ABSTRACT
We study the quenching of satellite galaxies by gradual depletion of gas due to star formation, by ram-pressure stripping and by tidally triggered starburst. Using progenitors constrained by the empirical model of Lu et al., in which outflow loading factor is low, we do not find an over-quenching problem in satellites even if there is no further cold gas supply from the cooling of the halo gas after a galaxy is accreted by its host. Gradual depletion alone predicts a unimodal distribution in specific star formation, in contrast to the bimodal distribution observed, and under-predicts the quenched fraction in low mass halos. Ram-pressure stripping nicely reproduces the bimodal distribution but under-predicts the quenched fraction in low-mass halos. Starbursts in gas-rich satellites triggered by tidal interactions with central galaxies can nicely reproduce the quenched satellite population in low-mass halos, but become unimportant for low-mass satellites in massive halos. The combined processes, together with the constrained progenitors, can reproduce the observed star formation properties of satellites in halos of different masses.

Key words: galaxies: general - galaxies: formation - galaxies: interstellar medium - dark matter - method: statistical

1 INTRODUCTION
The impact of the intra-halo environments on the star formation in satellite galaxies has been noticed for decades (see §15.5 in Mo et al. 2010), and has been investigated in detail using groups/clusters of galaxies. For example, using the group catalogue constructed from SDSS DR2, Weinmann et al. (2006a) examined how the fractions of late and early type satellite galaxies change with galaxy luminosity and the mass of the halo that host the satellites, and found that the early type fraction increases strongly with halo mass. More recently, Wetzel et al. (2012) carried out a similar but more thorough analysis. They found that the specific star formation rate (sSFR) of satellite galaxies in halos with different masses follows a bimodal distribution, but that the sSFR in the star forming sequence shows no strong dependence on halo mass. Wetzel et al. also investigated the excess in the quenched fraction of satellites relative to that of centrals, a quantity that reflects the strength of the impact of the intra-halo environments in quenching star formation, and found that this quantity increases with halo mass but does not depend strongly on the stellar mass of the satellites.

On the theory side, galaxy formation models are still struggling to reproduce the basic observational results regarding the quenching of star formation by environmental effects. By comparing results obtained from galaxy groups with the prediction of the semi-analytical model (SAM) of Croton et al. (2006), Weinmann et al. (2006b) suggested that the simple “strangulation” model adopted by Croton et al. (2006) and many other versions of SAM, in which the diffuse gaseous halo is assumed to be stripped immediately after a galaxy is accreted into a larger halo, significantly over-quenches the satellite population. Font et al. (2008) found that this “over-quenching” problem cannot be solved by a more physically motivated “strangulation” model, but may be solved by assuming that the ejected gas from a satellite by supernova feedback is recycled back to the satellite to sustain further star formation.

In a traditional SAM, a large number of prescrip-
tions and model parameters are adopted to describe a variety of physical processes relevant to the formation and evolution of the galaxy population. This comprehensive nature of the model makes it hard to put tight constraints on specific physical processes, because of the degeneracy among different parts of the model (e.g. Lu et al. 2011a). In particular, even the latest SAMs still have trouble in reproducing the bimodal distributions of the central galaxies in sSFR and color and in their evolution (e.g. Lu et al. 2014a). Since satellite galaxies are believed to be evolved from centrals at higher redshift, the failure of a model in correctly reproducing the properties of central galaxies can also hinder the interpretation of the model prediction for the quenching of satellites. Several investigations have attempted to isolate the problem of satellite quenching from the total evolutionary process of galaxy evolution. For example, Balogh et al. (2000) studied the halo-centric distribution of quenched satellites by assigning the SFRs of present-day centrals to satellites at the time of accretion and modeling the subsequent star formation of satellites with the Kennicutt law (Kennicutt 1998) and the cold gas reservoirs inherited from the progenitor centrals. Wetzel et al. (2013) used a set of observational constraints to initialize the SFRs of satellites, and modeled the subsequent evolution of star formation in the satellites by assuming that a rapid quenching (an exponential decline in SFR) of a satellite occurs at a certain time after it is accreted by a larger halo. These phenomenological models can be used to translate observational constraints into certain physically meaningful quantities, such as the quenching time scale, but they themselves are not linked directly to any particular physical processes.

In this paper, we study the quenching of satellites by intra-halo environments. Our approach is different from those of previous investigations in two aspects. First, the stellar masses, star formation rates and gas contents of the progenitors of satellite galaxies at the time of accretion are set by using the results obtained by Lu et al. (2015b) from an empirical model constrained by a broad range of observations. Second, the subsequent evolution of satellite galaxies is modeled with physically motivated processes. We test our model predictions by comparisons with results obtained from groups of galaxies. The paper is arranged as follows. The properties of the progenitors of satellite galaxies are described in §2. Our models for the evolution of satellite galaxies in dark matter halos are described in §3. Model predictions are presented in §4, and a summary of our main results is given in §5.

Throughout the paper, we use a ΛCDM cosmology with \(\Omega_{m,0} = 0.273\), \(\Omega_{\Lambda,0} = 0.727\), \(\Omega_{b,0} = 0.0455\), \(h = 0.704\), \(n = 0.967\) and \(\sigma_8 = 0.811\). This set of parameters is from the seven year WMAP observations (Komatsu et al. 2011). In addition, we adopt a Chabrier (2003) IMF.

## 2 PROGENITORS OF SATELLITE GALAXIES

We start to trace the evolution of a satellite at the time when it was first accreted into a bigger halo. The stellar mass and star formation rate (SFR) are initialized according to the empirical model of Lu et al. (2014b, 2015a) at the time of accretion. The empirical model links the galaxy stellar mass and SFR to dark matter halos of different masses at different redshifts in a self-consistent way, and is constrained by the observed stellar mass functions measured from various redshift surveys.

Following convention, we refer a halo that has accreted into a bigger host halo as a sub-halo. The distribution of sub-halos in accretion redshift \(z_{\text{acc}}\) and in initial mass \(M_{\text{i},i}\) is obtained from halo merger trees generated with the algorithm of Parkinson et al. (2008). This algorithm is based on the Extended Press-Schechter (EPS) formalism and tuned to match the conditional mass functions of halo merger trees (Cole et al. 2008) constructed from the Millennium Simulation (MS, Springel et al. 2005). As shown by Yang et al. (2011) and Jiang & van den Bosch (2014), this algorithm is in good agreement with simulations in many other halo properties, such as mass assembly history, merger rate and un-evolved sub-halo mass function. For our analysis, merger trees are constructed over the redshift range \(0 \leq z \leq 15\), with 100 snapshots evenly distributed in \(\ln(1 + z)\) space. The mass resolution is \(2 \times 10^9 h^{-1}\text{M}_\odot\).

Each sub-halo at the time of accretion is assigned a stellar mass and a star SFR according to its mass and accretion redshift based on the results of Lu et al. (2015a). A cold gas mass is derived from its star formation rate using a star formation law, as described in Lu et al. (2014b, 2015a) at the time of accretion. The empirical model of Lu et al. is in good agreement with simulations in many other halo properties, such as mass assembly history, merger rate and un-evolved sub-halo mass function. For our analysis, merger trees are constructed over the redshift range \(0 \leq z \leq 15\), with 100 snapshots evenly distributed in \(\ln(1 + z)\) space. The mass resolution is \(2 \times 10^9 h^{-1}\text{M}_\odot\).

where \(\Sigma_{\text{c}}(r) = \left\{ \begin{array}{ll} \Sigma_{\text{c},0} \exp \left(-\frac{r}{R_{\text{c}}}ight) & \text{if } r < R_{\text{tr},\text{c}} \\ 0 & \text{otherwise} \end{array} \right. \) (1)

where \(\Sigma_{\text{c},0}\) is the central surface density, \(R_{\text{c}}\) the scale radius, and \(R_{\text{tr},\text{c}}\) the truncation radius. At the time of accretion of a satellite, the truncation radius is set to \(\infty\). The subsequent evolution of the truncation radius is determined by different stripping processes, as described in the following sections. The total mass within the truncation radius is

\[ M_{\text{c}} = 2\pi R_{\text{c}}^2 \Sigma_{\text{c},0} \left[ 1 - \left(1 + \frac{R_{\text{tr},\text{c}}}{R_{\text{c}}}\right) \exp \left(-\frac{R_{\text{tr},\text{c}}}{R_{\text{c}}}\right) \right]. \] (2)

The stellar disks are assumed to follow a profile similar to the gas but with a different scale radius \(R_{\text{a}}\) and a different truncation radius \(R_{\text{tr},\text{a}}\). Both \(R_{\text{a}}\) and \(R_{\text{tr},\text{a}}\) are set using the model described in Lu et al. (2015b),
in which the ratio $\mathcal{L} \equiv R_g/R_*$ is assumed to be a constant larger than one.

The initial gas mass $M_{g,1}$ is obtained from the initial star formation rate (SFR) through an adopted star formation law. Specifically, we write

$$\text{SFR} = \int_0^\infty \Sigma_{\text{SF}} (\Sigma_g) 2\pi R dR,$$

where $\Sigma_{\text{SF}}$ is the SFR surface density, assumed to be determined by the cold gas mass surface density, $\Sigma_g$. In this paper we implement the Kennicutt-Schmidt law (Kennicutt 1998),\(^1\) which gives

$$\Sigma_{\text{SFR}} = \begin{cases} A_K \left( \frac{\Sigma_g}{M_{\odot} \text{pc}^{-2}} \right)^{N_K} & \text{if } \Sigma_g \geq \Sigma_c, \\ 0 & \text{if } \Sigma_g < \Sigma_c. \end{cases}$$

The power index $N_K \approx 1.4$, and the amplitude $A_K \approx 2.5 \times 10^{-4} \text{M}_{\odot} \text{yr}^{-1} \text{pc}^{-2}$. These parameters are constrained using the gas mass - stellar mass relation of local galaxies (Lu et al. 2015b, and references therein).

3 EVOLUTION OF SATELLITE GALAXIES

3.1 Orbits of Satellites

The orbits of satellite galaxies in host dark matter halos are modeled according to the orbits of sub-halos within which the satellites are located. The host halos are assumed to follow the NFW profile (Navarro et al. 1996):

$$\rho_D(r) = \frac{M_{\text{vir}}}{4\pi r^3_v} \frac{c^2}{\ln(1 + c) - c/(1 + c)} \frac{1}{x(1 + cx)^2},$$

where $M_{\text{vir}}$ is the halo mass, $r_v$ is the radius of the halo, and $x \equiv r/R_{\text{vir}}$. The concentration parameter, $c$, is modeled with the $c-M_{\text{vir}}$ relation given by Bullock et al. (2001). Within the mass and redshift ranges we are interested in, the model is a reasonably good approximation to the more recent models of the halo concentration - mass relation (e.g. Zhao et al. 2009; Dutton & Maccio 2014). The corresponding gravitational potential of this density profile is

$$\Phi_D(r) = -\frac{V_{\text{vir}}^2}{c^2} \frac{1}{\ln(1 + c) - c/(1 + c)} \frac{\ln(1 + cx)}{x},$$

where $V_{\text{vir}}^2 \equiv GM_{\text{vir}}/R_{\text{vir}}$ is the virial speed of the halo.

At the time of accretion, that is when a sub-halo crosses the virial radius ($R_{\text{vir}}$) of a larger halo, the infall velocity of the sub-halo is set to be exactly the virial velocity $V_{\text{vir}}$. This is a natural prediction of the spherical collapse model, and is used here as an approximation. The initial orbital angular momentum of the sub-halo is described by its orbital circularity, $\eta$, which is defined to be the ratio between the true orbital angular momentum of the sub-halo and the orbital angular momentum of a circular orbit with the same orbital energy as the sub-halo. The distribution of this quantity has been determined using high resolution simulations. Here we use the result of Zentner et al. (2005), which gives

$$P(\eta) \propto \eta^{1.2} (1 - \eta)^{1.2}.$$

The subsequent evolution of the sub-halo is traced by following the model of Taylor & Babul (2001), which includes dynamic friction and tidal stripping. As the sub-halo moves through the background of dark matter particles, a drag force (dynamic friction) acts to reduce its angular momentum and energy. This drag force is modeled by Chandrasekhar’s formula,

$$F_{df} = -4\pi G^2 M_s^2 \ln \Lambda \rho_D(r) \left[ \text{erf} \left( \frac{v_s}{V_c} \right) - \frac{2}{\sqrt{\pi}} \frac{v_s}{V_c} e^{-v_s^2/v_c^2} \right] \frac{v_s}{v_c^2} \frac{\omega}{2\pi},$$

where $M_s$ is the instantaneous mass of the sub-halo, $v_s$ is its velocity, and $V_c$ is the local circular velocity of the host halo. The Coulomb logarithm is estimated as $\ln \Lambda = \ln(M_{\text{vir}}/M_s)$, where $M_{\text{vir}}$ is the mass of the host halo.

The instantaneous tidal radius of the sub-halo is estimated using

$$R_{\text{tidal}} \approx \left( \frac{GM_s(< R_{\text{tidal}})}{\omega^2 - \sigma^2} \right)^{1/3},$$

where $\omega$ is the instantaneous orbital angular velocity. This tidal radius is calculated at each time step as the orbit of the satellite is integrated. Following Taylor & Babul (2001), the mass outside the tidal radius is assumed to be stripped at the rate given by

$$M_s = \frac{\omega}{2\pi} M_s (> R_{\text{tidal}}).$$

The sub-halo mass so obtained is then used in the dynamic friction calculation at each time step. Note that the disk contained in a sub-halo is also assumed to be tidally truncated, so that the disk truncation radius, $R_{\text{tr},s}$, introduced in §2, is at most $R_{\text{tidal}}$.

The orbit of each sub-halo (and of the galaxies it contains) is integrated starting from the time of accretion with the initial conditions specified above, taking into account both gravity and dynamical friction. We stop tracing the orbit of a sub-halo when it has lost all its orbital angular momentum. At this point, the satellite galaxy associated with the sub-halo is assumed to have merged with the central galaxy of the host halo.

3.2 High-order Substructures

When a group of galaxies merges with a more massive system, some of the galaxies and substructures

\(^1\) We have made a test using the star formation model of Krumholz et al. (2009). The results are qualitatively the same.
may still be associated with their parent halos, becoming high order substructures. Others may be tidally stripped, moving on independent orbits in the new, bigger halo. Whether a substructure will become a high-order substructure after merging depends on how they are gravitationally bound to each other as well as on the orbit of their parent halo. An accurate treatment of the high-order substructures, therefore, requires a recursive scheme; we need to trace not only the trajectories of the sub-halos but also the trajectories of the sub-sub-halos within them. A detailed treatment of this makes the model very complicated and is beyond the goal of this paper. Instead, we adopt the pruning scheme developed by Taylor & Babul (2004), in which the substructures that are loosely bound to the parent halo are assumed to be directly stripped when the merger happens. An empirical criterion suggested by the authors is that if a substructure has spent less than a period of $n_0 P_{\text{rad}}$ in the parent halo, it is considered loosely bound, where $n_0 = 2$ is a free parameter calibrated using more accurate models, and $P_{\text{rad}}$ is the mean orbital period. Each of the stripped sub-halos is assigned a new orbital parameter and is followed in the same way as the other sub-halos in the new halo. About 70% of the substructures are stripped from the parent halos according to this model. The remaining, tightly bound part is assumed to stay with their parent until the parent itself is destroyed.

### 3.3 Star Formation and Outflow

After a satellite is accreted by a larger halo, its cold gas can be depleted by different processes. The first is star formation and galactic outflows driven by the star formation. The gas mass depletion rate due to this process is

$$M_\text{g} = (1 - R + \epsilon_\text{w}) \int_0^{R_{\text{tr}, g}} \Sigma_{\text{SF}} (\Sigma_g) 2\pi RdR, \quad \text{(11)}$$

where $\Sigma_{\text{SF}} (\Sigma_g)$ is the star formation law, $R$ is the recycled fraction from stellar evolution and $\epsilon_\text{w}$ is the loading factor of the mass outflow driven by star formation. For simplicity, we assume that this process changes neither the scale radius nor the truncation radius of the disk, but only reduces the central surface density $\Sigma_{g, 0}$, so that

$$\frac{\Sigma_{g, 0}}{\Sigma_{g, 0}} = \frac{(1 - R + \epsilon_\text{w}) SFR}{M_\text{g}}. \quad \text{(12)}$$

The star formation in the gas disk is assumed to follow the same model as described in §2.

In addition to star formation and outflow, other processes, such as ram-pressure stripping and tidal interactions, may also deplete the disk gas outside a certain radius at which the external forces are balanced by the gravitational potential of the satellite itself. We assume that the stripping does not change the inner structure of the disk, so that $\Sigma_{g, 0}$ and $R_g$ are not affected. The detailed modeling of these processes are described in the following.

### 3.4 Ram Pressure Stripping

The host halos are assumed to be filled with diffuse hot gas. The total mass of the diffuse gas within the virial radius is modeled on the basis of the hot mode accretion presented in Keres et al. (2005) and Lu et al. (2011b), and the hot gas fraction is given by

$$f_{\text{hot}} = \frac{1}{2} \left( 1 + \exp \left[ \frac{\log_{10} (M_{\text{hot}}/M_{\text{tr}})}{\sigma_{\text{tran}}} \right] \right) \quad \text{(13)}$$

where $\log_{10} (M_{\text{tr}}) = 11.4$ is the transition halo mass, and $\sigma_{\text{tran}} = 0.4$.

The diffuse gas is assumed to be isothermal and has the virial temperature $\frac{1}{2} (\mu m_p/k_b) V_{\text{vir}}^2$, where $\mu$ is the mean molecular weight of the gas. Assuming hydrostatic equilibrium, the density profile of the diffuse gas is determined by the gravitational potential of the host halo,

$$\rho_X(r) = \rho_{X, 0} \exp \left[ -\frac{\mu m_p}{k_b T_{\text{vir}}} \Phi_D(r) \right] \quad \text{(14)}$$

The integral constant $\rho_{X, 0}$ is given by the total mass of the diffuse gas through

$$f_{\text{hot}} \frac{\Omega_{\text{b, 0}}}{\Omega_{\text{m, 0}}} M_{\text{diff}} = \int_0^{R_{\text{vir}}} \rho_X(r) 4\pi r^2 dr. \quad \text{(15)}$$

The gravitational restoring force per unit area of an infinitesimally thin disk is

$$P_{\text{gr}} = 2\pi G \Sigma_g, \quad \text{(16)}$$

where $\Sigma_g$ and $\Sigma_{g, 0}$ are the surface densities of stars and gas at the radius in question, respectively. For a face-on disk, the truncation radius owing to ram pressure stripping can be calculated by setting the gravitational restoring force to be equal to the ram pressure $\rho_X(r) v_g^2$, which gives

$$R_{\text{tp}} = \frac{R_g}{R_g + R_g} \ln \left( \frac{2\pi G \Sigma_{g, 0} \Sigma_{g, 0}}{\rho_X(r) v_g^2} \right). \quad \text{(17)}$$

In general, if the disk is inclined relative the moving direction, the ram pressure exerted on the disk depends on the inclination angle. Analytically deriving this inclination dependence is difficult, as the stripping becomes asymmetric when the disk in inclined. Using a set of hydrodynamic simulations, Roediger & Bruggen (2006) measured the mean truncation radius as a function of both the ram pressure and the inclination angle. It is found that the prediction of Eq. (17) only deviates from the simulations when the inclination angle $\theta \geq 60^\circ$. To take this into account, we add a correction term to Eq.(17) so that it is consistent with the simulation result:

$$R_{\text{tp}} = \frac{R_g}{R_g + R_g} \left[ \ln \left( \frac{2\pi G \Sigma_{g, 0} \Sigma_{g, 0}}{\rho_X(r) v_g^2} \right) - 0.8 \ln (\cos \theta) \right],$$

where $\cos \theta = |\mathbf{v}_s \cdot \mathbf{s}|/v_s$. The spin axis $\mathbf{s}$ of the disk
3.5 Tidally Triggered Starburst

Interaction between galaxies is another possible process that can lead to the quenching of star formation. As a satellite galaxy passes by the central galaxy, the tidal field from the central region can strip the outskirts of the disk of the satellite. This direct tidal stripping is modeled with the tidal radius described in §3.1. In addition, the tidal interaction can also exert a torque on the disk, causing the disk gas to lose angular momentum and funneling it into the nuclear region to trigger a star burst. This process may also quench the satellite by consuming the star forming gas quickly.

The enhancement of the star formation in flyby satellites has been found both observationally (Li et al. 2008) and theoretically (Di Matteo et al. 2007). In particular, Li et al. (2008) found that, for star-forming galaxies, (i) the enhancement of star formation in the flyby satellites occurs when the projected distance is less than $5r_{90}$, where $r_{90}$ is the radius that encloses 90% of the light of the central galaxies; (ii) the enhancement is a factor of about 3; (iii) the results depend only weakly on the mass of the companion.

Motivated by the findings of Li et al. (2008), we consider a simple phenomenological model to emulate the star burst triggered by tidal interaction. We assume that the tidal interaction becomes important only when a satellite is within some radius from the halo center. This radius is assumed to scale with the size of the central galaxy as $R_{sb} = x_{sb} R_{*}$, where $R_{*}$ is the scale radius of the stellar disk and is estimated following the redshift-dependent galaxy size - stellar mass relation as described in Lu et al. (2015b). When a satellite reaches $R_{sb}$, it is switched to a “star burst” mode, in which the SFR is assumed to be enhanced by a factor of $\mathcal{E}_{\text{enhance}}$ relative to its SFR immediately before it reaches $R_{sb}$. This enhanced SFR is assumed to last until the gas is completely exhausted or the satellite gets out of the sphere, whichever comes first. If the gas is not exhausted before the satellite moves out of the sphere, the galaxy will resume its pre-burst mode of star formation, but with SFR calculated from the reduced amount of cold gas. We also assume that the star burst generates an outflow with a loading factor $\epsilon_{w, sb}$, which may be different from $\epsilon_{w}$.

3.6 Environmental Quenching versus Self-Quenching

Quenched galaxies exist in a variety of environments, although satellites tend to have a higher fraction of quenched population than centrals of similar stellar masses. It is believed that the physical processes responsible for the quenching of centrals are not related to environments. A widely adopted scenario is quasar-mode feedback, which can eject the cold gas in a galaxy and quench its star formation in a very short time scale. Such mechanisms are usually referred to as “self-quenching”, to distinguish them from environmental processes. Self-quenching is expected to have also operated in the observed population of satellites before or after their accretions by larger halos. If one assumes that self-quenching and environmental quenching are independent of each other, the excess of quenched fraction in satellites relative to that in centrals can be used as an indicator of the efficiency of environmental quenching. This excess can be described by the ratio

$$f_{Q, \text{excess}} = \frac{f_{Q, \text{sat, now}} - f_{Q, \text{cen, now}}}{f_{Q, \text{cen, now}}},$$

where $f_{Q}$ denotes the quenched fraction, $f_{\Lambda}$ the total fraction, and the superscripts, ‘cen’ and ‘sat’, indicate centrals and satellites, respectively. This quantity has been estimated by Wetzel et al. (2012, 2013) using galaxy groups selected from the SDSS. In our investigation, we will focus on environmental quenching without modeling the mechanisms that are responsible for the quenching of centrals. In what follows, we will refer $f_{Q, \text{excess}}$ as the quenched fraction without including the population that is quenched before the galaxies become satellites.

4 MODEL PREDICTIONS

4.1 Quenching by Strangulation

In the literature, strangulation refers to a hypothesis that a galaxy in a sub-halo will stop accreting new gas as soon as the sub-halo is accreted into a larger halo. Because of strangulation, star formation in satellite galaxies will gradually slow down and eventually stop as the gas reservoir is consumed by star formation, even in the absence of any other additional processes. Galactic wind driven by star formation speeds up the depletion of the cold gas. This is generally how red satellites are produced in many semi-analytical models (SAMs) in the literature (e.g. Lu et al. 2014a). To understand how this process works, we consider a model in which all other environmental processes (ram pressure stripping and tidally triggered starburst) are switched off, so that we can focus on quenching by gradual gas consumption owing to star formation and the associated outflow.

Figure 1 shows the impact of the outflow loading...
factor, $\epsilon_w$, on the predicted quenched fraction of satellites. All the satellites stay as star forming galaxies if $\epsilon_w$ is set to be 0. This means that the cold gas reservoir left over from the time of accretion is sufficient to keep satellite galaxies active in star formation all the way to the present day if there is no other gas depletions than star formation. The observational constraints seem to favor $\epsilon_w = 1$ and adopting a loading factor $\epsilon_w > 1$ tends to over-quench the satellite population. Unfortunately, the loading factor of outflow is still poorly understood. In hydrodynamic simulations and SAMs, a large value seems to be required to lower the efficiency of star formation to reconcile the tension between the observed galaxy luminosity/stellar mass function (dominated by central galaxies) and the halo mass function at the low-mass end (e.g. Yang et al. 2003). For instance, in the Somerville model studied in Lu et al. (2014a), the loading factor is about 3 for a Milky-Way like halo and is more than 10 for a $10^{11} M_\odot$ halo. In many SAMs, the outflow loading factor assumed for satellite galaxies is the same (high value) as for centrals, and so the satellite population is expected to be over-quenched according to the results shown in Figure 1.

Indeed, satellite galaxies are predicted to be too red, with too little current star formation, in many SAMs (e.g. Lu et al. 2014a), and strangulation is usually blamed for this over-quenching problem (Wein-
mann et al. 2006b). However, detailed modeling of the stripping of the diffuse gas halo suggests that the problem cannot be solved by continuous accretion of fresh gas from the parent halo (Weinmann et al. 2010; Font et al. 2008). In order to solve this problem, Kang & van den Bosch (2008) and Weinmann et al. (2010) had to assume that ram pressure does not make any significant stripping of the diffuse halo gas. Indeed, Font et al. (2008) assumed that most of the gas re-heated by star formation - driven wind in a satellite will not be stripped, but will instead be re-accreted by the satellite, so that the gas reservoir originally associated with the satellite is always available for star formation, even though large amounts of gas is assumed to be in the wind.

The assumption of strong galactic winds for satellite galaxies needs to be revisited, however. In fact, to match theory with observation does not require a large loading factor for central galaxies. For instance, using empirically constrained star formation-halo mass relation and the gas phase metallicity, Lu et al. (2015b) found that the mass loading factor of galaxy wind required to suppress star formation in present-day central galaxies is no larger than unity if the accretion of pristine gas into a halo can be suppressed by preheating. If satellite galaxies inherit the same star formation and feedback from their progenitors, then there will be no over-quenching problem, as shown above.

As shown in Figure 2, the specific star formation rate (sSFR, defined to be the ratio between star formation rate and stellar of of a galaxy) distribution predicted by this gradual gas depletion model is always unimodal, and increasing the loading factor only broadens the distribution without changing the unimodal nature. This is in contrast with the observational results that the sSFR of satellite galaxies follows a bimodal distribution (shown by the smooth thick solid curve in the middle panel) and such a distribution persists in different environments (e.g. Wetzel et al. 2013). This demonstrates that strangulation alone cannot explain the observed quenching of satellite galaxies.

Figure 3 shows the quenched fraction as a function of host halo mass. Here results are only shown for the $\epsilon_w = 1$ model, as it correctly reproduces the overall quenched fraction. The quenched fraction increases with halo mass, because satellites of the same mass were accreted earlier in more massive halos and so a larger fraction of them have exhausted their gas by present day. The model under-predicts the quenched fraction in halos with masses $< 10^{13} \, M_\odot$ in comparison with the observational data, demonstrating again that strangulation alone cannot explain the observed quenching of satellite galaxies in different halos, even though the outflow loading factor is tuned so that the model correctly predicts the overall quenched fraction of the satellite population.

4.2 Quenching by Ram-Pressure Stripping

Next we include ram-pressure stripping in a way described in §3.4. Figure 4 shows the overall quenched fraction predicted by this model and how it changes with the assumed outflow loading factor. As one can see, when ram-pressure stripping is included, any loading factor $\epsilon_w > 0$ tends to over-quench the satellite population, and the model prefers $\epsilon_w = 0$, i.e. no outflow. Setting $\epsilon_w$ to 1, the model already overpredicts the quenched fraction by 80%. The two depletion mechanisms, outflow and ram-pressure stripping, do not act independently. Galactic wind depletes the gas at the central region of the disk, where most stars form. This reduces the gas surface density,
Figure 5. The distribution of sSFR of satellite galaxies predicted by the model of ram-pressure stripping. Different panels correspond to different assumed outflow mass loading factors. The stellar mass range of satellites is coded in different colors, as indicated in the right panel. The black solid line in the first panel is the observational distribution inferred from Wetzel et al. (2012, 2013).

Figure 6. The evolution history of three satellites on three different initial orbits (red: nearly radial; blue nearly circular; green: intermediate). The left panel shows the distance to the host center as a function of time, and the right panel shows the specific star formation rate as a function of time. The host halos are of $10^{13}\, h^{-1} M_\odot$ at present day.

thereby reducing the gravitational restoring force and making the ram pressure stripping more effective [see Eq. (18)].

Unlike gradual gas consumption, ram pressure stripping produces a bimodal distribution in the sSFR, as shown in Fig. 5. With $\epsilon_w = 0$, the model matches the observed distribution (the left panel) qualitatively. To understand how the bimodal distribution is produced, let us have a close look at how quenching proceeds in this model. Based on the model described in §3.4, a complete stripping of the gas disk occurs when the density of the surrounding medium reaches a critical value, which can be obtained from Eq. (17) by setting $R_{rp} = 0$,

$$\rho_{X,c} = \frac{2\pi G\Sigma_{*o}\Sigma_{g,o}}{V_{vir}^2}, \quad (20)$$

where we have replaced the speed of the satellite, $v_s$, by the virial speed $V_{vir}$ of the host halo. Only galaxies that have reached such a high density region can be completely quenched. Figure 6 shows the evolution of satellites on three typical orbits in a host halo with mass $10^{13}\, M_\odot$: a nearly circular orbit (blue), a nearly radial orbit (red) and an elliptical orbit in the middle (green). Although it has been accreted for about 6 Gyrs, the satellite on the circular orbit is still actively forming stars at present day. In contrast, the satellite on the radial orbit has almost completely lost its gas during the first approach to the peri-center, which happened less than 1 Gyr after accretion.

Compared with observational results, the predicted quenched fraction of satellites by ram-pressure stripping alone has much stronger dependence on host halo mass, quite independent of the assumed loading
Quenching of Satellite Galaxies

Figure 7. The fraction of quenched satellites as a function host halo mass. The colored solid lines are predictions by the model with ram pressure stripping and with $\epsilon = 0$, and the colored dashed lines are predictions of the model with $\epsilon = 1$. Data points with error bars are taken from Wetzel et al. (2012).

Figure 8. The predicted fractions of quenched galaxies by the model of tidally-triggered starburst. Different colors represent different outflow mass loading factors assumed for the starburst phase, as indicated in the panel. The data points are taken from Wetzel et al. (2013).

factor, as shown in Figure 7. There is a transition at a halo mass about $2 \times 10^{13} M_\odot$ (for $\epsilon = 0$). In more massive halos satellites are slightly over-quenched (by about 10%), while in halos with masses less than $5 \times 10^{12} M_\odot$ ram pressure by the halo gas seems unable to quench any satellites. This strong dependence on host halo mass is also responsible for the bimodal distribution seen in Figure 5: most of the star-forming satellites are from low-mass halos ($< 10^{13} M_\odot$) while most the quenched satellites are from more massive halos ($> 5 \times 10^{13} M_\odot$).

A simple analysis can be used to understand this strong dependence on halo mass. The critical density for complete stripping given by Eq. (20) is proportional to $\rho_{X,c} \propto V_{\text{vir}}^{-2}$. Assuming an isothermal sphere gas distribution for the host halo, this critical density corresponds to a critical radius,

$$x_c \equiv \frac{r_c}{R_{\text{vir}}} \propto V_{\text{vir}}.$$  \hspace{1cm} (21)

It is this dependence on the virial velocity that produces the strong dependence of the quenched fraction on the host halo mass shown in Figure 7.

4.3 Quenching by Tidally Triggered Starburst

Next, we switch off ram pressure stripping and investigate how the tidally-triggered starburst shapes the...
Figure 9. The distribution of the sSFR of satellite galaxies predicted by the model of tidally-triggered starburst. Different panels correspond to different outflow mass loading factors assumed for the starburst. The stellar mass range of satellites is coded in different colors, as indicated in the left panel. The black solid line in the last panel is the observational distribution inferred from Wetzel et al. (2012, 2013).

Figure 10. The fraction of quenched satellites as a function host halo mass. The colored solid lines are predictions by the model of tidally-triggered starburst with $\epsilon_{w, sb} = 5$. Data points with error bars are taken from Wetzel et al. (2012).

The overall quenched fractions as functions of stellar mass are shown in Figure 8, and the distributions of sSFR are shown in Figure 9. A modest outflow during the burst phase ($\epsilon_{w, sb} \leq 3$) produces a long tail in the quenched side, but only a limited number of galaxies are quenched after the star burst. A significant bimodal distribution in the sSFR and a reasonable overall quenched fraction can be produced if $\epsilon_{w, sb} \approx 5$ is assumed. This can be understood as follows. The time for a satellite to cross the region within $R_{sb}$ is about a few $10^8$ yr, which is also the typical starburst time scale assumed in our model, while the gas depletion time scale by star formation, $\tau_{dep} \equiv M_g/SFR$, is an order of magnitude longer. Thus, star formation alone can hardly consume a significant amount of gas during the star burst. Indeed, as shown in Di Matteo et al. (2007), the extra gas consumed by star burst is not significant. In order to get a complete quenching, $\epsilon_{\text{enhance}} (1 - R + \epsilon_{w, sb}) > 10$.

In contrast to the ram pressure stripping model,
4.4 Conspiracy between Ram-Pressure Stripping and Tidally Triggered Starburst

The existence of both the ram pressure stripping and tidally-triggered starburst have observational supports. Thus, both processes should be taken into account in a realistic model. The two processes do not act independently. Here we demonstrate how they conspire to quench satellites in different environments. Figure 11 shows the predictions of two models, one with $\epsilon_{w, sb} = 1$ and the other with $\epsilon_{w, sb} = 3$, both including the joint effects of ram pressure stripping and tidally-triggered starburst. For comparison, the predictions of models including only ram pressure or only starburst are shown as dashed lines and dot-dashed lines, respectively. The sSFR distribution is shown in Figure 12. As one can see, in halos with masses below $10^{13} \, M_\odot$, starburst with a modest outflow ($\epsilon_{w, sb} = 3$) alone can hardly quench any satellites. Ram pressure stripping acting alone is also too weak to strip the gas discs in such halos, as already discussed in §4.2. However, when both processes are turned on, the starburst can reduce the gas surface density significantly so as to make the ram pressure stripping more effective. In contrast, in more massive halos, ram pressure starts to

Figure 11. The fraction of quenched satellites as a function host halo mass. The dashed, dotted and solid lines correspond to models with only ram pressure, with only tidally triggered starburst, and with both processes, respectively. Left and right panels assume $\epsilon_{w, sb} = 1$ and 3, respectively.

Figure 12. The distribution of sSFR predicted by the model in which both ram-pressure stripping and tidally-triggered starburst (with $\epsilon_{w, sb} = 3$) are included.

Figure 13. The fraction of quenched satellites as a function of the redshift of accretion, predicted by the model in which both ram-pressure stripping and tidally-triggered starburst (with $\epsilon_{w, sb} = 3$) are included.
work long before the starburst is triggered, so that the inclusion of starburst has only a modest effect. Overall the prediction of the combined model matches roughly the observed quenched fraction as a function of halo mass and the sSFR distribution. The largest discrepancy occurs for low-mass galaxies in massive halos, where the model over-predict the quenched fraction by about 30%. For low-mass galaxies, the predicted halo mass dependence of the quenched fraction also appears to be stronger than in the data. Therefore, some halo mass-dependent prescriptions seem to be needed for the model to match the data better. For example, if the strength of tidally triggered starburst decreases with halo mass, or if the efficiency of ram pressure stripping in massive halos is lower than that given by the model, the match between model and data can be improved. The first possibility is actually likely. In halos with masses above $10^{13.5} \, M_\odot$, ram pressure stripping starts to work long before the starburst is triggered. By the time when a satellite galaxy reaches $R_{200}$, the cold gas mass in its disk has already been reduced by a large amount, and the interaction of the gas-poorer disk may not be able to trigger a strong starburst. In the extreme case where no starburst is triggered for satellites in these massive halos, the quenched fractions would be reduced to those represented by the dashed lines in Figure 11, which match the data better. This possibility is not in conflict with the result of Li et al. (2008), because the result is for the star-forming population, which is dominated by galaxies in relatively low-mass halos.

The wind loading factor, $\epsilon_{w,ab} = 3$, required for low-mass halos appears to be reasonable for starburst galaxies. Observations of galactic winds show that the mass in the outflow gas is typically a few times the star formation rate for starburst galaxies (e.g. Martin 1999). Once the gas is ejected from a satellite and becomes diffuse, ram-pressure and tidal force may strip it from the sub-halo of the satellite, even if the wind is not energetic enough to escape the sub-halo by itself.

The quenched fraction is a strong function of the accretion time of the satellites (Figure 13). It increases rapidly from 0.5 at $z \sim 0.5$ (corresponding to a look-back time of $\sim 5$ Gyr) to 1 at $z \sim 1$ (corresponding to a look-back time of 7.5 Gyr). This suggests that a satellite accreted at $z > 0.5$ is more likely to have already been quenched, while a satellite accreted at a later time is more likely to be observed as a star forming-galaxy. This is consistent with the finding of Wetzel et al. (2013), which shows that on average a galaxy can remain as star forming for $\approx 4$ Gyr after accretion before it is quenched abruptly.

5 SUMMARY

We investigated the star formation quenching of satellite galaxies using a set of simplified but physically motivated models. The models are built upon the dark halo merger histories of Parkinson et al. (2008) and the sub-halo model of Taylor & Babul (2001, 2004) that traces the the orbits, tidal stripping and disruption of sub-halos. Instead of following the whole histories of present-day satellites, we only keep track of the evolution of a satellite after it is accreted into a more massive system. The progenitors of satellites at the time of accretion, which are themselves central galaxies at that time, are initialized according to the dependences of stellar mass, star formation and cold gas content of star forming galaxies on halo mass as given in Lu et al. (2015b). Such dependences are constrained using current observations of the galaxy populations at different redshifts. This approach avoids some of the serious uncertainties in the assumptions made in previous models of satellite galaxies, and allows us to focus on the quenching of star formation by intra-halo environments and processes. Our model does not assume a diffuse gas reservoir in sub-halos, and as such it assumes instantaneous "strangulation".

We tested a series of models against the observational constraints derived from galaxy groups by Wetzel et al. (2012, 2013), including the sSFR distribution, its dependence on stellar mass of the satellite and on the mass of the halo in which the satellite resides. We found that these observational constraints provide important information regarding the process of environmental quenching.

We first considered a model in which the satellites are quenched by gradual gas consumption due to star formation and associated outflow. Without any outflow, all the survived satellites would lie in the star forming sequence. This means that the cold gas reservoir brought in at the time of accretion is large enough to sustain the star formation in satellites all the way to the present-day. A mass loading factor larger than unity produces an over-quenched satellite population. This critical loading factor is modest compared with what is traditionally assumed in SAMs of galaxy formation. The over-quenching of satellites in these SAMs has usually been attributed to the instantaneous "strangulation", and solutions to it have been focused on slowing down the stripping of the diffuse gas from sub-halos (Kang & van den Bosch 2008; Weinmann et al. 2010). For instance, Weinmann et al. (2010) suggested that no ram pressure of diffuse gas should happen in order to reproduce the correct quench fraction, in conflict with simple physical expectations. An alternative solution to this over-quenching problem is to assume a modest outflow, as is demonstrated by our calculation and supported by other studies of central galaxies (Lu et al. 2014a, 2015b). Thus, we believe that the so called "over-quenching" problem is primarily due to the assumption of a high loading factor in these models. However, gradual gas consumption by star formation and outflow alone cannot produce a bimodal distribution for sSFR, suggesting that additional quenching mechanisms are needed.
The second model we tested is the ram pressure stripping of the cold gas by hot halo gas. With realistic assumptions about the hot gas profiles in dark matter halos, quenching of star formation is found to occur in an abrupt manner when satellites reach deep enough into the halo center, where the ram pressure is sufficiently large to completely strip the cold gas. To obtain an overall quenched fraction that matches observation, this model does not need any net outflow powered by star formation. Ram pressure stripping is also able to reproduce a bimodal distribution in the sSFR. However, this bimodality is largely due to the strong dependence of the ram pressure stripping on halo mass: in massive halos (> $10^{13.5} \, M_\odot$), the satellite population is slightly over-quenched compared with observation, while in low mass groups (< $10^{13} \, M_\odot$), ram pressure is too weak to produce any quenched satellites. This is in conflict with the observational data in which a significant fraction of satellites in low-mass halos are also quenched.

In a third model, we proposed a phenomenological prescription to describe starbursts that are triggered by tidal interactions of satellites with central galaxies as they pass by the halo center. This model is motivated by and calibrated with the observational results of Li et al. (2008). The SFR is assumed to be enhanced by a constant factor $E_{\text{enhance}} = 3$ (1 means no enhancement) once a satellite reaches a critical radius $r_{\text{sb}} \sim 20 R_\odot$. The enhanced star formation alone cannot quench the flyby satellites, as the crossing time scale of the sphere $r_{\text{sb}}$ is too short. Without ram pressure, a strong outflow with a loading factor $w_{\text{out}} > 5$ during the starburst phase is required. The dependence of the quenched fraction on halo mass in this model is found to be much weaker than the prediction of the ram-pressure stripping model. In the halo mass range $< 5 \times 10^{13} \, M_\odot$, the prediction matches observation, while in more massive halos the critical radius is too small to produce a large enough population of quenched satellites to match observation.

Finally, we considered a model where both ram-pressure stripping and tidally-triggered starburst operate together. With ram pressure, the wind loading factor in the starburst required to match observation is reduced. In low mass halos, ram pressure and a starburst with $w_{\text{out}} \approx 3$ working together can produce a reasonable match to the observed quenched fraction. In high mass halos, quenching is always dominated by ram pressure.

In conclusion, our investigation demonstrates that, once the progenitors of satellites are properly modeled, the star formation properties of the satellite population can be understood in terms of physical processes that are expected to operate in galaxy groups/clusters. Further investigations are clearly needed. The assumptions regarding ram pressure stripping and tidally induced starburst need to be tested with numerical simulations. Detailed model predictions can be made by controlled simulations which follow the evolutions of satellites in different halos with properly-set initial conditions. Further observational consequences, such as spatial distributions and velocity dispersions of quenched and non-quenched satellites in their host dark matter halos, should be explored to test the model in more detail. We will come back to some of these problems in the future.

ACKNOWLEDGEMENTS

We thank Yu Lu and Frank van den Bosch for helpful discussions. HJM would like to acknowledge the support of NSF AST-1109354.

REFERENCES

Balogh M. L., Navarro J. F., Morris S. L., ApJ, 540, 113
Bullock J. S., Kolatt T. S., Sigad Y., et al., 2001, MNRAS, 321, 559
Chabrier G., 2003, PASP, 115, 763
Cole S., Helly J. C., Frenk C. S., Parkinson H., 2007, MNRAS, 383, 546
Croton D. J., et al., 2006, MNRAS, 365, 11
Di Matteo P.; Combes F.; Melchior A.-L.; Semelin B., 2007, A&A, 468, 61
Dutton A. & Maccio A., 2014, MNRAS 441, 3359
Font A. S. et al., 2008, MNRAS, 389, 1619
Jiang F., van den Bosch F.C., 2014, MNRAS in press (arXiv:1311.5225)
Kang X.; van den Bosch F. C., 2008, ApJ, 676, 101
Kennicutt R.C. Jr, 1998, ApJ, 498, 541
Keres D., Katz N., Weinberg D. H., & Dave R., MNRAS, 2005, 363, 2
Komatsu E., Smith K. M., Dunkley J., et al., 2011, ApJS, 192, 18
Kravtsov A.V., 2013, ApJ, 764, 31
Krumholz M. R., McKee C. F. & Tumlinson J., 2009, ApJ, 693, 216
Li C., Kauffmann G., Heckman T. M., Jing Y. P., White S. D. M., 2008, MNRAS, 385, 1903
Lu Y., Mo H. J., Weinberg M. D., & Katz N., 2011a, MNRAS, 416, 1949
Lu Y., Keres D., Katz N., Mo H. J., Fardal M. & Weinberg M. D., 2011b, MNRAS, 416, 660
Lu, Y., Wechsler R. H., Somerville R. S., et al., 2014a, ApJ, 795, 123
Lu Z., Mo H. J., Lu Y., et al., 2014b, MNRAS, 439, 1294
Lu Z., Mo H. J., Lu Y., et al., 2015a, MNRAS, 450, 1604
Lu Z., Mo H. J. & Lu Y., 2015b, MNRAS, 450, 606
Martin C. L. 1999, ApJ, 513, 156
Mo H. J., Mao S. & White S., 1998, MNRAS, 295, 319
Mo, H. J., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution, Cambridge, UK: Cambridge University Press, 2010.
Navarro, J. F., Frenk, C. S., & White, S. D. M., 1996, ApJ, 462, 563
Parkinson, H., Cole, S., & Helly, J., 2008, MNRAS, 383, 557
Roediger, E., & Brüggen, M., 2006, MNRAS, 369, 567
Roediger, E., & Brüggen, M., 2007, MNRAS, 380, 1399
Springel, V., White, S. D. M., Jenkins, A., et al., 2005, Nat, 435, 629
Taylor, J. E., & Babul, A., 2001, MNRAS, 348, 811
Taylor, J. E., & Babul, A., 2004, MNRAS, 348, 811
Weinmann, S. M., van den Bosch, F. C., Yang, X., Mo, H. J., 2006, MNRAS, 366, 2
Weinmann, S. M., van den Bosch, F. C., Yang, X., Mo, H. J., Croton, D. J., Moore, B., 2006, MNRAS, 372, 1161
Weinmann, S. M., Kauffmann, G., von der Linden, A., & De Lucia, G., 2010, MNRAS, 406, 2249
Wetzel, A. R., Tinker, J. L., & Conroy, C., 2012, MNRAS, 424, 232
Wetzel, A. R., Tinker, J. L., Conroy, C., & van den Bosch, F. C., 2013, MNRAS, 432, 336
Yang, X., Mo, H. J., & van den Bosch, F., 2003, MNRAS, 339, 1057
Yang, X., Mo, H. J., Zhang, Y. C., & van den Bosch, F., 2011, ApJ, 741, 13
Zentner, A. R., Berlind, A. A., Bullock, J. S., Kravtsov, A. V., & Wechsler, R. H., 2005, ApJ, 624, 505
Zhao, D. H., Jing, Y. P., Mo, H. J., Borner, G., 2009, ApJ, 707, 354