Probing the link between quenching and morphological evolution

I. Koutsouridou, A. Cattaneo

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **516,** 4194–4211 (2022) https://doi.org/10.1093/mnras/stac2240
Advance Access publication 2022 August 13

Probing the link between quenching and morphological evolution

I. Koutsouridou[1,2]★ and A. Cattaneo[3,4]★

[1]*Dipartimento di Fisica e Astronomia, Università degli Studi di Firenze, Via G. Sansone 1, I-50019 Sesto Fiorentino, Italy*
[2]*INAF/Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy*
[3]*Observatoire de Paris, LERMA, PSL University, 61 avenue de l'Observatoire, F-75014 Paris, France*
[4]*Institut d'Astrophysique de Paris, CNRS, 98bis Boulevard Arago, F-75014 Paris, France*

Accepted 2022 August 1. Received 2022 July 19; in original form 2022 March 2

ABSTRACT

We use a semi-analytic model of galaxy formation to compare the predictions of two quenching scenarios: halo quenching and black hole (BH) quenching. After calibrating both models so that they fit the mass function of galaxies, BH quenching is in better agreement with the fraction of passive galaxies as a function of stellar mass M_* and with the galaxy morphological distribution on a star formation rate versus M_* diagram. Besides this main finding, there are two other results from this research. First, a successful BH-quenching model requires that minor mergers contribute to the growth of supermassive BHs. If galaxies that reach high M_* through repeated minor mergers are not quenched, there are too many blue galaxies at high masses. Second, the growth of BHs in mergers must become less efficient at low masses in order to reproduce the $M_{\rm BH}$–M_* relation and the passive fraction as a function of M_*, in agreement with the idea that supernovae prevent efficient BH growth in systems with low escape speeds. Our findings are consistent with a quasar-feedback scenario in which BHs grow until they are massive enough to blow away the cold gas in their host galaxies and to heat the hot circumgalactic medium to such high entropy that its cooling time becomes long. They also support the notion that quenching and maintenance correspond to different feedback regimes.

Key words: galaxies: formation – galaxies: evolution – galaxies: structure.

1 INTRODUCTION

The star formation (SF) properties of galaxies are closely related to their morphologies (Sandage 1986) and display a bimodal distribution (Kauffmann et al. 2003; Baldry et al. 2004). Four scenarios have been proposed to explain these observations.

(i) Elliptical galaxies are passive because they converted all their gas into stars very rapidly at high redshift (Sandage 1986). Coupled to the hypothesis that elliptical galaxies formed through mergers (Toomre & Toomre 1972) and to the observation that mergers can drive starbursts (Sanders et al. 1988), this picture explains why rapid depletion of gas is associated with elliptical morphologies. Its main problem is that gas-rich mergers leave behind persistent blue cores (Cattaneo et al. 2005; Springel & Hernquist 2005). Elliptical galaxies with blue cores (E+A galaxies) are observed, but they are exceedingly rare.

(ii) Central black holes (BHs) grow until they are massive enough to heat and/or blow away all the gas in their host systems (BH quenching; Silk & Rees 1998; Fabian 1999; King 2003; Chen et al. 2020). In a version of this scenario (Springel, Di Matteo & Hernquist 2005; Hopkins et al. 2006, 2008a), gas-rich mergers feed the growth of supermassive BHs until quasar feedback blows away all the leftover gas.

(iii) The transition from star-forming (blue) to passive (red) galaxies is linked to the disruption of cold filamentary flows and the appearance of shock-heated gas in massive systems (halo quenching; Dekel & Birnboim 2006). This picture explains the galaxy colour–magnitude bimodality in quantitative detail (Cattaneo et al. 2006). The link to morphology is less direct (the shutdown of gas accretion prevents merger remnants from regrowing discs).

(iv) This scenario can be characterized as BH or halo quenching followed by BH maintenance. All the above scenarios share a common weakness: they do not explain what prevents the accretion of gas from restarting after the initial shutdown of SF (the so-called maintenance problem). Model (iv) is model (ii) or (iii) with the addition of mechanical heating by radio galaxies, which solves the maintenance problem by preventing the hot gas from cooling (Bower et al. 2006; Cattaneo et al. 2006; Croton et al. 2006; Somerville et al. 2008). In this picture, BHs constitute the link between quenching and morphology. Elliptical galaxies are passive because they are those with the most massive BHs.

The problem is complicated because all the mechanisms above play a role to some extent. The question is which ones dominate. An observational approach to it is to ask which observable is the best predictor of quiescence. Cheung et al. (2012), Fang et al. (2013), and Barro et al. (2017) showed that the surface density in the central kiloparsec, Σ_1, is the best predictor of quiescence in large surveys (galaxies are quiescent above a critical Σ_1 that depends on M_*). Fang et al. (2013) also mentioned a good correlation with the central stellar velocity dispersion $\sigma_{\rm c}$ and noted that this would follow naturally from that with Σ_1 since $\sigma_{\rm c}$ and Σ_1 are closely related. Terrazas et al. (2016) used observations of galaxies with direct measurements of the BH mass $M_{\rm BH}$ and showed that, for a same M_*, passive galaxies have higher $M_{\rm BH}$ than star-forming galaxies. Bluck et al. (2020) used the empirical $M_{\rm BH}$–$\sigma_{\rm c}$ relation (Ferrarese & Merritt 2000; Gebhardt et al.

★ E-mail: ioanna.koutsouridou@unifi.it (IK); andrea.cattaneo@obspm.fr (AC)

2000; Saglia et al. 2016; Sahu, Graham & Davis 2019) to conclude that quiescence correlates with $M_{\rm BH}$ more strongly that it does with stellar mass, halo mass, or any other observable property. An equally strong correlation with Σ_1 would not modify this conclusion to the extent that Σ_1 and $\sigma_{\rm c}$ are both tracers of $M_{\rm BH}$ (see Sahu, Graham & Davis 2022 for an observational analysis of the $M_{\rm BH}$–Σ_1 relation).

In this paper, we use a semi-analytic model (SAM) of galaxy formation to shed some more light on this topic and compare the observational implications of *BH quenching* versus *halo quenching* for

a) the galaxy stellar mass function (GSMF);
b) the fraction of passive galaxies as a function of M_*;
c) the $M_{\rm BH}$–M_* relation;
d) galactic morphologies.

Point (d) is important to break potential degeneracies of the SAM. A model with an overpropensity for early-type morphologies could reproduce the GSMF and the passive fraction correctly if this overpropensity were compensated by lower BH accretion rates, so that the product $M_{\rm BH}/M_{\rm bulge} \times M_{\rm bulge}/M_*$ stays the same (here $M_{\rm bulge}$ is the stellar mass of the bulge).

The paper starts by presenting the SAM used for this study (Section 2), which we call GALICS 2.2, since it is an updated version of GALICS 2.1 (Cattaneo et al. 2020). GALICS 2.2 contains two options for the quenching model: halo quenching (model A) and BH quenching (model B). It also includes two models for the morphologies of merger remnants. In model 1 (also referred to as the threshold model), there is a sharp separation between major and minor mergers. Model 2 follows numerical simulations by Kannan et al. (2015) and assumes a smoother transition. In Section 3, we compare the results of the four combinations A1, A2, B1, and B2. In Section 4, we summarize and discuss the conclusions of the paper.

2 THE MODEL

This section describes how GALICS 2.2 follows the hierarchical growth of dark matter (DM) haloes (Section 2.1), the accretion of gas on to haloes and galaxies (Section 2.2), SF (Section 2.3), stellar feedback (Section 2.4), disc instabilities (Section 2.5), mergers (Section 2.6), the growth of supermassive BHs (Section 2.7), quenching (Section 2.8), and ram-pressure stripping (RPS; Section 2.9). The presentation is more synthetic in Sections 2.1–2.4, where the differences with GALICS 2.1 (Cattaneo et al. 2020) are small (or inexistent, in Sections 2.1 and 2.2), and more detailed in Sections 2.5–2.9, where the new developments are.

2.1 *N*-body simulation

GALICS 2.2 uses the same DM merger trees as GALICS 2.1. They were extracted from a cosmological *N*-body simulation with $\Omega_{\rm M} = 0.308$, $\Omega_\Lambda = 0.692$, $\Omega_{\rm b} = 0.0481$, and $\sigma_8 = 0.807$ (Planck Collaboration XX 2014). The simulation has a volume of $(100\,{\rm Mpc})^3$ and contains 1024^3 particles, so that the halo mass functions are reproduced down to $M_{\rm vir} \simeq 4 \times 10^9\,{\rm M}_\odot$. It has been evolved from $z = 16.7$ to $z = 0$ storing 287 outputs equally spaced in the logarithm of the expansion factor.

The virial mass $M_{\rm vir}$, the virial radius $r_{\rm vir}$, and the virial angular momentum $\boldsymbol{J}_{\rm vir}$are computed at each time-step for each halo containing at least 100 particles, as in Cattaneo et al. (2017). The virial masses, orbits, and merging times of ghost subhaloes, i.e. subhaloes that fall below the resolution limit of the simulation, are computed as in Cattaneo et al. (2020).

2.2 Gas accretion

Gas accretion contains two separate aspects: the accretion of gas on to haloes and its later accretion (or reaccretion) on to galaxies. Let us start from the former. The gaseous mass that accretes on to a halo at each time-step is computed by requiring that

$$M_{\rm bar} = f_{\rm b} M_{\rm vir}, \tag{1}$$

where $M_{\rm bar}$ is the total baryonic mass that has accreted on to the halo (inclusive of any ejected material) and $f_{\rm b}$ is the halo baryon fraction (see below). If the left-hand side of equation (1) is smaller than the right-hand side, then gas accretes on to the halo until the validity of equation (1) is restored (see fig. 1 of Cattaneo et al. 2020 for the details of how this accretion scheme is implemented).

In massive haloes, $f_{\rm b}$ should approach the universal baryon fraction $\Omega_{\rm b}/\Omega_{\rm M}$. Photoionization heating (e.g. Efstathiou 1992; Gnedin 2000) and stellar feedback (e.g. Brook et al. 2012; Wang et al. 2017; Tollet et al. 2019), however, suppress gas accretion on to low-mass haloes, so that their baryon fraction is lower than the universal one. We follow Cattaneo et al. (2020) and assume that

$$f_{\rm b}(\upsilon_{\rm vir}) = \frac{\Omega_{\rm b}}{\Omega_{\rm M}} \left[{\rm erf}\left(\frac{\upsilon_{\rm vir}}{\sigma}\right) - \frac{2}{\sqrt{\pi}} \frac{\upsilon_{\rm vir}}{\sigma} {\rm e}^{-\left(\frac{\upsilon_{\rm vir}}{\sigma}\right)^2} \right], \tag{2}$$

where $\upsilon_{\rm vir} = \sqrt{{\rm G}M_{\rm vir}/r_{\rm vir}}$ is the virial velocity of the halo and $\sigma = 34\,{\rm km\,s^{-1}}$ is a parameter of the SAM that we fix in agreement with the results of the Numerical Investigation of a Hundred Astrophysical Objects (NIHAO) simulations (Tollet et al. 2019).

The gas that accretes on to a halo can stream on to the central galaxy in free-fall (cold-mode accretion) or it may be shock heated and become part of a hot atmosphere. The choice is based on a shock-stability criterion that compares the compression time with the post-shock cooling time on a halo-by-halo basis. This criterion, which improves previous work by Dekel & Birnboim (2006), has been discussed at length in Cattaneo et al. (2020) and Tollet et al. (2022), who have performed a detailed comparison with cosmological zoom simulations. Hence, there is no need for us to elaborate on it here. Our SAM also considers the effects of the Kelvin–Helmholtz instability in haloes where cold filamentary flows and a hot atmosphere coexist. The Kelvin–Helmholtz instability contributes to the disruption of the filaments in such situations (see Cattaneo et al. 2020 for details).

Shock-heated gas can accrete on to the central galaxy by cooling (hot-mode accretion), which we compute with the standard method of White & Frenk (1991). The cooling rate of the hot gas depends on its density distribution $\rho_{\rm h}(r)$. As in Cattaneo et al. (2020), we assume a density profile of the form

$$\rho_{\rm h}(r) = \rho_{\rm h,0} \left(1 + \frac{r}{\alpha_{\rm g} r_0}\right)^{-3}, \tag{3}$$

where r_0 is the core radius of the DM halo, fitted with a Navarro–Frenk–White (NFW; Navarro, Frenk & White 1997) profile, and $\rho_{\rm h,0}$ is the central density of the hot gas, determined by the condition that $\rho_{\rm h}/\rho_{\rm NFW} \to M_{\rm hot}/M_{\rm vir}$ for $r \to \infty$ ($M_{\rm hot}$ is the mass of the hot gas within the halo). We set $\alpha_{\rm g} = 0.59$ in agreement with adiabatic cosmological hydrodynamic simulations by Faltenbacher et al. (2007).

The black squares in Fig. 1 show the electron-density distribution $n_{\rm e} = \frac{14}{16}\frac{\rho_{\rm h}}{m_{\rm p}}$ predicted by equation (3) for a cluster halo, in which the gas is approximately adiabatic ($m_{\rm p}$ is the proton mass; the 14/16 coefficient is for a plasma that is three quarters hydrogen and one quarter helium in mass). The blue and the red curves are observational data for cool-core and non-cool-core clusters in the same range of masses, respectively. Real clusters have central electron densities

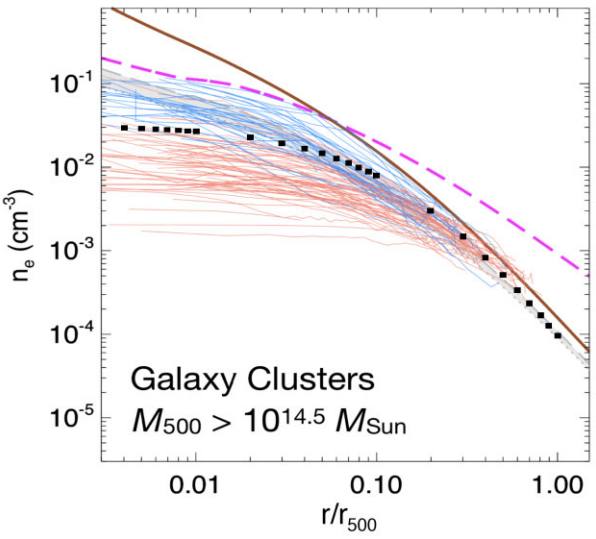

Figure 1. The hot-gas density profile of the biggest cluster in our computational volume ($M_{500} = 10^{14.6}$ M$_\odot$) as predicted by equation (3) (black squares), overplotted in fig. 2 of Voit et al. (2019). The blue and red curves show the density profiles of cool-core and non-cool-core clusters in the Archive of *Chandra* Cluster Entropy Profile Tables (ACCEPT) sample (Cavagnolo et al. 2009), respectively. The brown solid curve shows the Navarro–Frenk–White (NFW) density profile rescaled by the universal baryon fraction. The magenta dashed curve shows the precipitation limit. For densities above this limit, the cooling time is shorter than 10 free-fall times and the gas starts fragmenting into clouds (Voit et al. 2019). The grey shaded area shows hybrid profiles obtained by Voit (2019) combining the cosmological and precipitation limits (i.e. the constraints from the brown solid curve and the magenta dashed curve, respectively). The comparison with Voit (2019) is a minor point in relation to the goals of this paper. Interestingly, however, there is a very good agreement between his model and ours at $r \gtrsim 0.05\,r_{500}$. At $r < 0.05\,r_{500}$, the grey shaded area is not only above the black squares but also close to the upper boundary of the observationally permitted region. Not all clusters are close to the precipitation limit, but most cool-core clusters (blue curves) are.

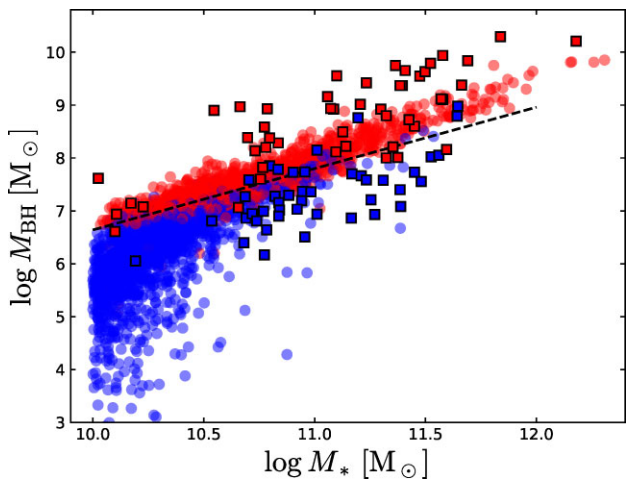

Figure 2. Local BH mass–stellar mass relation for star-forming [specific star formation rate (sSFR) $> 10^{-11}$ yr^{-1}; blue] and passive (sSFR $< 10^{-11}$ yr^{-1}; red) galaxies in model B2 (circles) and the observations (squares; Terrazas et al. 2016). The black dashed line shows the quenching boundary between the star-forming and the passive population (equation 12).

that span two orders of magnitude. Denser cores correspond to lower central entropies. Cool-core clusters are those in which the hot X-ray gas has cooled and condensed into the central region. The lowest central densities occur where heating has prevailed over cooling, so that the central entropy has gone up. In our model, neither cooling nor any non-gravitational heating mechanism, such as feedback from active galactic nuclei (AGN), has any effect on $n_e(r)$. Hence, it is remarkable that our predictions (the black squares) lie right at the boundary between the density profiles of cool-core and non-cool-core clusters.

We model a galaxy as the sum of four components: a disc, a bar/pseudo-bulge, a classical bulge, and a central cusp, composed of the gas that falls to the centre and undergoes starbursts during mergers (Cattaneo et al. 2017). We call it cusp because this material is responsible for the central light excesses in the photometric profiles of cuspy ellipticals, while the rest of the bulge is composed of the pre-existing stellar population (Kormendy et al. 2009).

All the gas that accretes on to galaxies is added to the disc component, the radius of which is computed assuming the conservation of angular momentum (Mo, Mao & White 1998). This assumption may be inaccurate (Jiang et al. 2019) but is highly standard (see Knebe et al. 2015; Somerville & Davé 2015 for on overview of SAMs) and gives mean disc sizes in good agreement with the observations, even though the scatter around the mean is larger in the SAM than in the observations because of the large intrinsic scatter in the halo spin parameter (Cattaneo et al. 2017).

2.3 Star formation and chemical enrichment

Star formation (SF), stellar evolution, chemical evolution, and feedback are computed separately for each component. In GALICS, there are two modes of SF: a quiescent mode, associated with the disc, and a starburst mode, associated with cusps (Hatton et al. 2003; Cattaneo et al. 2006, 2017). There is no SF in bulges, where stellar mass loss is the only source of gas. Bars are the most difficult component to fit in this schematic picture. We make the simplifying assumption of treating their SF as quiescent.

We define the SF time-scale of a component as the time in which it would consume its whole gas reservoir, were its star formation rate (SFR) constant. Many authors call this quantity the gas-depletion time-scale. We do not follow this convention here because SF is not the only process that depletes galaxies of gas. Feedback, too, plays a role.

Bigiel et al. (2008, 2011) advocated a constant SF time-scale of ∼2 Gyr in disc galaxies. Their study suggests that a similar conclusion applies to starbursts with a time-scale ∼10 times shorter. More complex models have been proposed based on both observations (Kennicutt 1998; Kennicutt & Evans 2012; Kennicutt & De Los Reyes 2021) and theory (Krumholz, Dekel & McKee 2012; Somerville, Popping & Trager 2015). We ourselves had considered a SF time-scale proportional to the dynamical time in our previous work (Hatton et al. 2003; Cattaneo et al. 2006, 2017). However, assuming a constant SF time-scale has the practical advantage that our SFRs do not depend on the sizes of discs. This prevents the possibility of artificial starbursts in a tail of discs with abnormally small exponential scale lengths (see Section 2.2, final paragraph).

In this paper, we have treated the SF time-scales for discs and starbursts as free parameters of the SAM. We have found good fits to the observations using 1 Gyr for the disc SF time-scale and 0.2 Gyr for the cusp SF time-scale. The time-scale for discs is shorter than the one found by Bigiel et al. (2011) but in agreement with Combes et al. (2013).

Chemical evolution is modelled as in Koutsouridou & Cattaneo (2019). The only novelty is that we no longer assume primordial metallicity for the intergalactic medium. The enrichment of the intergalactic medium is computed self-consistently as in Cattaneo et al. (2020).

2.4 Stellar feedback

In Cattaneo et al. (2020), we introduced a new model for stellar feedback that takes into account the multiphase structure of galactic winds by separating them into a cold component with mass-loading factor η_c and a hot component with mass-loading factor η_h, so that

$$\eta \equiv \frac{\dot{M}_{out}}{SFR} = \eta_c + \eta_h = 2\left(\frac{\epsilon_{SN_c}}{v_{w_c}^2} + \frac{\epsilon_{SN_h}}{v_{w_h}^2}\right) E_{SN} \Psi_{SN}. \quad (4)$$

Here $\dot{M}_{out}$ is the total mass outflow rate; $\Psi_{SN} = 1/140$ is the number of supernovae per unit stellar mass formed for a Chabrier (2003) initial mass function; $E_{SN} = 10^{51}$ erg is the energy released by one supernova; ϵ_{SN_c} and ϵ_{SN_h} are the fractions of this energy used to power the cold and the hot component of supernova-driven winds, respectively; v_{w_c} and v_{w_h} are the respective speeds.

Cattaneo et al. (2020) modelled the fraction of the total energy from supernovae deposited into the hot component with the functional form

$$\epsilon_{SN_h} = \min\left[\left(\frac{v_{vir}}{v_{SN}}\right)^{\alpha_v} (1+z)^{\alpha_z}, 1\right] \quad (5)$$

based on the results of cosmological hydrodynamic zoom simulations (NIHAO; Tollet et al. 2019), where α_v, α_z, and v_{SN} are free parameters of the SAM; v_{SN} is the virial velocity for which $\epsilon_{SN_h} = 1$ at $z = 0$. The same model applies here with $\alpha_v = -3.6$, $\alpha_z = 2.5$, and $v_{SN} = 55\,\mathrm{km\,s^{-1}}$ (see table 1 of Cattaneo et al. 2020 for the values of the same parameters in GALICS 2.1). The hot wind is assumed to flow out at the escape speed ($v_{w_h} = v_{esc}$).

The assumptions in the paragraph above are used to compute η_h; η_c is determined by the fountain fraction

$$\epsilon_{fount} \equiv \frac{\eta_c}{\eta_c + \eta_h}, \quad (6)$$

since it is the cold wind that falls back and forms the galactic fountain. In GALICS 2.1, ϵ_{fount} was a free parameter of the SAM and its value was the same for all galaxies ($\epsilon_{fount} = 0.7$). However, cosmological hydrodynamic simulations show that ϵ_{fount} decreases with increasing v_{vir} (fig. 10 of Christensen et al. 2016 and fig. 12 Tollet et al. 2019). Here we assume

$$\epsilon_{fount} = 0.7 \min\left[\frac{v_{break}}{v_{vir}}, 1\right], \quad (7)$$

where v_{break} is the virial velocity below which $\epsilon_{SN_h} = 1$ in equation (5).

Our model for what happens to the ejected gas is the same as in GALICS 2.1. The reaccretion time-scale for the cold gas in the galactic fountain is computed with equation (35) of Cattaneo et al. (2020). The fountain has a shorter thermal-evaporation time-scale when the virial temperature is higher (Cattaneo et al. 2020, equation 36). Hot winds escape from the halo entraining the hot circumgalactic medium with them if their momentum is $\dot{M}_{out}^{hot} v_{esc} > M_{hot} g_h$, where $\dot{M}_{out}^{hot}$ is the outflow rate of the hot component, M_{hot} is the mass of the hot circumgalactic medium (i.e. the hot gas within the halo), and g_h is the gravitational acceleration at the centre of the DM halo. They mix with the hot circumgalactic medium if this condition is not satisfied.

2.5 Disc instabilities

Since Mo et al. (1998) and van den Bosch (1998), all SAMs of disc instabilities have been based on the Efstathiou–Lake–Negroponte (ELN; Efstathiou, Lake & Negroponte 1982) criterion (GALACTICUS: Benson 2012; MORGANA: Lo Faro et al. 2009; SAG: Gargiulo et al. 2015; YSAM: Lee & Yi 2013; GALFORM: Gonzalez-Perez et al. 2014; SANTACRUZ: Porter et al. 2014; LGALAXIES: Henriques et al. 2015; SAGE: Croton et al. 2016; SHARK: Lagos et al. 2018).

The problem is that, while the ELN criterion determines not very accurately (Athanassoula 2008) but reasonably well whether a disc is stable or not, it does not by itself predict the mass of the bar or pseudo-bulge that forms in an unstable disc. Additional assumptions are needed.

Some SAMs make the extreme assumption that discs evolve into bulges when the ELN criterion is satisfied (e.g. Gonzalez-Perez et al. 2014; Gargiulo et al. 2015). Others adopt a more conservative approach and transfer to the bulge or pseudo-bulge just enough matter to make the disc stable again (e.g. Henriques et al. 2015; Cattaneo et al. 2017). Here, we adopt a new approach based on computer simulations.

Devergne et al. (2020) studied the growth of pseudo-bulges in isolated thin exponential stellar discs embedded in static spherical haloes. The novelty of their work was the size of the explored parameter space. They found that the final bulge-to-total stellar mass ratio (B/T) could be fitted to an accuracy of 30 per cent by the formula

$$\frac{B}{T} = 0.5 f_d^{1.8}(3.2 r_d), \quad (8)$$

where $f_d = v_d^2/v_c^2$ is the contribution of the disc to the total gravitational acceleration (v_c is the circular velocity at the optical radius $r = 3.2 r_d$, where r_d is the exponential scale length; v_d is the circular velocity considering only the disc's gravity).

We use equation (8) to compute the mass $M_{bar} = (B/T) M_{disc}$ of the bar developed by a thin exponential disc of mass M_{disc} when it is embedded in a spherical mass distribution such that v_c is the total circular velocity at $r = 3.2 r_d$. The gravitational potential of the spherical mass distribution is the sum of the contributions from the DM halo, the bulge, and the central cusp. If there is no bar, a bar of mass M_{bar} is created from stars and gas from the disc. Stars and gas are transferred from the disc to the bar in a proportion that reflects their contributions to the total disc mass.

The case with a pre-existing bar is more complicated because our guidance, equation (8), is based on simulations that did not contemplate that possibility. We compute $v_d(3.2 r_d)$ as if the bar were part of the disc, which is not unreasonable, since the formation of a bar should have no substantial impact on the rotation curve at the optical radius. If the M_{bar} computed with equation (8) is larger than the mass $M_{bar,0}$ of the pre-existing bar, we transfer a mass $M_{bar} - M_{bar,0}$ from the disc to the bulge.

2.6 Mergers

Decades of computer simulations have shown that nearly equal-mass (major) mergers result in the formation of elliptical galaxies and drive gas into the central region, where it fuels starbursts and can feed the growth of supermassive BHs (Toomre & Toomre 1972; Barnes 1988; Barnes & Hernquist 1996; Hopkins et al. 2006). The structure of merger remnants depends not only on the mass ratio of the merging galaxies but also on the initial conditions of the encounters (Negroponte & White 1983; Martin et al. 2018) and the gas content of the merging galaxies (discs can survive even equal-

mass mergers if the gas fraction is large enough; Hopkins et al. 2009). Minor mergers, in which a small satellite plunges into a much larger galaxy, heat the disc of the larger galaxy and can form a thick-disc component (Statler 1988; Quinn, Hernquist & Fullagar 1993). However, the heating efficiency scales as the square of the mass ratio. Hence, a Milky Way galaxy may experience ∼5–10 mergers with a mass ratio of 1:10 and still retain a spiral morphology (Hopkins et al. 2008b).

The challenge for SAMs is to find a simple analytic model that captures the essence of these complex dynamical transformations. Here, we consider two models for the morphologies of the merger remnants. The first is the same as in GALICS 2.0 (Cattaneo et al. 2017) and GALICS 2.1 (Cattaneo et al. 2020). It is the standard assumption in SAMs of galaxy formation. The second is a novelty of GALICS 2.2.

2.6.1 *Model 1*

In the so-called *threshold* model, a sharp threshold $\mu_{\rm m}$ in the mass ratio $\mu = \mathcal{M}_2/\mathcal{M}_1$ of the merging galaxies (with $\mathcal{M}_2 < \mathcal{M}_1$) distinguishes major and minor mergers. Bulges form in major mergers only and major mergers destroy discs completely. Variations of the threshold model remain the standard description of mergers in most SAMs to date (GALACTICUS: Benson 2012; MORGANA: Lo Faro et al. 2009; SAG: Gargiulo et al. 2015; YSAM: Lee & Yi 2013; GALFORM: Gonzalez-Perez et al. 2014; LGALAXIES: Henriques et al. 2015; SAGE: Croton et al. 2016; SHARK: Lagos et al. 2018). We now explain our implementation.

In our SAM, $\mathcal{M}_1$ and $\mathcal{M}_2$ are the total (baryonic plus DM) masses within the baryonic half-mass radii of the merging galaxies. In major mergers ($\mu > \mu_{\rm m}$), all the stars form a large bulge; all the cold galactic gas is transferred to the central cusp, where it fuels a starburst. In minor mergers ($\mu < \mu_{\rm m}$), the gas and the stars in the disc, the bar, the bulge, and the cusp of the smaller (secondary) galaxy are added to the disc, the bar, the bulge, and the cusp of the larger (primary) galaxy, respectively, without any changes in the structural properties of the latter (the sizes and characteristic speeds for the four components remain the same). We assume a threshold ratio of $\mu_{\rm m} = 0.25$ as in Cattaneo et al. (2017, 2020).

2.6.2 *Model 2*

Our second model is based on the results of Kannan et al. (2015), who used high-resolution hydrodynamical simulations to study the formation of classical bulges and the triggering of starbursts in mergers. Let $f_{\rm gas,\,disc}$ be the gas fraction in the disc of the primary galaxy. We assume that during a merger:

(i) a fraction μ of the stars in the disc of the primary galaxy is transferred to the central bulge;
(ii) another fraction 0.2μ of the same stars is scattered into the halo, which acquires in this way a stellar component in our SAM; the same applies to the stars of the secondary galaxy that accrete on to the disc of the merger remnant;
(iii) a fraction $\mu(1 - f_{\rm gas,\,disc})$ of the gas in the disc of the primary galaxy is transferred to the central cusp (gas falls to the centre less easily than stars if $f_{\rm gas,\,disc}$ is significant);
(iv) even the gas that remains in the disc undergoes a starburst in the case of a major merger ($\mu > 0.25$); we assume that this gas has the same SF time-scale as that in the central cusp;
(v) the stars and the gas in the bar of the primary galaxy are transferred to the bulge and the cusp of the merger remnant, respectively;
(vi) a fraction μ of all the stars of the secondary galaxy ends up in the bulge of the merger remnant; the rest is added to the disc;
(vii) all the gas of the secondary galaxy ends up in the central cusp.

Assumptions (i), (ii), (iii), and (vi) are based on the quantitative results of Kannan et al. (2015).[1] Assumption (iv) is in agreement with observations (Pan et al. 2019) and simulations (Powell et al. 2013; Kannan et al. 2015): most major mergers (including the Antennae; Wang et al. 2004) exhibit extended SF in their early stages, which then concentrates to the nuclear region as the galaxies approach coalescence. Our assumption that all the gas is exhausted during major mergers, whether it is transferred to the central cusp or remains in the disc, may be too strong. On the other hand, minor mergers (Kannan et al. 2015) and fly-by events (e.g. Lamastra et al. 2013) can induce starbursts, too, and we do not consider these effects in our SAM.

Kannan et al. (2015) did not consider the presence of bars and thus could not comment on their evolution (assumption v). Bournaud & Combes (2002) and Berentzen et al. (2007) found that the formation of a bulge and dynamical heating can prevent even gas-rich systems from retaining a bar after a major merger. The transfer of stars from the bar to the bulge during mergers is an assumption that has no effect on our results because it changes neither the total stellar masses nor the SFRs of galaxies, but what happens to the gas in the bar is important. Our assumption that it moves to the cusp is based on an observationally motivated consideration: bars funnel gas into the central region, where it settles into a nuclear minispiral, which the interaction with a companion can easily destabilize (see e.g. NGC 1097; Prieto et al. 2019). The latter assumption is crucial for BH growth and quenching in our model, since our SAM does not currently include a model for BH accretion via disc instabilities.

Kannan et al. (2015) did not comment on the fate of the gas in the secondary galaxy either (assumption vii), but we have tested different models, and we have found, in agreement with Hopkins et al. (2009), that out treatment of the gas in the secondary galaxy has no statistically significant effect on the evolution of galaxies.

2.7 Supermassive black hole growth

We assume that BHs are fed by the cold gas that accumulates in the central cusp during mergers. The formation of a BH is triggered when the mass of this gas exceeds our minimum BH mass $M_{\rm seed} = 10^2\,{\rm M}_\odot$ (consistent with stellar mass BHs formed from the collapse of massive Population III stars; e.g. Latif & Ferrara 2016; Haemmerlé et al. 2020). If both galaxies have pre-existing BHs, we assume that they instantaneously coalesce, so that any accretion occurs on to the remnant.

The BH growth rate is given by

$$\dot{M}_{\rm BH} = (1 - \epsilon_{\rm rad})\,\min(\epsilon_{\rm BH}{\rm SFR}_{\rm cusp},\, \dot{M}_{\rm Edd}), \tag{9}$$

[1] Kannan et al. (2015) find $f_{\rm db} = 0.3\mu$–0.4μ, where $f_{\rm db}$ is the fraction of the primary's stellar disc that ends up in the bulge of the merger remnant. Using high-resolution hydrodynamic simulations, Hopkins et al. (2009) find $f_{\rm db} = \mu$, instead. We adopt the latter value here (assumption i), as we find it to be in better agreement with the observed B/T.

where $\epsilon_{\rm rad} = 0.1$ is the radiative efficiency of BH accretion, $\epsilon_{\rm BH}$ is a parameter of the SAM that sets the efficiency of BH accretion relative to that of SF in the cusp component, $\rm SFR_{cusp}$ is the SFR in the central cusp, and $\dot{M}_{\rm Edd} = M_{\rm BH}\, c^2/t_{\rm S}$ is the Eddington rate, where c is the speed of light and $t_{\rm S}$ is the Salpeter (1964) time-scale.

In the previous versions of our SAM, $\epsilon_{\rm BH}$ was a constant (Cattaneo et al. 2017 used $\epsilon_{\rm BH} = 0.01$). However, the observations show that $\dot{M}_{\rm BH}/\rm SFR$ depends on M_* (e.g. Delvecchio et al. 2015; Rodighiero et al. 2015). Hydrodynamic simulations reproduce this behaviour and show that it can be explained as a consequence of supernova feedback. Supernovae blow the gas out and thus prevent its accretion on to BHs in systems with low escape speeds (Dubois et al. 2015; Anglés-Alcázar et al. 2017; Bower et al. 2017; Habouzit, Volonteri & Dubois 2017; Davies et al. 2019b; Dekel, Lapiner & Dubois 2019; Oppenheimer et al. 2020).

Here, we follow Kauffmann & Haehnelt (2000) and assume that

$$\epsilon_{\rm BH} = \frac{\epsilon_{\rm BH,max}}{1 + (\upsilon_{\rm BH}/\upsilon_{\rm vir})^2}, \tag{10}$$

where $\epsilon_{\rm BH,\,max}$ and $\upsilon_{\rm BH}$ are free parameters that control the maximum BH accretion efficiency and the virial velocity at which the efficiency saturates, respectively. This choice prevents the formation of overly massive BHs in low-mass, gas-rich galaxies.

Kauffmann & Haehnelt (2000) and Croton et al. (2006) fitted the observations with $\epsilon_{\rm BH,\,max} = 0.03$ and $\upsilon_{\rm BH} = 280$ km s^{-1}. We adopt the same $\upsilon_{\rm BH}$ as theirs, but a higher $\epsilon_{\rm BH,\,max} = 0.05$, since our efficiency is relative to $\rm SFR_{cusp}$ and not to the total SFR.[2] The case for $\epsilon_{\rm BH,\,max} = 0.05$ can be justified a posteriori, with the argument that it fits the $M_{\rm BH}$–M_* relation at $z = 0$, but it can also be made a priori (directly from observations, without running the SAM). Harris et al. (2016) have shown that bright quasars with bolometric luminosities $\gtrsim 3 \times 10^{13}\,\rm L_\odot$ (and thus accretion rates $\gtrsim 20\,\rm M_\odot\,yr^{-1}$) at $2 < z < 3$ live in galaxies with SFRs of 400–500 $\rm M_\odot\,yr^{-1}$. It is reasonable to presume that, in such systems, most of the SF comes from the starburst mode (e.g. fig. 3 of Cattaneo et al. 2013). That gives a BH accretion rate-to-SFR ratio of $\epsilon_{\rm BH} \sim 20/400 = 0.05$, in agreement with our best-fitting value for $\epsilon_{\rm BH,\,max}$.

In GALICS 2.2, mergers are the only channel for BH growth. We do not consider accretion via disc instabilities or direct accretion of gas from the hot halo. However, the role of disc instabilities is indirectly accounted for through the assumption (v) of model 2 in Section 2.6. The accretion of hot gas is important to explain the population of weak radio galaxies, i.e. Fanaroff–Riley Class I radio sources (Fanaroff & Riley 1974), but it makes a small contribution to BH masses (e.g. Cattaneo & Best 2009).

2.8 Quenching

Observationally, the probability that a central galaxy is passive correlates strongly with both $M_{\rm vir}$ and $M_{\rm BH}$ (e.g. Bluck et al. 2020). We thus consider two different models for the quenching of galaxies.

[2]In Kauffmann & Haehnelt (2000), all the cold gas present after a merger was available for BH growth. In Croton et al. (2006), the mass of the available gas was reduced by a factor equal to the merger ratio μ, as in our model, but there was no additional factor $1 - f_{\rm gas,\,disc}$ (assumption iii)). Our BHs have a smaller accretion reservoir. They need to accrete more efficiently to reach the same final masses.

2.8.1 *Model A*

The notion of halo quenching has its origin in the work of Dekel & Birnboim (2006) and Cattaneo et al. (2006). In our current implementation (model A), cold-mode accretion, hot-mode accretion, and reaccretion from the galactic fountain are shut down when $M_{\rm vir} > M_{\rm crit}$, the latter being a free parameter of the SAM. Dekel & Birnboim (2006) identified $M_{\rm crit}$ with the critical halo mass above which a stable shock heats the infalling cold intergalactic gas, giving rise to a hot quasi-hydrostatic atmosphere. However, shock heating does not necessarily cause quenching if it is not accompanied by a mechanism that prevents the shock-heated gas from cooling down again (e.g. Cattaneo et al. 2020).

Cattaneo et al. (2006) proposed that BHs couple to the hot gas and that this coupling provides this mechanism. As soon as a hot atmosphere appears, the growth of BHs becomes self-regulated. In the self-regulated regime, BHs accrete slowly and release most of their power mechanically, via radio jets. The power released is just enough to compensate the radiative losses to X-rays, so that the hot gas is kept in thermodynamic equilibrium (Cattaneo & Teyssier 2007; Cattaneo et al. 2009, and references therein).

Halo quenching can thus act as an implicit model for radio-mode AGN feedback. Cattaneo et al. (2006) reproduced the colour–magnitude distributions of galaxies in the Sloan Digital Sky Survey (SDSS) in quantitative detail with a shutdown of cooling above $M_{\rm crit} = 2 \times 10^{12}\,\rm M_\odot$.

2.8.2 *Model B*

Silk & Rees (1998) proposed a model in which BHs grow until the energy deposited in the surrounding gas is larger than the one required to unbind the entire baryonic content of their host haloes. Energy injection is not the only mechanism through which AGN can drive winds and blow away the gas in their host galaxies (momentum can drive them, too; Fabian 1999; King 2003). Energy driving is, however, the most common assumption (Ostriker, Bode & Babul 2005; Booth & Schaye 2010, 2011; Bower et al. 2017; Oppenheimer 2018; Davies et al. 2019a, 2020; Oppenheimer et al. 2020) and appears in better agreement with recent observational developments (see below).

Chen et al. (2020) derived a similar model, which they called the binding-energy model, from an observational rather than theoretical perspective. Using a simple analytic model, they found that they could reproduce several scaling relations for star-forming and passive galaxies if galaxies cross into the green valley when the time-integrated energy output of their central BHs exceeds a multiple $f_{\rm bind}$ of the gravitational binding energy of the baryons within the halo, i.e. when

$$\epsilon_{\rm BH\,fb} M_{\rm BH}\, c^2 > f_{\rm bind} E_{\rm bind} = f_{\rm bind} \times \frac{1}{2} f_{\rm b} M_{\rm vir} \upsilon_{\rm vir}^2, \tag{11}$$

where $\epsilon_{\rm BH\,fb}$ is the efficiency for converting BH accretion into the useful feedback energy and c is the speed of light. The important point here is that although $\epsilon_{\rm BH\,fb}$ and $f_{\rm bind}$ are considerably uncertain, only their ratio $\epsilon_{\rm eff} = \epsilon_{\rm BH\,fb}/f_{\rm bind}$ matters to determine the critical BH mass,

$$M_{\rm BH}^{\rm crit} = \frac{f_{\rm b}}{2\epsilon_{\rm eff}} M_{\rm vir} \left(\frac{\upsilon_{\rm vir}}{c}\right)^2, \tag{12}$$

above which a galaxy is quenched. Equation (12) translates into a relation between $M_{\rm BH}^{\rm crit}$ and M_* because of the tight M_*–$M_{\rm vir}$ relation at a given z. We treat its normalization $\epsilon_{\rm eff}$ as a free parameter of

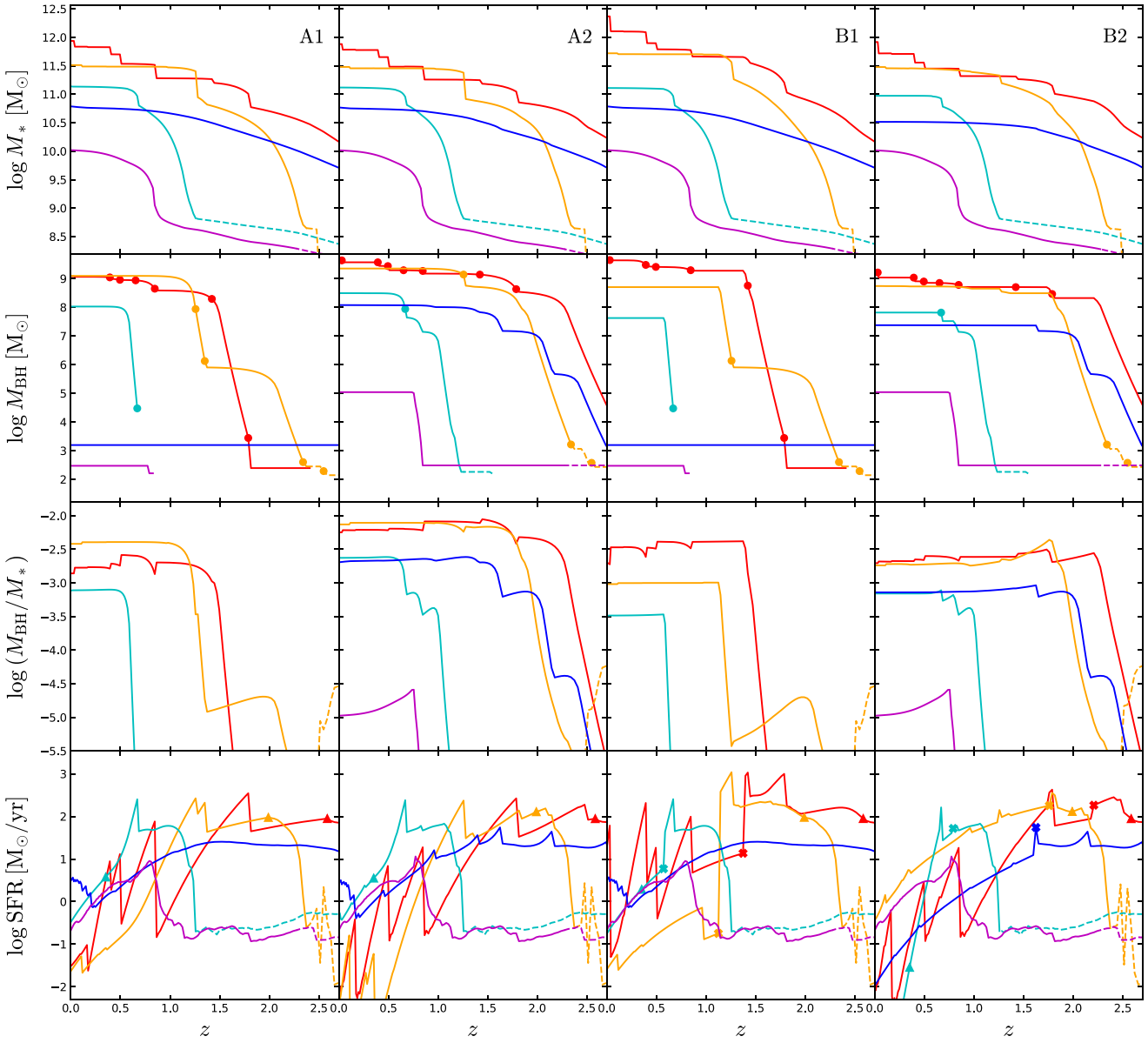

Figure 3. Evolutionary tracks of five galaxies in the four different models considered in this work (from left to right: A1, A2, B1, and B2), on the M_*–z, M_{BH}–z, M_{BH}/M_*–z, and SFR–z plane. Dashed lines indicate periods when galaxies are satellites, while solid lines indicate periods when galaxies are centrals (a galaxy that was a satellite at some point in the past and is central at $z = 0$ is a backsplash galaxy, i.e. a galaxy that has passed through a group or a cluster and come out to the other side). Points in the second row indicate major mergers, which correspond to the strongest bursts of SF visible in the fourth row. Minor mergers are not noted due to their high frequency. Coloured triangles in the fourth row show the point when each galaxy is quenched in the halo quenching model A, i.e. when $M_{vir} > 2 \times 10^{12}$ $M_\odot$, while X marks show when a galaxy is quenched in model B1 (third panel) and B2 (fourth panel), i.e. when it satisfies the BH-quenching criterion (12).

the SAM and set it to $\epsilon_{eff} = 0.00115$ to reproduce the quenching boundary $M_{BH}^{crit}(M_*)$ that separates passive and star-forming galaxies on the M_{BH}–M_* diagram at $z = 0$ (Fig. 2; the dashed curve has the same normalization as in fig. 2c of Chen et al. 2020).

As soon as the condition in equation (12) is satisfied, all the gas within the central components of the galaxy (the central cusp, the bulge, and the bar, but not the disc) is reheated and transferred instantaneously to the hot atmosphere. So is the gas within the cold filaments and the galactic fountain, after which the cooling of hot gas is permanently shut down. The subsequent evolution of the galaxy is characterized by a decreasing SFR as the cold gas left over in the disc is depleted, unless gas-rich mergers reactivate SF and BH growth (Fig. 3).

Our assumption that the cold gas heated by quasar feedback mixes with the hot atmosphere was made to conserve mass (we had to put it somewhere, be it within the halo or in the intergalactic medium) but has no bearing on the properties of galaxies, since we do not let it reaccrete. It would matter, however, if we extended the scope of our analysis from galaxies to the circumgalactic medium by comparing to X-ray observations of the hot gas in early-type galaxies or if we addressed the maintenance problem instead of just assuming that, after the initial quenching, accretion is permanently shut down.

Physically, it is reasonable to suppose that the effects of quasar feedback should depend on the depth of the gravitational potential well. In a low-mass galaxy, quasar feedback may empty the halo of its entire gaseous content if the condition in equation (12) is satisfied. In a group, that is a more difficult proposition. In a cluster, it is downright impossible (we see the gas in X-rays).

Observationally, a recent study (Hou et al. 2021) has analysed archival *Chandra* X-ray data for 57 field early-type galaxies in the mass range $10^9 \lesssim M_* \lesssim 10^{11}\,\mathrm{M}_\odot$. Only in seven of those (all at $M_* > 10^{11}\,\mathrm{M}_\odot$), there is a firm detection of extended X-ray emission and only one (NGC 3193) has an X-ray halo that extends beyond the starlight distribution. Retrospectively and without consequences for the conclusions of this paper, it thus seems that a complete expulsion of all the gas in the halo would have been a more realistic assumption, at least for $M_* < 10^{11}\,\mathrm{M}_\odot$.

To facilitate a better understanding of the models considered in this work, Fig. 3 compares the evolution of five galaxies (two giants, two with masses comparable to that of the Milky Way, and one of low mass) in models A1, A2, B1, and B2. For halo quenching (A), the merger model has little impact on the evolution of M_*. For BH quenching (B), quenched galaxies end up with higher masses in merger model 1 than in merger model 2 (in model B1, quenching occurs later because one has to wait for a major merger). The final stellar masses in model B2 are similar to those in models A1 and A2 for giant galaxies but slightly lower at intermediate masses.

The merger model affects $M_{\rm BH}$ more than the quenching mechanism does. In model 1, BHs form and grow in major mergers only (circles), although a galaxy can also inherit a BH from a smaller companion in a minor merger (magenta curve in models A1 and B1). In model 2, minor mergers, too, play a role; BHs grow faster; galaxies are quenched earlier and there is less SF at $z = 0$.

In giant galaxies (red and orange curves), halo quenching usually occurs earlier than BH quenching. At intermediate masses (cyan and blue curves), the opposite is true. The galaxy corresponding to the blue curve is never quenched in model A. At lower masses ($M_* \approx 10^{10}\,\mathrm{M}_\odot$), galaxies are commonly not quenched in any model (see also Fig. 10). Even in that case, however, the SFR can decline at low z if the accretion rate on to the halo declines (magenta curve). In the same way that the SFR can decline without quenching, the opposite is also possible. Quenching does not always mark the end of SF, which can be reactivated by gas-rich mergers.

2.9 Environmental effects: ram-pressure stripping

Stripping is the removal of matter from satellite galaxies or their circumgalactic media through environmental effects, i.e. tides and ram pressure. Tides can take place due to interactions with the gravitational potential of the host halo (tidal stripping; Merritt 1984; Byrd & Valtonen 1990; Mayer et al. 2006) or to smaller scale galaxy–galaxy interactions (harassment; Farouki & Shapiro 1981; Moore et al. 1996; Duc & Bournaud 2008). Tidal stripping affects not only gas but also DM and stars, but ram pressure strips the gas more effectively than tides do (e.g. Simpson et al. 2018).

Independently of its physical mechanism (tides or ram pressure), stripping manifests differently depending on its intensity. Weak stripping deprives satellite galaxies of their circumgalactic media. Without an accretion reservoir, their SFR declines gradually until all the cold gas is exhausted. Proposed by Larson, Tinsley & Caldwell (1980) to explain the high ratio of S0 to spiral galaxies in clusters, this evolutionary picture has often been referred to as 'strangulation' (Balogh, Navarro & Morris 2000) or 'starvation' (Bekki, Couch & Shioya 2002). Strong stripping removes matter from within galaxies, starting from their outer parts (Gunn & Gott 1972; Abadi, Moore & Bower 1999). The jellyfish galaxies (Ebeling, Stephenson & Edge 2014) are the most extreme observational evidence of this phenomenon.

In early SAMs, there was no accretion on to satellite galaxies. This assumption is equivalent to an extreme form of strangulation and leads to predicting that nearly all satellites are passive, at odds with observations (Baldry et al. 2006; Weinmann et al. 2006; Kang & van den Bosch 2008; Kimm et al. 2009) and cosmological simulations (Cattaneo et al. 2007). Font et al. (2008) and Guo et al. (2011, 2013) improved the agreement with the observations by considering that the circumgalactic media of satellite galaxies are stripped gradually. Henriques et al. (2015) found that the ram-pressure stripping (RPS) model of Guo et al. (2013) worked well for clusters of galaxies but was too efficient and thus produced too many quiescent satellites in groups.

In the standard versions of GALICS 2.0 (Cattaneo et al. 2017) and GALICS 2.1 (Cattaneo et al. 2020), there was no explicit model for tides or RPS, but tidal stripping was partially accounted for through its effects on the DM (haloes stop growing and accreting baryons once they become subhaloes; e.g. Tollet et al. 2017). In GALICS 2.0, this tidally induced starvation was more than sufficient to shut down gas accretion on to satellites because galaxies grew through cold accretion only and cold filaments disappear quickly (on a free-fall time) if they are not replenished by accretion from the intergalactic medium. Too many satellites were passive. Adding stripping (as we did in one of the versions of GALICS 2.0 considered in Koutsouridou & Cattaneo 2019) would only make the problem worse. In GALICS 2.1, satellite galaxies can accrete through cooling. Hence, RPS is important to avoid that mergers with gas-rich satellites reactivate SF in massive central galaxies. In contrast, the effects of tidal stripping on the gas content of satellites are negligible compared to those of RPS (e.g. Simpson et al. 2018).

Our model for RPS is the same as the one described in Koutsouridou & Cattaneo (2019) and very similar to that by Cora et al. (2018). It is based on the notion that stripping is a two-stage process. The hot gas is stripped first. The removal of cold galactic gas does not begin until the hot atmosphere has been stripped down to the size of the galaxy (Bekki 2009).

The RPS of hot gas is computed following McCarthy et al. (2008), who showed that a simple analytic model for a spherical gas distribution reproduces remarkably well the stripping measured in hydrodynamic simulations. The RPS radius $R_{\rm s,\,h}$ for the hot gas is determined by requiring that the ram pressure exerted by the hot gas of the host on the satellite should balance the subhalo's gravitational restoring force:

$$p_{\rm ram}(r) = \rho_{\rm host}(r) v_{\rm sat}^2(r) = \frac{\pi}{2} \frac{G M_{\rm sat}(R_{\rm s,h}) \rho_{\rm sat}(R_{\rm s,h})}{R_{\rm s,h}}. \quad (13)$$

Here r is the distance of the satellite from the centre of the host system, $\rho_{\rm host}(r)$ is the density of the hot atmosphere of the host at the satellite's location, $v_{\rm sat}(r)$ is the orbital speed of the satellite in the host's reference frame, G is the gravitational constant, $\rho_{\rm sat}(R_{\rm s,\,h})$ is the density of the hot atmosphere of the satellite at the stripping radius, and $M_{\rm sat}(R_{\rm s,\,h})$ is the mass of the subhalo within that radius; $\rho_{\rm host}$ and $\rho_{\rm sat}$ are obtained by applying equation (3) to the host halo and the subhalo, respectively.

The hot gas retained by the satellite is that within $R_{\rm s,\,h}$. Its mass is given by the volume integral of $\rho_{\rm sat}$ within a sphere of radius $R_{\rm s,\,h}$. If this integral returns a value lower than the mass of the hot gas within the subhalo, the difference is stripped from the subhalo and transferred to the host. If $R_{\rm s,\,h}$ is smaller than the optical radius of

the satellite (i.e. the radius that contains 83 per cent of the baryonic mass of the galaxy), we assume that the whole gaseous halo of the satellite is stripped. In our SAM, this is also a necessary condition for the stripping of cold galactic gas.

Our model for RPS of cold gas in the discs of satellite galaxies is based on Gunn & Gott (1972). The equation for the stripping radius $R_{\rm s,\,c}$ of the cold gaseous disc is

$$p_{\rm ram} = \rho_{\rm h}(r)\upsilon^2_{\rm sat}(r)\cos\theta = 2\pi\, G\,\Sigma_{\rm disc}(R_{\rm s,c})\Sigma_{\rm gas}(R_{\rm s,c}). \quad (14)$$

Equation (14) is analogous to equation (13) with two differences. First, the left-hand side depends on the angle θ between the disc's rotation axis and the orbital velocity of the satellite galaxy (i.e. the direction of the hot wind that blows the cold gas away if we examine the problem in the reference frame of the satellite). Second, the restoring force at the right-hand side is now assumed to be due to the disc's rather than the halo's gravity. $\Sigma_{\rm disc}$ and $\Sigma_{\rm gas}$ are the total surface density (stars plus gas) and the surface density of the gaseous disc, respectively (we assume that they have the same exponential scale length). Once we have determined the RPS radius $R_{\rm s,\,c}$ for the cold mass, the calculation of the cold gas stripped from the satellite proceeds in the same manner as we did for the hot gas. We compute the mass of the cold gaseous disc within $R_{\rm s,\,c}$. If this mass is lower than the gas mass of the disc, the difference is removed from the disc and transferred to the hot atmosphere of the host halo.

Several studies have shown that the stripping radii from equation (14) are similar to those found in hydrodynamical simulations, albeit somewhat smaller (e.g. Roediger & Brüggen 2007; Tonnesen & Bryan 2009; Steinhauser, Schindler & Springel 2016). This small discrepancy is understandable because the inner halo should make some contribution to the disc's restoring force and also because of the clumpiness of the cold gas, which makes it more resilient to RPS.

3 RESULTS

In this section, we compare the predictions of our two quenching models (the halo-quenching model A and the BH/binding-energy-quenching model B) when combined with the merger models 1 and 2 described in Section 2.6. That makes four models in total: A1, A2, B1, and B2.

Fig. 4 focuses on the GSMF at $0 < z < 2.5$. With halo quenching, the GSMF does not depend on the assumed merger model. Models A1 and A2 are almost indistinguishable. For $M_{\rm crit} = 2 \times 10^{12}\,{\rm M}_\odot$, they are both in good agreement with the observations (data points with error bars) at all redshifts. The only discrepancy is at the knee of the GSMFs, where halo quenching predicts a sharp decrease in the number of massive galaxies in contrast to the smoother downturn seen in observations. This feature arises from the abrupt shutdown of all accretion modes at $M_{\rm crit}$. We have run tests and we have verified that we could improve the agreement with the observations by using three different critical masses for the shutdown of cooling, cold accretion, and reaccretion from the galactic fountain, but here we have tried to keep the model as simple as possible to make our point clearer.

With BH/binding-energy quenching, the merger model has a significant impact on the GSMF. Model B2 clearly fits the observations much better than model B1. Its only problem is a slight excess in the number densities of very massive galaxies but this could be attributed to the lack of tidal stripping in our SAM (compare the green curve with tidal stripping and the blue curve without it in Tollet et al. 2017, fig. 10, right-hand panel).

The fraction of passive galaxies as a function of M_* (Fig. 5) display a similar behaviour. Models A1 and A2 are, once again, almost indistinguishable. They both fail to reproduce the observed passive fractions of central galaxies at all but the highest masses, i.e. at $\log(M_*/{\rm M}_\odot) \lesssim 11.5$. Model B1 is even more catastrophic. Model B2 is the only one in reasonable agreement with the passive fractions of both central and satellite galaxies, if we exclude its difficulty at explaining the existence of a small population of passive central dwarf galaxies. Notice that all models reproduce the fraction of passive satellites below $\sim 10^{10}\,{\rm M}_\odot$, stressing the environmental origin of their quenching mechanisms (i.e. strangulation/RPS). Fig. 6 (BH masses) and Figs 8 and 9 (morphologies) explain why model B2 works much better than model B1.

The red, green, and blue curves in Fig. 6 show the median $M_{\rm BH}$–M_* relation for galaxies of different B/T ratios ($B/T > 0.7$, $0.3 < B/T < 0.7$, and $B/T < 0.3$, respectively). In all models, $M_{\rm BH}/M_*$ grows with B/T but the trend is far more pronounced in models A1 and B1 than it is in models A2 and B2. The reason is as follows. Merger model 1 assumes that, in minor mergers, the gas in the disc and the bar of the secondary galaxy is moved to the disc and the bar of the merger remnant, respectively, and there it stays unless the minor merger is followed by a major one. So does the gas in the disc and the bar of the primary galaxy. In merger model 2, even minor mergers can transfer gas to the central cusp (Section 2.6) and contribute to the growth of the central BH. Hence, the quenching condition in equation (12) is more easily satisfied.

Furthermore, if spheroids form through major mergers only, as model 1 assumes, then repeated minor mergers create a route for the formation of giant spirals. The fraction of massive galaxies ($M_* \sim 10^{11}\,{\rm M}_\odot$) with $B/T < 0.1$ is almost 20 per cent in model A1 and more than 20 per cent in model B1, but only a few per cent in models A2 and B2, in much better agreement with the observations (Fig. 9; Mendel et al. 2014). Model A1 has fewer massive galaxies with $B/T < 0.1$ than model B1 because halo quenching prevents merger remnants from reaccreting a disc at $M_{\rm vir} > M_{\rm crit}$. In model B1, it is $M_{\rm BH}$ that counts. Hence, the low BH-to-stellar mass ratios of giant spirals result in too much SF at high masses (Fig. 5) and too many massive galaxies in general (Fig. 4).

Having analysed and understood how the four models behave, we now we wish to look at the comparison with the observational data in closer detail and especially at the BH masses, which are crucial for model B. The observational data (Davis et al. 2018; Sahu et al. 2019) show that, for a same stellar mass, galaxies with higher B/T host BHs that are systematically more massive (Fig. 6). This finding is in qualitative agreement with our theoretical predictions.

The quantitative difference between the $M_{\rm BH}$ of different morphological types in the observations is smaller than our predictions for models A1 and B1 but larger than our predictions for models A2 and B2. This finding suggests that the role of minor mergers and disc instabilities in the growth of supermassive BHs is underestimated in merger model 1 and overestimated in merger model 2. GALICS 2.2 does not contain any explicit model for the growth of BHs via disc instabilities but disc instabilities form bars. The assumption (v) of merger model 2 that the gas in the bar of the primary galaxy falls to the centre during mergers is the back door through which disc instabilities affect the growth of BHs in our SAM. Assuming that all this gas falls to the centre even in very minor mergers is too extreme. Replacing this assumption with a more realistic model for BH growth via disc instabilities is likely to widen the gap between the $M_{\rm BH}$–M_* relations for spiral and elliptical galaxies and thus improve the agreement with the data points.

Despite the above remark, models A2 and B2 fit the data much better than models A1 and B1, especially for late-type galaxies. In merger model 2, the $M_{\rm BH}$–M_* relation for galaxies with $B/T < 0.3$

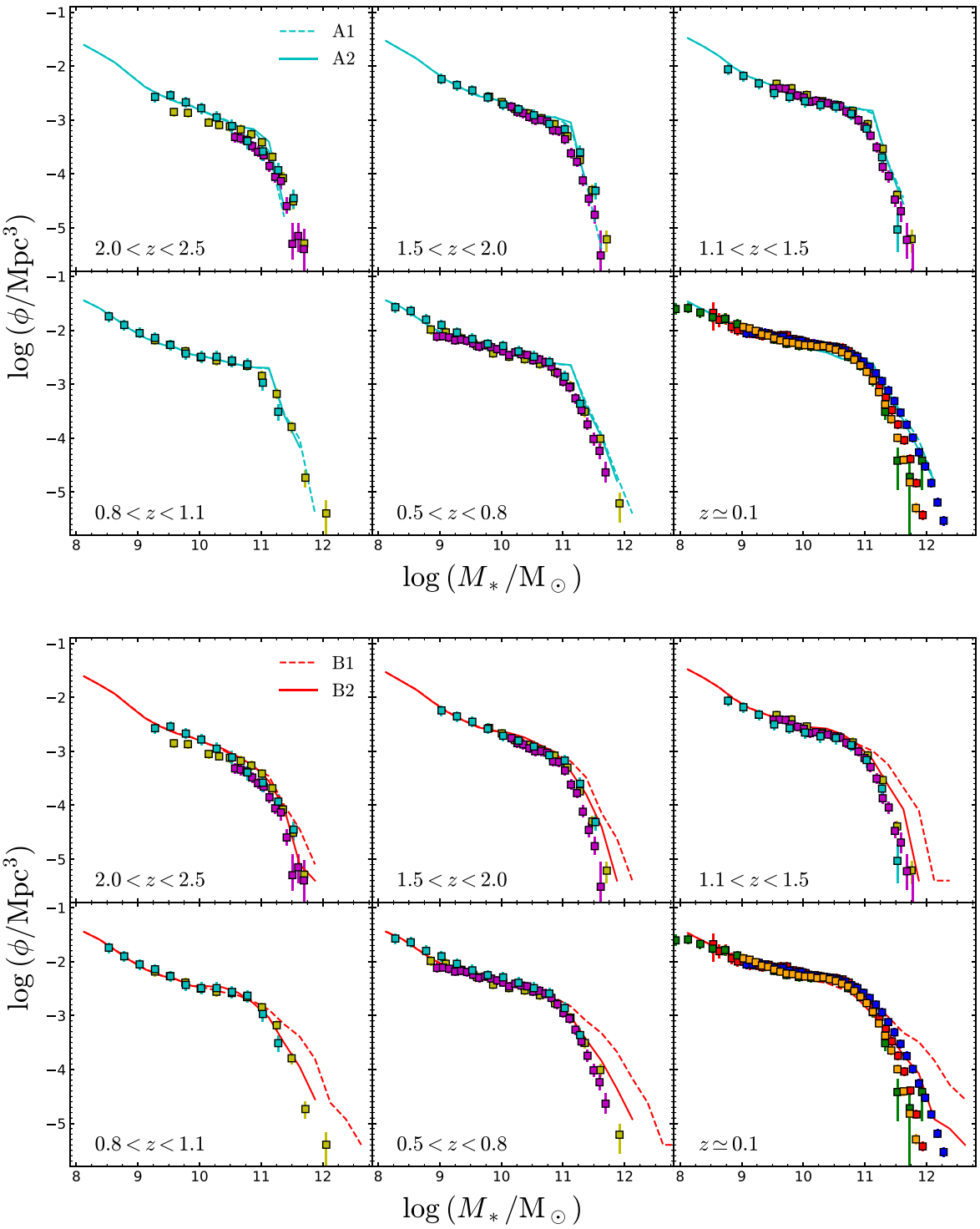

Figure 4. Galaxy stellar mass functions (GSMFs) at $0 < z < 2.5$ for models A1 and A2 (above; cyan dotted and solid curves, respectively) and models B1 and B2 (below; red dotted and solid curves, respectively). The data points show the observations from Ilbert et al. (2013; yellow), Muzzin et al. (2013; magenta), Tomczak et al. (2014; cyan), Yang et al. (2009; red), Baldry et al. (2012; green), Bernardi et al. (2013; blue), and Moustakas et al. (2013; orange).

overlaps with the observations of Davis et al. (2018, blue circles), while merger model 1 is completely off.

For $B/T > 0.3$ (red and green curves), model A2 fits the data points (circles in the colour range from red to green; Sahu et al. 2019) better than model B2 because, in model A2, BHs keep growing until SF depletes the accretion reservoir even if the quenching criterion is satisfied. In model B2, the gas in the central cusp is instantaneously ejected once the quenching criterion is satisfied (some BH accretion is nevertheless possible after quenching if mergers with gas-rich satellites replenish the cusp; Fig. 3).

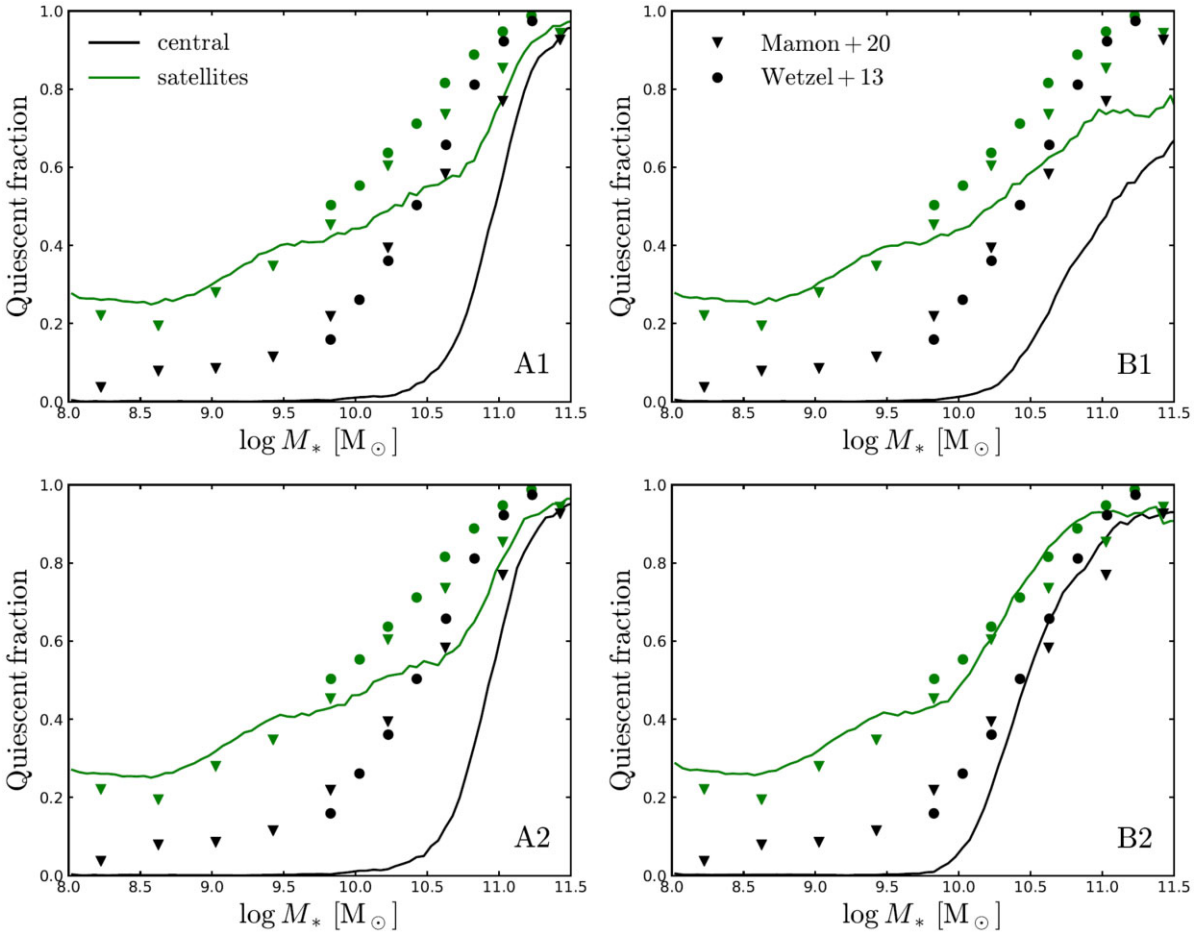

Figure 5. Fraction of passive central (black) and satellite (green) galaxies as a function of stellar mass at $z \sim 0.1$. The left-hand panels show models A1 (top) and A2 (bottom) and right-hand panels show models B1 (top) and B2 (bottom). Data points in all panels show the observations of Wetzel et al. (2013; circles) and Mamon et al. (2020; triangles). Passive galaxies are defined as those with sSFR = SFR/M_* < 10^{-11} yr^{-1}.

We could bring model B2 in better agreement with the BH masses of early-type galaxies by increasing the normalization $\epsilon_{\rm BH,\,max}$ of the efficiency of BH accretion while lowering the efficiency $\epsilon_{\rm eff}$ of BH feedback, so that the fraction of passive galaxies and the GSMF stay the same. However, that would have two drawbacks. First, the dashed line in Fig. 2 would no longer be at the boundary that separates the red and the blue squares. Second, the change would bring model B2 in conflict with direct measurements of the BH accretion rate-to-SFR ratio.

Delvecchio et al. (2015) have studied ∼8600 star-forming galaxies selected from surveys in the far-infrared with *Herschel* and measured their BH accretion rates from stacked X-ray data (*Chandra*). The data points with error bars in Fig. 7 show their findings. The BH accretion rate-to-SFR ratio increases with M_*. GALICS 2.2 (model B2, coloured curves) reproduces this trend. In both the model and the observations, the highest stellar masses drop out of the lowest redshift panels because the galaxies with the highest masses are the first to become passive (downsizing). Quantitatively, the BH accretion rates for galaxies with $10^{10.5} < M_* < 10^{11}\,{\rm M}_\odot$ are in reasonable agreement with the observations at all z. At $10^{11} < M_* < 10^{11.5}\,{\rm M}_\odot$, the model and the observations are in reasonably good agreement at $z \geq 0.65$. At $z \leq 0.35$, however, BH accretion rates are *overpredicted* both at $M_* > 10^{11}\,{\rm M}_\odot$ (red, orange curves) and at $10^{10} < M_* < 10^{10.5}\,{\rm M}_\odot$ (cyan curves). That is the opposite of what one expects if the efficiency of BH accretion is underestimated.

A bias in the observed BH masses may explain the tension between Figs 2 and 7. Dynamical mass estimates, such as those used by Sahu et al. (2019), require that the gravitational sphere of influence of the BH should be resolved. Shankar et al. (2016) argued that this requirement introduces a bias in favour of galaxies with higher BH masses. Model B2 is in reasonably good agreement with Shankar et al. (2016)'s debiased $M_{\rm BH}$–M_* relation (Fig. 6, black dashed curve). Alternatively, the mean BH accretion rate may be underestimated because of the relatively small size of Delvecchio et al. (2015)'s sample compared to the rarity of the quasar phenomenon (in the local Universe, there is one quasar every $\sim 10^4$ early-type galaxies; Wisotzki, Kuhlbrodt & Jahnke 2001).

Another factor that can alter $M_{\rm BH}$ and thus the final galaxy stellar masses is the assumed initial BH mass $M_{\rm BH,\,seed}$ (models in which AGN quenching depends on $\dot{M}_{\rm BH}$ rather than $M_{\rm BH}$ are less sensitive on $M_{\rm BH,\,seed}$; e.g. Somerville et al. 2008). BHs with $M_{\rm BH,\,seed} \simeq 100\,{\rm M}_\odot$ are expected to have formed at high redshift ($z \sim 15$–30) from the collapse of massive Population III stars (stellar mass BHs; Latif & Ferrara 2016; Haemmerlé et al. 2020). Observations of extremely bright quasars at $z \sim 6$–7.5 hosting supermassive BHs with $M_{\rm BH} \sim 10^8$–$10^{10}\,{\rm M}_\odot$ (e.g. Fan 2006; Wu et al. 2015; Bañados et al. 2018) are, however, in tension with this BH formation scenario if subsequent

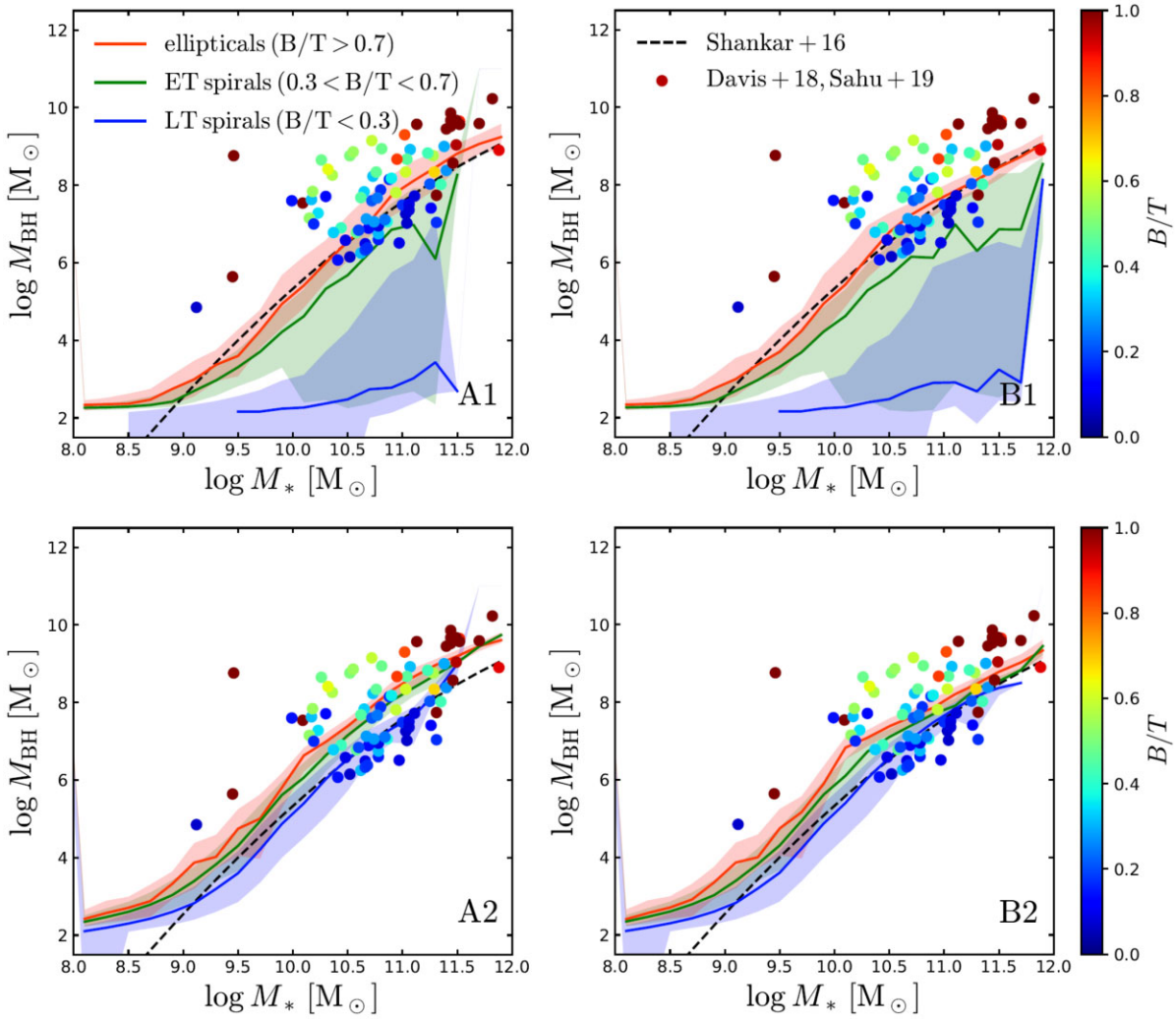

Figure 6. Median central BH mass as a function of stellar mass, as predicted by the halo-quenching models A1 and A2 (top and bottom left-hand panels) and the binding-energy quenching models B1 (top right-hand panel) and B2 (bottom right-hand panel). The shaded areas indicate the 16th and 84th percentiles of the predicted distributions. Red, green, and blue colour refer to elliptical galaxies [bulge-to-total stellar mass ratios (B/T) > 0.7], early-type spirals (0.3 <B/T < 0.7), and late-type spirals (B/T < 0.3), respectively. Observations of individual galaxies from Davis, Graham & Cameron (2018) and Sahu et al. (2019) are shown with points, colour coded according to the galaxies' B/T. The black dashed line shows the M_{BH}–M_* relation for early-type galaxies (ellipticals and lenticulars) corrected for selection biases from Shankar et al. (2016).

BH growth is dominated by Eddington-limited accretion. Models aiming at resolving this issue invoke super-Eddington accretion or the formation of massive BH seeds ($\sim 10^5$–10^6 $M_\odot$) either via direct collapse of metal-poor gas in protogalaxies (Woods et al. 2019; Inayoshi, Visbal & Haiman 2020) or during major mergers of disc-dominated galaxies (Mayer et al. 2010).

Adopting a higher $M_{BH,\,seed}$ for all galaxies in our SAM results in a dearth of galaxies with $M_* \lesssim 10^{10}$ $M_\odot$. However, if massive seed BHs dominate mostly the massive end of the mass function (as proposed by e.g. Bonoli, Mayer & Callegari 2014), then our adopted BH/binding-energy quenching model would be in even better agreement with the observed GSMF, as that would solve the problem at the high-mass end.

Galactic morphologies are important both as a test of the merger model (model 1 versus model 2) and because of the M_*–morphology relation. In Fig. 9, we compare the results of our four models to the observations of Mendel et al. (2014; circles) and Dimauro et al. (2018; stars) for the B/T ratios and the morphological breakdown of the galaxy population. In this figure, but not in the rest of the article, we classify galaxies using B/T ratios that consider only the classical bulge. The reason is that Mendel et al. (2014) fit bulges with a de Vaucouleurs (1948) model, which is a poor fit to pseudobulges, and assign $B/T = 0$ to discs hosting pseudobulges where 'no de Vaucouleurs component is a fit'. Once again, the agreement is best for merger model 2. Since merger model 1 does not allow for any bulge growth during minor mergers, models A1 and B1 predict too many late-type spirals with $B/T < 0.1$ (Fig. 9) and average B/T ratios that are too low (Fig. 8, blue curves).

In contrast to halo quenching, BH quenching also affects galaxies at intermediate masses ($10^{10} \lesssim M_* \lesssim 10^{11}$ $M_\odot$) by preventing gas accretion on to the disc component. This is why, in this mass range, the median B/T is higher in model B2 than in model A2 (Fig. 8, black curves). In Fig. 9, the fraction of galaxies with $0.1 < B/T < 0.7$ (green curve) are lower in model B2 than in model A2 at intermediate masses, while the fraction of galaxies with $B/T > 0.7$ (red curve) are larger, because BH quenching has prevented merger remnants from

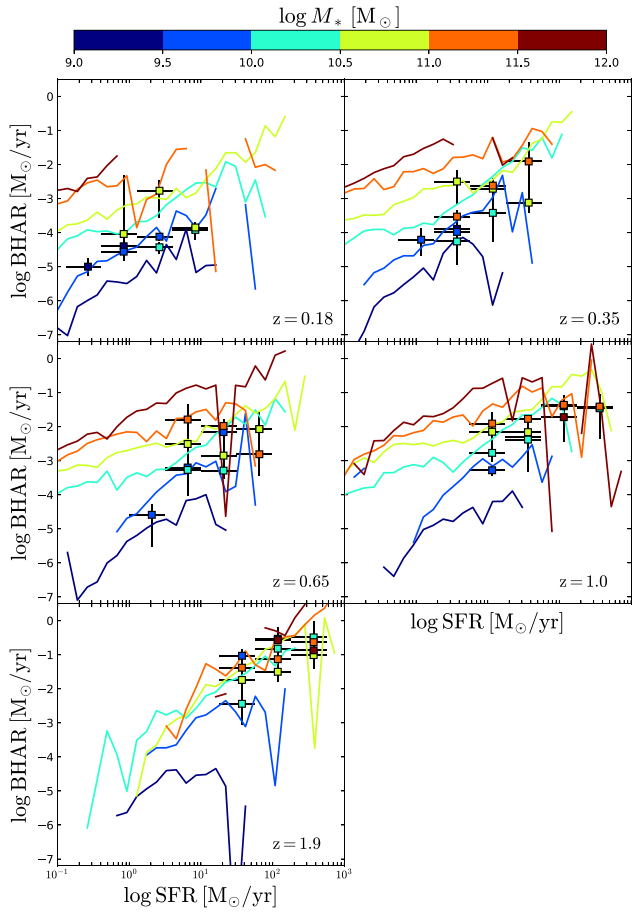

Figure 7. Mean BH accretion rate as a function of the total SFR in GALICS 2.2, model B2. The coloured curves correspond to six bins of stellar mass. The panels to five redshifts. The points with error bars are the observations from Delvecchio et al. (2015), who selected galaxies in the far-infrared and measured BH accretion rates from stacked X-ray data.

regrowing a disc. In both cases, model B2 is in better agreement with the observations than model A2. These considerations do not apply when comparing models A1 and B1 because in model B1 there are not enough passive galaxies at $M_* < 10^{11}$ M$_\odot$ for them to affect the statistical properties of the galaxy population (Fig. 5).

Fig. 10 uses the morphological distribution of galaxies on the SFR–M_* plane to compare the evolution in the halo-quenching and BH-quenching scenarios. Triangles and circles represent unquenched and quenched galaxies, respectively. Quenched galaxies are those that satisfy or have satisfied our quenching criterion, which is $M_{\rm vir} > 2 \times 10^{12}$ M$_\odot$ for models A1 and A2 and equation (12) for models B1 and B2. We note that a galaxy may be quenched but still star forming if it has not consumed all its gas yet (there are circles on the main sequence of star-forming galaxies). It is also possible that an unquenched galaxy may be passive (there are a few triangles below the main sequence of star-forming galaxies). This rare phenomenon occurs when a galaxy ceases to accrete and runs out of gas because its halo has stopped growing.

Models A1 and B1 are once again the most catastrophic, showing too few late-type spirals and a red sequence populated almost completely by galaxies with $B/T \sim 1$. Furthermore, in model B1 the main sequence of star-forming galaxies extends too far at high masses because, in merger model 1, there are too many massive spirals with BH masses that are not large enough to induce quenching.

In model A2, disc instabilities and mergers cause B/T to increase gradually towards high masses along the main sequence of star-forming galaxies until galaxies reach the stellar mass-scale at which $M_{\rm vir} = M_{\rm crit} = 2 \times 10^{12}$ M$_\odot$. Above $M_* = 10^{11}$ M$_\odot$, nearly all galaxies are quenched and most of them are passive (there is a time lag between quenching and the actual end of SF). In model B2, B/T increases not only towards higher masses on the main sequence but also towards lower SFRs at fixed M_*. The green valley is not only a region of intermediate SFRs but also a region of intermediate morphologies. Model B2 is in better agreement with the observations (e.g. Wuyts et al. 2011; Whitaker et al. 2015; Morselli et al. 2017; Dimauro et al. 2022).

There are two reasons why model A2 fails on this point and model B2 is successful. First, to have a transition from spiral to elliptical galaxies along a vertical line of fixed M_*, there has to be a spiral population in the main sequence of star-forming galaxies at that M_*. In model A2, quenching starts at $M_* \sim 10^{11}$ M$_\odot$, where the main sequence is dominated by lenticular ($B/T \sim 0.5$) rather than spiral morphologies. In model B2, galaxies with $M_* > 3 \times 10^{10}$ M$_\odot$ often contain BHs that are massive enough for the BH-quenching criterion (equation 12) to be satisfied. Hence, the green valley starts at lower masses, where there is a significant spiral population. Lowering $M_{\rm vir}^{\rm crit}$ in model A2 would result in a more extended green valley, in better agreement with the observations, but would also give too few massive galaxies at all redshifts. However, the extent of the green valley is not the only problem and here comes the second point. In model A2 there is no causal connection between quenching and morphology. The only connection is that they are both related to mass. Hence, no correlation should be expected between SFR and B/T at constant M_* beyond the one introduced by mergers, which may considerably accelerate the depletion of gas in an already quenched galaxy (this mechanism explains why, at a fixed M_*, the blue circles are statistically above the red circles in all models). In model B2, quenching is linked to BH mass. For a same M_*, galaxies with larger B/T have higher BH masses (Fig. 6). Therefore, it is natural that higher B/T should carry a higher quenching probability.

4 DISCUSSION AND CONCLUSION

We have investigated the mechanisms that quench SF in galaxies by comparing the results of two competing scenarios: (A) halo quenching and (B) BH/binding-energy quenching. In model A, quenching occurs when $M_{\rm vir}$ grows above the critical halo mass $M_{\rm vir}^{\rm crit}$, a free parameter whose order of magnitude is determined by Dekel & Birnboim (2006)'s critical mass for shock heating. In model B, quenching occurs when $M_{\rm BH}$ grows above the critical BH mass $M_{\rm BH}^{\rm crit}(v_{\rm vir})$ determined by Chen et al. (2020)'s binding-energy criterion.

Model B is more complicated than model A because BH masses, unlike halo masses, depend on the assumptions of the galaxy-evolution model, notably those for morphological evolution, since the growth of supermassive BHs is related to the formation of bulges. Some of these assumptions are degenerate: a same galaxy can have the same BH mass with a lower B/T and a higher BH-to-bulge mass ratio. Hydrodynamic simulations can assist us to make reasonable assumptions, but, even if one doubted their guidance, there are enough observational constraints to break the aforementioned degeneracy.

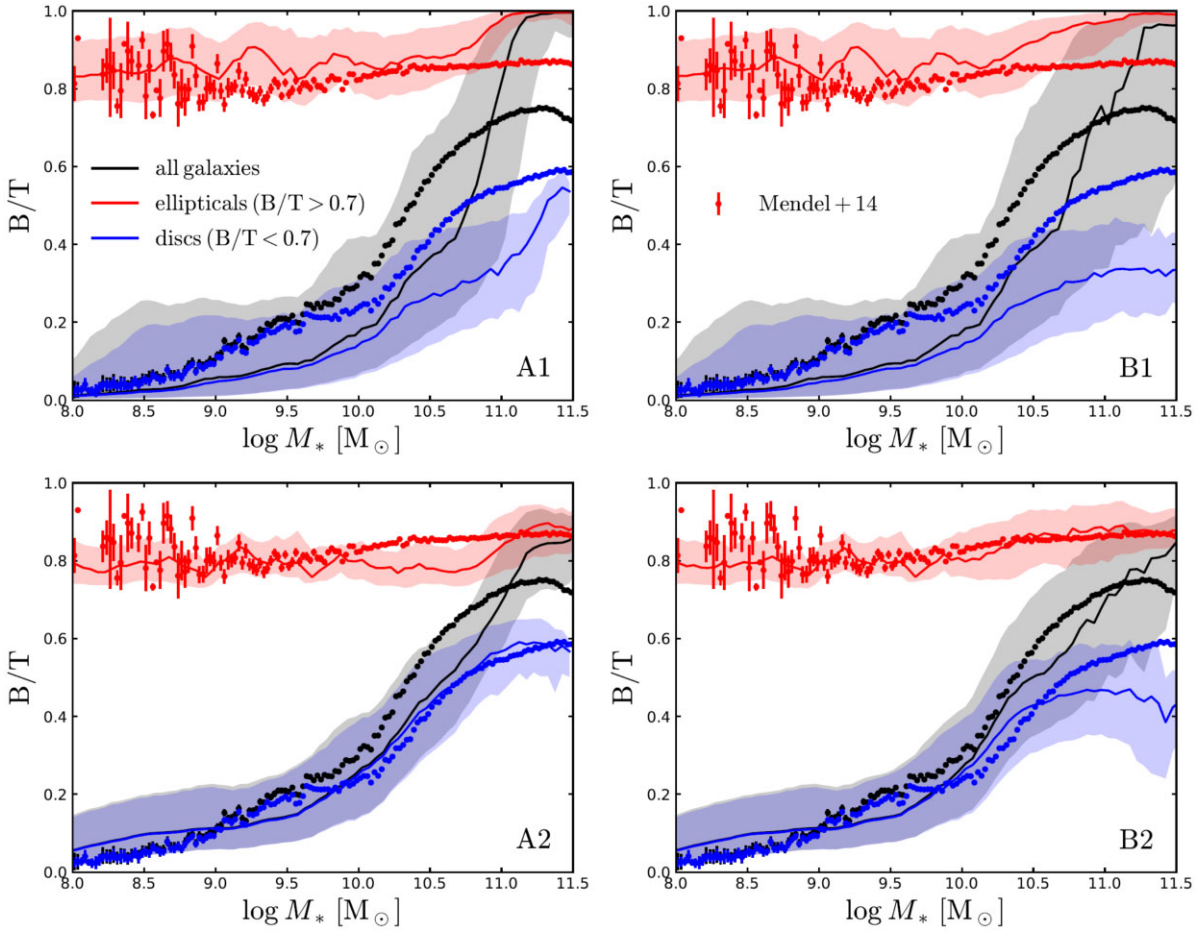

Figure 8. Median B/T as a function of galaxy stellar mass for the two halo-quenching models (left-hand panels) and the two binding-energy quenching models (right-hand panels), at z 0.1. The shaded areas indicate the 25th and 75th percentiles of the predicted distribution. Points with error bars show the observations by Mendel et al. (2014). Black, red, and blue colour refer to all galaxies in the sample, ellipticals ($B/T > 0.7$), and disc galaxies ($B/T < 0.7$), respectively.

We have considered two versions of models A and B: one in which only *major* mergers lead to the growth of BHs and bulges (models A1 and B1), the other in which the effects of major and minor mergers are modelled following the numerical results of Kannan et al. (2015, models A2 and B2). Model B2 is in better agreement with the observations than the other three. For BH accretion rates equal to ∼5 per cent of the SFR in the central starburst, model B2 reproduces the observed BH masses, GSMFs, quiescent fractions, and morphological properties of galaxies.

Three conclusions follow from these findings.

(i) In agreement with Terrazas et al. (2016) and Bluck et al. (2020), quenching is more closely related to the growth of $M_{\rm BH}$ than it is to that of $M_{\rm vir}$. Halo quenching fails to produce enough passive galaxies at $M_* < 10^{11.5}\,{\rm M}_\odot$ (Fig. 5, panels A1 and A2). Lowering $M_{\rm vir}^{\rm crit}$ alleviates the problem but destroys the agreement with the GSMF at high masses. BH-quenching model B2 does not present this problem because, even at $M_* < 10^{11.5}\,{\rm M}_\odot$, there are some galaxies with $M_{\rm BH} > M_{\rm BH}^{\rm crit}$. Halo quenching also fails to reproduce the SFR–morphology relation. With halo quenching, the transition to higher B/T occurs along the main sequence of star-forming galaxies; the green valley is composed of early-type galaxies. With BH quenching, the transition on the SFR–M_* diagram is much more vertical; at constant M_*, B/T increases in the direction of lower SFRs.

(ii) Morphological evolution leads to quenching because mergers transform discs into bulges while feeding the growth of supermassive BHs. Restricting the formation of bulges and supermassive BHs to major mergers is incompatible with the fraction of passive galaxies at high masses. There are always galaxies that reach high masses through a series of minor mergers. If these galaxies are not quenched, there are too many star-forming galaxies at high masses.

(iii) If the gas fraction fed to the central BHs in mergers is the same for massive and low-mass systems, then we have too many passive galaxies at low masses. The efficiency with which gas is fed to the central BH must be lower at lower $v_{\rm vir}$. This finding is consistent with the proposal that supernova feedback prevents the efficient growth of supermassive BHs in systems with low escape speeds (Dubois et al. 2015; Anglés-Alcázar et al. 2017; Bower et al. 2017; Habouzit et al. 2017; Davies et al. 2019b; Dekel et al. 2019; Oppenheimer et al. 2020).

We end with a couple of comments on how these findings fit into our broader understanding of BH feedback. The first is that our BH-quenching criterion (model B) is a criterion for $M_{\rm BH}$, while previous SAMs (Croton et al. 2006; Somerville et al. 2008; Henriques et al. 2015) had assumed that $\dot{M}_{\rm BH}$ is the relevant quantity to determine whether gas accretion is shut down or not. This assumption is largely based on the insight from cool-core clusters. In these systems, the intracluster medium has a short cooling time but is not able to cool

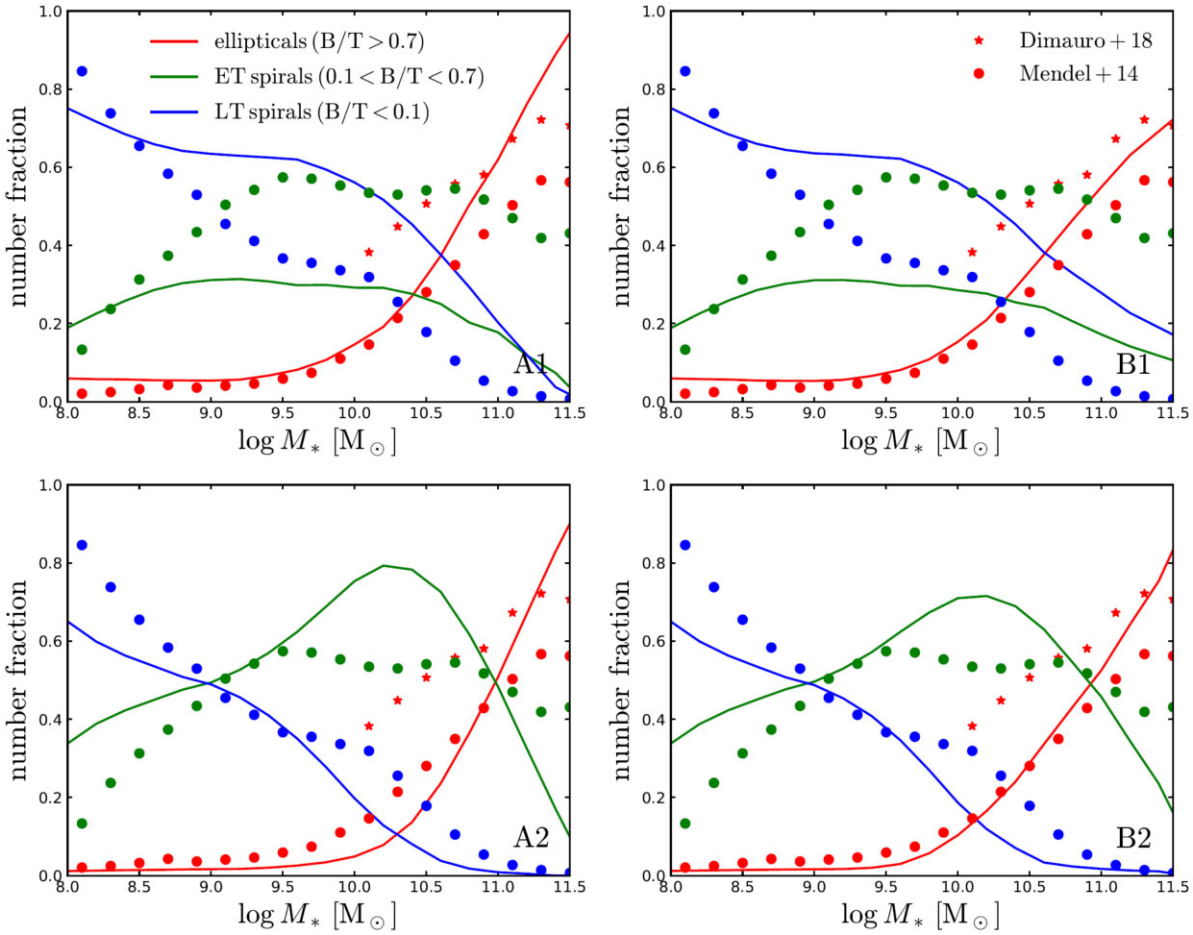

Figure 9. Fraction of elliptical galaxies ($B/T > 0.7$; red lines), earlier-type spirals ($0, 1 < B/T < 0.7$; green lines), and late-type spirals ($B/T < 0.1$; blue lines) as a function of stellar mass, for the two halo-quenching models (left-hand panels) and the two binding-energy quenching models (right-hand panels), at $z \sim 0.1$. The coloured points show the observations by Mendel et al. (2014; circles) and Dimauro et al. (2018; stars).

and flow efficiently on to the central galaxy because mechanical heating by radio jets from AGN compensates the radiative losses to X-rays (Peterson & Fabian 2006; McNamara & Nulsen 2007, 2012; Cattaneo et al. 2009; Hlavacek-Larrondo, Li & Churazov 2022). The requirement that the jet power should compensate the X-ray luminosity L_X translates into a criterion for the BH accretion rate required to shut down cooling: $\epsilon_{jet}\dot{M}_{BH}c^2 \sim L_X$, where ϵ_{jet} is the efficiency with which the rest-mass energy of the gas that accretes on to the BH is converted into jet kinetic energy.

As a good theory should explain as many observations as possible with the fewest possible assumptions, it is logical that SAMs should try and extend the lesson from cool-core clusters to lower masses. In this case, however, Occam's razor does not cut it. Chen et al. (2020) and we have shown that we need a different criterion to explain the quenching of SF in galaxies. We have not modelled the processes through which BHs quench SF. Hence, we cannot comment on the physics of BH feedback. We remark, however, that the criterion required for the initial quenching of SF is different from the one required to maintain it shut down afterwards. We can, therefore, make the argument that quenching and maintenance correspond to different feedback regimes. In fact, a criterion for M_{BH} arises naturally in models where BHs accrete at the Eddington limit, so that the accretion power grows with M_{BH} until it is large enough to blow away all the gas in the host system (Silk & Rees 1998; Fabian 1999; King 2003).

The second comment is on the entropy excesses on group scales (the X-ray-emitting gas has higher entropies in groups than in clusters when all entropies are rescaled to the virial one; Edge & Stewart 1991; Evrard & Henry 1991; Kaiser 1991; Ponman et al. 1996; Valageas & Silk 1999; Lloyd-Davies, Ponman & Cannon 2000). Already Evrard & Henry (1991) made an explicit argument that these excesses have no reason to be present if shock heating from gravitational collapse is the dominant mechanism driving the evolution of the hot gas, while quasar feedback provides a possible explanation for their existence. Hence, their observation disfavours model A, in which gravitational shock heating quenches SF. It also disfavours any model in which the only BH feedback is maintenance feedback: if BHs provided just enough heat to maintain the X-ray-emitting gas in thermal equilibrium, then BHs would prevent the hot gas from dropping below the entropy at which it was shock heated but would not raise its entropy above it either. The entropy excesses on group scales are evidence of a past epoch during which the gas was heated violently. Violent heating implies rapid BH growth. Hence, it is reasonable to associate it with the quasar-mode feedback that quenched SF.

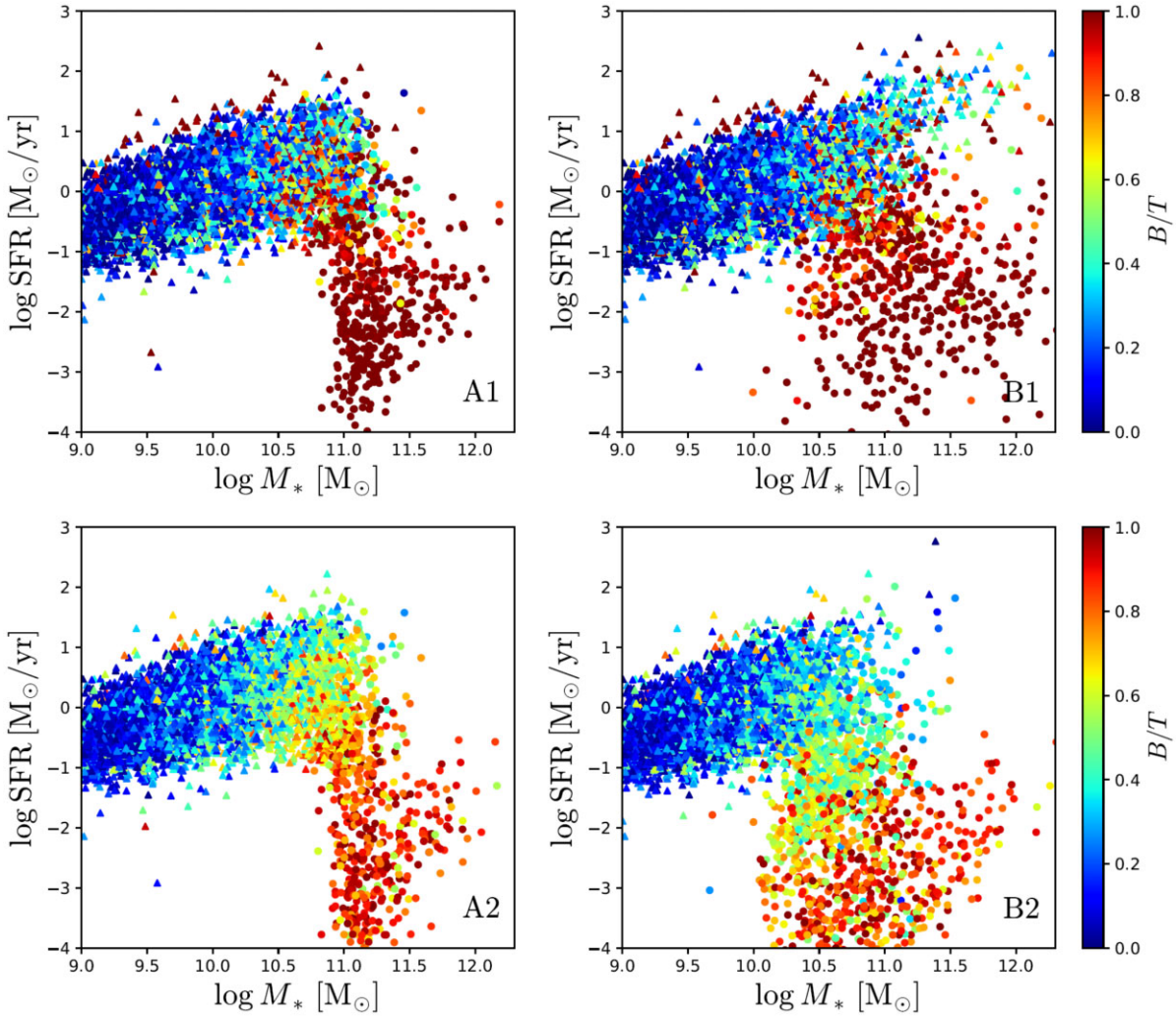

Figure 10. Star formation rate (SFR) as a function of stellar mass for individual central galaxies in the four models considered, colour coded according to the galaxies' B/T at $z \sim 0.1$. Circles denote galaxies that have already been quenched, i.e. have satisfied the $M_{\rm vir} > 2 \times 10^{12}$ M$_\odot$ criterion for models A1 and A2 and the binding-energy criterion (equation 12) for models B1 and B2. Triangles denote galaxies that have not been quenched yet.

In conclusion, our findings fit naturally in a scenario in which BH feedback is bimodal. In star-forming galaxies, BHs grow rapidly (quasar mode) until they are massive enough to blow away all the gas (most early-type galaxies with $10^9 \lesssim M_* \lesssim 10^{11}$ M$_\odot$ lack extended X-ray emission; Hou et al. 2021) or heat it to high entropy. In this second regime, the accretion is self-regulated. BHs grow just enough to maintain the hot gas in thermal equilibrium.

ACKNOWLEDGEMENTS

The authors acknowledge interesting conversation with Paola Dimauro and Gary Mamon. IK acknowledges support from the ERC Starting Grant NEFERTITI H2020/808240.

DATA AVAILABILITY

The data underlying this paper will be shared on reasonable request to the corresponding author.

REFERENCES

Abadi M. G., Moore B., Bower R. G., 1999, MNRAS, 308, 947
Anglés-Alcázar D., Faucher-Giguère C.-A., Quataert E., Hopkins P. F., Feldmann R., Torrey P., Wetzel A., Kereš D., 2017, MNRAS, 472, L109
Athanassoula E., 2008, MNRAS, 390, L69
Baldry I. K. et al., 2012, MNRAS, 421, 621
Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, MNRAS, 373, 469
Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681
Balogh M. L., Navarro J. F., Morris S. L., 2000, ApJ, 540, 113
Bañados E. et al., 2018, Nature, 553, 473
Barnes J. E., 1988, ApJ, 331, 699
Barnes J. E., Hernquist L., 1996, ApJ, 471, 115
Barro G. et al., 2017, ApJ, 840, 47
Bekki K., 2009, MNRAS, 399, 2221
Bekki K., Couch W. J., Shioya Y., 2002, ApJ, 577, 651
Benson A. J., 2012, New Astron., 17, 175
Berentzen I., Shlosman I., Martinez-Valpuesta I., Heller C. H., 2007, ApJ, 666, 189

Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697
Bigiel F. et al., 2011, ApJ, 730, L13
Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846
Bluck A. F. L., Maiolino R., Sánchez S. F., Ellison S. L., Thorp M. D., Piotrowska J. M., Teimoorinia H., Bundy K. A., 2020, MNRAS, 492, 96
Bonoli S., Mayer L., Callegari S., 2014, MNRAS, 437, 1576
Booth C. M., Schaye J., 2010, MNRAS, 405, L1
Booth C. M., Schaye J., 2011, MNRAS, 413, 1158
Bournaud F., Combes F., 2002, A&A, 392, 83
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Bower R. G., Schaye J., Frenk C. S., Theuns T., Schaller M., Crain R. A., McAlpine S., 2017, MNRAS, 465, 32
Brook C. B., Stinson G., Gibson B. K., Wadsley J., Quinn T., 2012, MNRAS, 424, 1275
Byrd G., Valtonen M., 1990, ApJ, 350, 89
Cattaneo A. et al., 2007, MNRAS, 377, 63
Cattaneo A. et al., 2009, Nature, 460, 213
Cattaneo A. et al., 2017, MNRAS, 471, 1401
Cattaneo A., Best P. N., 2009, MNRAS, 395, 518
Cattaneo A., Blaizot J., Devriendt J., Guiderdoni B., 2005, MNRAS, 364, 407
Cattaneo A., Dekel A., Devriendt J., Guiderdoni B., Blaizot J., 2006, MNRAS, 370, 1651
Cattaneo A., Koutsouridou I., Tollet E., Devriendt J., Dubois Y., 2020, MNRAS, 497, 279
Cattaneo A., Teyssier R., 2007, MNRAS, 376, 1547
Cattaneo A., Woo J., Dekel A., Faber S. M., 2013, MNRAS, 430, 686
Cavagnolo K. W., Donahue M., Voit G. M., Sun M., 2009, ApJS, 182, 12
Chabrier G., 2003, PASP, 115, 763
Chen Z. et al., 2020, ApJ, 897, 102
Cheung E. et al., 2012, ApJ, 760, 131
Christensen C. R., Davé R., Governato F., Pontzen A., Brooks A., Munshi F., Quinn T., Wadsley J., 2016, ApJ, 824, 57
Combes F., García-Burillo S., Braine J., Schinnerer E., Walter F., Colina L., 2013, A&A, 550, A41
Cora S. A. et al., 2018, MNRAS, 479, 2
Croton D. J. et al., 2006, MNRAS, 365, 11
Croton D. J. et al., 2016, ApJS, 222, 22
Davies J. J., Crain R. A., McCarthy I. G., Oppenheimer B. D., Schaye J., Schaller M., McAlpine S., 2019b, MNRAS, 485, 3783
Davies J. J., Crain R. A., Oppenheimer B. D., Schaye J., 2020, MNRAS, 491, 4462
Davies L. J. M. et al., 2019a, MNRAS, 483, 5444
Davis B. L., Graham A. W., Cameron E., 2018, ApJ, 869, 113
de Vaucouleurs G., 1948, Annales d'Astrophysique, 11, 247
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dekel A., Lapiner S., Dubois Y., 2019, preprint (arXiv:1904.08431)
Delvecchio I. et al., 2015, MNRAS, 449, 373
Devergne T. et al., 2020, A&A, 644, A56
Dimauro P. et al., 2018, MNRAS, 478, 5410
Dimauro P. et al., 2022, MNRAS, 513, 256
Dubois Y., Volonteri M., Silk J., Devriendt J., Slyz A., Teyssier R., 2015, MNRAS, 452, 1502
Duc P.-A., Bournaud F., 2008, ApJ, 673, 787
Ebeling H., Stephenson L. N., Edge A. C., 2014, ApJ, 781, L40
Edge A. C., Stewart G. C., 1991, MNRAS, 252, 414
Efstathiou G., 1992, MNRAS, 256, 43P
Efstathiou G., Lake G., Negroponte J., 1982, MNRAS, 199, 1069
Evrard A. E., Henry J. P., 1991, ApJ, 383, 95
Fabian A. C., 1999, MNRAS, 308, L39
Faltenbacher A., Hoffman Y., Gottlöber S., Yepes G., 2007, MNRAS, 376, 1327
Fan X., 2006, New Astron. Rev., 50, 665
Fanaroff B. L., Riley J. M., 1974, MNRAS, 167, 31P
Fang J. J., Faber S. M., Koo D. C., Dekel A., 2013, ApJ, 776, 63
Farouki R., Shapiro S. L., 1981, ApJ, 243, 32
Ferrarese L., Merritt D., 2000, ApJ, 539, L9
Font A. S. et al., 2008, MNRAS, 389, 1619
Gargiulo I. D. et al., 2015, MNRAS, 446, 3820
Gebhardt K. et al., 2000, ApJ, 539, L13
Gnedin N. Y., 2000, ApJ, 542, 535
Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, MNRAS, 439, 264
Gunn J. E., Gott J. R., III, 1972, ApJ, 176, 1
Guo Q. et al., 2011, MNRAS, 413, 101
Guo Q., Cole S., Eke V., Frenk C., Helly J., 2013, MNRAS, 434, 1838
Habouzit M., Volonteri M., Dubois Y., 2017, MNRAS, 468, 3935
Haemmerlé L., Mayer L., Klessen R. S., Hosokawa T., Madau P., Bromm V., 2020, Space Sci. Rev., 216, 48
Harris K. et al., 2016, MNRAS, 457, 4179
Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Hlavacek-Larrondo J., Li Y., Churazov E., 2022, preprint (arXiv:2206.00098)
Hopkins P. F., Cox T. J., Younger J. D., Hernquist L., 2009, ApJ, 691, 1168
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, ApJS, 163, 1
Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008a, ApJS, 175, 356
Hopkins P. F., Hernquist L., Cox T. J., Younger J. D., Besla G., 2008b, ApJ, 688, 757
Hou M., Li Z., Jones C., Forman W., Su Y., 2021, ApJ, 919, 141
Ilbert O. et al., 2013, A&A, 556, A55
Inayoshi K., Visbal E., Haiman Z., 2020, ARA&A, 58, 27
Jiang F. et al., 2019, MNRAS, 488, 4801
Kaiser N., 1991, ApJ, 383, 104
Kang X., van den Bosch F. C., 2008, ApJ, 676, L101
Kannan R., Macciò A. V., Fontanot F., Moster B. P., Karman W., Somerville R. S., 2015, MNRAS, 452, 4347
Kauffmann G. et al., 2003, MNRAS, 346, 1055
Kauffmann G., Haehnelt M., 2000, MNRAS, 311, 576
Kennicutt R. C. J., 1998, ApJ, 498, 541
Kennicutt R. C. J., De Los Reyes M. A. C., 2021, ApJ, 908, 61
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Kimm T. et al., 2009, MNRAS, 394, 1131
King A., 2003, ApJ, 596, L27
Knebe A. et al., 2015, MNRAS, 451, 4029
Kormendy J., Fisher D. B., Cornell M. E., Bender R., 2009, ApJS, 182, 216
Koutsouridou I., Cattaneo A., 2019, MNRAS, 490, 5375
Krumholz M. R., Dekel A., McKee C. F., 2012, ApJ, 745, 69
Lagos C. D. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, MNRAS, 481, 3573
Lamastra A., Menci N., Fiore F., Santini P., 2013, A&A, 552, A44
Larson R. B., Tinsley B. M., Caldwell C. N., 1980, ApJ, 237, 692
Latif M. A., Ferrara A., 2016, Publ. Astron. Soc. Aust., 33, e051
Lee J., Yi S. K., 2013, ApJ, 766, 38
Lloyd-Davies E. J., Ponman T. J., Cannon D. B., 2000, MNRAS, 315, 689
Lo Faro B., Monaco P., Vanzella E., Fontanot F., Silva L., Cristiani S., 2009, MNRAS, 399, 827
Mamon G. A., Trevisan M., Thuan T. X., Gallazzi A., Davé R., 2020, MNRAS, 492, 1791
Martin G., Kaviraj S., Devriendt J. E. G., Dubois Y., Pichon C., 2018, MNRAS, 480, 2266
Mayer L., Kazantzidis S., Escala A., Callegari S., 2010, Nature, 466, 1082
Mayer L., Mastropietro C., Wadsley J., Stadel J., Moore B., 2006, MNRAS, 369, 1021
McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, MNRAS, 383, 593
McNamara B. R., Nulsen P. E. J., 2007, ARA&A, 45, 117
McNamara B. R., Nulsen P. E. J., 2012, New J. Phys., 14, 055023

Mendel J. T., Simard L., Palmer M., Ellison S. L., Patton D. R., 2014, ApJS, 210, 3
Merritt D., 1984, ApJ, 276, 26
Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319
Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, Nature, 379, 613
Morselli L., Popesso P., Erfanianfar G., Concas A., 2017, A&A, 597, A97
Moustakas J. et al., 2013, ApJ, 767, 50
Muzzin A. et al., 2013, ApJ, 777, 18
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Negroponte J., White S. D. M., 1983, MNRAS, 205, 1009
Oppenheimer B. D. et al., 2020, MNRAS, 491, 2939
Oppenheimer B. D., 2018, MNRAS, 480, 2963
Ostriker J. P., Bode P., Babul A., 2005, ApJ, 634, 964
Pan H.-A. et al., 2019, ApJ, 881, 119
Peterson J. R., Fabian A. C., 2006, Phys. Rep., 427, 1
Planck Collaboration XX, 2014, A&A, 571, A20
Ponman T. J., Bourner P. D. J., Ebeling H., Böhringer H., 1996, MNRAS, 283, 690
Porter L. A., Somerville R. S., Primack J. R., Johansson P. H., 2014, MNRAS, 444, 942
Powell L. C., Bournaud F., Chapon D., Teyssier R., 2013, MNRAS, 434, 1028
Prieto M. A., Fernandez-Ontiveros J. A., Bruzual G., Burkert A., Schartmann M., Charlot S., 2019, MNRAS, 485, 3264
Quinn P. J., Hernquist L., Fullagar D. P., 1993, ApJ, 403, 74
Rodighiero G. et al., 2015, ApJ, 800, L10
Roediger E., Brüggen M., 2007, MNRAS, 380, 1399
Saglia R. P. et al., 2016, ApJ, 818, 47
Sahu N., Graham A. W., Davis B. L., 2019, ApJ, 876, 155
Sahu N., Graham A. W., Davis B. L., 2022, ApJ, 927, 67
Salpeter E. E., 1964, ApJ, 140, 796
Sandage A., 1986, A&A, 161, 89
Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z., 1988, ApJ, 325, 74
Shankar F. et al., 2016, MNRAS, 460, 3119
Silk J., Rees M. J., 1998, A&A, 331, L1
Simpson C. M., Grand R. J. J., Gómez F. A., Marinacci F., Pakmor R., Springel V., Campbell D. J. R., Frenk C. S., 2018, MNRAS, 478, 548
Somerville R. S., Davé R., 2015, ARA&A, 53, 51
Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481
Somerville R. S., Popping G., Trager S. C., 2015, MNRAS, 453, 4337
Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776
Springel V., Hernquist L., 2005, ApJ, 622, L9
Statler T. S., 1988, ApJ, 331, 71
Steinhauser D., Schindler S., Springel V., 2016, A&A, 591, A51
Terrazas B. A., Bell E. F., Henriques B. M. B., White S. D. M., Cattaneo A., Woo J., 2016, ApJ, 830, L12
Tollet É., Cattaneo A., Macciò A. V., Dutton A. A., Kang X., 2019, MNRAS, 485, 2511
Tollet E., Cattaneo A., Macciò A. V., Kang X., 2022, MNRAS, 515, 3453
Tollet É., Cattaneo A., Mamon G. A., Moutard T., van den Bosch F. C., 2017, MNRAS, 471, 4170
Tomczak A. R. et al., 2014, ApJ, 783, 85
Tonnesen S., Bryan G. L., 2009, ApJ, 694, 789
Toomre A., Toomre J., 1972, ApJ, 178, 623
Valageas P., Silk J., 1999, A&A, 350, 725
van den Bosch F. C., 1998, ApJ, 507, 601
Voit G. M., 2019, ApJ, 880, 139
Voit M. et al., 2019, BAAS, 51, 405
Wang L., Dutton A. A., Stinson G. S., Macciò A. V., Gutcke T., Kang X., 2017, MNRAS, 466, 4858
Wang Z. et al., 2004, ApJS, 154, 193
Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., Croton D. J., Moore B., 2006, MNRAS, 372, 1161
Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, MNRAS, 432, 336
Whitaker K. E. et al., 2015, ApJ, 811, L12
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
Wisotzki L., Kuhlbrodt B., Jahnke K., 2001, in Márquez I., Masegosa J., del Olmo A., Lara L., García E., Molina J., eds, QSO Hosts and Their Environments. Kluwer, Dordrecht, p. 83
Woods T. E. et al., 2019, Publ. Astron. Soc. Aust., 36, e027
Wu X.-B. et al., 2015, Nature, 518, 512
Wuyts S. et al., 2011, ApJ, 742, 96
Yang X., Mo H. J., van den Bosch F. C., 2009, ApJ, 693, 830

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