 7; see also similar figures in Appendix C.

[5] We estimate atomic gas masses as described in Crain et al. (2017), using the prescriptions from Rahmati et al. (2013) and Blitz & Rosolowsky (2006).
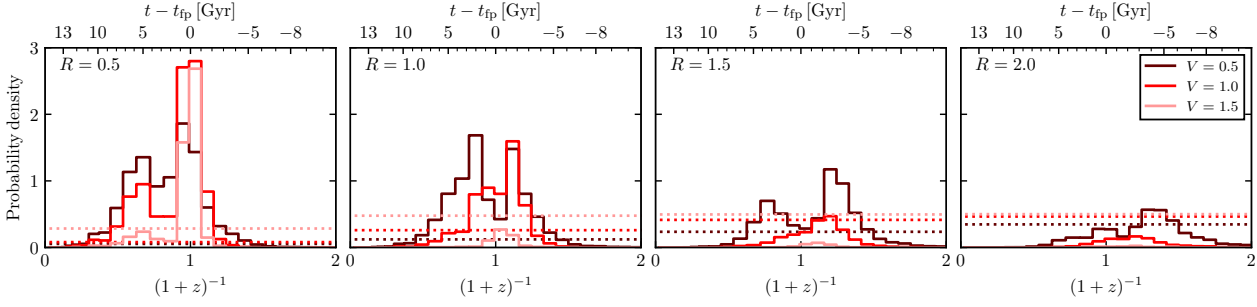
Figure 4. Sample probability distributions for the scale factor $(1+z)^{-1}$ of first pericentre or, equivalently, the time since first pericentre $t - t_{\rm fp}$. These distributions allow for the possibility that the first pericentric passage is in the future, in this case encoding information about how far in the future it will occur. All examples are for hosts with $10^{14} < M_{\rm host}/{\rm M}_\odot < 10^{15}$ and satellites with $10^{11} < M_{\rm sat}/{\rm M}_\odot < 10^{12}$. Each panel corresponds to a different radius R (in units of $r_{\rm vir}$) from the host, as labelled on the panels, and different colours correspond to different velocity offsets V (in units of $\sigma_{\rm 3D}$), as labelled in the legend. The dotted lines illustrate the relative number of interlopers: the integrals of the solid and dotted curves are proportional to the number of satellites and interlopers, respectively.

are defined as those with $M_{\rm HI}/M_\star > 10^{-3}$. The close correspondance with the centre panels is striking – in Hydrangea, satellites clearly experience substantial stripping near pericentre, often enough to immediately shut down star formation (see also Stevens et al. 2019, who find a similar trend in the IllustrisTNG simulations).

We draw our inspiration for a simple analytic model for quenching (or gas stripping) from previous work (OH16) and from the results presented in Fig. 5.

4.2 Definition and fitting method

The model which we adopt relates the fraction of satellite galaxies which are actively star-forming[6] to their time since first pericentre $t - t_{\rm fp}$. We explicitly handle galaxies which are still on their first approach and have not yet reached pericentre, these simply have a negative value of $t - t_{\rm fp}$. The model is intended to capture the relative effect of a host system on its satellites by comparing the properties of satellites of the host with the galaxy population immediately surrounding the host – in practice any survey of a host also covers foreground/background galaxies which have projected positions and velocities consistent with the satellite populations. Our model therefore does not separately handle 'pre-processing' in sub-groups falling into target hosts: members of such sub-groups contribute to the average properties of interlopers. We consider this a benefit rather than a drawback, as it means that we are sensitive only to the differential effect of the final host system. This formulation also ensures that our reference ('field') sample is exactly compatible with our satellite sample: a single selection on a parent catalogue yields both the reference and satellite populations together.

The model has four free parameters: $(f_{\rm before}, f_{\rm after}, t_{\rm mid}, \Delta t)$. $f_{\rm before}$ and $f_{\rm after}$ describe the fraction of galaxies which are actively star forming before the effect of the host system begins to be felt, and after processing by the host is complete, respectively. We model the transition between these two states as a linear decline with time, with the reference time measured relative to the time of first pericentre

---

[6] Or the fraction which are blue, or gas-rich; for brevity in this section we will use language which assumes an application to observations of sSFR.

$t_{\rm fp}$. $t_{\rm mid}$ sets the time at which half of the satellite population has been processed, while Δt fixes the total width of the linear decline. These parameters are schematically illustrated in Fig. 6. When constraining model parameters, we adopt a flat prior probability distribution for each parameter, allowing values in the ranges: $0 < f_{\rm before} < 1$, $0 < f_{\rm after} < 1$, $-5 < t_{\rm mid}/{\rm Gyr} < 10$ and $0 < \Delta t/{\rm Gyr} < 10$. We also impose the constraint that $f_{\rm after} \le f_{\rm before}$.

The likelihood function $\mathcal{L}$ for our model is summarized as:

$$p_a(t - t_{\rm fp}) = \begin{cases} 0 & \text{if } t - t_{\rm fp} < t_{\rm mid} - \frac{\Delta t}{2} \\ \frac{1}{2} + \frac{t - t_{\rm fp} - t_{\rm mid}}{\Delta t} & \text{if } |t - t_{\rm fp} - t_{\rm mid}| \le \frac{\Delta t}{2} \\ 1 & \text{if } t - t_{\rm fp} > t_{\rm mid} + \frac{\Delta t}{2} \end{cases}$$

$$(1)$$

$$p_{a,i} = \frac{\int_{t=0}^{t_{\rm max}} p_a(t - t_{\rm fp}) p_{{\rm peri},i}(R_i, V_i, t - t_{\rm fp}) {\rm d}t}{\int_{t=0}^{t_{\rm max}} {\rm d}t}$$

$$(2)$$

$$p_{A,i} = f_{\rm before} - p_{a,i}(f_{\rm before} - f_{\rm after}) \quad (3)$$

$$P_i = \begin{cases} p_{A,i} & \text{if active(/blue/ gas-rich)} \\ 1 - p_{A,i} & \text{if passive(/red/ gas-poor)} \end{cases} \quad (4)$$

$$\log \mathcal{L} = \sum_i \log P_i \quad (5)$$

Briefly, Eq. 1 encodes the progress of the host in processing the satellites which it affects as a function of time, with a linear drop from 'none' to 'all'. Eq. 2 is the convolution of Eq. 1 with the probability distribution for the pericentre time of the $i^{\rm th}$ satellite $p_{{\rm peri},i}$, evaluated given its observable properties (R_i, V_i). This weights the processing probability at a given time-to-pericentre by the probability that the satellite actually has that time-to-pericentre. In practice the integrals are evaluated as discrete sums, since the pericentre time probability distribution functions (see Sec. 3.2) are discrete. Note that these probability distributions sum to ≤ 1. In the case where the sum is less than one, the remaining probability budget corresponds to the probability that the 'satellite' is in fact an interloper. Eq. 3 simply scales Eq. 2 by the fractions of active galaxies, $f_{\rm after}$ and $f_{\rm before}$. Eq. 4 expresses that the probability for a given galaxy to appear in the sample is $p_{A,i}$ if that galaxy is active, or $1 - p_{Q,i}$ if it is passive (see Fig. 2). Finally, Eq. 5 simply multiplies the
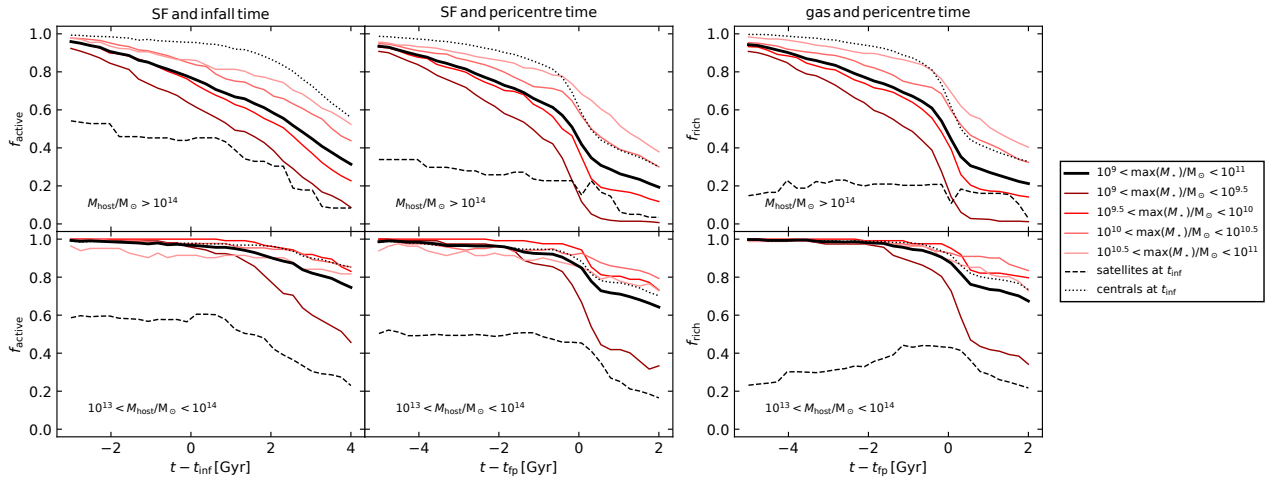
Figure 5. Fraction of star forming satellites, $f_{\rm active}$, defined as those with $\mathrm{sSFR} > 10^{-11}\,\mathrm{yr}^{-1}$ (columns 1 and 2), or of gas-rich satellites (column 3), $f_{\rm rich}$, defined as those with $M_{\rm HI}/M_\star > 10^{-3}$, as a function of orbital time around Hydrangea clusters (upper row) and groups (lower row). In the first column, the orbital phase is aligned to the infall time $t_{\rm inf}$; in the second and third columns the reference time is $t_{\rm fp}$, the time of first pericentre. The heavy black line shows the trend for a fiducial sample. Coloured lines subdivide this sample by peak stellar mass, while broken lines subdivide it by those satellites which were centrals or satellites of another group at the time of infall.
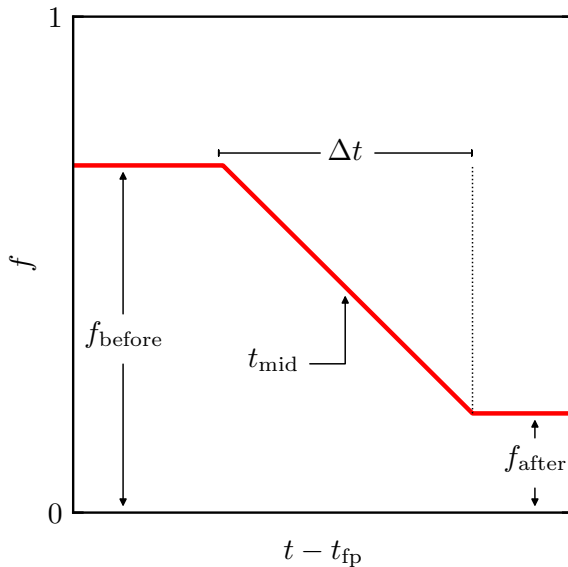

Figure 6. Schematic representation of the free parameters of the environmental processing model encoded in Eqs. 1–5. The fraction f of the galaxy population which is in the un-processed state (e.g. $f_{\rm blue}$, $f_{\rm active}$, $f_{\rm HI\,detected}$) decreases linearly from $f_{\rm before}$ to $f_{\rm after}$ on a timescale Δt. The time when the drop is half complete is $t_{\rm mid}$. The reference time for a given galaxy is the time of its first pericentric approach to its final host, $t_{\rm fp}$.

probabilities for all galaxies, with the usual use of the logarithm to turn the sum into a product and keep the value of $\log \mathcal{L}$ within the realm of practical floating-point computation.

We estimate the posterior probability distribution of the model parameters by Markov chain Monte Carlo (MCMC) sampling using the likelihood function and priors described above, and the EMCEE implementation (Foreman-Mackey et al. 2013)

of the affine-invariant ensemble sampler for MCMC of Goodman & Weare (2010).

4.3 Tests of the model

We perform two tests to check the accuracy of our model and fitting process. In the first, we use the same library of orbits drawn from our N-body simulation which was used to tabulate the probability distributions for $t - t_{\rm fp}$ (i.e. those illustrated in Fig. 4). We tabulate the projected phase space coordinates for each object in the library at $z = 0$ (arbitrarily assuming a line of sight along the simulation z-axis). We assign each a stellar mass based on the Behroozi et al. (2013a) abundance matching prescription, the same which we use (inverted) to assign halo masses to observed galaxies. We then choose fiducial values for the model parameters ($f_{\rm before}$, $f_{\rm after}$, $t_{\rm mid}$, Δt) and randomly flag each object as active or quenched with a probability defined by the model parameters and the objects $t - t_{\rm fp}$ at $z = 0$ as determined from its orbit. In this way we obtain a sample of data which is exactly described by the model which we wish to fit, and additionally has a distribution of orbits exactly consistent with those which will be used to infer $t - t_{\rm fp}$ from the projected phase space coordinates. This is therefore a 'best case scenario' data sample.

From this sample, we draw a random subsample of 2000 objects (in a narrow range of stellar mass) and draw a MCMC sample of the posterior distribution for the model parameters as described in Sec. 4.2. We repeat this exercise 5000 times for different random subsamples and find that we recover an unbiased estimate of the input model parameters $f_{\rm before}$, $f_{\rm after}$ and $t_{\rm mid}$. The parameter Δt, however, tends to be underestimated, with a probability density peaking at 0 even when the input Δt is > 0, and a median value typically underestimated by up to $\sim 1\,\mathrm{Gyr}$. However, we find that in all cases the confidence intervals are representative of the uncertainties. The true parameter values are without exception within the 68, 95 and 99 per cent confidence intervals of the estimates at least 68,
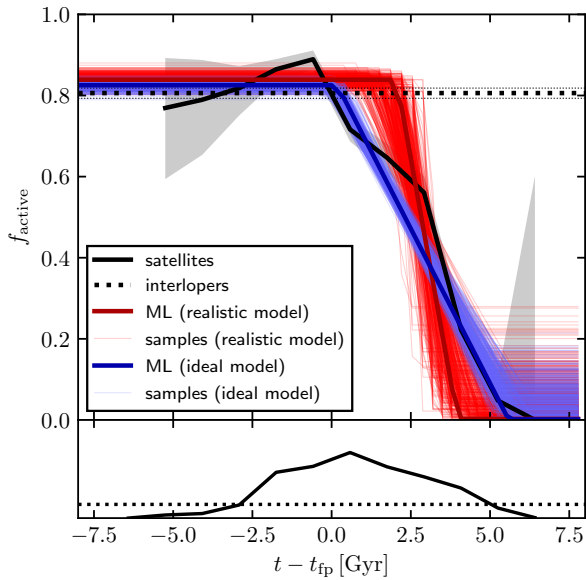
Figure 7. Illustration of our model constrained by Hydrangea data at $z = 0.64$, where the solution is known, for the stellar mass bin centered at $M_\star \sim 10^{10}$ M$_\odot$ (see Fig. 9). The active fraction f_{active} as a function of time to first pericentre $t - t_{\text{fp}}$, calculated in 0.5 dex bins, is shown with the black solid line – the shaded band marks the 1σ confidence interval, estimated as proposed in Gehrels (1986). The horizontal dashed lines mark the active fraction and 1σ confidence interval for the interloper population. The solid line in the lower panel illustrates the relative counts in each bin; the dotted line is for the interlopers, normalized such that the integrals of the two curves are proportional to the relative abundance of interlopers and satellites. The blue curves are individual samples from the Markov chain computed using a model which is given the exact value of $t - t_{\text{fp}}$ for each satellite, while the red curves are similar but for a model where $t - t_{\text{fp}}$ is estimated from the observed location in phase space of each satellite (see Sec. 4.3 for details). The heavier lines of each colour mark the sample from the chain with the highest likelihood.

95 and 99 per cent of the time (sometimes slightly more, suggesting that the widths of the confidence intervals are modestly overestimated). We also repeat this exercise with smaller subsamples, down to a minimum of 100 objects. We find that the confidence intervals, though wider, continue to accurately represent the uncertainty in the estimates.

The second test which we perform uses mock data drawn from the Hydrangea simulations. In order to allow for a scenario in which the value of $t - t_{\text{fp}}$ is known for each object in the sample, we make our mock observations on the $z \sim 0.64$ snapshot of the simulation (approximately the midpoint in lookback time) such that we can track the orbits forward for objects with negative $t - t_{\text{fp}}$. Again, we arbitrarily choose the simulation z-axis as the line of sight and tabulate the projected phase space coordinates of satellites within $R < 2.5$ and $V < 2.0$ around each cluster with $M_{\text{host}} > 10^{14}$ M$_\odot$, i.e. including interlopers. We estimate pre-infall halo masses from the stellar masses using the abundance matching prescription[7] for $z = 0.64$ of Behroozi et al. (2013a). Objects with sSFR $> 10^{-11}$ yr^{-1} are flagged as

---

[7] We also repeated the same test using the exact maximum virial masses from any time $z \geq 0.64$ of the satellites and found no significant change in the parameter estimates.

active, and those below this threshold as quenched. This definition differs from those used for observed galaxies, but we note (i) that all we require for our model is a binary split of the galaxy population, so simply assuming that the bimodal sSFR distribution in the simulations and the observed bimodal $(g - r)$ colour and sSFR distributions broadly reflect the same active/passive populations seems reasonable, and (ii) we do not attempt to draw comparisons between the parameters estimated for the simultions and those estimated for the observations, rather using the simulations only as a test for our methodology. For consistency, we also compute new $t - t_{\text{fp}}$ probability distributions at $z = 0.64$ from our N-body simulation. We exclude poorly resolved galaxies, retaining only those with $M_\star > 2 \times 10^9$ M$_\odot$ ($\gtrsim 10^3$ star particles). This leaves ~ 6500 satellites and interlopers.

We rank the simulated galaxies by stellar mass and split the sample into 4 bins with even counts. For each of these subsamples, we estimate the parameter values of the model described in Sec. 4.2 by MCMC sampling the posterior probability distribution. In each case, we run two fits. In the first, we replace the probability distribution for $t - t_{\text{fp}}$ for each object with the exact value as determined by tracking its orbit (we also inform the model as to which 'satellites' are actually interlopers). This allows us to quantify the behavior of the model given perfect knowledge of the orbits. In Fig. 7, we compare the distribution of model realizations from this Markov chain (blue lines) with the active fraction as a function of $t - t_{\text{fp}}$ (black line) for the stellar mass bin centered at $M_\star \sim 5 \times 10^{10}$ M$_\odot$ (see Fig. 9), demonstrating that our model achieves a good description of the underlying data when given optimal information. We also show the one- and two-dimensional marginalized posterior probability distributions for the 4 model parameters with open blue histograms/contours in Fig. 8. Figures similar to Figs. 7 & 8 for the other stellar mass bins are included in Appendices C and E. In all cases, we find that this model fit is a fair representation of the underlying data. The parameter constraints as a function of stellar mass are summarized in Fig. 9, drawn with the lighter tone of each colour/symbol type, and dotted lines.

We fit the model a second time in each stellar mass bin, this time using the probability distributions for $t - t_{\text{fp}}$ to infer this quantity based on the 'observable' properties of the satellites/interlopers. This represents treating the simulations as closely as possible as an observed data sample. The corresponding model realizations are shown with red lines in Fig. 7 and filled red contours/histograms in Fig. 8. The parameter constraints as a function of stellar mass are shown with the darker symbols and solid lines in Fig. 9. As in the first test described above, we find that the quenching timescale Δt is systematically underestimated (the solid line in the lower panel of Fig. 9 lies well below the dotted line). In contrast with the test using the model 'painted onto' the N-body orbit library, however, this time the 'true' values fall significantly outside the 68 per cent confidence intervals in all cases. The timing of quenching, i.e. t_{mid}, is still well-recovered (centre panel of Fig. 9). The active fractions f_{before} and f_{after} are generally well-recovered. The overestimate at low $M_\star$ arises because the 'zoom-in' regions of the Hydrangea simulations is not large enough to enclose the entire interloper population within the $V/\sigma_{\text{3D}} < 2$ aperture around the cluster. Interestingly, this does not seem to impact the accurate recovery of t_{mid}, and does not seem to be related to the bias in Δt (e.g. there is no visible degeneracy between f_{before} and Δt in Fig. 8 or similar figures in Appendix C).
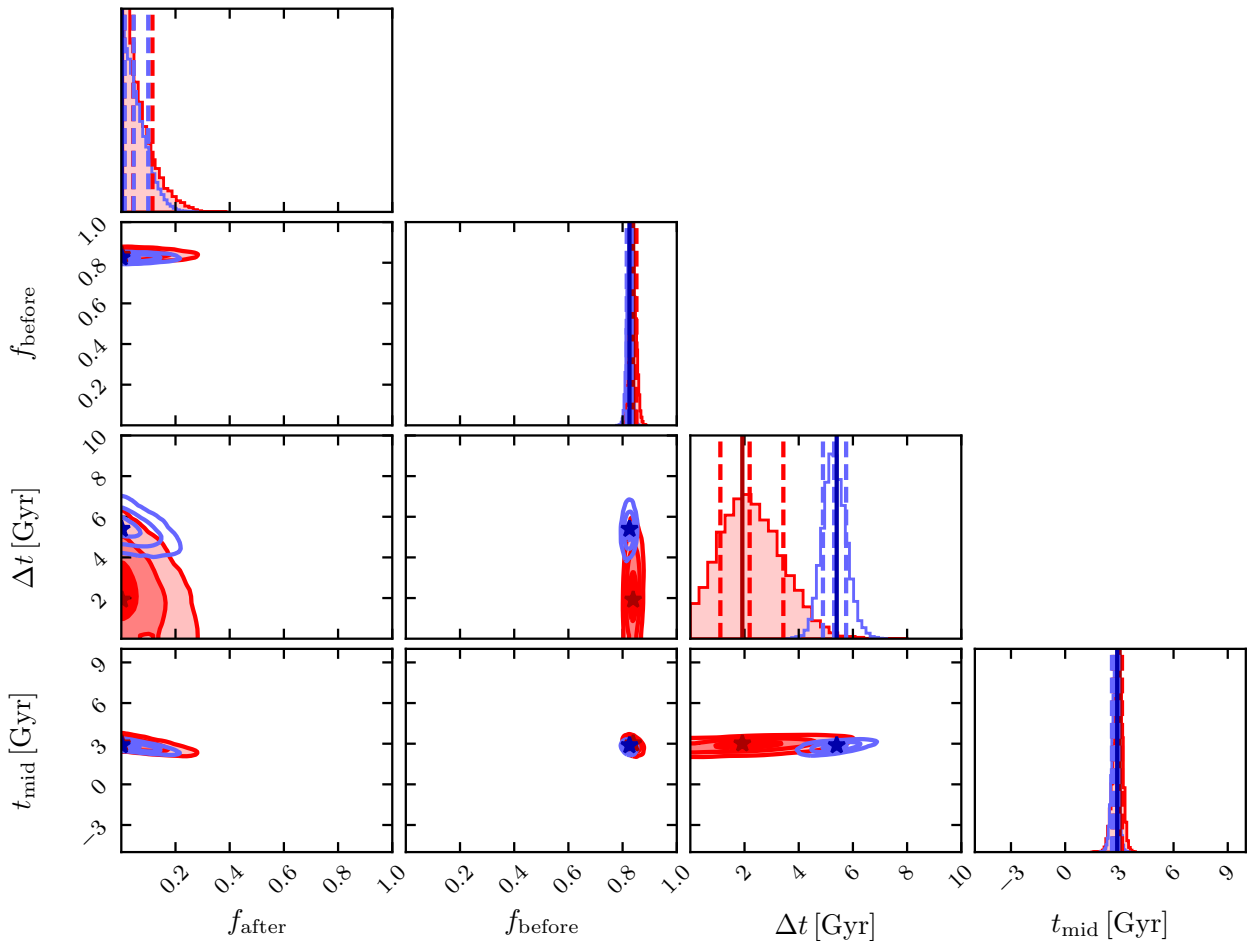
Figure 8. Example one- and two-dimensional marginalized posterior probability distributions for the parameters of the model defined in Sec. 4 constrained by mock sSFR data from the Hydrangea simulations at $z \sim 0.64$. The example shown here corresponds to the stellar mass bin at $\sim 5 \times 10^{10}$ in Fig. 9. Open blue contours/histograms correspond to a fit where full knowledge of $t - t_{\rm fp}$ for each object is provided to the model (the 'truth'), while filled red contours/histograms correspond to fits where $t - t_{\rm fp}$ is estimated from projected phase space coordinates. Contours are drawn at 68, 95 and 99 per cent confidence intervals, and dashed lines are drawn at the $16^{\rm th}$, $50^{\rm th}$ and $84^{\rm th}$ percentiles. The stars and solid lines mark the position of the maximum likelihood parameter sample drawn in the Markov chain. Similar figures for all fits presented in Figs. 9 and 10 are included in the Appendices C and E.

We note the presence of two 'peaks' of significant probability density for the parameters corresponding to the second ($M_{\star} \sim 5 \times 10^9 \,{\rm M}_{\odot}$) stellar mass bin, visible as two separate bulges in the 'violins' in Fig. 9. Inspecting the pairwise marginalized probability distributions for the parameters (Appendix C, Fig. C6), we find that the lower $f_{\rm after}$ peak is associated to the higher $t_{\rm mid}$ peak, and the higher Δt peak. Such degenerate solutions occured occasionally while we experimented with tests of our model, and we found that the lower $f_{\rm after}$ peak (usually consistent with $f_{\rm after} = 0$) invariably corresponded to the 'true' parameter values. This observation will be used in Sec. 5 to motivate fixing $f_{\rm after} = 0$ in our fiducial parameter estimates.

Together, these two tests suggest that when we apply the same method to observational data, our estimates of $f_{\rm before}$, $f_{\rm after}$ and $t_{\rm mid}$ are likely to be accurate. Δt seems to be much more difficult to constrain – the results to the two tests considered together suggest that constraints for this parameter should be taken as lower limits, which will motivate our treatment of Δt as a 'nuisance parameter' in Sec. 5.

5   CHARACTERISTIC TIMING OF QUENCHING AND STRIPPING

We now turn to the constraints on our model parameters using the observational inputs from SDSS and ALFALFA (Secs. 2.1 and 2.2). After experimenting with these data, we have made two choices in the presentation of our fiducial results in this section. The first is to omit the parameter Δt from the discussion in this section. The probability distributions for this parameter tend to be very broad and, given the biases seen in tests in Sec. 4.3, are difficult to interpret. We still allow this parameter to vary with the same prior described above (flat in the interval 0–10 Gyr), but treat it as a nuisance parameter which we marginalize over in the discussion below. Details on the constraints for Δt are, however, included in Appendix B.
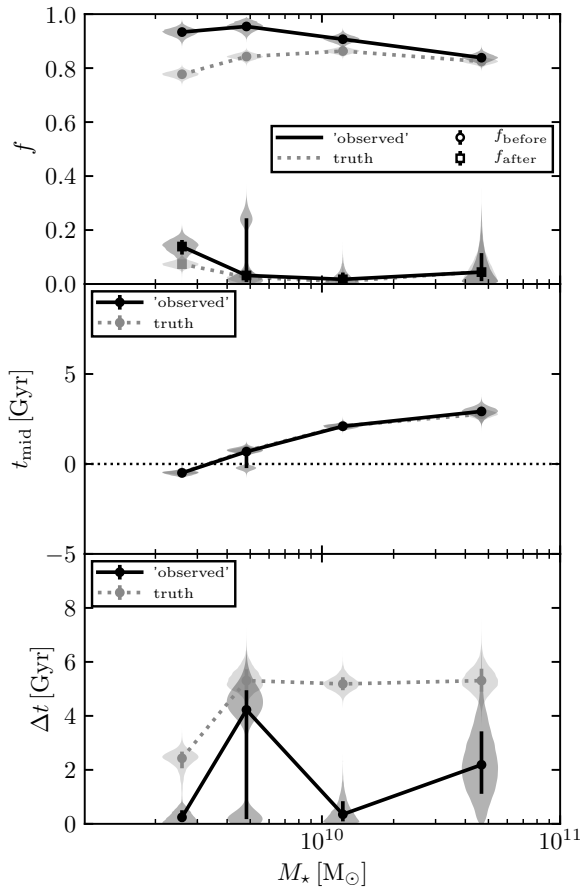
Figure 9. The marginalized median and 16th–84th percentile confidence interval for the parameters of the model defined in Sec. 4 constrained by mock sSFR data from the Hydrangea simulations at $z \sim 0.64$ are shown with points and error bars; the transparent 'violins' show the full marginalized posterior probability distribution for each parameter. The lighter symbols of each type correspond to fits where full knowledge of $t - t_{\mathrm{fp}}$ for each object is provided to the model (the 'truth'), while darker symbols correspond to fits where $t - t_{\mathrm{fp}}$ is estimated from projected phase space coordinates. The upper panel shows the active fractions before (f_{before}, circles) and after (f_{after}, squares) quenching by the host. The centre panel shows the quenching timing parameter t_{mid}, and the lower panel the quenching timescale Δt. The sample is selected to have $M_\star > 2 \times 10^9 \, \mathrm{M}_\odot$ and is split into 4 stellar mass bins with equal counts. The symbols are plotted at the median stellar mass in each bin, offset by a small amount to ensure that the error bars are legible.

The second fiducial choice which we make is to fix the parameter f_{after} to 0. In the majority of cases this is the preferred value when the parameter is left free in any case[8]. Furthermore, while $f_{\mathrm{after}} > 0$ is mathematically straightforward, its physical interpretation is not. In principle is represents the fraction of galaxies which are blue/active/gas-rich once the group/cluster environment has had 'long enough' to exert its influence, where 'long enough' is encoded in Δt. This leads to a degeneracy between the two parameters: if one waits longer (higher Δt), more

---

---

galaxies are quenched/stripped (lower f_{after}) – this is visible in the Δt vs. f_{after} panel of Fig. 8. This, in conjunction with Δt being poorly constrained as explained above, leaves the interpretation of f_{after} somewhat ambiguous as well. Finally, we note that allowing f_{after} to vary does not change our qualitative conclusions, and makes only small quantitative differences. Neither t_{mid} nor f_{before} exhibit any apparent degeneracy with f_{after} (see Fig. 8, and Appendices C and E). For completeness, the probability distributions including f_{after} as a free parameter are included in Appendix B.

Lastly, before moving on to the actual parameter constraints, we re-iterate the physical interpretation of the two parameters which are the focus of our discussion below:

- f_{before} is the fraction of blue/active/gas-rich galaxies 'outside' the cluster. This is determined by the combination of the galaxies which are in the group/cluster but have not yet felt its effects, and those interlopers within $R < 2.5$ and $V < 2.0$. These relatively small apertures around the hosts mean that f_{before} is a measure of the blue/active/gas-rich fraction *just* outside the hosts, i.e. including pre-processed galaxies. Our measurement is therefore sensitive to the *differential* effect of the final host, rather than the cumulative effect of all hosts for galaxies which fall into larger hosts while already being members of smaller groups.
- t_{mid} is the characteristic time along their orbits when galaxies transition from blue/active/gas-rich to red/passive/gas-poor. Put another way, if a randomly selected blue/active/gas-rich satellite is dropped into a cluster, at time t_{mid} (recalling that $t_{\mathrm{mid}} = 0$ corresponds to the time of first pericentre) there is a 50 per cent chance that it has become red/passive/gas-poor. Put yet another way, assuming the blue-red/active-passive/gas-rich-to-poor transition for an individual galaxy is rapid – motivated by the bimodal distributions[9] of Fig. 2 – t_{mid} is a measure of the typical time within the population when this rapid transition occurs.

An individual, independent Markov chain of model parameters is evaluated for each combination input galaxy sample properties: host mass (3 bins with edges at 10^{12}, 10^{13}, 10^{14} and $10^{15} \, \mathrm{M}_\odot$), satellite stellar mass (galaxies are ranked by $M_\star$ and separated into 6 bins with equal counts), and tracer property (($g - r$) colour, sSFR, or H I content).

In the upper panels of Fig. 10, we show the marginalized posterior probability distributions for f_{before} and t_{mid} derived from each of these chains, focusing on the differences between the different tracer properties. f_{before} is a declining function of stellar mass for both the blue fraction (circles connected by solid lines) and the active fraction (triangles connected by dashed lines), as may be expected in the stellar mass interval in question (e.g. Peng et al. 2010). The gas-rich fraction (squares connected by dotted lines) are much lower and flatter – due to the limited depth of the ALFALFA survey these are certainly underestimates (see Sec. 2.2), but this incompleteness is not expected to bias the corresponding estimates of t_{mid} (see Sec. 6 for further details).

We note that the input sample of galaxies for the gas analysis is a subset of that used for the colour and sSFR analyses, corresponding to the overlap region between the ALFALFA and

---

[8] In cases where it is not the preferred value, there are usually multiple peaks – it is likely that the $f_{\mathrm{after}} \sim 0$ peak corresponds to the more correct set of parameter estimates, as discussed in Sec. 4.3.

[9] It is less clear that the gas-fraction distribution is bimodal, motivating some caution when discussing the stripping t_{mid}, below.

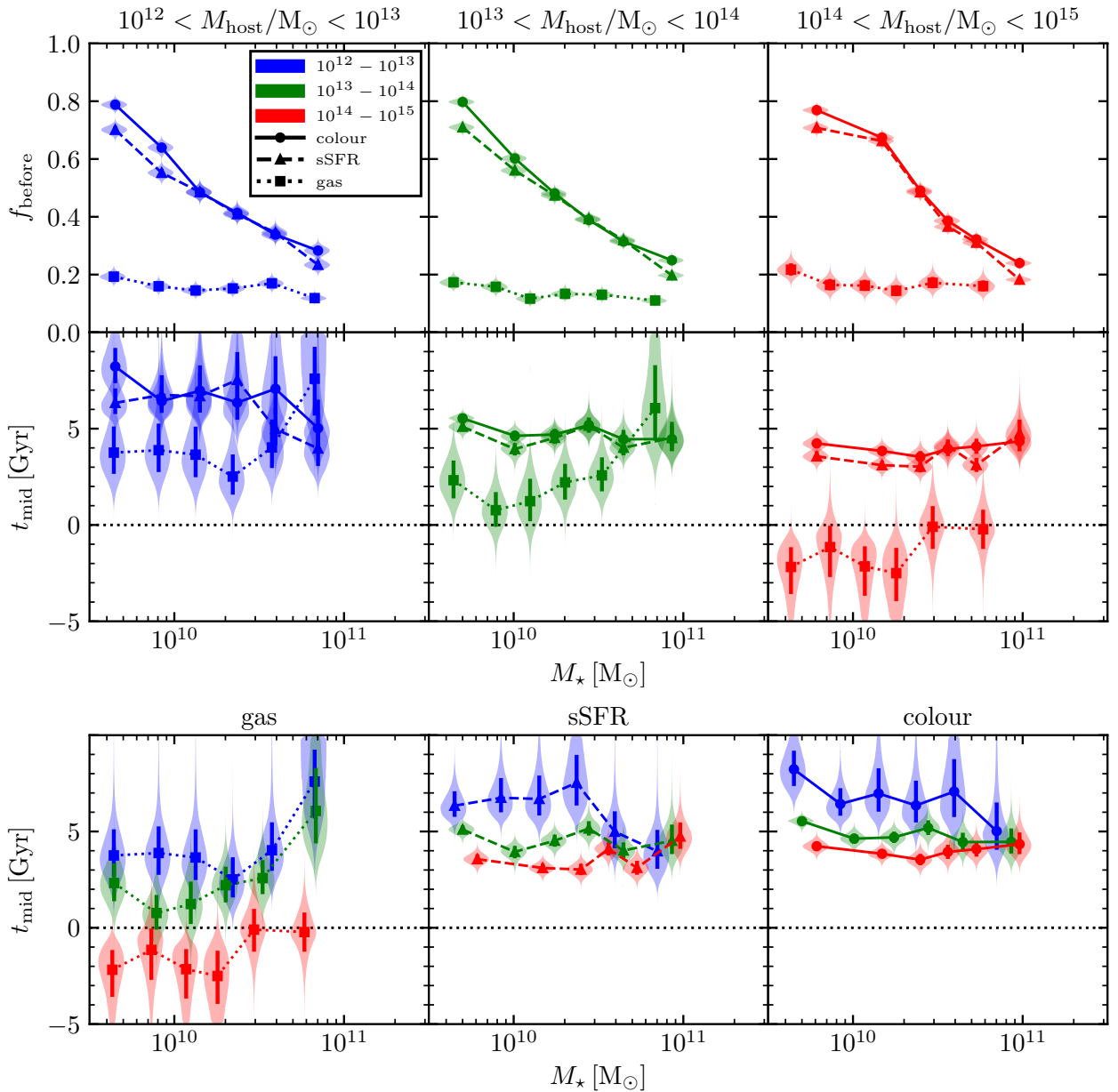**Figure 10.** *Upper panels*: Marginalized posterior probability distributions for the f_{before} (upper row) and t_{mid} (lower row) model parameters as a function of stellar mass around hosts of $10^{12} < M_{\text{host}}/\text{M}_\odot < 10^{13}$ (left column, blue), $10^{13} < M_{\text{host}}/\text{M}_\odot < 10^{14}$ (centre column, green), and $10^{14} < M_{\text{host}}/\text{M}_\odot < 10^{15}$ (right column, red). The parameters are estimated for two tracers of star formation quenching – $(g - r)$ colour (circles connected with solid lines) and sSFR derived from Balmer emission line strength (triangles connected with dashed lines) – and for gas stripping as traced by detection in the ALFALFA survey (squares connected by dotted lines). Points mark the median value of each probability distribution, and error bars the 16–84th percentile confidence intervals; the transparent 'violins' show the full marginalized posterior probability distributions. (The parameter estimates corresponding to the rightmost blue and green squares are likely spurious, see Sec. 6.1.3.) *Lower panels*: Exactly as the lower row in the upper panels, but re-organized to highlight trends with host mass: each column is a for a single tracer (from left to right: H I, Balmer emission lines, broadband colour), rather than a single host mass interval.

SDSS surveys in redshift (see Fig. 1) and sky coverage. This is the reason for the horizontal offset between the square symbols and the triangles and circles in Fig. 10. We have verified that using exactly the same input galaxies for all three analyses does not change the results of the colour and sSFR analyses, other than somewhat widening the confidence intervals.

5.1 Quenching lags stripping

The central result of our analysis is shown in the second row of Fig. 10. The characteristic time t_{mid} when galaxies transition from blue to red within their host (circles connected by solid lines) is consistently found to be well after the first pericentric passage, by ~ 4–5 Gyr in the highest mass hosts (lower right panel) up to perhaps ~ 7–8 Gyr in lower mass hosts (lower

left panel), although here the confidence intervals are somewhat wider. The characteristic time when star formation activity ceases (as traced by the disappearance of Balmer emission lines) is coincident with or slightly (a few hundred Myr to a Gyr) earlier than the colour transition. Such a short delay is not unexpected as the time taken for the stellar population to age and redden once star formation ceases is somewhat longer than that for the emission lines to disappear (by about $\sim 300\,\mathrm{Myr}$, e.g. Balogh et al. 2004a). The characteristic time when galaxies are stripped of neutral hydrogen[10], on the other hand, is well before the quenching time (whether traced by sSFR or colour), by 2-5 Gyr. In the most massive hosts (upper panels, right column of Fig. 10), stripping seems to be well underway even $\gtrsim 1\,\mathrm{Gyr}$ before the first pericentric passage, while around lower mass hosts (centre and left columns) satellites appear to keep the bulk of their H I until up to several Gyr after the first pericentric passage. We regard this difference between the t_{mid} values for star formation quenching (traced by colour or emission lines) and neutral gas stripping as strong evidence for continued star formation well after the onset of ram-pressure stripping of H I. This is consistent with a 'starvation' quenching scenario, although from our measurements we cannot discriminate between the molecular gas directly fueling star formation eventually being depleted, or alternatively being stripped on a subsequent pericentric passage. We will discuss this interpretation further in Sec. 6.3 below.

We repeat the same information shown in the second row of Fig. 10 in the lower panels, but re-arrange the curves to highlight differences between hosts of different masses. There is a clear trend for satellites of a given stellar mass to be stripped (left), become passive (centre), and be redden (right) earlier around more massive hosts, reflecting the generally harsher nature of higher density environments.

6 DISCUSSION

We first consider the reliability of our results in the context of various statistical and systematic effects (Sec. 6.1) before comparing with the results of other studies (Sec. 6.2), and discussing the inferences that can be drawn regarding the processes governing the evolution of satellite galaxies based on our measurements (Sec. 6.3).

6.1 Robustness of parameter constraints

6.1.1 Completeness of input catalogues

We next turn our attention to various biases and systematic effects which could influence the parameter estimates presented in Sec. 5.

The main systematic biases which are of concern for our statistical analysis are any which cause galaxies with a given property to be preferentially included in our sample. Biases which are tied to the PPS coordinates are of particular concern

---

[10] More properly, when they are sufficiently stripped to fall below the ALFALFA detection threshold. We note that, when including only galaxies below the ALFALFA redshift limit of $z = 0.06$, the redshift distributions of satellites around low-, intermediate- and high-mass hosts are reasonably similar (see Fig. 1), so a bias in distance is unlikely to be driving this result.
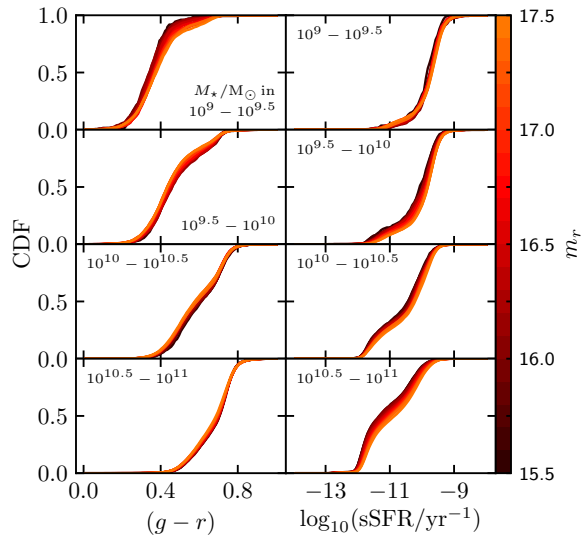
Figure 11. Assessment of colour and sSFR biases in the SDSS input catalogue. The cumulative distribution of galaxy colours (left column) and sSFRs (right column) are shown for 4 intervals in stellar mass (rows, as labelled), and as a function of the apparent r-band magnitude m_r, with fainter galaxies corresponding to lighter-coloured curves. For galaxies with $9 < \log_{10}(M_\star/\mathrm{M}_\odot) < 9.5$, there is a bias toward faint, red galaxies (top left panel), while more massive galaxies are biased toward faint, active galaxies (right panels, rows 2-4).

as these can affect the timescale t_{mid}; PPS-independent biases will primarily affect f_{after} and f_{before}. We consider as an example a single Markov chain, corresponding to a given interval in M_{host} and $M_\star$. We now suppose that in the input galaxy sample red galaxies are preferentially included relative to blue ones. If this occurs uniformly across the (R, V) PPS plane, the result will be a lowering of the estimates for f_{before} and f_{after}. The t_{mid} parameter, on the other hand, is tied to how long each satellite has been orbiting its host, and information about orbital phase comes exclusively from the PPS coordinates, so t_{mid} can only be affected by a bias if it is not uniform across PPS.

In the context of the quenching analysis, we have checked whether galaxies of a given stellar mass are preferentially included in the sample as a function of either their $(g - r)$ colour or sSFR by comparing the distributions of these two quantities change as a function of apparent r-band magnitude m_r, shown in Fig. 11. We find that for stellar masses $M_\star > 10^{9.5}\,\mathrm{M}_\odot$, the $(g - r)$ colour distribution is very close to independent of m_r, however for lower mass galaxies there is a relative overabundance of faint red galaxies in the catalogue. This is what motivates our stellar mass threshold. Curiously, the sSFR distributions show an opposite behaviour: there is no apperent bias as a function of m_r for low mass galaxies, but there is an overabundance of massive, faint active galaxies. Wishing to use the same galaxy sample for both the colour and sSFR analysis, we could find no way of mitigating both biases simultaneously and so have accepted that the active fractions (f_{after} and f_{before}) may be slightly too high overall. We see no reason that, and find no evidence that, these biases should vary with PPS position, so our quenching t_{mid} estimates should be unaffected. We have not identified any biases which depend on the PPS coordinates in the SDSS input catalogue.

The situation for the stripping analysis is somewhat less

clear, and more difficult to assess. Whereas in the optical catalogues there are numerous galaxies detected outside our adopted stellar mass and apparent magnitude limits which are useful in assessing possible biases in the catalogue, since we are directly using detection in ALFALA as a tracer of gas content, a similar assessment is more challenging. This is further compounded by the relatively small total number of detections (within the $2.5r_{\rm vir}$ and $2\sigma_{\rm 3D}$ aperture around our sample of host systems). In a simplified scenario where the presence of a galaxy in the ALFALFA source catalogue depends only on its H I mass and its distance our approach is robust: a galaxy of given stellar mass and in a given group (i.e. at a given distance) has the same probability of being detected regardless of its R and V coordinates, making our measurement of $t_{\rm mid}$ for neutral gas stripping reliable[11]. In reality, however, other factors influence the detection of sources in the ALFALFA survey. As an example, we consider the possible fates of H I gas which ceases to be detected in a satellite galaxy – it may have simply been removed, or it may be removed and subsequently ionized, or it may be ionized in place, or it may condense into the molecular phase. Given the poor spatial resolution of the measurements, it is plausible that displaced gas which remains neutral could keep a galaxy above the detection threshold even when the gas is no longer 'inside' the satellite. Since the thermodynamic properties of the ambient gas affect the details of how H I disappears from detectability, and vary as a function of PPS and $M_{\rm host}$, our measurements must be affected at some level. As another example, source confusion in the ALFALFA survey can cause neutral gas in neighbouring galaxies to overlap within the (rather large) beam. This could push a gas-poor galaxy above the detection threshold. The overall confusion rate in ALFALFA is ≤ 5 per cent (Jones et al. 2015), but crowding of H I-bearing galaxies is likely to be more severe in the denser regions, especially in the gas-rich group environment.

The above discussion of possible biases in ALFALFA serves to highlight some of the trends which should ideally be considered. However, given the limitation of the data, we find ourselves unable to fully explore these issues. Our overall assessment is that none of these is likely to drive the several-Gyr differences in $t_{\rm mid}$ needed to affect our main qualitative conclusions. Nevertheless, we stress that our measurements are a first attempt which can be revisited and improved as larger, more sensitive, and better resolved H I surveys become available.

6.1.2   *Compatibility of orbit libraries*

One of the key assumptions underpinning our analysis is that the ensemble of orbits drawn from the N-body simulation (see Sec. 3.2) is representative of the ensemble of orbits occupied by the observed galaxies. For instance, satellites heavily stripped of dark matter may still appear as SDSS detections – the stellar component of galaxies is centrally concentrated and more tightly bound than the bulk of the dark matter – however, the analogous objects in the N-body simulation (where there is no stellar component) may have their dark matter haloes fully disrupted and thus fail to appear in our list of orbits. This effect is

more important for low mass galaxies. In our N-body simulation, haloes of $\log_{10}(M_{\rm vir}/{\rm M}_\odot) < 10.5$ are made up of < 50 particles; below this limit the halo finder begins to struggle to identify them, and of course once only a few particles remain a halo will be 'artificially disrupted', even though a bound core would remain in a realization with higher numerical resolution (van den Bosch & Ogiya 2018, but see Bahé et al. 2019). A satellite halo which falls in with a mass of $\log_{10}(M_{\rm vir}/{\rm M}_\odot) = 11.5$ can therefore be stripped of ~ 90 per cent of its mass before disappearing from the catalogue, while a more massive $\log_{10}(M_{\rm vir}/{\rm M}_\odot) = 12.5$ halo would continue to appear until ~ 99 per cent of its initial mass is stripped. We have examined the distribution of stripped mass fractions as a function of maximum halo mass for satellites in our N-body simulation, and find that[12] at $\log_{10}(M_{\rm sat}/{\rm M}_\odot) = 12.5$, only 5 per cent of (surviving) satellites have been stripped of > 90 per cent of their peak mass (the median galaxy with $M_{\rm sat} = 10^{12}$ M$_\odot$ has been stripped of ~ 35 per cent of its mass at $z = 0$), even though any stripped of more than this will continue to be tracked while they lose another decade in mass. It is therefore only in the lowest stellar mass bin in our analysis of SDSS galaxies ($M_\star \sim 5 \times 10^9$ M$_\odot$, corresponding to $M_{\rm sat} \sim 3 \times 10^{11}$ M$_\odot$) that a significant number (~ 20 per cent) of orbits will be missing due to artificial disruption. These will of course preferentially be the orbits which have the earliest $t_{\rm fp}$'s, which occupy a very biased region of PPS, at low R and low V. The net effect on the parameters of our model is to bias $t_{\rm mid}$ to earlier times (lower values). We assess the magnitude of this bias by artificially degrading the resolution of our orbit catalogue by $1\,{\rm dex}$ in mass and repeating our analysis. We find that $t_{\rm mid}$ is underestimated by up to $3\,{\rm Gyr}$ at $M_\star \lesssim 5 \times 10^{10}$ M$_\odot$. We stress that this bias only significantly affects the leftmost point on each curve in Fig. 10, and, encouragingly, these points do not seem to be significantly or systematically offset from those at higher $M_\star$.

As a crude upper bound on the systematic error budget due to the compatibility of the orbit libraries with the orbital distribution of observed galaxies, we report the result of our analysis when we match to simulated haloes based on their $z = 0$, rather than their peak, mass. This results in essentially all satellites being assigned orbits which should belong to higher mass satellites. Since these preferentially fall in at later times, this results in a large systematic underestimate of $t_{\rm mid}$. However, even this gross mis-assignment of orbits caused an offset of $\lesssim 3\,{\rm Gyr}$ in $t_{\rm mid}$, and occurred nearly uniformly across the entire range in both host and stellar mass, leaving our qualitative conclusions unchanged and lending some additional confidence in their robustness against these types of biases in orbit assignment.

6.1.3   *Statistical considerations*

Moving on to statistical, rather than systematic, considerations, we performed an additional set of model parameter estimates to check for a 'preferred solution' to which the model parameters might converge, for instance due to the mathematical formulation of the model or the choice of priors, rather than being driven by the evidence in the data. We check this by repeating the parameter constraints for the $10^{14} < M_{\rm host}/{\rm M}_\odot < 10^{15}$

---

[11] We recall that by using detection as a proxy for gas-richness, we must abandon the meaning of $f_{\rm before}$ in the absolute sense: many 'gas-rich' galaxies will appear in the catalogue as non-detections simply because they are distant, which will bias $f_{\rm before}$ to lower values.

[12] We recall that $M_{\rm sat}$ is defined as the maximum mass which a satellite halo has had at any (past) time.

galaxy sample for the $(g - r)$ colour analysis. However, before evaluating the Markov chains, we randomly 'shuffle' the colours of the galaxies within each stellar mass bin, such as to destroy any correlation between galaxy colour and PPS coordinates while preserving all other properties of the galaxy sample. The parameter constraints in this case are characterized by a distribution for t_{mid} which is very broad and prefers large values, specifically extending all the way to the upper limit of the prior distribution ($t_{\mathrm{mid}} = 10\,\mathrm{Gyr}$) with either a flat shape at large values, or a peak at the prior bound. This is intuitive: if colour (or any other property) and PPS position are uncorrelated, there is no evidence that the host environment impacts the colour of the satellites – the very late t_{mid} represents satellites orbiting for a long time within their host without changing their colour. This 'preferred solution' seems to be the one reached for the highest stellar mass bin in the $10^{12} < M_{\mathrm{host}}/\mathrm{M}_{\odot} < 10^{13}$ and $10^{13} < M_{\mathrm{host}}/\mathrm{M}_{\odot} < 10^{14}$ galaxy samples (e.g. rightmost blue and green points in lower left panel of Fig. 10). This interpretation is corroborated by the $f_{\mathrm{HI\,detected}}$ distribution in PPS for these two sets of galaxies (see figures in Appendix A) which do not show a clear gradient in PPS. We therefore do not consider these two t_{mid} estimates reliable.

6.1.4 Realism of the model

Our model, as summarized in Fig. 6, clearly cannot be a perfect description of the real time-evolution of the blue/active/gas-rich fraction in dense environments. It does not have enough freedom in shape to accomodate the full spectrum of possibilities: it assumes that the fraction does not evolve outside the time interval defined by t_{mid} and Δt, and that the times when individual galaxies make the transition from blue/active/gas-rich to red/passive/gas-poor are uniformly distributed over the Δt interval, leading to a linear decline. We showed in Sec. 4.3 that we are able to recover all parameters perfectly, within the statistical uncertainties, when the model is an exact description of the data. However, in the more 'realistic' test using the Hydrangea clusters, the Δt parameter, in particular is not accurately recovered (the differences in f_{before} are accounted for by the limited size of the zoom-in region of these simulations, as explained in Sec. 6.1.1 above). We have not found any other plausible explanation in the course of the various tests and method variations which we have carried out, so we tentatively attribute this failure to reliably recover the Δt timescale to the inevitable mismatch between the form of the model and the underlying 'truth' encoded in the data. The mismatch cannot be too severe, however, as evinced by the excellent recovery of the other parameters illustrated in Fig. 9. We note that this is in part due to a careful formulation of the model – a mathematically equivalent formulation which replaces t_{mid} and Δt with alternative parameters $t_{\mathrm{start}} = t_{\mathrm{mid}} - \frac{1}{2}\Delta t$ and $t_{\mathrm{end}} = t_{\mathrm{mid}} + \frac{1}{2}\Delta t$ introduces a strong degeneracy between the two 'time' parameters and allows the uncertainty in the width of the transition (Δt) to wash out the tight constraint on its timing (t_{mid}).

6.1.5 Summary

Taken together, our assessment of the overall implications of the various biases and uncertainties discussed in this section are:

• There may be significant systematic offsets in t_{mid}, of up to $\sim 3\,\mathrm{Gyr}$, but we find that most such possible offsets tend to apply approximately uniformly at all stellar masses and host halo masses. This implies that our recovery of the ordering of the transitions – gas-rich to gas-poor, followed significantly later by active to passive, and then almost immediately by blue to red – is most likely a robust result. We are similarly confident in our conclusion that gas stripping and quenching proceed somewhat more quickly in more massive host haloes.

• The highest stellar mass bins in the gas stripping analysis for the low and intermediate host mass bins appear to correspond to the 'preferred solution' of the model in the absence of evidence and are unlikely to represent reliable measurements of the model parameters. This is corroborated by the absence of a visible gradient in $f_{\mathrm{HI\,detected}}$ in PPS.

• The limited resolution of the N-body simulation, which causes low-mass satellite haloes to be disrupted too early, is likely to be driving an underestimate of t_{mid} in the lowest stellar bin of each host halo mass bin. This could mask a decreasing trend of t_{mid} with increasing stellar mass, a point to which we will return in Sec. 6.2, below.

6.2 Comparison with prior studies

6.2.1 Oman & Hudson (2016, OH16)

In terms of methodology, the previous study most similar to ours is that of OH16. The first crucial difference between the two analyses is that they use the infall time (defined as the first inward crossing of $2.5r_{\mathrm{vir}}$) as a reference '$t = 0$' from which to measure the timescale for quenching, while we use the time of the first pericentric passage. We would therefore expect our t_{mid} values, specifically for the sSFR input, to be smaller than theirs by approximately the time taken for a satellite to orbit from infall to first pericentre, about 3–$4\,\mathrm{Gyr}$. However, comparing their measurements of $t_{1/2}$, also the time when half of the satellite population has 'transitioned', to our t_{mid} (the dashed and solid red lines in their fig. 9 correspond closely in terms of galaxy selection to the green and red dashed lines in our Fig. 10), our measurements are only $\sim 1\,\mathrm{Gyr}$ shorter. In order to unambiguously determine the origin of this apparent discrepancy, we have repeated our analysis with four changes, implemented one at a time, which together lead to a quantitative reproduction of the result of OH16. The first two, which change our measurement very little, are to adopt their stellar mass binning, and replace our linear decline model (Eq. 1) with their exponential decline model (their eq. 12). Next, we replace the probability distributions used to estimate t_{fp} with similar distributions to estimate the infall times of satellites. This results in the expected $\sim 3.5\,\mathrm{Gyr}$ upward shift in t_{mid}. Finally, we reproduce their use of the $z = 0$ satellite halo masses in the N-body simulation to associate possible orbits to observed satellites (whereas we use the maximum past halo mass of satellite haloes), which results in a $\sim 2.5\,\mathrm{Gyr}$ downward shift in t_{mid}. These changes together bring the two analyses into agreement at $\log 10(M_{\star}/\mathrm{M}_{\odot}) \gtrsim 10.5$; the remaining discrepancy at the low mass end is consistent with being due to their (erroneous) use of a lower resolution N-body simulation (see Sec. 6.1.2). In summary, our finding that the characteristic time for quenching, t_{mid}, is several Gyr after the first pericentric passage, rather than around or just after this even as reported by OH16, is due to improvements in model assumptions, and that the present study should be taken as superseding this earlier measurement.

6.2.2   Wetzel et al. (2013, W13)

We next compare our measurement to the conclusions of W13, who presented the first empirical evidence for the 'delayed-then-rapid' quenching scenario. We focus in particular on comparison with this study (i) because the differences we explain below serve to highlight many important systematic dependencies on model assumptions and (ii) because it is recognized as a landmark result of the field. Our methodology differs from theirs on a few key points. First, whereas the 'rapid' nature of the blue-to-red/active-to-passive/gas-rich-to-gas-poor transition is as assumption[13] in our method, W13 constrain the 'fading timescale' ($\tau_{\rm Q,fade}$ in their notation), on which the sSFRs of individual satellites decline, directly. $\tau_{\rm Q,fade}$ should not be confused with our Δt, which represents the interval over which a collection of satellites each cross our sharp division of the population. Second, in their analysis, 'infall' is defined as the time when a galaxy first becomes a satellite (defined via friends-of-friends association with a more massive system) of *any* host, in contrast to our use of the first pericentric passage within the (progenitor of) the *current* host system. Our initial expectation is then that our measurements of $t_{\rm mid}$ (for our sSFR analysis), loosely comparable to their $t_{\rm Q}$ (not $t_{\rm Q,delay}$), should be uniformly shorter, as their delay time includes the time in any previous hosts, and the time to orbit from infall to first pericentre.

The fact that our $t_{\rm mid}$ measurements (Fig. 10, bottom centre) are similar to or larger than their $t_{\rm Q}$ measurements (their fig. 8, upper panel) therefore merits careful consideration. First, we note that the host-mass ordering of the curves in the two figures differs: whereas we find a monotonic increase in $t_{\rm mid}$ with decreasing $M_{\rm host}$ at low $M_\star$, W13 find the longest $t_{\rm Q}$ in intermediate mass hosts; both figures agree that the delay time is approximately constant with $M_{\rm host}$ at high $M_\star$. The reason for this difference is clear from their fig. 2: in higher mass hosts, the first infall is much earlier than the most recent infall, while in low mass hosts this difference is very small. An approximate correction for this would involve shifting their low/intermediate/high host mass measurements down by $\sim 0.5/2/3\,{\rm Gyr}$, respectively, which results in the same host-mass ordering as for our measurement. We therefore agree that the *isolated* environmental influence (i.e. excluding group pre-processing) of more massive hosts is felt more quickly by their satellites (for further discussion see also W13, sec. 4.3.1).

The next clear difference is the slope of the curves – whereas we find generally near-flat dependences of $t_{\rm mid}$ on $M_\star$, W13 find a strongly declining slope. The most plausible explanation for this apparent discrepancy lies in the differences between our respective treatments of group pre-processing. Whereas W13 'start the clock' for the quenching delay time when a galaxy first becomes a satellite of any host, we use the first pericentric passage in the present-day host as a reference time. For massive satellite galaxies, these two definitions turn out to be at least roughly equivalent (except for the time interval required to orbit from infall to pericentre): the majority of more massive satellites fall into their hosts at later times (by about $1.5\,{\rm Gyr}$ between $\log_{10}(M_\star/{\rm M}_\odot) = 9.7$ and 11.3) and, where

they are quenched as satellites, are quenched in their present-day host (W13, especially their fig. 10). Lower mass galaxies, however, tend to have earlier 'first infall' times (e.g. into a group that will later fall into a cluster), such that even if they reach their final host un-quenched, they are already partially 'processed' and more vulnerable to the environmental influence of the present-day host. This 'pre-processed' population will have preferentially shorter quenching times (as defined in terms of time spent in the present-day host), flattening out the trend with $M_\star$ relative to the measurement of W13.

Differences in the trend with $M_{\rm host}$ and $M_\star$ now having been explained, we are left with any differences in the absolute normalization of $t_{\rm mid}$ (or, in the notation of W13, $t_{\rm Q}$) left to be accounted for. The easiest sub-sample of galaxies to use for this is the high-$M_\star$ end of the satellites with $10^{14} < M_{\rm host}/{\rm M}_\odot < 10^{15}$. Here, offsets due to the different definitions of infall time, and different handling of group pre-processing, should be minimal, as outlined above. We would then expect our measurement to be perhaps $2\,{\rm Gyr}$ shorter than that of W13, to account for the time to orbit from infall (defined in terms of a friend-of-friends membership criterion) to first pericentre. However, the difference runs in the opposite sense, with our typical quenching time lagging theirs by $\sim 2\,{\rm Gyr}$ (for this particular $M_{\rm host}$ and $M_\star$). We have found no entirely satisfactory explanation for this discrepancy, and will return to this point in the summary (Sec. 6.2.4) below.

6.2.3   Rhee et al. (2020)

Rhee et al. (2020) use the SFR distribution of disc galaxies as a function of their PPS coordinates to derive a quenching timescale, under the assumption that the earliest infalling galaxies have the lowest star formation rates. As in our analysis, they consider the environmental influence of the present-day host independent of (or, perhaps more accurately, over-and-above) pre-processing. However, rather than an explicit definition of quenching, they adopt a simple model for the star formation history as a function of time since infall in order to constrain two exponential decay timescales for the SFR, corresponding to outside ($\tau_{\rm ex-situ}$) and inside ($\tau_{\rm cluster}$) the final host, and a delay timescale ($t_{\rm d}$) representing the time since infall before the transition between the two decay timescales occurs[14]. Rhee et al. (2020) helpfully provide their measurements following the definitions of OH16 in their fig. 11; the $t_{\rm Q}$ given there is directly comparable to our $t_{\rm mid}$, except for an offset to higher $t_{\rm Q}$ corresponding to the typical difference between the infall (defined at $2.5 r_{\rm vir}$) and first pericentre times, which is $3.8\,{\rm Gyr}$ (independent of $M_\star$), for a host mass range $5 \times 10^{13} < M_{\rm host}/{\rm M}_\odot < 10^{15}$ – most similar to the red line in the bottom centre panel of Fig. 10. It is immediately clear that our $t_{\rm mid}$ well exceeds their (adjusted) $t_{\rm Q}$ at all stellar masses – by $\sim 2\,{\rm Gyr}$ at the low stellar mass end, up to $\sim 5\,{\rm Gyr}$ at the high stellar mass end. This is further exacerbated by the fact that our $t_{\rm mid}$ likely corresponds to that for galaxies just now quenching[15], whereas Rhee et al. (2020) estimate the $z \sim 0$ value of

---

[13] Actually, we only assume that galaxies cross our sharp deliniation of blue/red (or similar) rapidly, so we are not sensitive to the equivalent of the 'fading time' of W13, which is constrained mostly by the shape of the sSFR distribution, which we have reduced to the fraction $f_{\rm active}$.

[14] They also constrain a parameter α describing the redshift evolution of the cluster quenching timescale, but this turns out to be consistent with no evolution in all cases (although the confidence intervals are rather broad).

[15] Suppose that active galaxies in our sample which are just now falling

t_Q assuming a $t_Q(z_{inf}) \propto (1 + z_{inf})^{-1.5}$ dependence on the infall redshift[16], motivated by the redshift dependence of the crossing timescale of host systems. We might therefore expect our quenching timescale to be about half of their measurement, though this is hard to ascertain precisely given the substantial ambiguity in defining the reference time from which any delay should be measured.

6.2.4 Summary

At a glance, the apparent quantitative agreement – once differences due to the various definitions adopted are accounted for – between the quenching timescales measured by W13, OH16 and Rhee et al. (2020) would suggest that our measurements are outliers and perhaps suspect. However, as explained in Sec. 6.2.1, by systematically adjusting one aspect of our analysis at a time until we adopt an identical set of assumptions to OH16, we can reproduce their measurements in quantitative detail. This approach reveals that the discrepancy can be explained in terms of (i) their assignment of orbits to satellites neglecting tidal stripping of dark matter, and (ii) their (erroneous) use of an N-body simulation with inadequate numerical resolution, suggesting that our measurements are not 'simply outliers'. We are further encouraged by the excellent agreement between the 'true' t_{mid} values – which are fully independent of the t_{fp} probability distributions derived from the N-body simulation – and the estimates derived from 'observed' quantities, as shown in Fig. 9.

Nevertheless, making the link between the stellar mass of satellites and which candidate haloes' orbits should be selected based on their mass in the N-body simulation remains the most challenging aspect of our approach, and directly impacts the absolute calibration of t_{mid}. W13 and Rhee et al. (2020) also cite difficulties in this area. W13 discuss at length (e.g. their appendix A) the assumptions required to account for the continued stellar mass growth of satellites while their dark matter haloes are being tidally stripped, and Rhee et al. (2020) need to work around a factor of ~ 2 mismatch in the satellite stellar mass function between their sample of observed satellites and their cosmological hydrodynamical simulations of clusters. A fully self-consistent treatment requires simultaneously accounting for: (i) the dark matter mass loss after infall; (ii) stellar mass growth due to continued star formation between infall and quenching; (iii) stellar mass loss due to tides (though this is likely to be a minor effect, see Bahé et al. 2017a) and supernovae/winds from massive stars. In addition to varying strongly as a function of the (unknown) stellar (or halo) mass at infall, and the (unknown) infall time via the redshift-dependence of the stellar-to-halo mass relation, the stellar mass growth in particular also depends on the quenching timescale. These strong couplings between unknown parameters require either additional assumptions, or an increase in the dimensionality of the parameter space through the introduction of additional 'nuisance parameters'. Realistic implementations of both approaches in our

framework would be associated in a significant increase in computational cost for each model evaluation, motivating us to proceed with our simplified approach, for the present.

With the above in mind, we focus our interpretation below on what we perceive as the strengths of our analysis: a uniform input galaxy sample across multiple tracers (colour, Balmer emission lines and H I) yielding robust estimates of the relative timescales associated to quenching and stripping as traced by each, and the use of the first pericentre as a reference time offering better discrimination between processes which occur at or well away from first pericentre than an 'infall' reference time.

6.3 Physical interpretation

To guide our interpretation, we begin with an illustrative example. According to our parameter constraints, a galaxy of $M_\star \sim 10^{10}\,\mathrm{M_\odot}$ in a massive host system ($M_{host} > 10^{14}\,\mathrm{M_\odot}$) typically continues star formation for $\sim 5\,\mathrm{Gyr}$ after it is no longer detected in ALFALFA. To remain above our threshold sSFR of $10^{-11}\,\mathrm{yr^{-1}}$ over this interval, it must grow its stellar mass by at least $\Delta M_\star \sim 5 \times 10^8\,\mathrm{M_\odot}$. While a typical galaxy with this stellar mass that is detected in ALFALFA has $M_{HI} \sim 10^{10}\,\mathrm{M_\odot}$, by the time it is undetected it likely has $M_{HI} \lesssim 5 \times 10^9\,\mathrm{M_\odot}$, so the star formation 'efficiency' during this time interval must be $\Delta M_\star / M_{HI} \gtrsim 0.1$. At a glance, this value suggests that the gas supply is ample to fuel the required star formation. However, the sSFR is unlikely to remain just at the threshold value – more likely the average over the $5\,\mathrm{Gyr}$ interval is a factor of a few higher. Furthermore, star formation is not a highly efficient process: as stars are formed, some gas is launched as a wind with a mass loading $\eta = \dot{M}_{wind}/\dot{M}_\star$. Within a cluster, the gas in the wind is unlikely to be able to return to fuel later star formation as any wind launched beyond the ISM is exposed to the harsh intra-cluster medium. The value of η is debated, but is likely ~ 1 for galaxies of this mass (e.g. McGee et al. 2014, and references therein), so that at most $\sim \frac{1}{2}$ the available gas mass can ultimately end up in stars. Together, these considerations leave rather little room for H I gas to be removed by other processes, such as ram pressure or tides, over the same time interval.

It is clear that the ISM gas feels the influence of such massive host systems relatively early: there are numerous examples of satellites with prominent gas tails generally agreed to be on their first approach to their host (e.g. Chung et al. 2009; Smith et al. 2010; Yoon et al. 2017; Jaffé et al. 2018; Roberts & Parker 2020). One physical picture consistent with this, and our measurements, is that as a satellite falls into a cluster, it is stripped of H I by ram pressure on its initial approach – not completely, but enough that it is no longer detected in ALFALFA by the time it reaches pericentre. RPS then ceases as soon as the satellite passes pericentre and the ram-pressure force rapidly drops off. The remaining, centrally concentrated H I then continues to fuel star formation, driving winds which carry away some of the gas. It is unclear whether it is the eventual depletion of the gas supply by this process, or a later episode of stripping, which finally quenches the galaxy. Indeed, based on the timescales involved and the scatter in galaxy and orbital properties, both of these possibilities probably occur with non-negligible frequency. This picture is also consistent with the observations discussed in Owers et al. (2019), who found a significant population of cluster satellites, likely on their first orbital passage, with recently

---

into groups will become passive (arbitrarily) 1 Gyr earlier than would be expected from our measured t_{mid}. Since the delay time has not yet elapsed for these objects, there is no evidence for this change in timescale contained in the data – our measurement cannot be sensitive to it.

[16] This is very close to $t_Q(t_{inf}) \propto t_{inf}$, where t is the age of the Universe.

quenched outer discs but continued star formation in their central regions.

We have so far neglected another reservoir of fuel for star formation to which the observations we have used in our analysis are blind: the molecular gas phase. Around the time of infall, this reservoir is unlikely to contain enough additional gas to make much difference to the picture outlined above – the molecular-to-atomic gas ratio of our example galaxy would be of about 0.15 (Catinella et al. 2018). However, Moretti et al. (2020) suggest that atomic hydrogen may be highly efficiently converted to H_2 by ram pressure-driven compression in 'jellyfish' galaxies. This leaves the detailed evolution of the satellite galaxies somewhat ambiguous – if a substantial amount of H I is fed into the molecular reservoir in this way, then the long delay before quenching becomes less surprising, and subsequent RPS or tidal stripping may even be *required* to truncate star formation. The ambiguity could be alleviated by (i) spatially resolved observations of the H I gas sufficiently deep to reveal the extent to which RPS removes atomic gas, e.g. by measuring how much is present in a stripped tail of gas (see Deb et al. 2020, for one example where the majority of H I is contained in a tail), and (ii) observations of molecular gas tracers, such as dust or CO, to determine when along satellite orbits the content of this reservoir increases or is depleted.

Turning our attention next to lower mass hosts, we consider the same example satellite galaxy around a 10^{13} to 10^{14} M$_\odot$ host. Such satellites quench slightly (~ 1 Gyr) later than those around more massive hosts, but become undetected in ALFALFA much later (~ 3 to 4 Gyr, and thus *after* the first pericentric passage[17]), such that the interval between these two events is shorter, about 3 Gyr. Thogether, this suggests that RPS is insufficient to push these galaxies below the ALFALFA detection threshold on their first passage through their host. That the transition from gas-rich to gas-poor occurs near the first apocentric passage, when the RPS rate is likely to be near zero – any gas that could be stripped should already have been removed when the ram pressure peaked during the first pericentric passage – suggests that it is more likely consumption of gas by star formation, rather than removal through stripping, which carries these satellites across the ALFALFA detectability threshold. Conversely, that these galaxies quench only ~ 3 Gyr later leaves perhaps too little time to consume *all* the remaining gas in forming stars, suggesting that the second peak in ram pressure and tidal forces at the second pericentric passage may be the ultimate driver of the shutdown of star formation in these satellites.

The situation in the lowest mass hosts is more ambiguous. The statistical errors are larger, and t_{mid} becomes very long (~ 7 Gyr for the transition in colour). We recall that our methodology is only sensitive to transitions which have actually occured by the time of observation – essentially, it finds the time along the orbits of satellites *actually present in their hosts* when there is a transition in the gas content/SFR/colour of the satellites. If this transition has yet to occur, our choice to fix $f_{\mathrm{after}} = 0$ is incorrect. In the case of 10^{12} to 10^{13} M$_\odot$ hosts, when f_{after} is left free to vary (see Appendix B), f_{after} shows some preference for non-zero values, suggesting that the satellite population has not yet had time to be completely 'pro-

---

[17] Though this point is difficult to establish conclusively, as it is sensitive to the treatment of f_{after}, see Appendix B.

cessed' by these hosts. In this case, the t_{mid} values drop by 1 to 2 Gyr across all tracers. In any case, the statistical errors on all parameters remain large, likely due to the increased uncertainty in the centre (in position and velocity) of these groups, and their velocity dispersions, along with the generally weaker influence of these hosts on relatively massive satellites (as opposed to e.g. dwarf spheroidals), causing us to remain cautious in our physical interpretation of these systems.

Finally, we comment briefly on the generally flat trend as a function of $M_\star$ of our t_{mid} measurements, across all host masses and tracers. We recall that this is intimately connected to our treatment of 'pre-processing' (see Sec. 6.2.2). This flatness leads us to two conclusions. First, it suggests that, although lower-mass hosts can eventually quench their satellites, if these fall into a more massive host, the influence of the new, denser environment will usually be sufficiently harsher that it will itself set the timescales for stripping and quenching, regardless of the earlier processing. Second, it seems that it is ultimately the external influence of the host system which determines these timescales, independent of the mass of the satellite, unless it arises as a result of some fine-tuning between the scalings of sSFR, gas fraction, and other galaxy properties with stellar mass.

7   SUMMARY AND CONCLUSIONS

We have used a combination of galaxy photometry and spectroscopy from the SDSS, H I fluxes from the ALFALFA survey, and orbital information inferred from tracking haloes in an N-body simulation to constrain a simple model linking the star formation and gas properties of satellites with their orbital histories. In the context of discriminating between the various physical processes contributing to the shutdown of star formation in satellites, the t_{mid} and Δt parameters of our model are the most informative. t_{mid} is the median time relative to the first pericentric passage within the present-day host when satellites transition from blue to red, from active to passive, or from gas-rich (defined as detected in the ALFALFA survey) to gaspoor, while Δt describes the (full width) scatter in this transition time; this latter parameter is poorly constrained in our analysis and treated as a 'nuisance parameter'. Built into our methodology is the assumption that, provided it orbits within its host for 'long enough' (as expressed by the parameter combination $t_{\mathrm{mid}} + \frac{1}{2}\Delta t$), all satellites are ultimately stripped of gas and are quenched. Our main results are summarized as follows:

• We clearly detect the separation in time between the three stages of environmental processing probed by our analysis: first, neutral gas ceases to be detected, followed a few Gyr later by a drop in sSFR (traced by the disappearance of Balmer emission lines associated to star formation), and in turn $\lesssim 1$ Gyr later by reddening in $(g - r)$. This ordering is ubiquitous across hosts from $M_{\mathrm{host}} = 10^{12}$ to 10^{15} M$_\odot$, though the time intervals between the stages are somewhat longer in more massive hosts.

• At fixed host mass, t_{mid} associated to each tracer (H I, sSFR, colour) is remarkably independent of satellite stellar mass. We note that this is intimately connected to our treatment of 'pre-processing' (see Sec. 6.2.2).

• In massive hosts ($M_{\mathrm{host}} \gtrsim 10^{14}$ M$_\odot$), neutral hydrogen disappears before or around the first pericentric passage, while in lower mass hosts it remains detectable up to several Gyr later.

This difference is not driven by a distance or similar bias in the sample of host systems.

• Star formation persists in a typical satellite for up to ∼ 5 Gyr after the first pericentric passage, by which time most satellites will be somewhere between second pericentre and having completed multiple orbits within the host.

Our measurements are broadly consistent with the 'delayed-then-rapid' quenching scenario (De Lucia et al. 2012), however we infer a delay timescale which is much longer than most other studies (W13; OH16; Rhee et al. 2020, in particular). Detailed tests of our methodology on a mock galaxy sample drawn from the Hydrangea cosmological hydrodynamical simulation suite suggest that our recovery of t_{mid} is accurate (see Sec. 4.3), but the absolute calibration of this timescale, and comparison across studies employing various zero-points from which the delay is measured, remains challenging.

Our measurement of a long quenching timescale contrasts with results from contemporary galaxy formation simulations, which predominantly find that hosts – especially rich clusters – quench their satellites rather efficiently on the first orbital passage (e.g. Bahé et al. 2017b with the Hydrangea suite – see also Fig. 5, Lotz et al. 2019 with the Magneticum suite, Arthur et al. 2019 with the 'TheThreeHundred' suite), but the leading reason for this is likely numerical: the cold ISM is notoriously difficult to properly resolve in such simulations (Bahé et al. 2017b).

It may be possible to further improve the general approach of constraining models for stripping and quenching via inference of the orbits of satellites based on simulation 'orbit libraries'. For instance, the distribution of satellite orbits around host systems in different dynamical states differ (Haggar et al. 2020); this could be incorporated into the model constraints based on observable properties of the host systems. Alternatively, or in addition, a more sophisticated model describing the redshift evolution of the satellite and/or host properties could be used, and additional constraints such as the full colour or sSFR distributions could replace our simple, binary classifications – this would amount to an attempt to combine the strengths of several existing studies. However, our impression is that the required sophistication of the models begins to be cumbersome, and that it may be more fruitful to adopt a more direct simulation approach. While the details of hydrodynamical interactions in dense environments remain out of reach of the resolving power of current cosmological hydrodynamical simulations, a semi-analytic or hybrid (e.g. a semi-analytic model implemented on top of a hydrodynamical rather than an N-body simulation) approach may be a more straightforward framework in which to attempt to capture the full breadth of environmental physics affecting the evolution of satellite galaxies in a fully self-consistent manner.

REFERENCES

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Arthur, J., Pearce, F. R., Gray, M. E., et al. 2019, MNRAS, 484, 3968

Bahé, Y. M., & McCarthy, I. G. 2015, MNRAS, 447, 969

Bahé, Y. M., McCarthy, I. G., Balogh, M. L., & Font, A. S. 2013, MNRAS, 430, 3017

Bahé, Y. M., Schaye, J., Crain, R. A., et al. 2017a, MNRAS, 464, 508

Bahé, Y. M., Crain, R. A., Kauffmann, G., et al. 2016, MNRAS, 456, 1115

Bahé, Y. M., Barnes, D. J., Dalla Vecchia, C., et al. 2017b, MNRAS, 470, 4186

Bahé, Y. M., Schaye, J., Barnes, D. J., et al. 2019, MNRAS, 485, 2287

Balogh, M., Eke, V., Miller, C., et al. 2004a, MNRAS, 348, 1355

Balogh, M. L., Baldry, I. K., Nichol, R., et al. 2004b, ApJ, 615, L101

Balogh, M. L., Morris, S. L., Yee, H. K. C., Carlberg, R. G., & Ellingson, E. 1999, ApJ, 527, 54

Balogh, M. L., Navarro, J. F., & Morris, S. L. 2000, ApJ, 540, 113

Balogh, M. L., McGee, S. L., Mok, A., et al. 2016, MNRAS, 456, 4364

Barnes, D. J., Kay, S. T., Bahé, Y. M., et al. 2017, MNRAS, 471, 1088

Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013a, ApJ, 770, 57

Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013b, ApJ, 762, 109

Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013c, ApJ, 763, 18

Blitz, L., & Rosolowsky, E. 2006, ApJ, 650, 933

Boselli, A., & Gavazzi, G. 2006, PASP, 118, 517

Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MN-RAS, 351, 1151

Bryan, G. L., & Norman, M. L. 1998, ApJ, 495, 80

Catinella, B., Saintonge, A., Janowiecki, S., et al. 2018, MN-RAS, 476, 875

Chung, A., van Gorkom, J. H., Kenney, J. D. P., Crowl, H., & Vollmer, B. 2009, AJ, 138, 1741

Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, ApJ, 647, 201

Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, MNRAS, 450, 1937

Crain, R. A., Bahé, Y. M., Lagos, C. d. P., et al. 2017, MNRAS, 464, 4204

De Lucia, G., Weinmann, S., Poggianti, B. M., Aragón-Salamanca, A., & Zaritsky, D. 2012, MNRAS, 423, 1277

Deb, T., Verheijen, M. A. W., Gullieuszik, M., et al. 2020, MN-RAS, 494, 5029

Dressler, A. 1980, ApJ, 236, 351

Eckert, K. D., Kannappan, S. J., Stark, D. V., et al. 2015, ApJ, 810, 166

Fillingham, S. P., Cooper, M. C., Wheeler, C., et al. 2015, MN-RAS, 454, 2039

Fillingham, S. P., Cooper, M. C., Kelley, T., et al. 2019, arXiv e-prints, arXiv:1906.04180

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306

Fossati, M., Wilman, D. J., Mendel, J. T., et al. 2017, ApJ, 835, 153

Gavazzi, G., Consolandi, G., Gutierrez, M. L., Boselli, A., & Yoshida, M. 2018, A&A, 618, A130

Gehrels, N. 1986, The Astrophysical Journal, 303, 336

Goodman, J., & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, Vol. 5, No. 1, p. 65-80, 2010, 5, 65

Gouin, C., Aghanim, N., Bonjean, V., & Douspis, M. 2019, arXiv e-prints, arXiv:1912.00655

Haggar, R., Gray, M. E., Pearce, F. R., et al. 2020, MNRAS, 492, 6074

Haines, C. P., Pereira, M. J., Smith, G. P., et al. 2013, ApJ, 775, 126

—. 2015, ApJ, 806, 101

Haynes, M. P., Giovanelli, R., Martin, A. M., et al. 2011, AJ, 142, 170

Haynes, M. P., Giovanelli, R., Kent, B. R., et al. 2018, ApJ, 861, 49

Hernández-Fernández, J. D., Haines, C. P., Diaferio, A., et al. 2014, MNRAS, 438, 2186

Hogg, D. W., Blanton, M. R., Brinchmann, J., et al. 2004, ApJ, 601, L29

Hou, A., Parker, L. C., & Harris, W. E. 2014, MNRAS, 442, 406

Jaffé, Y. L., Smith, R., Candlish, G. N., et al. 2015, MNRAS, 448, 1715

Jaffé, Y. L., Poggianti, B. M., Moretti, A., et al. 2018, MN-RAS, 476, 4753

Jones, M. G., Papastergis, E., Haynes, M. P., & Giovanelli, R. 2015, MNRAS, 449, 1856

Kauffmann, G., White, S. D. M., Heckman, T. M., et al. 2004, MNRAS, 353, 713

Kelkar, K., Gray, M. E., Aragón-Salamanca, A., et al. 2019, MNRAS, 486, 868

Koribalski, B. S., Staveley-Smith, L., Westmeier, T., et al. 2020, arXiv e-prints, arXiv:2002.07311

Lemaux, B. C., Tomczak, A. R., Lubin, L. M., et al. 2019, MNRAS, 490, 1231

Lim, S. H., Mo, H. J., Lu, Y., Wang, H., & Yang, X. 2017, MNRAS, 470, 2982

Lotz, M., Remus, R.-S., Dolag, K., Biviano, A., & Burkert, A. 2019, MNRAS, 488, 5370

Lu, T., Gilbank, D. G., McGee, S. L., Balogh, M. L., & Gallagher, S. 2012, MNRAS, 420, 126

Mahajan, S., Mamon, G. A., & Raychaudhury, S. 2011, MN-RAS, 416, 2882

Maier, C., Ziegler, B. L., Haines, C. P., & Smith, G. P. 2019, A&A, 621, A131

Mamon, G. A., Sanchis, T., Salvador-Solé, E., & Solanes, J. M. 2004, A&A, 414, 445

Maraston, C. 1999, in Star Formation in Early Type Galaxies, ed. P. Carral & J. Cepa, ASP Conf. Ser. 163, Astron. Soc. Pac., San Francisco, p. 28

McGee, S. L., Bower, R. G., & Balogh, M. L. 2014, MNRAS, 442, L105

Mendel, J. T., Simard, L., Palmer, M., Ellison, S. L., & Patton, D. R. 2014, ApJS, 210, 3

Miyoshi, T., & Chiba, M. 2020, arXiv e-prints, arXiv:2003.07006

Mok, A., Balogh, M. L., McGee, S. L., et al. 2013, MNRAS, 431, 1090

—. 2014, MNRAS, 438, 3070

Moretti, A., Paladino, R., Poggianti, B. M., et al. 2020, arXiv e-prints, arXiv:2006.13612

Muzzin, A., van der Burg, R. F. J., McGee, S. L., et al. 2014, ApJ, 796, 65

Noble, A. G., Webb, T. M. A., Yee, H. K. C., et al. 2016, ApJ, 816, 48

Oman, K. A. 2013, Master's thesis, Univ. Waterloo

Oman, K. A., & Hudson, M. J. 2016, MNRAS, 463, 3083

Oman, K. A., Hudson, M. J., & Behroozi, P. S. 2013, MNRAS, 431, 2307

Owers, M. S., Hudson, M. J., Oman, K. A., et al. 2019, ApJ, 873, 52

Pallero, D., Gómez, F. A., Padilla, N. D., et al. 2019, MNRAS, 488, 847

Pasquali, A., Smith, R., Gallazzi, A., et al. 2019, MNRAS, 484, 1702

Peng, Y.-j., Lilly, S. J., Kovač, K., et al. 2010, ApJ, 721, 193

Ponomareva, A. A., Verheijen, M. A. W., Peletier, R. F., & Bosma, A. 2017, MNRAS, 469, 2387

Rahmati, A., Pawlik, A. H., Raičević, M., & Schaye, J. 2013, MNRAS, 430, 2427

Rasmussen, J., Mulchaey, J. S., Bai, L., et al. 2012, ApJ, 757,

122

Rhee, J., Smith, R., Choi, H., et al. 2020, ApJS, 247, 45

Roberts, I. D., & Parker, L. C. 2020, arXiv e-prints, arXiv:2004.12033

Roberts, I. D., Parker, L. C., Brown, T., et al. 2019, ApJ, 873, 42

Salim, S., Rich, R. M., Charlot, S., et al. 2007, ApJS, 173, 267

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521

Smith, R., Pacifici, C., Pasquali, A., & Calderón-Castillo, P. 2019, ApJ, 876, 145

Smith, R. J., Lucey, J. R., Hammer, D., et al. 2010, MNRAS, 408, 1417

Stevens, A. R. H., Diemer, B., Lagos, C. d. P., et al. 2019, MNRAS, 483, 5334

Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, AJ, 124, 1810

Taranu, D. S., Hudson, M. J., Balogh, M. L., et al. 2014, MN-RAS, 440, 1934

van de Voort, F., Bahé, Y. M., Bower, R. G., et al. 2017, MN-RAS, 466, 3460

van den Bosch, F. C., & Ogiya, G. 2018, MNRAS, 475, 4066

von der Linden, A., Best, P. N., Kauffmann, G., & White, S. D. M. 2007, MNRAS, 379, 867

von der Linden, A., Wild, V., Kauffmann, G., White, S. D. M., & Weinmann, S. 2010, MNRAS, 404, 1231

Wang, J., Bose, S., Frenk, C. S., et al. 2019, arXiv e-prints, arXiv:1911.09720

Wetzel, A. R., Tinker, J. L., & Conroy, C. 2012, Monthly Notices of the Royal Astronomical Society, 424, 232

Wetzel, A. R., Tinker, J. L., Conroy, C., & van den Bosch, F. C. 2013, MNRAS, 432, 336

—. 2014, MNRAS, 439, 2687

Wheeler, C., Phillips, J. I., Cooper, M. C., Boylan-Kolchin, M., & Bullock, J. S. 2014, MNRAS, 442, 1396

Yoon, H., Chung, A., Smith, R., & Jaffé, Y. L. 2017, ApJ, 838, 81

APPENDICES

We include as supplementary materials, available from the publisher:

• Appendix A: Figures similar to Fig. 3 for all bins in host mass and satellite stellar mass.

• Appendix B: Expanded version of Fig. 10 including Δt, and alternative version of Fig. 10 where f_{after} is a free parameter.

• Appendix C: Figures similar to Figs. 7 and 8 for the other stellar mass bins included in Fig. 9.

• Appendix D: Tables with the median and 68 per cent confidence intervals for all the marginalized posterior probability distributions shown in the supplementary Figs. B1–B4 (i.e. including those from Fig. 10).

• Appendix E: Figures showing the individual and pairwise marginalized probability distributions (i.e. similar to Fig. 8) for the parameters summarized in Fig. 10, and similar constraints when f_{after} is a free parameter.