

The origin of the atomic and molecular gas contents of early-type galaxies – I. A new test of galaxy formation physics

Claudia del P. Lagos,¹★ Timothy A. Davis,¹ Cedric G. Lacey,² Martin A. Zwaan,¹ Carlton M. Baugh,² Violeta Gonzalez-Perez¹ and Nelson D. Padilla³

¹European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching, Germany

²Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

³Departamento Astronomía y Astrofísica, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Stgo., Chile

Accepted 2014 June 16. Received 2014 June 9; in original form 2014 April 29

ABSTRACT

We study the atomic (H I) and molecular hydrogen (H₂) contents of early-type galaxies (ETGs) and their gas sources using the GALFORM model of galaxy formation. This model uses a self-consistent calculation of the star formation rate, which depends on the H₂ content of galaxies. We first present a new analysis of H I Parkes All-Sky Survey and ATLAS^{3D} surveys, with special emphasis on ETGs. The model predicts H I and H₂ contents of ETGs in agreement with the observations from these surveys only if partial ram pressure stripping of the hot gas is included, showing that observations of neutral gas in ‘quenched’ galaxies place stringent constraints on the treatment of the hot gas in satellites. We find that ≈ 90 per cent of ETGs at $z = 0$ have neutral gas contents supplied by radiative cooling from their hot haloes, 8 per cent were supplied by gas accretion from minor mergers that took place in the last 1 Gyr, while 2 per cent were supplied by mass-loss from old stars. The model predicts neutral gas fractions strongly decreasing with increasing bulge fraction. This is due to the impeded disc regeneration in ETGs, resulting from both active galactic nuclei feedback and environmental quenching by partial ram pressure stripping of the hot gas.

Key words: stars: formation – galaxies: evolution – galaxies: formation – galaxies: ISM.

1 INTRODUCTION

The classic picture of early-type galaxies (ETGs), which include elliptical and lenticular galaxies, is that they are ‘red and dead’, without any significant star formation (SF), and contain mainly old stellar populations (Bower, Lucey & Ellis 1992). ETGs have also been long connected to the red sequence in the colour–magnitude relation, establishing a strong connection between quenching of SF and morphological transformation (e.g. Bower et al. 1992; Strateva et al. 2001; Baldry et al. 2004; Balogh et al. 2004; Bernardi et al. 2005; Schiminovich et al. 2007). By analysing the Sérsic index of galaxies in the star formation rate (SFR)–stellar mass plane, Wuyts et al. (2011) showed that galaxies in the star-forming sequence (the so called ‘main sequence’ of galaxies) are typically disc-like galaxies (with Sérsic indices close to 1), while passive galaxies tend to have higher Sérsic indices (typically > 3). Wuyts et al. also showed that these trends are observed in galaxies from $z = 0$ to ≈ 2 , suggesting that the relation between morphology and quenching is fundamental and that is present over most of the SF history of galaxies.

Although this simple paradigm of ‘red and dead’ ETGs is qualitatively sufficient to explain their location on the red sequence of galaxies, it is far from being quantitatively correct. High quality, resolved observations of the different components of ETGs, mainly from the ATLAS^{3D} multiwavelength survey (Cappellari et al. 2011), showed that this paradigm is too simplistic. This survey showed that at least 20 per cent of ETGs have molecular and atomic hydrogen contents large enough to be detected (Young et al. 2011; Serra et al. 2012; see also Welch, Sage & Young 2010 for similar results from an independent survey). The approximate detection limits for molecular hydrogen (H₂) and atomic hydrogen (H I) masses in the ATLAS^{3D} are $\approx 10^7$ – 10^8 M_⊙. Large amounts of cold gas in ETGs are frequently found when SF is observed (e.g. Davis et al. 2014). Some of these galaxies with ongoing SF lie on the red sequence of galaxies in the colour–magnitude relation (Kaviraj et al. 2007; Smith et al. 2012; Young et al. 2013). All of this evidence points to a large fraction of ETGs, which were before seen as ‘passive’ in terms of their colours, having SFRs and cold interstellar medium (ISM) contents that can be rather large. From this it is reasonable to conclude that the quenching of galaxies is

★ E-mail: clagos@eso.org

¹ <http://www-astro.physics.ox.ac.uk/atlas3d/>

indeed more complex than the simple picture of ‘passive, red and dead’ ETGs.

This shift of paradigm in ETGs poses new questions regarding how we understand the formation of this galaxy population. For instance, how do we understand the presence of a non-negligible cold ISM in ETGs and their location on the red sequence of galaxies in the colour–magnitude diagram? Was the cold gas accreted recently or does it come from internal processes, such as recycling of old stars? These questions are at the core of the understanding of the quenching of galaxies and the decline of the SF activity with time.

Simulations of galaxy formation have long explored the origin of galaxy morphologies in the context of the hierarchical growth of structures. Pioneering ideas about the formation of galaxy discs and bulges were presented by Toomre (1977) and White & Rees (1978). Toomre proposed for the first time that galaxy mergers could lead to the formation of spheroids, which was implemented in early semi-analytic models of galaxy formation (e.g. Baugh, Cole & Frenk 1996; Kauffmann 1996; Cole et al. 2000). However, with the advent of large area surveys, and more sophisticated cosmological N -body simulations, it became clear that major mergers (mergers between galaxies with mass ratios $\gtrsim 0.3$) could not be the only formation mechanism of spheroids (e.g. see Le Fèvre et al. 2000 for an observational example and Naab & Burkert 2003 for a theoretical work) because of their expected rareness, which is incompatible with the large numbers of ETGs observed (Bernardi et al. 2003; Lintott et al. 2008). Theoretical work on the formation mechanisms of spheroids led to the conclusion that minor mergers (e.g. Malbon et al. 2007; Parry, Eke & Frenk 2009; Hopkins et al. 2010; Bournaud et al. 2011; Naab et al. 2013) and disc instabilities (e.g. Mo, Mao & White 1998; Gammie 2001; Bournaud, Elmegreen & Martig 2009; Elmegreen & Burkert 2010; Krumholz & Burkert 2010) can also play a major role. Many studies exploited numerical simulations and semi-analytic models to study the formation of ETGs and their mass assembly with interesting predictions, for example that massive ETGs assemble their mass relatively late but have stellar populations that are very old (e.g. Baugh et al. 1996; Kauffmann 1996; De Lucia et al. 2006; Parry et al. 2009), and that the formation paths for ETGs can be many, going from having had one or more major mergers, to having had no mergers at all (e.g. Naab et al. 2013).

Despite all this progress, little attention has been paid to the study of the neutral gas content of the ETG population. The atomic and molecular gas contents of ETGs may provide strong constraints on the recent accretion history. Lagos et al. (2011a), for example, show that the neutral gas content of galaxies is very sensitive to short term variations in the accretion history, while the stellar mass and optical colours are not. Similarly, Serra et al. (2014) show that although simulations can reproduce the nature of slow and fast rotators in the early-type population, the atomic hydrogen content predicted by the same simulations is too low. Another reason to believe that the neutral gas content of ETGs will provide strong constraints on galaxy formation models, is that they show different correlations between their gas and stellar contents. For example, for normal star-forming galaxies, there is a good correlation between the $H\text{I}$ mass and the stellar mass (e.g. Catinella et al. 2010; Cortese et al. 2011; Huang et al. 2012; Wang et al. 2014), while ETGs show no correlation between these two quantities (e.g. Welch et al. 2010; Serra et al. 2012). Different physical mechanisms are then driving the $H\text{I}$ content of ETGs. Similar conclusions were reached for molecular hydrogen (e.g. Lisenfeld et al. 2011; Saintonge et al. 2011; Young et al. 2011; Boselli et al. 2014b).

The motivation behind this paper is to investigate the neutral gas content of ETGs in hierarchical galaxy formation models and relate

them to the formation and quenching mechanisms of ETGs. We explore the question of the origin of the atomic and molecular gas contents of ETGs and attempt to connect this to their observed $H\text{I}$ and H_2 contents and their stellar mass content. In Lagos et al. (in preparation (Paper II)), we will explore the question of the alignments between the angular momenta of the gas disc and the stellar contents of ETGs. For the current study, we use three flavours of the semi-analytical model *GALFORM* in a Λ cold dark matter (Λ CDM) cosmology (Cole et al. 2000), namely those of Lagos et al. (2012), Gonzalez-Perez et al. (2014), and Lacey et al. (in preparation). The three models include the improved treatment of SF implemented by Lagos et al. (2011a). This extension splits the hydrogen content of the ISM into $H\text{I}$ and H_2 . In addition, these three models allow bulges to grow through minor and major galaxy mergers and through global disc instabilities. The advantage of using three different flavours of *GALFORM* is the ability to characterize the robustness of the trends found. The outputs of the three models shown in this paper will be made publicly available through the Millennium data base.²

This paper is organized as follows. In Section 2, we present the observations of the $H\text{I}$ and H_2 content of the entire galaxy population and of ETGs, along with the gas mass functions and gas fraction distribution functions for these two populations. In Section 3, we describe the galaxy formation model and the main aspects which relate to the growth of bulges: SF, disc and bulge build-up, recycling of intermediate- and low-mass stars and the treatment of the partial ram pressure stripping of the hot gas. We also describe the main differences between the three *GALFORM* flavours and the dark matter (DM) simulations used. In Section 4, we compare the model predictions with observations of the neutral gas content of ETGs and show the impact of including partial ram pressure stripping of the hot gas. In Section 5, we analyse the connection between the neutral gas content of ETGs, their bulge fraction and quenching and explain the physical processes behind this connection. In Section 6, we analyse all the sources that contribute to the neutral gas content of ETGs and their environmental dependence. Finally, our main conclusions are presented in Section 7.

2 THE OBSERVED NEUTRAL HYDROGEN CONTENT OF LOCAL GALAXIES

Our aim is to study the neutral gas content of ETGs and show how this compares to the overall galaxy population. We first need to define the observational data sets that we will use as the main constraints on the galaxy formation simulations.³ We focus on the $H\text{I}$ and H_2 gas contents of galaxies and how these compare to the stellar content. We do this through comparisons with the K -band luminosity, which is closely related to the stellar mass in galaxies but directly measured by observations. We extensively use the $H\text{I}$ Parkes All-Sky Survey (HIPASS; Meyer et al. 2004), the Arecibo Legacy Fast ALFA Survey (ALFALFA; Giovanelli et al. 2005), the Five College Radio Astronomy Observatory CO(1–0) survey of Keres, Yun & Young (2003) and the ATLAS^{3D} survey (Cappellari et al. 2011).

The ATLAS^{3D} survey is of particular interest as it is a volume limited sample of ETGs, where all ETGs within a volume of $1.16 \times 10^5 \text{ Mpc}^3$ and with K -band rest-frame luminosities of $L_K > 6 \times 10^9 L_\odot$ were studied in detail. The total number of ETGs in the

² <http://gavo.mpa-garching.mpg.de/Millennium>

³ The units in the observations are expressed in physical units. Note that we normalized all the observational data sets to the choice of $h = 0.73$.

ATLAS^{3D} catalogue is 260. The morphological classification was obtained through careful visual inspection. The ATLAS^{3D} survey contains multiwavelength information, such as broad-band photometry in the *B*, *r* and *K* bands (Cappellari et al. 2011), as well as 21 cm interferometry (presented in Serra et al. 2012), from which the H I mass is derived, and CO(1–0) single dish observations, presented in Young et al. (2011), from which the H₂ mass is derived, along with detailed stellar kinematic information. This survey provides an in depth view of ETGs in the local Universe and shows how the neutral gas content correlates with other galaxy properties.

We constructed H I and H₂ mass functions for the ATLAS^{3D} objects. The H I survey presented in Serra et al. (2012) includes every ATLAS^{3D} ETG visible with the Westerbork Synthesis Radio Telescope, a total of 166 objects. We took H I masses from Serra et al. (2012) to construct a H I mass function, which we plot in Fig. 1. This H I survey is incomplete in depth due to observations being performed using a fixed 12 h integration. This means that H I

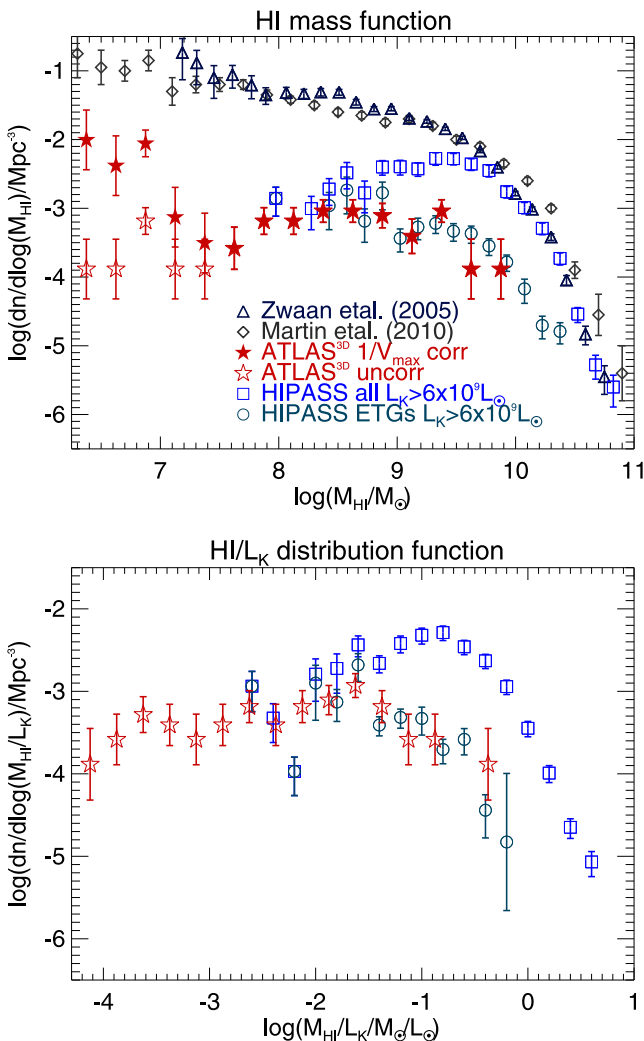


Figure 1. Top panel: the H I mass function from Zwaan et al. (2005) and Martin et al. (2010) at $z = 0$, which include both early- and late-type galaxies; the H I mass function of ETGs from HIPASS and from the ATLAS^{3D} survey (with and without $1/V_{\max}$ correction), calculated in this work. Bottom panel: distribution function of the ratio between the H I mass to the *K*-band luminosity for all and ETGs for the HIPASS and the ATLAS^{3D} surveys, calculated in this work. Note that we express densities and masses in physical units.

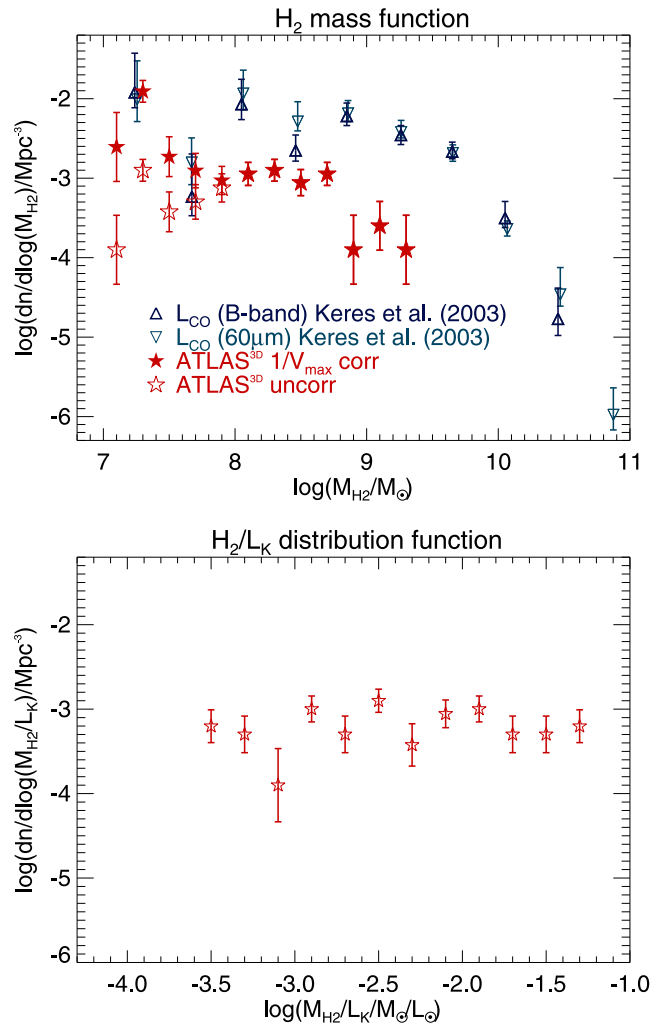


Figure 2. Top panel: the H₂ mass function from Keres et al. (2003) using CO(1–0) observations of parent samples selected in *B* band and 60 μm , as labelled, and adopting a Milky Way H₂-to-CO conversion factor, $N_{\text{H}_2}/\text{cm}^{-2} = 2 \times 10^{-20} I_{\text{CO}}/\text{K km s}^{-1}$. Here N_{H_2} is the column density of H₂ and I_{CO} is the integrated CO(1–0) line intensity per unit surface area. Also shown is the H₂ mass function of ETGs from the ATLAS^{3D} survey (stars), with and without the $1/V_{\max}$ correction, also for a Milky Way H₂-to-CO conversion factor. Bottom panel: distribution function of the H₂ gas fraction for the ATLAS^{3D} survey.

masses $< 10^{7.5} M_{\odot}$ could only be detected in nearby objects. Thus, we correct the lower H I mass bins for incompleteness using the standard $1/V_{\max}$ method (Schmidt 1968). Note that since ATLAS^{3D} is complete above a *K*-band luminosity of $L_K > 6 \times 10^9 L_{\odot}$, we use a V_{\max} calculated for the H I mass only. Total H₂ masses are available from Young et al. (2011) for all galaxies in the full ATLAS^{3D} survey volume. We constructed H₂ mass functions from this data in the same way (see Fig. 2). The IRAM (Institute for Radio Astronomy in the Millimeter Range)-30 m telescope observations have a fixed noise limit, and Fig. 2 shows both the original and a corrected mass function (where the lower H₂ mass bins have been corrected using the $1/V_{\max}$ method).

Blind H I surveys, such as ALFALFA and HIPASS provide information on the H I content of galaxies in larger volume. H I mass functions were derived from these two surveys and presented in Martin et al. (2010) and Zwaan et al. (2005), respectively. These H I mass functions are shown in the top panel of Fig. 1. However,

because we are interested in how ETGs differ from the overall galaxy population and in isolating the physical processes leading to such differences, we need a proxy for the stellar mass of these H I-selected galaxies. To this end, we use the HIPASS survey cross-matched with the Two Micron All-Sky Survey (Jarrett et al. 2000) to obtain K -band luminosities for the HIPASS galaxies (see Meyer et al. 2008). We limit the analysis to the southern HIPASS sample (Meyer et al. 2004), because the completeness function for this sample is well described and the H I mass function is determined accurately (Zwaan et al. 2005). We find that 86 per cent of the southern HIPASS galaxies have K -band counterparts, and all of these have morphological classifications which are described in Doyle et al. (2005). Galaxy morphologies are taken from the SuperCOSMOS Sky Survey (Hambly et al. 2001), and are obtained by visual inspection, predominantly in the b_J band.

With the subsample of HIPASS galaxies with K -band luminosities and assigned morphological types, we calculate the H I mass function for ETGs, take to be those identified as ‘E-Sa’. For this, we use the maximum likelihood equivalents of the $1/V_{\max}$ values determined by Zwaan et al. (2005). These values represent the maximum volume over which each of the HIPASS galaxies could have been detected, but they are corrected so as to remove the effects of large scale inhomogeneities in the HIPASS survey (see Zwaan et al. 2005 for details). We add an additional selection in the K -band luminosity to match the selection used in ATLAS^{3D} and re-calculate the H I mass function of the overall galaxy population using the same V_{\max} values determined by Zwaan et al. Note that we do not need to recalculate V_{\max} due to the K -band survey being much deeper than the H I survey (i.e. it is able to detect small galaxies further out than the H I survey). The results of this exercise are shown in the top panel of Fig. 1. We also show the estimated H I mass function of the ATLAS^{3D} ETGs. There is very good agreement between the H I mass function of ETGs from the HIPASS and ATLAS^{3D} surveys in the range where they overlap (despite the inclusion of Sa galaxies in the HIPASS sample, which are absent in ATLAS^{3D}).

ATLAS^{3D} provides an important insight into the H I mass function of ETGs in the regime of low H I masses that are not present in HIPASS. In the case of the H I mass function of all galaxies that have K -band luminosities above $6 \times 10^9 L_{\odot}$, we find that the H I mass function is fully recovered down to $M_{\text{H I}} \approx 5 \times 10^9 M_{\odot}$, with a drop in the number density of lower H I masses due to the K -band luminosity limit. Note that this drop is not because of incompleteness but instead is a real feature connected to the minimum H I mass that normal star-forming galaxies with $L_K > 6 \times 10^9 L_{\odot}$ have. By normal star-forming galaxy we mean those that lie on the main sequence of galaxies in the plane of SFR versus stellar mass. The HIPASS matched sample is complete for K -band luminosities $L_K > 6 \times 10^9 L_{\odot}$. The turnover observed at $M_{\text{H I}} \approx 5 \times 10^9 M_{\odot}$ is simply the H I mass expected for a normal star-forming galaxy with $L_K \approx 6 \times 10^9 L_{\odot}$.

Similarly, we calculate the distribution function of the ratio of H I mass to the K -band luminosity, which we refer to as the H I gas fraction. We do this for all HIPASS galaxies with $L_K > 6 \times 10^9 L_{\odot}$ in the cross-matched catalogue and for the subsample of ETGs. This is shown in the bottom panel of Fig. 1. Also shown is the distribution of $M_{\text{H I}}/L_K$ for the ETGs in the ATLAS^{3D} survey. Note that these distribution functions provide higher order constraints on galaxy formation simulations than the more commonly used scaling relations between the H I mass or H₂ mass and stellar mass (e.g. Catinella et al. 2010; Saintonge et al. 2011). This is because this distribution allows us to test not only if the amount of neutral gas in model galaxies is in the expected proportion to their stellar

mass, but also that the number density of galaxies with different gas fractions is correct. Kauffmann et al. (2012) show that this is a stronger constraint on semi-analytic models of galaxy formation, as the distribution of the neutral gas fraction depends on the quenching mechanisms included in the models, and the way they interact. So far, no such comparison has been presented for cosmological hydrodynamical simulations.

We perform the same exercise we did for H I masses but now for H₂ masses. In this case there are no blind surveys of carbon monoxide or any other H₂ tracer, and therefore the data for large samples of galaxies is scarce. Keres et al. (2003) reported the first and only attempt to derive the local luminosity function (LF) of CO(1–0). This was done using B band and a 60 μm selected samples, and with follow up using the Five College Radio Astronomy Observatory. This is shown in the top panel of Fig. 2. Also shown is the H₂ mass function of ETGs from ATLAS^{3D}. Here, we adopt a Milky Way H₂-to-CO conversion factor, $N_{\text{H}_2}/\text{cm}^{-2} = 2 \times 10^{-20} I_{\text{CO}}/\text{K km s}^{-1}$, for both the Keres et al. sample and the ATLAS^{3D} [but see Lagos et al. (2012) for more on conversions]. Here N_{H_2} is the column density of H₂ and I_{CO} is the integrated CO(1–0) line intensity per unit surface area. In the bottom panel of Fig. 2, we show the distribution function of the M_{H_2}/L_K ratios (we which refer to as H₂ gas fractions) for the ATLAS^{3D} sources.

Throughout this paper, we compare different flavours of GALFORM to the set of observations presented in Figs 1 and 2.

3 MODELLING THE EVOLUTION OF THE MORPHOLOGY, NEUTRAL GAS CONTENT AND STAR FORMATION IN GALAXIES

Here we briefly describe the GALFORM semi-analytical model of galaxy formation and evolution (introduced by Cole et al. 2000), focusing on the aspects that are relevant to the build-up of ETGs.

The GALFORM model takes into account the main physical processes that shape the formation and evolution of galaxies. These are (i) the collapse and merging of DM haloes, (ii) the shock heating and radiative cooling of gas inside DM haloes, leading to the formation of galactic discs, (iii) quiescent SF in galaxy discs, (iv) feedback from supernovae (SNe), from heating by active galactic nuclei (AGN) and from photoionization of the intergalactic medium, (v) chemical enrichment of stars and gas and (vi) galaxy mergers driven by dynamical friction within common DM haloes which can trigger bursts of SF, and lead to the formation of spheroids (for a review of these ingredients see Baugh 2006; Benson 2010). Galaxy luminosities are computed from the predicted SF and chemical enrichment histories using a stellar population synthesis model (see Gonzalez-Perez et al. 2014 to see the impact of using different models).

In the rest of this section, we describe the SF law used and how this connects to the two-phase ISM (Section 3.1), the recycled fraction and yield from newly formed stars and how that gas fuels the ISM (Section 3.2), the physical processes that give rise to discs and bulges in GALFORM (Section 3.3) and the modelling of the ram pressure stripping of the hot gas (Section 3.4) in satellite galaxies. We put our focus into these processes because we aim to distinguish the contribution of each of them to the neutral gas content of ETGs: galaxy mergers, hydrostatic cooling and recycling from stars. These processes are the same in the three variants of GALFORM we use in this paper: the models of Lagos et al. (2012, hereafter Lagos12), Gonzalez-Perez et al. (2014, hereafter Gonzalez-Perez14) and Lacey et al. (in preparation, hereafter Lacey14), albeit

with different parameters. In Section 3.5, we describe the differences between these variants. We finish the section with a short description of the N -body cosmological simulation used and the parameters adopted (Section 3.6).

3.1 Interstellar medium gas phases and the star formation law

In GALFORM the SF law developed in [Lagos et al. \(2011b\)](#), hereafter ‘L11’) is adopted. In this SF law the atomic and molecular phases of the neutral hydrogen in the ISM are explicitly distinguished. L11 found that the SF law that gives the best agreement with the observations without the need for further calibration is the empirical SF law of Blitz & Rosolowsky (2006; hereafter BR). Given that the SF law has been well constrained in spiral and dwarf galaxies in the local Universe, L11 decided to implement this molecular-based SF law only in the quiescent SF mode (SF following gas accretion on to the disc), keeping the original prescription of Cole et al. (2000) for starbursts (driven by galaxy mergers and global disc instabilities).

Quiescent star formation. The empirical SF law of Blitz & Rosolowsky has the form

$$\Sigma_{\text{SFR}} = \nu_{\text{SF}} \Sigma_{\text{mol}}, \quad (1)$$

where Σ_{SFR} and Σ_{mol} are the surface densities of SFR and molecular gas, respectively, and ν_{SF} is the inverse of the SF time-scale for the molecular gas. The molecular gas mass includes the contribution from helium. The ratio between the molecular and total gas mass, f_{mol} , depends on the internal hydrostatic pressure through $\Sigma_{\text{H}_2}/\Sigma_{\text{H}_1} = f_{\text{mol}}/(f_{\text{mol}} - 1) = (P_{\text{ext}}/P_0)^\alpha$. Here H I and H_2 only include hydrogen (which in total corresponds to a fraction $X_{\text{H}} = 0.74$ of the overall cold gas mass). To calculate P_{ext} , we use the approximation from Elmegreen (1989), in which the pressure depends on the surface density of gas and stars. The parameters ν_{SF} , P_0 and α are given in Section 3.5 for each of the three GALFORM variants.

Starbursts. For starbursts the situation is less clear than in SF in discs mainly due to observational uncertainties, such as the conversion between CO and H_2 in starbursts, and the intrinsic compactness of star-forming regions, which have prevented a reliable characterization of the SF law (e.g. Genzel et al. 2010). For this reason, we choose to apply the BR law only during quiescent SF (fuelled by accretion of cooled gas on to galactic discs) and retain the original SF prescription for starbursts (see Cole et al. 2000; L11 for details). In the latter, the SF time-scale is proportional to the bulge dynamical time-scale above a minimum floor value and involves the whole cold gas content of the galaxy, $\text{SFR} = M_{\text{cold}}/\tau_{\text{SF}}$ (see Granato et al. 2000; Lacey et al. 2008 for details). The SF time-scale is defined as

$$\tau_{\text{SF}} = \max(\tau_{\text{min}}, f_{\text{dyn}} \tau_{\text{dyn}}), \quad (2)$$

where τ_{dyn} is the bulge dynamical time-scale, τ_{min} is a minimum duration adopted for starbursts and f_{dyn} is a free parameter. The values of the parameters τ_{min} and f_{dyn} are given in Section 3.5 for each of the three GALFORM variants.

3.2 Recycled fraction and yield

In GALFORM we adopt the instantaneous mixing approximation for the metals in the ISM. This implies that the metallicity of the cold gas mass instantaneously absorbs the fraction of recycled mass and newly synthesized metals in recently formed stars, neglecting the time delay for the ejection of gas and metals from stars.

The recycled mass injected back to the ISM by newly born stars is calculated from the initial mass function (IMF) as

$$R = \int_{m_{\text{min}}}^{m_{\text{max}}} (m - m_{\text{rem}}) \phi(m) dm, \quad (3)$$

where m_{rem} is the remnant mass and the IMF is defined as $\phi(m) \propto dN(m)/dm$. Similarly, we define the yield as

$$p = \int_{m_{\text{min}}}^{m_{\text{max}}} m_i(m) \phi(m) dm, \quad (4)$$

where $m_i(m)$ is the mass of newly synthesized metals ejected by stars of initial mass m . The integrations limits are taken to be $m_{\text{min}} = 1 M_\odot$ and $m_{\text{max}} = 120 M_\odot$. Stars with masses $m < 1 M_\odot$ have lifetimes longer than the age of the Universe, and therefore they do not contribute to the recycled fraction and yield. The quantities $m_{\text{rem}}(m)$ and $m_i(m)$ depend on the initial mass of a star and are calculated by stellar evolution theory. The stellar evolution model we use for intermediate stars ($1 M_\odot < m \lesssim 8 M_\odot$) is Marigo (2001) (i.e. which provides $m_{\text{rem}}(m)$ and $m_i(m)$ for those types of stars), while for massive stars, $m \gtrsim 8 M_\odot$, we use Portinari, Chiosi & Bressan (1998).

We describe in Section 3.5 the IMF adopted in the three variants of GALFORM.

3.3 Morphological transformation of galaxies

3.3.1 The build-up of discs

Galaxies form from gas which cools from the hot halo observing conservation of angular momentum. As the temperature decreases, thermal pressure stops supporting the gas which then settles in a rotating discs (Fall & Efstathiou 1980).

We model the gas profile of the hot gas with a β profile (Cavaliere & Fusco-Femiano 1976),

$$\rho_{\text{hot}}(r) \propto (r^2 + r_{\text{core}}^2)^{-3\beta_{\text{fit}}/2}, \quad (5)$$

The simulations of Eke et al. (1998) show that $\beta_{\text{fit}} \approx 2/3$ and that $r_{\text{core}}/R_{\text{NFW}} \approx 1/3$. In GALFORM, we adopt the above values and use $\rho_{\text{hot}}(r) \propto (r^2 + R_{\text{NFW}}^2/9)^{-1}$, where R_{NFW} is the scale radius of the Navarro, Frenk & White (1997) profile of the DM haloes.

During a time step δt in the integration of galaxy properties, we calculate the amount of gas that cools and estimate the radius at which $\tau_{\text{cool}}(r_{\text{cool}}) = t - t_{\text{form}}$, where t_{form} corresponds to the time at which the halo was formed and t is the current time. The gas inside r_{cool} is cool enough to be accreted on to the disc. However, in order to be accreted on to the disc, the cooled gas should have had enough time to fall on to the disc. Thus, the gas that has enough time to cool and be accreted on to the disc is that within the radius in which the free-fall time and the cooling time are smaller than $(t - t_{\text{form}})$, defined as r_{ff} and r_{cool} , respectively. The mass accreted on to the disc simply corresponds to the hot gas mass enclosed within $r = \min[r_{\text{cool}}, r_{\text{ff}}]$.

We calculate r_{cool} from the cooling time, which is defined as

$$\tau_{\text{cool}}(r) = \frac{3}{2} \frac{\mu m_{\text{H}} k_{\text{B}} T_{\text{hot}}}{\rho_{\text{hot}}(r) \Lambda(T_{\text{hot}}, Z_{\text{hot}})}. \quad (6)$$

Here, $\Lambda(T_{\text{hot}}, Z_{\text{hot}})$ is the cooling function that depends on the gas temperature, T_{hot} , which corresponds to the virial temperature of the halo ($T_{\text{hot}} = T_{\text{v}}$) and the metallicity Z_{hot} (i.e. the ratio between the mass in metals heavier than Helium and total gas mass). The cooling rate per unit volume is $\epsilon_{\text{cool}} \propto \rho_{\text{hot}}^2 \Lambda(T_{\text{hot}}, Z_{\text{hot}})$. In GALFORM we adopt the cooling function tabulated of Sutherland & Dopita (1993).

3.3.2 The formation of spheroids

Galaxy mergers and disc instabilities give rise to the formation of spheroids and elliptical galaxies. Below we describe both physical processes.

Galaxy mergers. When DM haloes merge, we assume that the galaxy hosted by the most massive progenitor halo becomes the central galaxy, while all the other galaxies become satellites orbiting the central galaxy. These orbits gradually decay towards the centre due to energy and angular momentum losses driven by dynamical friction with the halo material.

Eventually, given sufficient time satellites spiral in and merge with the central galaxy. Depending on the amount of gas and baryonic mass involved in the galaxy merger, a starburst can result. The time for the satellite to hit the central galaxy is called the orbital time-scale, τ_{merge} , which is calculated following Lacey et al. (1993) as

$$\tau_{\text{merge}} = f_{\text{df}} \Theta_{\text{orbit}} \tau_{\text{dyn}} \left[\frac{0.3722}{\ln(\Lambda_{\text{Coulomb}})} \right] \frac{M}{M_{\text{sat}}}. \quad (7)$$

Here, f_{df} is a dimensionless adjustable parameter which is $f_{\text{df}} > 1$ if the satellite's halo is efficiently stripped early on during the infall, Θ_{orbit} is a function of the orbital parameters, $\tau_{\text{dyn}} \equiv \pi R_v/V_v$ is the dynamical time-scale of the halo, $\ln(\Lambda_{\text{Coulomb}}) = \ln(M/M_{\text{sat}})$ is the Coulomb logarithm, M is the halo mass of the central galaxy and M_{sat} is the mass of the satellite, including the mass of the DM halo in which the galaxy was formed.

Lagos12 and Gonzalez-Perez14 used the Θ_{orbit} function calculated in Lacey et al. (1993),

$$\Theta_{\text{orbit}} = \left[\frac{J}{J_c(E)} \right]^{0.78} \left[\frac{r_c(E)}{R_v} \right]^2, \quad (8)$$

where J is the initial angular momentum and E is the energy of the satellite's orbit, and $J_c(E)$ and $r_c(E)$ are, respectively, the angular momentum and radius of a circular orbit with the same energy as that of the satellite. Thus, the circularity of the orbit corresponds to $J/J_c(E)$. The function Θ_{orbit} is well described by a lognormal distribution with median value $\langle \log_{10} \Theta_{\text{orbit}} \rangle = -0.14$ and dispersion $(\langle \log_{10} \Theta_{\text{orbit}} \rangle - \langle \log_{10} \Theta_{\text{orbit}} \rangle^2)^{1/2} = 0.26$. These values are not correlated with satellite galaxy properties. Therefore, for each satellite, the value of Θ_{orbit} is randomly chosen from the above distribution. Note that the dependence of Θ_{orbit} on J in equation (8) is a fit to numerical estimates. Lacey14 use the updated dynamical friction function of Jiang et al. (2008), which slightly changes the dependence on the mass ratio of the satellite to the central galaxy. The net effect of such a change is that minor mergers occur faster, while major mergers occur slower when compared to the Lacey et al. (1993) prescription.

If the merger time-scale is less than the time that has elapsed since the formation of the halo, i.e. if $\tau_{\text{merge}} < t - t_{\text{form}}$, we proceed to merge the satellite with the central galaxy at t . If the total mass of gas plus stars of the primary (largest) and secondary galaxies involved in a merger are $M_p = M_{\text{cold},p} + M_{*,p}$ and $M_s = M_{\text{cold},s} + M_{*,s}$, the outcome of the galaxy merger depends on the galaxy mass ratio, M_s/M_p , and the fraction of gas in the primary galaxy, $M_{\text{cold},p}/M_p$.

(i) $M_s/M_p > f_{\text{ellip}}$ drives a major merger. In this case, all the stars present are rearranged into a spheroid. In addition, any cold gas in the merging system is assumed to undergo a burst of SF and the stars formed are added to the spheroid component. We typically take $f_{\text{ellip}} = 0.3$, which is within the range found in simulations (e.g. see Baugh et al. 1996 for a discussion).

(ii) $f_{\text{burst}} < M_s/M_p \leq f_{\text{ellip}}$ drives minor mergers. In this case all the stars in the secondary galaxy are accreted on to the primary galaxy spheroid, leaving the stellar disc of the primary intact. In minor mergers the presence of a starburst depends on the cold gas content of the primary galaxy, as set out in the next bullet point.

(iii) $f_{\text{burst}} < M_s/M_p \leq f_{\text{ellip}}$ and $M_{\text{cold},p}/M_p > f_{\text{gas, burst}}$ drives a starburst in a minor merger. The perturbations introduced by the secondary galaxy are assumed to drive all the cold gas from both galaxies to the new spheroid, producing a starburst. There is no starburst if $M_{\text{cold},p}/M_p < f_{\text{gas, burst}}$. The Bau05 and Bow06 models adopt $f_{\text{gas, burst}} = 0.75$ and $f_{\text{gas, burst}} = 0.1$, respectively.

(iv) $M_s/M_p \leq f_{\text{burst}}$ results in the primary disc remaining unchanged. As before, the stars accreted from the secondary galaxy are added to the spheroid, but the overall gas component (from both galaxies) stays in the disc, and the stellar disc of the primary is preserved. The Bau05 and Bow06 models adopt $f_{\text{burst}} = 0.05$ and $f_{\text{burst}} = 0.1$, respectively.

Disc instabilities. If the disc becomes sufficiently massive that its self-gravity is dominant, then it is unstable to small perturbations by satellites or DM substructures. The criterion for instability was described by Efstathiou, Lake & Negroponte (1982) and Mo et al. (1998) and introduced into GALFORM by Cole et al. (2000),

$$\epsilon = \frac{V_{\text{circ}}(r_d)}{\sqrt{G M_d/r_s}}. \quad (9)$$

Here, $V_{\text{circ}}(r_d)$ is the circular velocity of the disc at the half-mass radius, r_d is the scale radius of the disc and M_d is the disc mass (gas plus stars). If $\epsilon < \epsilon_{\text{disc}}$, where ϵ_{disc} is a parameter, then the disc is considered to be unstable. In the case of unstable discs, stars and gas in the disc are accreted on to the spheroid and the gas inflow drives a starburst. Lagos12 and Gonzalez-Perez14 adopt $\epsilon_{\text{disc}} = 0.8$, while Lacey14 adopt a slightly higher value, $\epsilon_{\text{disc}} = 0.9$.

3.4 Gradual ram pressure stripping of the hot gas

The standard treatment of the hot gas in accreted satellites in GALFORM is usually referred to as ‘strangulation’⁴ of the hot gas. In this extreme case, the ram pressure stripping of the satellite's hot gas reservoir by the hot gas in the main halo is completely efficient and is assumed to occur as soon as a galaxy becomes a satellite. This treatment has shown to drive redder colour of satellite galaxies (Font et al. 2008). As we are studying ETGs, which tend to be found more frequently in denser environments, we test the impact of a more physical and gradual process, the partial ram pressure stripping of the hot gas, on the neutral gas content of ETGs. Simulations show that the amount of gas removed from the satellite's hot reservoir depends upon the ram pressure experienced which is turn is determined by the pericentre of the orbit (McCarthy et al. 2008).

Here we briefly describe the more physical partial ram pressure stripping model introduced by Font et al. (2008). The partial ram pressure stripping of the hot gas is applied to a spherical distribution of hot gas. The model considers that all the host gas outside the stripping radius, r_{str} , is removed from the host gas reservoir and transferred to the central galaxy halo. The stripping radius is defined as the radius where the ram pressure, P_{ram} , equals the gravitational

⁴ Another way this is referred to in the literature is ‘starvation’, but both terms refer to the same process: complete removal of the hot gas reservoir of galaxies when they become satellites.

restoring force per unit area of the satellite galaxy, P_{grav} . The ram pressure is defined as

$$P_{\text{ram}} \equiv \rho_{\text{gas,p}} v_{\text{sat}}^2, \quad (10)$$

and the gravitational pressure as

$$P_{\text{grav}} \equiv \alpha_{\text{rp}} \frac{G M_{\text{tot,sat}}(r_{\text{str}}) \rho_{\text{gas,s}}(r_{\text{str}})}{r_{\text{str}}}. \quad (11)$$

Here, $\rho_{\text{gas,p}}$ is the gas density of the parent halo, v_{sat} is the velocity of the satellite with respect to the parent halo gas medium, $M_{\text{tot,sat}}(r_{\text{str}})$ is the total mass of the satellite galaxy (stellar, gas and DM components) enclosed within r_{str} and $\rho_{\text{gas,s}}(r_{\text{str}})$ is the hot gas density of the satellite galaxy at r_{str} . In this model, r_{str} is measured from the centre of the satellite galaxy subhalo. The coefficient α_{rp} is a geometric constant of order unity. In this paper we use $\alpha_{\text{rp}} = 2$ which is the value found by McCarthy et al. (2008) in their hydrodynamical simulations. The hot gas of the parent halo follows the density profile of equation (5).

This model assumes that the hot gas of the satellite galaxy inside r_{str} remains intact while the hot gas outside is stripped on approximately a sound crossing time. In GALFORM, r_{str} is calculated at the time a galaxy becomes satellite solving equation (10) and setting the ram pressure to its maximum value, which occurs at the pericentre of the orbit of the satellite galaxy. The hot gas outside r_{str} is instantaneously stripped once the galaxy crosses the virial radius of the parent halo. This simplified modelling overestimates the hot gas stripped between the time the satellite galaxy crosses the virial radius and the first passage. Font et al. (2008) argue that this is not a bad approximation as the time-scale for the latter is only a small fraction of the time a satellite galaxy spends orbiting in the parent halo. Font et al. also argue that in terms of hot gas removal, ram pressure is the major physical mechanism, while tidal heating and stripping are secondary effects.

The infall velocity of the satellite galaxy is randomly sampled from the 2D distribution of infalling velocity of the DM substructures, measured by Benson (2005) from a large suite of cosmological simulations. Then, the pericentre radius and velocity at the pericentre are computed by assuming that the orbital energy and angular momentum are conserved and by treating the satellite as a point mass orbiting within a Navarro–Frenk–White gravitational potential with the same total mass and concentration as the parent halo.

The remaining hot gas in the satellite galaxy halo can cool down and feed the satellite's disc. The cooling of this remaining hot gas is calculated by assuming that the mean density of the hot gas of the satellite is not altered by the stripping process, and using a nominal hot halo mass that includes both the current hot gas mass and the hot gas mass that has been stripped. The difference with the standard calculation described in Section 3.3.1 is that the cooling radius cannot be larger than r_{str} . As SF continues to take place in satellite galaxies, there will be an additional source of hot gas which corresponds to the winds escaping the galaxy disc that mix or evaporate to become part of the hot halo gas. Most of this SF takes place when the satellite galaxy is on the outer parts of its orbit, where the ram pressure is small. Font et al. (2008) then suggested that a fraction, ϵ_{strip} (less than unity) of this gas is actually stripped from the hot halo of the satellite. Font et al. discussed the effects of different values for ϵ_{strip} and adopted $\epsilon_{\text{strip}} = 0.1$ to reproduce the colours of satellite galaxies. Throughout this paper we adopt the same value for ϵ_{strip} , but we discuss in Appendix A the effect of varying it.

Finally, in order to account for the growth of the parent halo and the effect this has on the ram pressure, the ram pressure is recalculated for each satellite galaxy every time the parent halo doubles its mass compared to the halo mass at the instant of the initial stripping event.

We test the effect of partial ram pressure stripping of the hot gas by including the above modelling into the three variants of GALFORM, and we refer to the variants with partial ram pressure stripping of the hot gas as Lagos12+RP, Gonzalez-Perez14+RP and Lacey14+RP.

3.5 Differences between the Lagos12, Gonzalez-Perez14 and Lacey14 models

The Lagos12 model is a development of the model originally described in Bower et al. (2006), which was the first variant of GALFORM to include AGN feedback as the mechanism suppressing gas cooling in massive haloes. The Lagos12 model assumes a universal IMF, the Kennicutt (1983) IMF.⁵ Lagos12 extend the model of Bower et al. by including the self-consistent SF law described in Section 3.1, and adopting $v_{\text{SF}} = 0.5 \text{ Gyr}^{-1}$, $\log(P_0/k_B[\text{cm}^{-3}\text{K}]) = 4.23$, where k_B is Boltzmann's constant, and $\alpha = 0.8$, which correspond to the values of the parameters reported by Leroy et al. (2008) for local spiral and dwarf galaxies. This choice of SF law greatly reduces the parameter space of the model and also extends its predictive power by directly modelling the atomic and molecular hydrogen content of galaxies. All of the subsequent models that use the same SF law have also the ability to predict the H I and H₂ gas contents of galaxies. Lagos12 adopt longer duration starbursts (i.e. larger f_{dyn}) compared to Bower et al. to improve the agreement with the observed LF in the rest-frame ultraviolet (UV) at high redshifts. Lagos12 adopts $\tau_{\text{min}} = 100 \text{ Myr}$ and $f_{\text{dyn}} = 50$ in equation (2). The Lagos12 model was developed in the Millennium simulation, which assumed a *Wilkinson Microwave Anisotropy Probe 1* (WMAP1) cosmology (Spergel et al. 2003).

The Gonzalez-Perez14 model updated the Lagos12 model to the WMAP7 cosmology (Komatsu et al. 2011). A small number of parameters were recalibrated to recover the agreement between the model predictions and the observed evolution of both the UV and K-band LFs. These changes include a slightly shorter starburst duration, i.e. $\tau_{\text{min}} = 50 \text{ Myr}$ and $f_{\text{dyn}} = 10$, and weaker SNe feedback. See Gonzalez-Perez14 for more details.

The Lacey14 model is also developed in the WMAP7 cosmology but it differs from the other two flavours in that it adopts a bimodal IMF. The IMF describing SF in discs (i.e. the quiescent mode) is the same as the universal IMF in the other two models, but a top-heavy IMF is adopted for starbursts (i.e. with an IMF slope $x = 1$). This choice motivated by Baugh et al. (2005) who used a bimodal IMF to recover the agreement between the model predictions and observations of the number counts and redshift distribution of submillimetre galaxies. We note, however, that Baugh et al. adopted a more top-heavy IMF for starbursts with $x = 0$. The stellar population synthesis model used for Lacey14 is also different. While both Lagos12 and Gonzalez-Perez14 use Bruzual & Charlot (2003), the Lacey14 model uses Maraston (2005). Another key difference between the Lacey14 model and the other two GALFORM flavours considered here, is that Lacey14 adopt a slightly larger value of the

⁵ The distribution of the masses of stars formed follows $dN(m)/d \ln m \propto m^{-x}$, where N is the number of stars of mass m formed, and x is the IMF slope. For a Kennicutt (1983) IMF, $x = 1.5$ for masses in the range $1 \text{ M}_{\odot} \leq m \leq 100 \text{ M}_{\odot}$ and $x = 0.4$ for masses $m < 1 \text{ M}_{\odot}$.

SF efficiency rate, $\nu_{\text{SF}} = 0.74 \text{ Gyr}^{-1}$, still within the range allowed by the most recent observation compilation of Bigiel et al. (2011), making SF more efficient.

3.6 The N -body simulations and cosmological parameters

We use halo merger trees extracted from the Millennium cosmological N -body simulation (adopting *WMAP1* cosmology; Springel et al. 2005) and its *WMAP7* counterpart. The Millennium simulation⁶ has the following cosmological parameters: $\Omega_m = \Omega_{\text{DM}} + \Omega_{\text{baryons}} = 0.25$ (giving a baryon fraction of 0.18), $\Omega_\Lambda = 0.75$, $\sigma_8 = 0.9$ and $h = 0.73$. The resolution of the N -body simulation is fixed at a halo mass of $1.72 \times 10^{10} h^{-1} M_\odot$. Lagos et al. (2014) show that much higher resolution merger trees are needed to fully resolve the H I content of galaxies from $z = 0$ to $z = 10$. However, in this work we are concerned about galaxies with $L_K \gtrsim 10^9 L_\odot$, which are well resolved in the Millennium simulations. The Lacey14 and Gonzalez-Perez14 were developed in the *WMAP7* version of the Millennium simulation, where the cosmological parameters are $\Omega_m = \Omega_{\text{DM}} + \Omega_{\text{baryons}} = 0.272$ (with a baryon fraction of 0.167), $\Omega_\Lambda = 0.728$, $\sigma_8 = 0.81$ and $h = 0.704$ (*WMAP7* results were presented in Komatsu et al. 2011).

Throughout this work, we show gas masses in units of M_\odot , luminosities in units of L_\odot and number densities in units of $\text{Mpc}^{-3} \text{ dex}^{-1}$. This implies that we have evaluated the h factors. The largest difference driven by the different cosmologies is in the number density, but this is only 0.05 dex, which is much smaller than the differences between the models or between model and observations.

4 THE NEUTRAL GAS CONTENT OF LOCAL EARLY-TYPE GALAXIES: MODELS VERSUS OBSERVATIONS

In this section, we compare the predicted properties of ETGs in the three variants of *GALFORM* with and without the inclusion of partial ram pressure stripping of the hot gas. This modification has little effect on the local b_J - and K -band LFs in the three variants of *GALFORM*. These observables are usually considered to be the main constraints for finding the best set of model parameters (see for example Bower et al. 2010 and Ruiz et al. 2013). Other $z = 0$ properties, such as half-mass radii, gas and stellar metallicity, are also insensitive to the inclusion of partial ram pressure stripping. We therefore conclude that the partial ram pressure stripping versions of the three *GALFORM* models provide a representation of the local Universe as good as the standard models.

The first comparison we perform is the fraction of galaxies that are ETGs as a function of galaxy luminosity. This is a crucial step in our analysis, as we aim to characterize the neutral gas content of the ETG population. Throughout we will refer to ETGs in the model as those having bulge-to-total stellar mass ratios, B/T , > 0.5 . Although this selection criterion is very sharp and has been analysed in detail in the literature (see e.g. Weinzirl et al. 2009; Khochfar et al. 2011), we find that our results are not sensitive to the threshold B/T selecting ETGs, as long as this threshold is $B/T > 0.3$. We analyse this selection criterion in more detail in Section 5.

Fig. 3 shows the fraction of ETGs, f_{early} , as a function of the r -band absolute magnitude at $z = 0$ for the three *GALFORM* models described

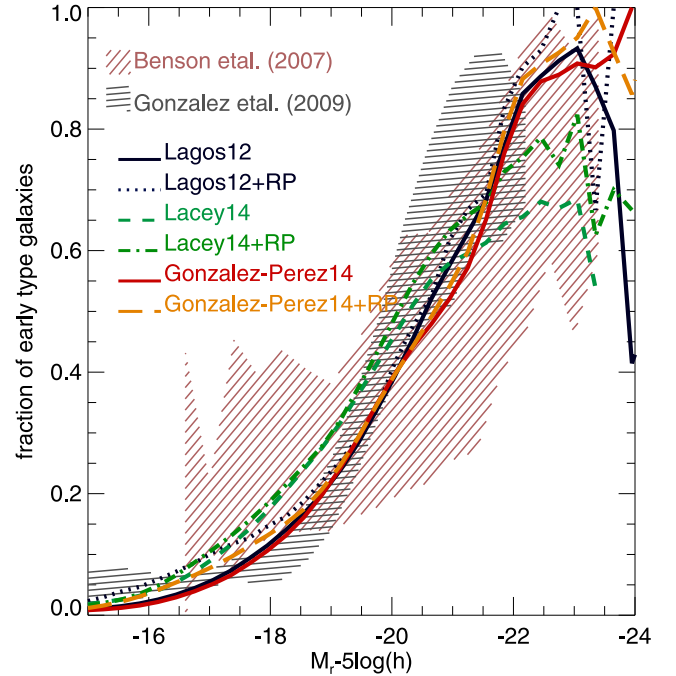


Figure 3. Fraction of ETGs as a function of the rest-frame r -band absolute magnitude, for the Lagos12, Lacey14, Gonzalez-Perez14 and the variants including partial ram pressure of the hot gas ('+RP'), as labelled. ETGs in the models are those with a bulge-to-total stellar mass ratio $B/T > 0.5$. The shaded regions correspond to the observational estimates of Benson et al. (2007) and González et al. (2009) using the SDSS, as labelled. In the case of González et al. (2009) the upper limit of the shaded region is given by the Sérsic index selection, while the lower limit is given by the concentration selection (see the text for details).

in Section 3.5 and their variants including partial ram pressure of the hot gas. Observational estimates of f_{early} are also shown in Fig. 3, for three different ways of selecting ETGs. The first one corresponds to Benson et al. (2007), in which a disc and a bulge component were fitted to r -band images of 8839 bright galaxies selected from the Sloan Digital Sky Survey (SDSS) Early Data Release. The free parameters of the fitting of the disc and bulge components of each galaxy are the bulge ellipticity and disc inclination angle, i . The second corresponds to González et al. (2009), where the r -band concentration, c , and Sérsic index, n , of the SDSS were used to select ETGs: $c > 2.86$ or $n > 2.5$. The upper and lower limits in the shaded region for the Gonzalez et al. measurements correspond to the two early-type selection criteria.

All the models predict a trend of increasing f_{early} with increasing r -band luminosity in good agreement with the observations within the error bars. Note that the inclusion of partial ram pressure stripping of the hot gas leads to a slightly larger f_{early} in galaxies with $M_r - 5 \log(h) > -18$ in the three variants of *GALFORM*. The same happens for the brightest galaxies, $M_r - 5 \log(h) < -22$. Both trends are due to the higher frequency of disc instabilities in the models when partial ram pressure stripping is included; the continuous fuelling of neutral gas in satellite galaxies due to cooling from their hot haloes (which in the case of partial ram pressure stripping is preserved to some extent) drives more SF in discs, lowering the stability parameter of equation (9). These lower stability parameters result in more disc instabilities, driving the formation of spheroids. The space allowed by the observations is large enough so that we cannot discriminate between models.

⁶ Data from the Millennium simulation is available on a relational data base accessible from <http://galaxy-catalogue.dur.ac.uk:8080/Millennium>.

4.1 The H_2 -to- $H\text{I}$ mass ratio dependence on morphology

It has been shown observationally that the ratio between the H_2 and $H\text{I}$ masses correlates strongly with morphological type, with ETGs having higher $H_2/H\text{I}$ mass ratios than late-type galaxies (e.g. Young & Knezek 1989; Bettoni et al. 2003; Lisenfeld et al. 2011; Boselli et al. 2014a). Fig. 4 shows the $H_2/H\text{I}$ mass ratio in the models as a function of the bulge-to-total luminosity ratio in the B band, B/T_B , for all galaxies with B -band absolute magnitude of $M_B - 5 \log(h) < -19$. This magnitude limit is chosen as it roughly corresponds to the selection criteria applied to the observational data shown in Fig. 4. The observational data have morphological types derived from a visual classification of B -band images (de Vaucouleurs et al. 1991), and have also been selected in blue bands (e.g. Simien & de Vaucouleurs 1986; Weinzierl et al. 2009).

The models predict a relation between the $H_2/H\text{I}$ mass ratio and B/T_B that is in good agreement with the observations. Note that, for $B/T_B < 0.2$, the models slightly overpredict the median $H_2/H\text{I}$ mass ratio. The latter has been also observed in the recent *Herschel* Reference Survey (HRS; Boselli et al. 2014a; stars in Fig. 4) for galaxies of morphological types later than Sd (including irregular galaxies). In GALFORM there is a monotonic relation between $H_2/H\text{I}$ mass ratio and stellar mass in a way that $H_2/H\text{I}$ decreases with decreasing stellar mass (see L11 for a detailed discussion). On

the other hand, the relation between stellar mass and B/T is not monotonic, in a way that the median stellar mass in the lowest B/T bins ($B/T < 0.1$) is higher than at $B/T \sim 0.2$. This drives the slight increase of $H_2/H\text{I}$ at the lowest B/T . The physical reason why B/T is not monotonically correlated to stellar mass is because environment plays an important role in the morphology (for example in the number of galaxy mergers, and disc instabilities), which makes it a more transient property of galaxies, while stellar mass is not necessarily correlated to environment but to halo mass. This will be discussed in more detail in Paper II.

Of the three GALFORM variants, the model that predicts the steepest slope for the relation between $H_2/H\text{I}$ mass ratio and B/T is the Lacey14 model. However, when including partial ram pressure stripping of the hot gas, all the models show a slight increase in this slope, with ETGs having higher $H_2/H\text{I}$ mass ratios. This comes from the higher gas surface densities that ETGs have on average when including partial ram pressure, which drive higher hydrostatic pressure. The strangulation scenario, which removes the hot gas instantaneously, drives a quick depletion of the cold gas reservoir in galaxies as SF continues, while in the partial ram pressure scenario the cold gas reservoir is still replenished due to continuous inflow of gas from the satellite's remaining hot halo.

4.2 The $H\text{I}$ content of early-type galaxies

Fig. 5 shows the model predictions for the $H\text{I}$ mass function for all galaxies (top panel), for the subsample of galaxies with K -band luminosities $L_K > 6 \times 10^9 L_\odot$ (middle panel) and for ETGs with the same K -band luminosity cut (bottom panel). The observational results in Fig. 5 are described in Section 2. For the overall galaxy population (top panel of Fig. 5), the six models provide a good description of the $H\text{I}$ mass function, and the inclusion of partial ram pressure stripping of the hot gas has little effect.

For the galaxy population with $L_K > 6 \times 10^9 L_\odot$ (middle panel of Fig. 5), the Lagos12 and Gonzalez-Perez14 models predict a slightly lower number density at the peak of the $H\text{I}$ mass distribution compared to the observations, while the Lacey14 model predicts a $H\text{I}$ mass distribution in good agreement with the observations throughout the full $H\text{I}$ mass range. Note that the inclusion of partial ram pressure stripping of the hot gas has the effect of increasing the number density of galaxies with $10^8 M_\odot < M_{H\text{I}} < 3 \times 10^9 M_\odot$, and decreasing the number density of galaxies with $M_{H\text{I}} < 10^8 M_\odot$. The reason for this is that many of the galaxies with low $H\text{I}$ masses ($M_{H\text{I}} < 10^8 M_\odot$) become more gas rich when partial ram pressure stripping is included compared to the case of strangulation of the hot gas, and move to higher $H\text{I}$ masses ($10^8 M_\odot < M_{H\text{I}} < 3 \times 10^9 M_\odot$). This slight change improves the agreement with the observations in the three GALFORM models, particularly around the turnover at $M_{H\text{I}} \approx 5 \times 10^9 M_\odot$ in the mass function shown in the middle panel of Fig. 5.

In the case of ETGs with $L_K > 6 \times 10^9 L_\odot$ (bottom panel of Fig. 5), the Lagos12 model predicts a number density of galaxies with $10^9 M_\odot < M_{H\text{I}} < 10^{10} M_\odot$ lower than observed, while the predictions from the Gonzalez-Perez14 and Lacey14 models agree well with the observations. The inclusion of partial ram pressure stripping of the hot gas in the three models has the effect of increasing the number density of ETGs with $H\text{I}$ masses $M_{H\text{I}} > 10^8 M_\odot$. This increase allows the models to get closer to the observations, and particularly the Lacey14+RP model predicts a $H\text{I}$ mass function of ETGs in very good agreement with the observations.

The $H\text{I}$ mass function has become a standard constraint on the GALFORM model since L11. However, Lemonias et al. (2013) show

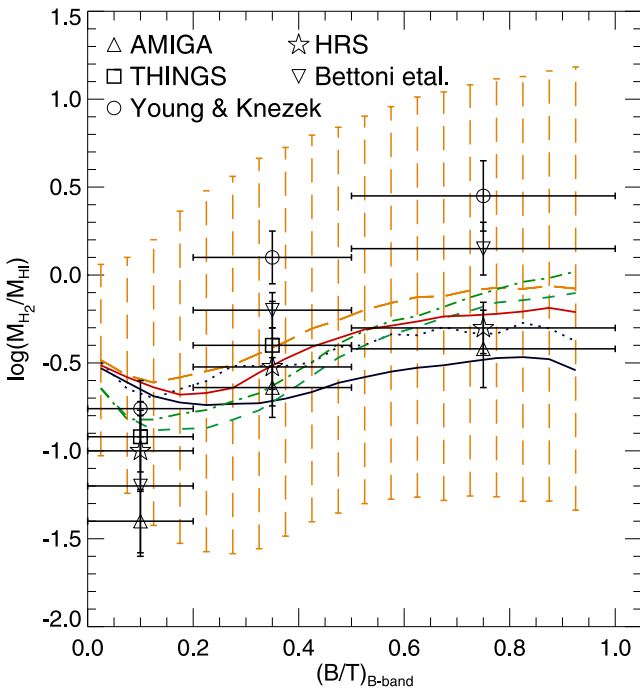


Figure 4. Molecular-to-atomic hydrogen mass ratio, $M_{H_2}/M_{H\text{I}}$, as a function of the bulge-to-total luminosity in the B band, $(B/T)_B$, in the same models of Fig. 3 (lines as labelled in Fig. 3), for galaxies with absolute B -band magnitudes, $M_B - 5 \log(h) < -19$. Lines show the median of the relations for each model and the 10 and 90 percentiles are only shown for the Gonzalez-Perez14+RP model, as an illustration of the dispersion. The other model shows very similar 10 and 90 percentiles. Observational results from Young & Knezek (1989), Bettoni, Galletta & García-Burillo (2003), Leroy et al. (2008) (THINGS sample), Lisenfeld et al. (2011) (AMIGA sample) and Boselli et al. (2014a) (HRS sample) are shown as symbols, as labelled, and we combine them so that $B/T < 0.2$ corresponds to Irr, Sm, Sd galaxies; $0.2 < B/T < 0.5$ corresponds to Sc, Sb, Sa galaxies; $B/T > 0.5$ corresponds to E and S0 galaxies (see de Vaucouleurs et al. 1991 for a description of each morphological type).

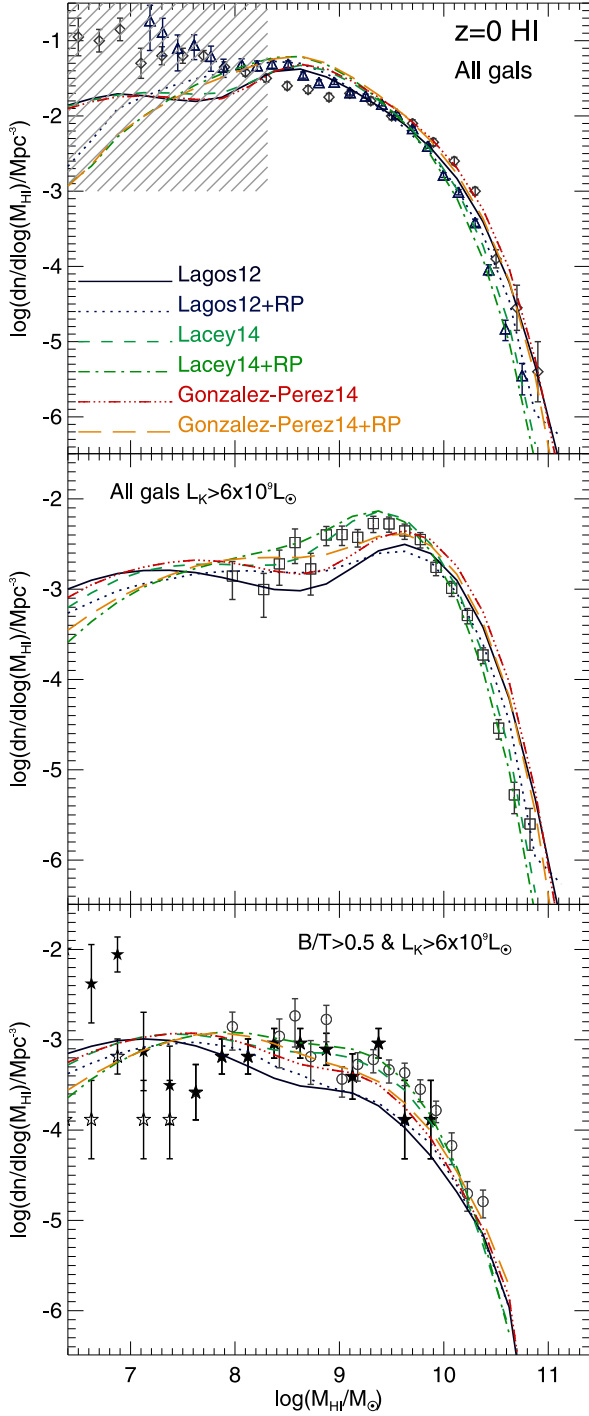


Figure 5. Top panel: the H I mass function of all galaxies at $z = 0$ for the models Lagos12, Lacey14 and Gonzalez-Perez14 and its variants including partial ram pressure stripping of the hot gas ('+RP'), as labelled. Observations correspond to Zwaan et al. (2005) (triangles) and Martin et al. (2010) (diamonds). The shaded region shows the range where the number density of galaxies decline due to halo mass resolution effects. Middle panel: as in the top panel, but here the H I mass function is shown for galaxies with K -band luminosities $L_K > 6 \times 10^9 L_\odot$. Observations correspond to the analysis of HIPASS presented in Section 2. Bottom panel: as in the top panel but for ETGs (those with a bulge-to-total stellar mass ratio > 0.5) and K -band luminosities $L_K > 6 \times 10^9 L_\odot$. Observations correspond to the analysis of HIPASS (circles) and the ATLAS^{3D} (with and without volume correction as filled and empty stars, respectively) surveys containing galaxies with $L_K > 6 \times 10^9 L_\odot$ and described in Section 2.

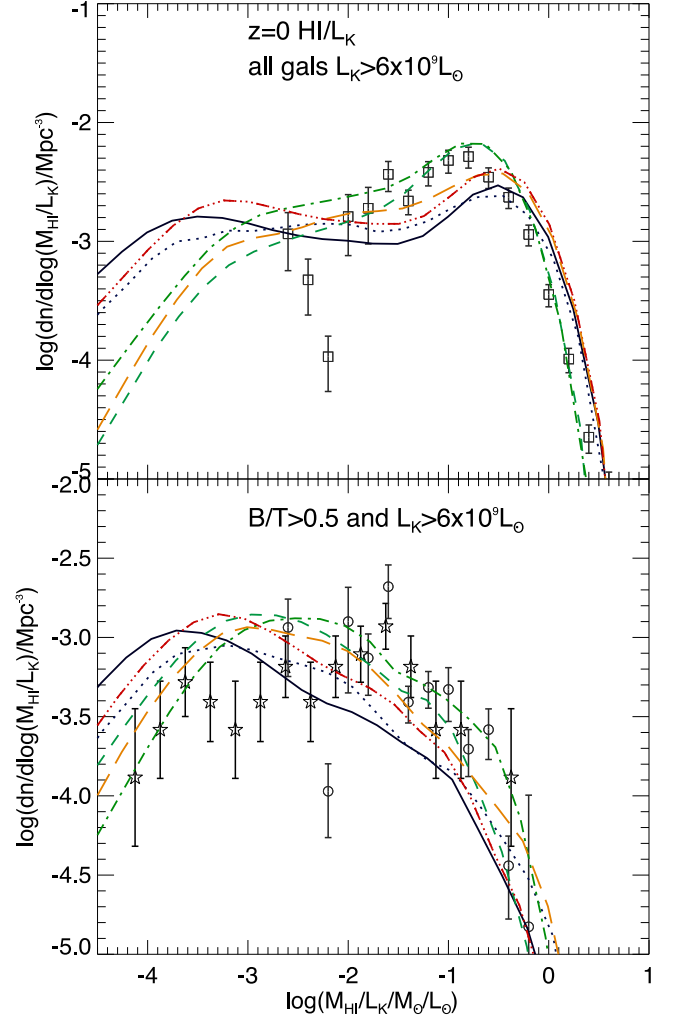


Figure 6. Top panel: the distribution function of the ratio between the H I mass and the K -band luminosity for the same models shown in Fig. 5 for galaxies with $L_K > 6 \times 10^9 L_\odot$. Observations correspond to the analysis of HIPASS presented in Section 2. Bottom panel: as in the top panel but for ETGs ($B/T > 0.5$) with $L_K > 6 \times 10^9 L_\odot$. Observations correspond to the analysis of HIPASS (circles) and ATLAS^{3D} (stars) presented in Section 2.

that a stronger constraint on simulations of galaxy formation is provided by the conditional mass function of gas, or similarly, the gas fraction distribution. Here, we compare the predictions for the H I gas fraction distribution function with observations in Fig. 6. The H I gas fraction is taken with respect to the K -band luminosity to allow direct comparisons with the observations without the need of having to convert between different adopted IMFs, which would be the case if stellar mass was used. Mitchell et al. (2013) show that when simulations and observations adopt different IMF, the comparison between the stellar masses predicted by the models and observations is misleading. Instead, a full SED fitting needs to be performed to make a fair comparison in such a case. Note that the same applies when the SF histories adopted in the observations differ significantly from the simulated galaxies.

The top panel of Fig. 6 shows the predicted H I gas fraction for galaxies with $L_K > 6 \times 10^9 L_\odot$. The observations are described in Section 2. The Lagos12 and Gonzalez-Perez14 models predict a peak of the gas fraction distribution at higher gas fractions than observed, while the Lacey14 predicts a peak closer to the observed

one (i.e. $M_{\text{H I}}/L_K \approx 0.15 M_\odot/L_\odot$). The galaxies at the peak of the H I gas fraction distribution also lie in the main sequence of galaxies in the SFR–stellar mass plane (Lagos et al. 2011a). The inclusion of partial ram pressure of the hot gas increases the number density of galaxies with H I gas fractions $M_{\text{H I}}/L_K > 10^{-3} M_\odot/L_\odot$ and reduces the number density of galaxies with lower H I gas fractions. The physical reason behind these trends is that the inclusion of partial ram pressure stripping increases the H I gas fraction of gas poor galaxies, compared to the strangulation scenario, due to replenishment of their cold gas reservoir. The bottom panel of Fig. 6 shows the H I gas fraction distribution of ETGs ($B/T > 0.5$) with $L_K > 6 \times 10^9 L_\odot$. The H I gas fraction distribution of ETGs is very different from that of all galaxies, showing a much broader distribution with a tail to very low H I gas fractions in all six models. The Lagos12 and Gonzalez-Perez14 models predict lower H I gas fractions for ETGs than observed, while the predictions from Lacey14 are closer to the observations. By including partial ram pressure stripping of the hot gas, the H I gas fractions increase in all the models due to the replenishment of the cold gas reservoirs in satellite galaxies. The predictions of the Lacey14+RP model are a very good description of the observations, with a peak in the number density of ETGs in the range $M_{\text{H I}}/L_K \approx 0.002\text{--}0.02 M_\odot/L_\odot$. Overall, the H I content of ETGs is moderately sensitive to the treatment of the hot gas content of satellite galaxies, while the overall galaxy population does not show the same sensitivity to this physical process due to the dominance of central galaxies.

4.3 The H₂ content of early-type galaxies

The top panel of Fig. 7 shows the predicted H₂ mass functions for all galaxies. The observations are from Keres et al. (2003) and are described in Section 2. All the models provide a good description of the observations. The largest differences between the models are at the high-mass end. The Gonzalez-Perez14 model predicts the highest number densities of galaxies with $M_{\text{H}_2} > 5 \times 10^9 M_\odot$, although it is still in agreement with the observations within the error bars. Note that the inclusion of partial ram pressure stripping of the hot gas leads to very little change. This is due to the dominance of central galaxies in the H₂ mass function, which are only indirectly affected by the treatment of partial ram pressure stripping. In the bottom panel of Fig. 7, we show the H₂ mass function of ETGs ($B/T > 0.5$) with K -band luminosities $L_K > 6 \times 10^9 L_\odot$. The differences between the models are significant: the Lagos12, Gonzalez-Perez14 and Lacey14 models predict a number density of ETGs with H₂ masses $M_{\text{H}_2} > 10^7 M_\odot$ much lower than is observed, by factors of 10, 6 and 4, respectively. It is only when partial ram pressure stripping of the hot gas is included that their predictions get closer to the observations. In particular, the Lacey14+RP model predicts a number density of ETGs with $M_{\text{H}_2} > 10^7 M_\odot$ that is very close to the observations. The Lagos12+RP and Gonzalez-Perez14+RP models are still a factor of 3–4 lower than the observations. At the high-mass end of the H₂ mass function for all galaxies (top panel of Fig. 7), the contribution from ETGs is significant (although not dominant), while for the H I, ETGs are only a minor contribution. The reason for this is simply the higher H₂-to-H I mass ratios in ETGs compared to late-type galaxies (see Fig. 4).

Fig. 8 shows the H₂ gas fraction distribution for ETGs with $L_K > 6 \times 10^9 L_\odot$ for the same six models of Fig. 7. Similarly to the H₂ mass function, the Lagos12 and Gonzalez-Perez14 models predict a number density of ETGs with $M_{\text{H}_2}/L_K > 10^{-3} M_\odot/L_\odot$ lower than observed, while their variants including partial ram pressure stripping of the hot gas predict higher number densities, in better

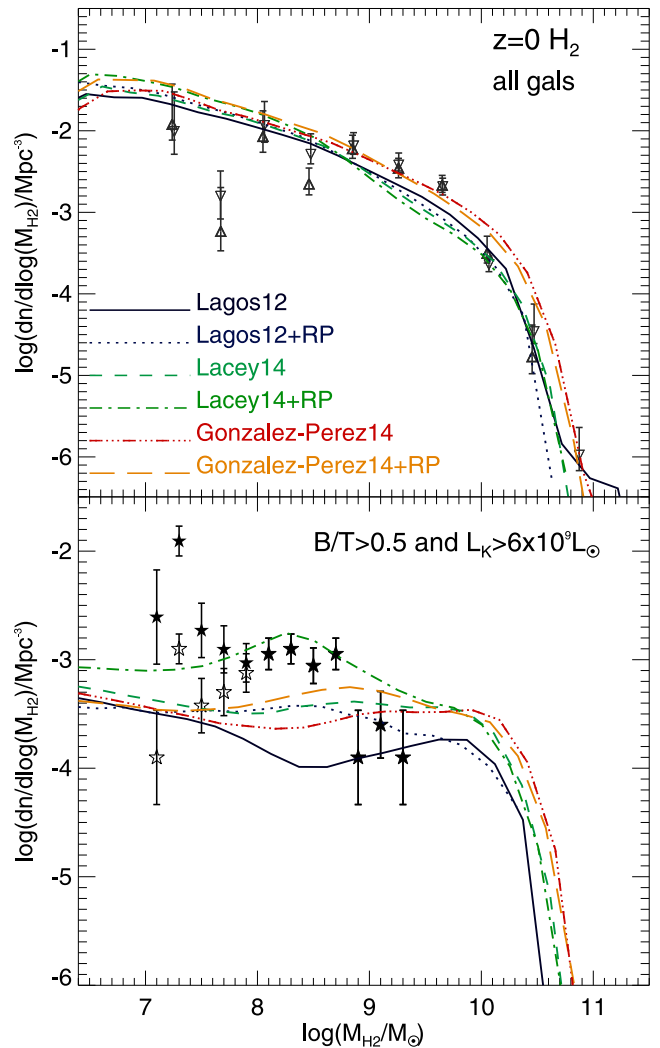


Figure 7. Top panel: the H₂ mass function at $z = 0$ for all galaxies in the Lagos12, Lacey14 and Gonzalez-Perez14 models and their variants including partial ram pressure stripping of the hot gas ('+RP'). Observations from the 60 μm (downwards pointing triangles) and B -band (triangles) samples of Keres et al. (2003) are also shown (see Section 2). Bottom panel: as in the top panel but for ETGs ($B/T > 0.5$) with $L_K > 6 \times 10^9 L_\odot$. Observations correspond to ETGs from the ATLAS^{3D} survey with (filled stars) and without (open stars) volume correction (see Section 2 for details).

agreement with the observations. The Lacey14+RP model provides the best description of the observed H₂ gas fractions. The physical reason for the higher number density of H₂ 'rich' ETGs in the models including partial ram pressure stripping of the hot gas is that the replenishment of the cold gas reservoir leads to an increase in the surface density of gas. Since the H I saturates at $\Sigma_{\text{H I}} \approx 10 M_\odot \text{pc}^{-2}$, due to H₂ self-shielding at higher densities, the effect of cold gas replenishment in the ISM has a stronger effect on the H₂ reservoir than on the H I.

The incorporation of partial ram pressure stripping of the hot gas brings the models into better agreement with the observed gas fractions of galaxies, and particularly of ETGs. This indicates that partial ram pressure of the hot gas is relevant in a wide range of environments. Note that do not include any description of the ram pressure stripping of the cold gas, which has been shown to take place in clusters through observations of the H I and H₂ contents of galaxies in the Virgo cluster (e.g. Cortese et al. 2011; Boselli et al.

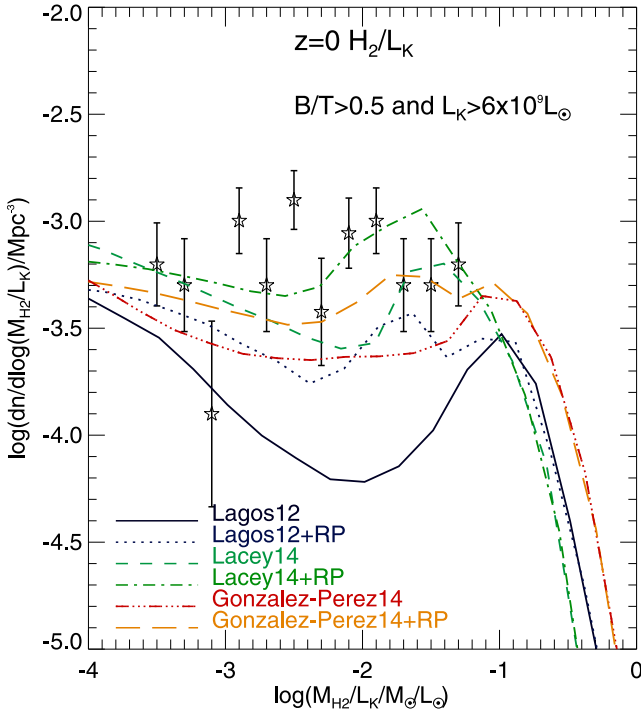


Figure 8. The distribution function of the ratio between the H_2 mass and the K -band luminosity for ETGs ($B/T > 0.5$) with $L_K > 6 \times 10^9 L_\odot$, for the models as labelled. Observations correspond to ETGs from the ATLAS^{3D} survey (see Section 2 for details of the observational data set).

2014b). However, no deficiency of H_1 or H_2 has been observed in galaxies in environments other than clusters. Tecce et al. (2010) using galaxy formation models show that ram pressure stripping of the cold gas is relevant only in haloes with $M_{\text{halo}} > 3 \times 10^{14} h^{-1} M_\odot$. Since most of the galaxies in the ATLAS^{3D} and HIPASS surveys are not cluster galaxies, we expect the effect of the ram pressure stripping of the cold gas to be insignificant in our analysis.

The study of the neutral gas content of ETGs offers independent constraints on the modelling of the ram pressure stripping of the hot gas. Font et al. (2008) used the fraction of passive to active galaxies as the main constraint on the satellite’s hot gas treatment. Here we propose that the exact levels of activity or cold gas content of galaxies that are classified as passive offer new, independent constraints.

The Lacey14+RP model agrees the best with the observations of H_1 and H_2 in different galaxy populations. This is the first time such a successful model is presented. Serra et al. (2014) use a sample of hydrodynamical simulations of galaxies and find that the simulations have difficulties reproducing the H_1 masses of ETGs. This may partially be due to the small sample of simulated ETGs analysed by Serra et al. (50 in total). Here, by taking the full simulated galaxy population, we can make a statistically robust comparison with the observed ETG population.

4.4 Expectations for the evolution of the H_1 and H_2 mass functions

In the near future, the Australian Square Kilometre Array and the South African Karoo Array Telescope will be able to trace the evolution of the H_1 gas content of ETGs towards redshifts higher than $z = 0.1$, while current millimetre telescopes, such as the Plateau de Bure Interferometer and the Atacama Large Millimetre Array

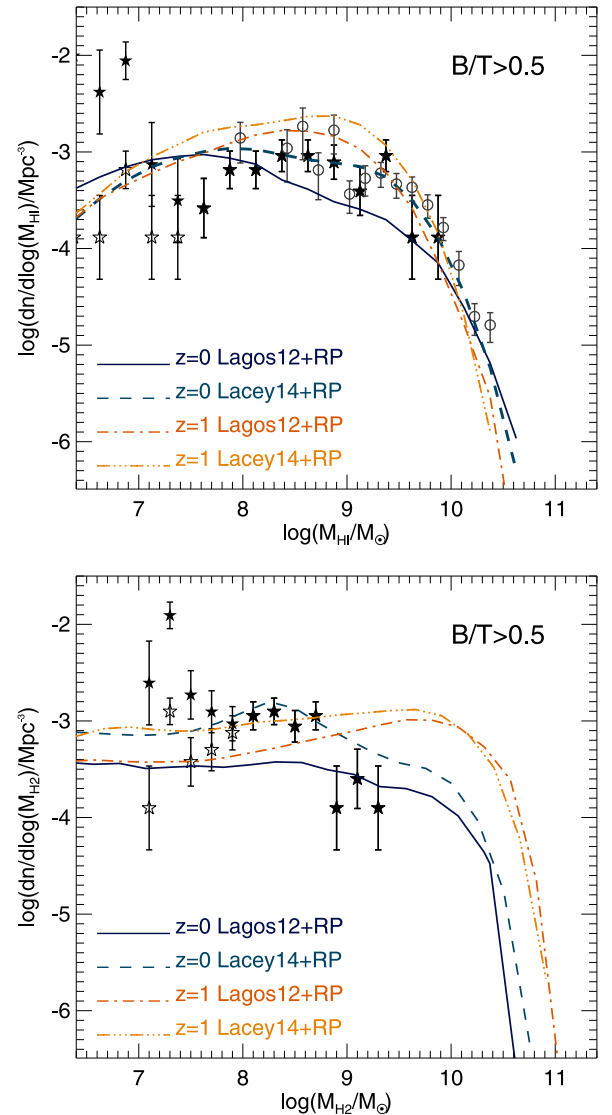


Figure 9. Top panel: the H_1 mass function of ETGs ($B/T > 0.5$) at $z = 0$ and 1 for the Lagos12+RP and Lacey14+RP models. Observations are as in the bottom panel of Fig. 5. Bottom panel: as in the top panel, but for H_2 . Observations are as in the bottom panel of Fig. 7.

(ALMA), can already trace H_2 in ETGs. To provide insights into the expected redshift evolution of the H_1 and H_2 mass functions of ETGs, we show in Fig. 9 the predictions for the mass functions at $z = 0$ and 1 for the Lagos12+RP and Lacey14+RP models, which give the lowest and highest number densities of ETGs (see Figs 5 and 7).

The top panel of Fig. 9 shows that both models predict an ETG H_1 mass function weakly evolving with redshift. On the contrary, a large increase in the number density of ETGs with large H_2 masses, $M_{H_2} > 10^{10} M_\odot$, from $z = 0$ to 1 is predicted by both models (bottom panel of Fig. 9). This is driven by the predicted increase of the H_2/H_1 mass ratio as well as an increase in the overall gas content of ETGs with increasing redshift. L11 and Lagos et al. (2014) present detailed studies which unveil the physics behind this evolution. In short, this is due to a combination of higher gas contents and more compact galaxies at high redshift, which increase the hydrostatic pressure in galaxies, and therefore the H_2/H_1 mass ratio.

5 THE MORPHOLOGICAL TRANSFORMATION AND QUENCHING OF GALAXIES

A key aspect in the analysis performed here is the morphological selection of ETGs. Observationally, morphologies are derived from visual inspection of optical images of galaxies (see e.g. Cappellari et al. 2011 for the ATLAS^{3D} survey sample selection). This means that among the galaxies selected as early type there could be contamination by edge-on spirals that are gas poor, and that therefore would appear red with no spiral arms. Following this argument, Khochfar et al. (2011) suggested that to select ETGs in the models that are comparable to the observed ones one needs to include truly bulge dominated objects, which are selected by their bulge-to-total mass ratio, and late-type galaxies, which are gas poor. Khochfar et al. suggest the following selection to select early-type looking galaxies: $B/T > 0.5$ and $B/T < 0.5$ with gas fractions < 0.1 . Note that of the latter subpopulation, only the fraction that are edge-on oriented will be confused as early types, and therefore only a small fraction will contribute to the ETG population. We assume random inclinations for late-type galaxies, select those that have inclination angles $> 45^\circ$ and add them to the sample of galaxies with $B/T > 0.5$. The H I mass function and gas fractions of ETGs obtained using this selection criterion are shown in Figs 10 and 11, respectively, for the Lacey14+RP model. The focus on the latter model as it predicts H I and H₂ masses of ETGs in best agreement with the observations. The effect of including the contamination from gas poor late-type galaxies is very small and therefore does not change the results presented earlier.

Another interesting question is how much the B/T threshold to select ETGs in the model affects our results. In order to answer this question, we show in Figs 10 and 11 different B/T thresholds

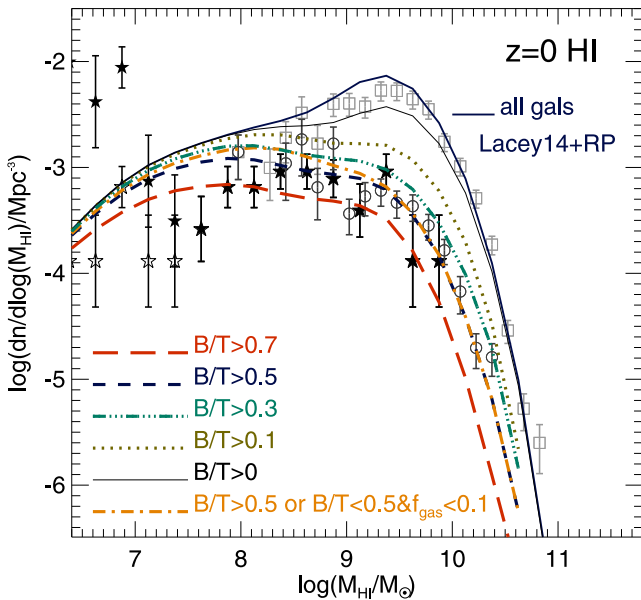


Figure 10. The H I mass function for galaxies with $L_K > 6 \times 10^9 L_\odot$ in the Lacey14+RP model. Galaxy populations with different bulge-to-total mass ratios are shown: all galaxies (solid thick line), $B/T > 0$ (solid thin line), $B/T > 0.1$ (dotted line), $B/T > 0.3$ (triple dot-dashed line), $B/T > 0.5$ (dashed line) and $B/T > 0.7$ (long dashed line). We also show the H I mass function for ETGs selected in an alternative way: in addition to those with $B/T > 0.5$, galaxies with $B/T < 0.5$ that are gas poor, $M_{\text{gas}}/M_{\text{stellar}} = f_{\text{gas}} < 0.1$, can also appear as early types. The latter values are consistent with those presented in Khochfar et al. (2011).

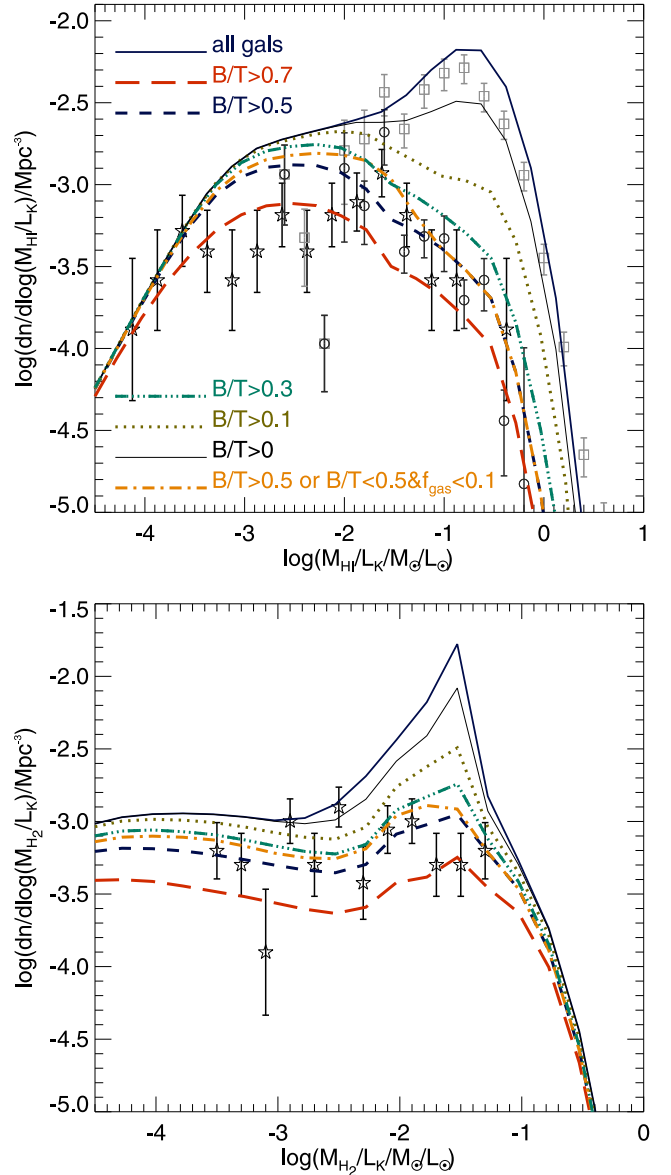


Figure 11. Same as Fig. 10 but for the H I gas fraction (top panel) and H₂ gas fraction (bottom panel) distribution functions.

to select ETGs in the Lacey14+RP model. B/T thresholds lower than 0.5 have the expected impact of increasing the number density of galaxies compared to the canonical value of 0.5, particularly at $M_{\text{HI}} > 10^7 M_\odot$, $M_{\text{HI}}/L_K > 5 \times 10^{-3} M_\odot/L_\odot$ and throughout the whole H₂ gas fraction range. This increase is of a factor of 2 for $B/T = 0.3$ and 7 for $B/T = 0.1$. In addition, about 40 per cent of the galaxies with $L_K > 6 \times 10^9 L_\odot$ are pure discs (no bulges; see difference between thick and thin solid lines in Figs 10 and 11), which are mainly located around the peaks of the H I and H₂ gas fraction distributions of all galaxies with $L_K > 6 \times 10^9 L_\odot$ (see Fig. 11). This shows that the development of a small bulge is connected with an important gas depletion in galaxies. Spheroids in the model are formed when galaxies undergo a starburst, either driven by a galaxy merger or a global disc instability. The fact that these galaxies remain bulge dominated is because large discs fail to regrow after the formation of the spheroid. To adopt a higher B/T threshold has the expected effect of lowering the number density of ETGs.

The number density of galaxies with low H I content is only weakly dependent on the B/T threshold. This is due to a connection between low gas fractions and large B/T ; i.e. if a galaxy has a large bulge fraction it will also be gas poor. This has been observed in the ATLAS^{3D} (Cappellari et al. 2013), where ETGs with the largest velocity dispersions (a tracer of bulge fraction) have the lowest gas fractions. This implies that the modelling of the morphological transformation in GALFORM is able to capture the processes that lead to the relation between bulge fraction and gas depletion. Thresholds of B/T used to select ETGs which are in the range 0.3–0.6 produce similar results and therefore does not affect the analysis presented here.

The key question is why do galaxies fail to rebuild large discs and remain spheroid dominated with low gas fractions? First of all, ETGs have relatively old stellar populations as we will see in Section 6, and therefore we need to understand why after the formation of bulges, galaxy discs fail to reaccrete significant quantities of gas to keep forming stars at the level of the main sequence of galaxies in the $\text{SFR}-M_{\text{stellar}}$ plane. This question is then related to the physical mechanisms quenching SF in ETGs. In order to understand why large discs fail to regrow, we first look into the nature of galaxies with $B/T > 0$. The top panel of Fig. 12 shows the H I mass function of galaxies with $B/T > 0$ and $L_K > 6 \times 10^9 L_\odot$ separated into centrals and satellites. The satellite galaxy population makes up most of the tail of low H I masses, $M_{\text{H I}} < 5 \times 10^8 M_\odot$. These galaxies have little cold gas replenishment after they become satellites as they continuously lose part of their hot gas reservoir due to the continuous action of ram pressure stripping. This has the consequence of lowering the cooling rates. However, satellite galaxies do preserve some neutral gas reservoir; i.e. they hardly completely deplete their gas content. The mechanism for this is connected to the dependence of the SF time-scale with the gas surface density. Low gas surface densities produce low $\text{H}_2/\text{H I}$ ratios and low SFRs. As the gas is being depleted, the SF time-scale becomes longer and longer, allowing satellite galaxies to retain their gas reservoir. This mechanism drives the population of satellite galaxies with low gas fractions. This is consistent with observations, where there is a non-negligible fraction of ETGs with H_2 and/or H I contents detected in high-mass groups or clusters (e.g. Young et al. 2011).

In the case of central galaxies, there is a population with H I masses $M_{\text{H I}} > 10^9 M_\odot$ and H I gas fractions of $\approx 0.15 M_\odot/L_\odot$ that have cooling rates large enough to replenish their cold gas contents. The tail of central galaxies with low H I masses, i.e. $M_{\text{H I}} < 10^9 M_\odot$, fail to replenish their gas reservoirs and rebuild a new disc due to the effect of AGN heating their hot halo (see dashed line in the top panel of Fig. 12). In GALFORM, AGN feedback acts in haloes where the cooling time is larger than the free fall time at the cooling radius (‘hot accretion’ mode; Fanidakis et al. 2012). In these haloes, the AGN power is examined and if it is greater than the cooling luminosity, the cooling flows are switched off (see Bower et al. 2006). This means that in central galaxies under the action of AGN feedback, there will be no further gas accretion on to the galaxy, driving a low H I and H_2 gas contents. Note that, in the model, the close connection between bulge fraction and gas depletion naturally arises in galaxies where AGN feedback operates. This is because the black hole grows together with the bulge (Fanidakis et al. 2012). The consequence of this is that large black hole masses, hosted by large bulges, are capable of large mechanical luminosities which can more easily affect their hot halo. These large bulges are also connected to large bulge-to-total stellar mass ratios due to the impeded disc regrowth.

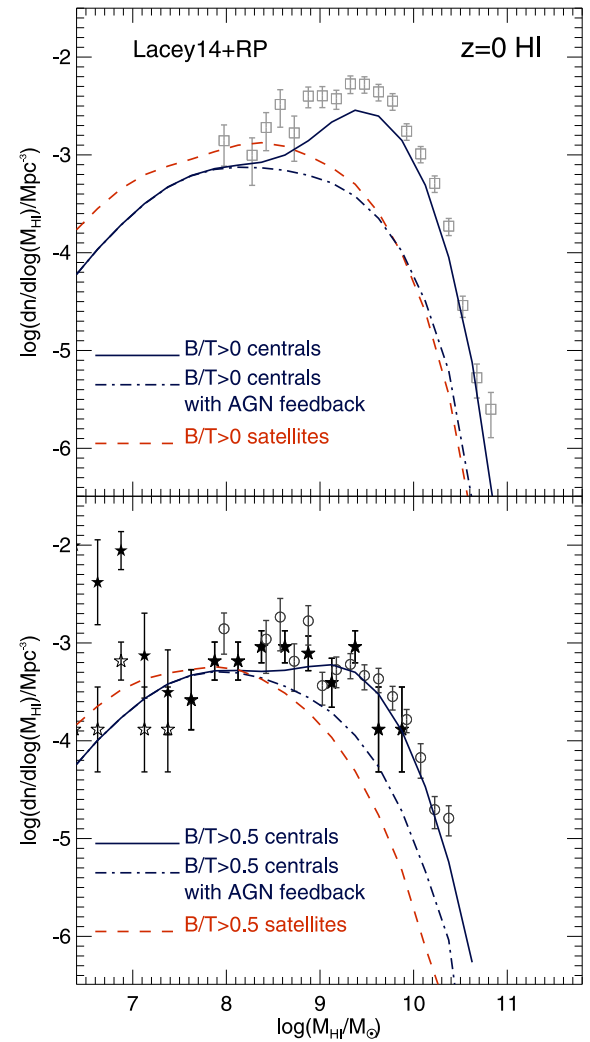


Figure 12. Top panel: the H I mass function for galaxies with $B/T > 0$ and $L_K > 6 \times 10^9 L_\odot$ separated into central (solid line) and satellite galaxies (dashed line) in the Lacey14+RP model. The subpopulation of central galaxies with AGN heating the hot halo of galaxies is also shown as dot-dashed line. Observations correspond to the analysis of HIPASS galaxies with $L_K > 6 \times 10^9 L_\odot$. Bottom panel: same as in the top panel but for galaxies with $B/T > 0.5$. Observations here correspond to the ATLAS^{3D} and HIPASS, described in Section 2.

We show the H I gas fraction as a function of B/T in Fig. 13 with the background colour scheme showing the mean bulge mass in 2D bins of H I gas fraction and B/T . There is a clear anticorrelation between the H I gas fraction and B/T . In addition, at a fixed B/T there is a trend of increasing H I gas fraction with decreasing bulge mass. The latter is related to the stronger AGN feedback in higher mass bulges that lead to gas depletion. Recent high-redshift observations show evidence for the strong connection between quenching and bulge mass (Lang et al. 2014). Lang et al. point to the bulge mass as a fundamental property related to SF quenching, rather than the bulge fraction, which in our model is also understood due to the effect of AGN feedback.

Focusing on the population of galaxies with $B/T > 0.5$ (bottom panel of Fig. 12), one can find similar trends. Most of the galaxies with $M_{\text{H I}} < 10^8 M_\odot$ correspond to satellite galaxies, while the central galaxy population is responsible for the massive end of the H I mass function. The tail of central galaxies with $M_{\text{H I}} < 5 \times 10^8 M_\odot$

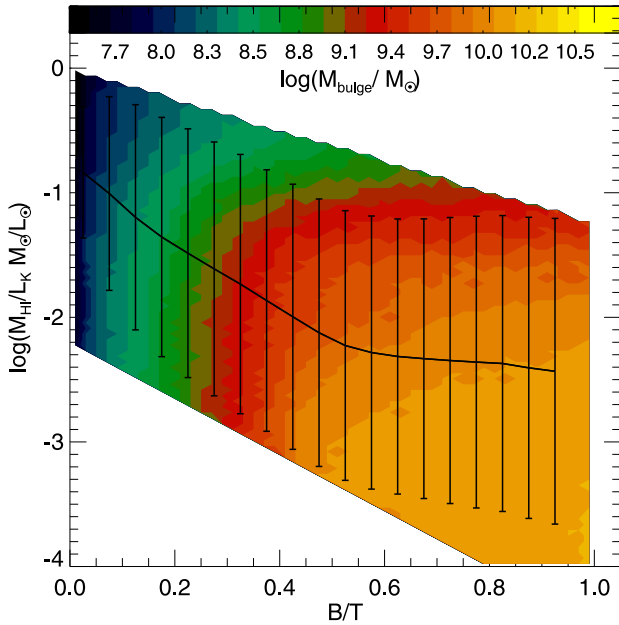


Figure 13. H I gas fraction as a function of the bulge-to-total stellar mass for galaxies in the Lacey14+RP model with $L_K > 6 \times 10^9 L_\odot$. The solid lines and error bars represent the median and 10 and 90 percentiles of the distribution. In colours we show the mean bulge mass in 2D bins of H I gas fraction and B/T , in an arbitrary region encompassing the error bars. The mean bulge masses are as labelled in the bar at the top of the figure. The wiggle features in the coloured region are an artefact of the binning.

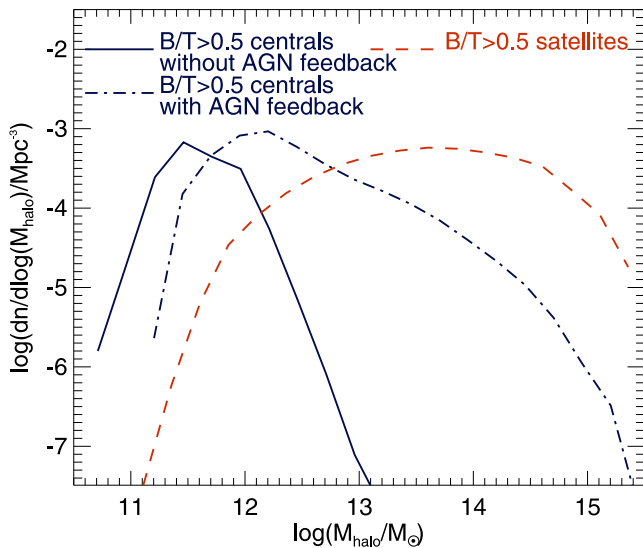


Figure 14. Host halo mass distribution for ETGs ($B/T > 0.5$) in the Lacey14+RP model that are satellites (dashed line), centrals under the action of AGN feedback (dot-dashed line) and centrals without AGN feedback (solid line).

is under the influence of AGN feedback which explains the low H I contents. There is a subpopulation of ETGs with large H I masses and gas fractions, $M_{H I} > 10^9 M_\odot$ and $M_{H I}/L_K \gtrsim 0.1 M_\odot/L_\odot$, respectively. The latter population is also the one living in low-mass haloes. Fig. 14 shows the distribution of masses of the haloes hosting ETGs in the Lacey14+RP model. ETGs with the highest H I gas contents, which correspond to centrals without AGN feedback (see Fig. 12), are hosted by low-mass haloes, with a median host halo

mass of $3 \times 10^{11} M_\odot$; centrals with AGN feedback on are hosted by higher mass haloes, with median masses of $2 \times 10^{12} M_\odot$, while satellites are distributed throughout a wider range of halo masses, with a median mass of $3 \times 10^{13} M_\odot$. In the observations similar trends are seen. Young et al. (2011) show that the ETGs with the highest H_2 masses also reside in the lowest density environments, which in our model correspond to ETGs that are central galaxies and are not undergoing AGN feedback. Similarly, Serra et al. (2012) reported that the H I mass as well as the ratio $M_{H I}/L_K$ decrease with increasing density of the environment. We find that these trends are simply a reflection of the halo masses in which either AGN feedback switches on or environmental quenching acts more effectively (where ram pressure stripping of the hot gas is effective enough to remove significant amounts of hot gas).

Our conclusion is that most ETGs with low gas fractions correspond to satellite galaxies, which due to environmental quenching (partial ram pressure stripping of the hot gas), are unable to regrow a significant disc. The rest of the ETGs with low gas fractions are central galaxies under the action of AGN feedback. The break at the high end of the H I gas fraction distribution, traced by HIPASS, is due to a population of ETGs with large H I masses that are still forming stars at the same level of spiral galaxies.

6 THE COMPETING SOURCES OF THE NEUTRAL GAS CONTENT OF EARLY-TYPE GALAXIES

One the aims of this paper is to answer the question: what is the source of the H I and H_2 gas contents in ETGs? We follow all the gas sources throughout the SF history of galaxies identified as ETGs today: radiative cooling from hot haloes (which we refer to as ‘cooling’), mass-loss from old stars (which we refer to as ‘recycling’) and galaxy mergers (see Appendix B for details of how we do this). Note that in our model the recycled mass from old stars does not incorporate into the hot halo and therefore it can be distinguished from the gas coming from cooling.

We study the sources of the gas in ETGs in two of the models shown in Section 4 to find general trends present in the different models as well as variations between them. We focus here on the Lagos12+RP and Lacey14+RP models as, after including partial ram pressure stripping of the hot gas, they give the lowest and highest number densities of ETGs, respectively.

We first estimate the fraction of ETGs that have neutral gas masses ($M_{H I} + M_{H_2}$) $> 10^7 M_\odot$. We find that 58 per cent of ETGs with K -band luminosities $L_K > 6 \times 10^9 L_\odot$ in the Lacey14+RP model and 65 per cent in the Lagos12+RP model have $M_{H I} + M_{H_2} > 10^7 M_\odot$. We analyse these subsamples of ETGs and estimate the fraction of the ETGs with neutral gas contents supplied mainly by mergers, recycling or cooling (summarized in Table 1). Most ETGs have neutral gas contents supplied predominantly by cooling. A smaller percentage have neutral gas contents supplied mainly by mergers (≈ 8 per cent for the Lacey14+RP and 17 per cent for the Lagos12+RP model) or by recycling (≈ 1.5 per cent for the Lacey14+RP and 0.8 per cent for the Lagos12+RP model). The latter percentages are not sensitive to the K -band luminosity or stellar mass of ETGs. However, they are sensitive to the current neutral gas content and halo mass. In order to get an insight into the properties of ETGs that have different gas suppliers, we show in Table 1 the fraction of ETGs that had a minor merger in the last 1 Gyr, the fraction of these that increased the neutral gas content by at least a factor of 2, and the fraction of ETGs that had a starburst

Table 1. The percentage of ETGs with $L_K > 6 \times 10^9 L_\odot$ in the Lacey14+RP and Lagos12+RP models under different selection criteria, which we group in four categories: neutral gas content, neutral gas sources, mergers and disc instability.

Selection	ETGs ($L_K > 6 \times 10^9 L_\odot$) Lacey14+RP	ETGs ($L_K > 6 \times 10^9 L_\odot$) Lagos12+RP
Neutral gas content.		
ETGs with neutral gas content $M_{H1} + M_{H2} > 10^7 M_\odot$	58 per cent	65 per cent
Neutral gas sources of the sample of ETGs with $M_{H1} + M_{H2} > 10^7 M_\odot$.		
ETGs with current gas content dominated by mergers	7.5 per cent	17 per cent
ETGs with current gas content dominated by recycling	1.5 per cent	0.8 per cent
ETGs with current gas content dominated by cooling	91 per cent	82 per cent
Mergers of ETGs with $M_{H1} + M_{H2} > 10^7 M_\odot$.		
ETGs that had a merger in the last 1 Gyr	11 per cent	25 per cent
ETGs that had a merger-driven starburst in the last 1 Gyr	1 per cent	1 per cent
Mergers in ETGs that took place in $M_{\text{halo}} < 10^{14} M_\odot h^{-1}$	95 per cent	94 per cent
Mergers in ETGs that increased the neutral gas content by a factor of >2	69 per cent	66 per cent
Disc instabilities on ETGs with $M_{H1} + M_{H2} > 10^7 M_\odot$.		
ETGs that had a disc instability in the last 1 Gyr	4 per cent	2 per cent

driven by either a galaxy merger or a disc instability in the last 1 Gyr. The main conclusions we draw from Table 1 are as follows.

(i) Minor mergers took place in a tenth of the ETG population with $L_K > 6 \times 10^9 L_\odot$ and $M_{H1} + M_{H2} > 10^7 M_\odot$ in the last 1 Gyr in the Lacey14+RP model. Only 10 per cent of these resulted in a starburst, although none of these starbursts made a significant contribution to the stellar mass build-up (mass weighted stellar ages are usually >7 Gyr). The large percentage of minor mergers that did not drive starbursts in the last 1 Gyr is due to the very low mass ratios between the accreted galaxy and the ETG, which is on average ≈ 0.05 . Such small mass ratios are not considered to drive starbursts in GALFORM unless they are very gas rich (see Section 3.3.2). For the Lagos12+RP model the fraction of galaxies that had a minor merger in the last 1 Gyr is higher, 25 per cent, with a smaller fraction (0.05) of these driving starbursts. The mass ratios of these minor merger events are also very small, which explains the small percentage of merger driven starbursts.

(ii) ≈ 68 per cent of minor mergers in ETGs in both the Lacey14+RP model and Lagos12+RP models increased the neutral gas content significantly (at least by a factor of 2). The frequency of minor mergers times the percentage of those which increased the gas content significantly explains the percentages of ETGs with neutral gas contents supplied by minor merger accretion in the models.

(iii) Of these minor merger accretion episodes, ≈ 95 per cent in both the Lacey14+RP and Lagos12+RP models took place in haloes with masses $< 10^{14} M_\odot$, which implies that this source of neutral gas accretion is negligible in cluster environments. This agrees with the observations of Davis et al. (2011).

(iv) There is only a small percentage, ≈ 3 per cent, of ETGs with $L_K > 6 \times 10^9 L_\odot$ and $M_{H1} + M_{H2} > 10^7 M_\odot$, that had a starburst driven by disc instabilities in the last 1 Gyr. These galaxies have very small discs (usually the bulge half-mass radius is larger than the disc half-mass radius) and $B/T \gtrsim 0.9$. In the model, we use the properties of galaxy discs to determine whether they are unstable under small perturbations (see equation 9). In reality, one would expect that such large bulges dominate over the gravity of the disc, stabilizing it. Martig et al. (2013) show this to happen in hydrodynamical simulations of individual galaxies: self-gravity of the disc is reduced when it is embedded in a bulge, preventing

gas fragmentation. This results in an overall lower efficiency of SF in ETGs. Our model does not capture this physics showing that it needs improvement to account for these cases.

The largest differences found between the two models is in the fraction of ETGs with $M_{H1} + M_{H2} > 10^7 M_\odot$ and the percentage of those with current neutral gas contents dominated by merger accretion. These differences are due to a combination of the different threshold values to examine disc instabilities and the different dynamical friction prescriptions used by the models (Section 3.3.2). In the Lacey14+RP model more disc instabilities take place due to the higher ϵ_{disc} , which drives a more rapid gas exhaustion in the galaxies that are prone to disc instabilities. Many of the galaxies that go through disc instabilities in the Lacey14+RP model do not do so in the Lagos12+RP model due to lower value of ϵ_{disc} . In the case of the dynamical friction, the prescription used by the Lagos12+RP model produces more minor mergers at lower redshifts than the prescription used in the Lacey14+RP model.

Another interesting dichotomy between cluster environments and lower mass haloes is that the amount of ETGs that have neutral gas contents supplied mainly by mass-loss from intermediate- and low-mass stars increases with increasing halo mass. This is shown in Fig. 15 for the models Lagos12+RP and Lacey14+RP. In cluster environments, we expect 7 per cent of ETGs to have neutral gas contents mainly supplied by recycling in the Lacey14+RP model and 3 per cent in the Lagos12+RP model, while that percentage drops dramatically at halo masses $M_{\text{halo}} < 10^{14} M_\odot$. Note that both models show a minimum contribution from recycling at $M_{\text{halo}} \approx 2 \times 10^{12} M_\odot$, which is connected to the halo mass in which feedback, either by stellar feedback in the lower halo masses or AGN in the higher halo masses, is the least effective. At this halo mass we have the most efficient accretion of newly cooled gas, which minimizes the contribution from mass-loss from old stars.

The higher frequency of ETGs with neutral gas contents dominated by internal origin in higher mass haloes takes place together with the aging of the bulge. In Fig. 16, we show the look-back time to the last time ETGs had a bulge-to-total stellar mass ratio < 0.5 (a late-type morphology), as a function of the current halo and stellar mass. There is a positive correlation between the current halo and stellar mass and the last time these ETGs were late-type: ETGs

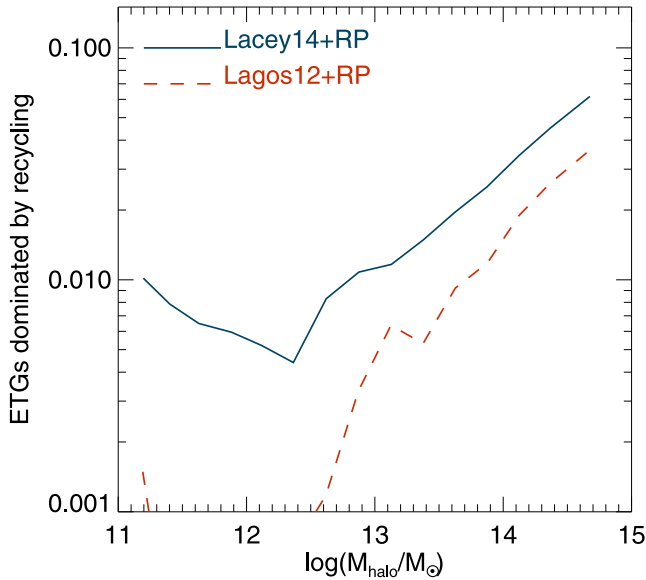


Figure 15. Fraction of ETGs ($B/T > 0.5$) with $L_K > 6 \times 10^9 L_\odot$ and $M_{H1} + M_{H2} > 10^7 M_\odot$ at $z = 0$ in the Lacey14+RP and Lagos12+RP models that have most of their cold gas supplied by mass-loss from intermediate- and low-mass stars, as a function of the host halo mass.

residing in high-mass halo have had an early-type morphology for longer time than those residing in lower mass haloes, and similarly, the most massive ETGs have a tendency of having had an early-type morphology for longer than lower mass ETGs. Note, however, that the dispersion around these relations is very high, showing that there is no single path for the formation of spheroids and that the SF history of ETGs can be quite complex (see also Naab et al. 2013). The Lagos12+RP model predicts systematically lower times at halo masses $M_{\text{halo}} < 10^{13} M_\odot$ than the Lacey14+RP model, which is due to the higher disc instability threshold in the latter model, which drove ETGs to undergo disc instabilities on average earlier than in the Lagos12+RP model.

7 CONCLUSIONS

We have studied the current neutral gas content of ETGs and its origin in the context of hierarchical galaxy formation. We first use the HIPASS and ATLAS^{3D} surveys to quantify the H I and H₂ gas fraction distribution functions for the overall galaxy population and ETGs observationally. We then explored the predictions for the neutral gas content of galaxies in three flavours of the GALFORM semi-analytic model of galaxy formation, the Lagos12, Gonzalez-Perez14 and Lacey14 models and performed a thorough comparison with observations.

For quiescent SF, the three models use the pressure-based SF law of Blitz & Rosolowsky (2006), in which the ratio between the surface density of H₂ and H I is derived from the radial profile of the hydrostatic pressure of the disc. The SFR is then calculated from the surface density of H₂. The advantage of this SF law is that the atomic and molecular gas phases of the ISM of galaxies are explicitly distinguished, which allows us to compare the predictions for the H I and H₂ contents of ETGs directly with observations. Other physical processes in the three models are different, such as the IMF adopted and the strength of both the SNe and the AGN feedback, as well as the cosmological parameters. We also tested the importance of the modelling of the processing of the hot gas of galaxies

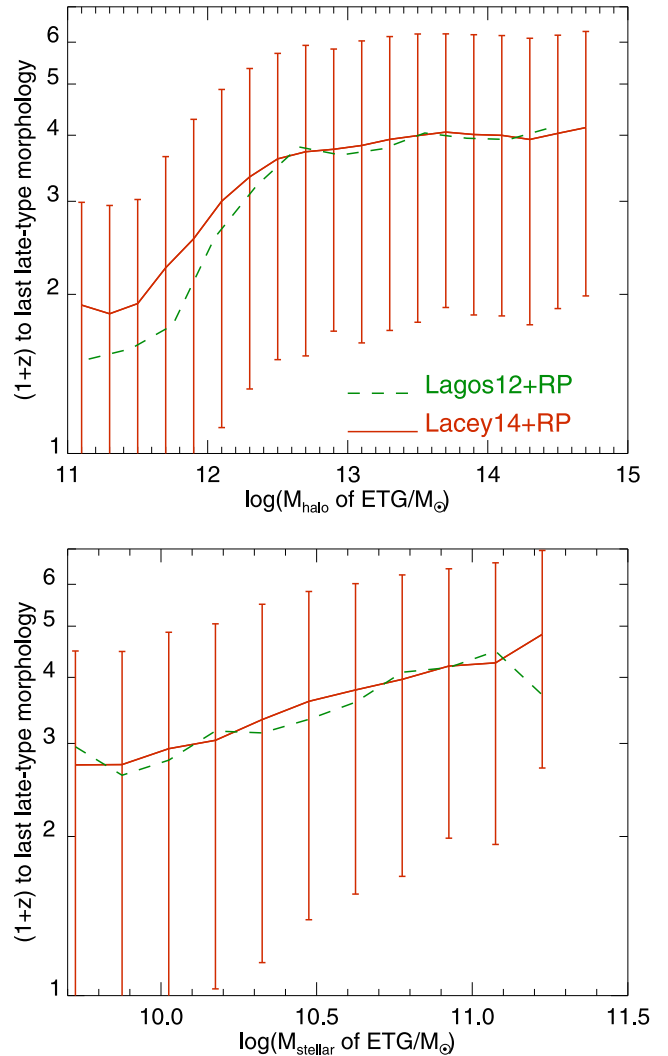


Figure 16. Top panel: look-back time to the last time ETGs ($B/T > 0.5$) selected as $L_K > 6 \times 10^9 L_\odot$, $M_{H1} + M_{H2} > 10^7 M_\odot$ at $z = 0$ had a $B/T < 0.5$ (late-type morphology), expressed in $1 + z$, as a function of the current host halo mass, for the Lacey14+RP and Lagos12+RP models. Lines with error bars show the median and 10 and 90 percentiles of the distributions, respectively. For clarity, error bars are only shown for the Lacey14+RP model, but the ones in the Lagos12+RP model are of similar magnitude. Bottom panel: as in the top panel but here the look-back time to the last time ETGs had a $B/T < 0.5$ is shown as a function of stellar mass.

once they become satellites. The original GALFORM flavours include a strangulation treatment of the hot gas: once galaxies become satellites they immediately lose all of their hot gas reservoir. We run the three GALFORM flavours with a different hot gas processing: the partial ram pressure stripping of the hot gas of satellite galaxies, which depends upon the orbit followed by the satellite galaxy, with cooling continuing on to the satellite galaxies.

Our conclusions are as follows.

- (i) The three flavours of GALFORM predict overall H I and H₂ mass functions in good agreement with the observations, regardless of the treatment of the hot gas of galaxies once they become satellites. However, when focusing exclusively on the ETG population, the inclusion of partial ram pressure stripping of the hot gas (as opposed to the strangulation scenario) results in the models predicting ETGs with higher contents of H I and H₂, improving the agreement

with the observations. This shows that the H I and H₂ gas contents of ETGs are a great test for the modelling of the hot gas stripping in simulations of galaxy formation. Moreover, the gas fraction distribution is a statistical measurement which is particularly good for placing constraints on models.

(ii) The presence of a bulge in galaxies is strongly correlated with depleted H I and H₂ gas contents in the three GALFORM models tested. This close correspondence between the bulge fraction and the depleted neutral gas contents in ETGs has been observed and here we provide a physical framework to understand it. We show that this is due to AGN feedback in central galaxies, and the environmental quenching due to partial ram pressure stripping of the hot gas in satellite galaxies. In the former, the black hole mass is correlated with the bulge mass, which implies that feedback can be stronger in larger bulges (as the Eddington luminosity increases with black hole mass). Galaxies experiencing AGN feedback do not accrete significant amounts of newly cooled gas, which impedes the regeneration of a prominent disc. In the case of satellites, the lower accretion rates due to depletion of the hot gas prevent the regrowth of a substantial disc, or drive the exhaustion of the gas in the disc, leaving a (close to) gas-free disc. There is a fraction of ETGs though that are neither experiencing AGN feedback, nor environmental quenching, that have normal H I and H₂ contents which are comparable to those obtained for late-type galaxies of the same mass.

(iii) We find that about ≈ 90 per cent of ETGs accreted most of their neutral gas from the hot halo through radiative cooling. A lower fraction have current H I and H₂ contents supplied by accretion from minor galaxy mergers (ranging from 8 to 17 per cent, depending on the model). An even smaller fraction (0.5–2 per cent) have their neutral gas content supplied mass-loss from intermediate- and low-mass stars. Interestingly, most of those galaxies are hosted by large mass haloes ($M_{\text{halo}} > 10^{14} M_{\odot}$; clusters of galaxies), while most of those dominated by minor merger accretion are in non-cluster environments ($M_{\text{halo}} < 10^{14} M_{\odot}$). We find that the source of the H I and H₂ gas in ETGs has strong consequences for the expected alignment between the gas disc and the stellar component, which we discuss in depth in paper II (Paper II).

(iv) We find a general trend of increasing look-back time to the last time ETGs were late-types ($B/T < 0.5$) with increasing host halo mass and stellar mass. However, these trends are characterized by a very large dispersion around the median, suggesting that the paths for the formation of ETGs of a given stellar mass are variable and non-self-similar. The latter is due to the stochastic nature of galaxy mergers and disc instabilities.

Our analysis shows the power of studying the gas contents of galaxies, and how subsamples of them are affected by different physical processes. In particular, our work points to the need for improved modelling of ram pressure stripping of the hot gas, which has an important effect in a wide range of environments. Although we show how the model we include, originally developed by Font et al. (2008), works well with the observational constraints we currently have, it may be too simplistic. For example, this model does not explicitly take into account the 3D position and velocity of galaxies in the simulation, which means that we do not consider their specific position in the halo to calculate the ram pressure stripping throughout its transit. This will be possible with high-resolution simulations, given that for such a detailed analysis it is necessary to resolve all haloes with few hundreds particles at least. In the future, we suggest that the study of the H I and H₂ gas contents of galaxies classified as ‘passive’ will provide stringent constraints on the details of the ram pressure stripping modelling.

ACKNOWLEDGEMENTS

We thank Martin Meyer, Diederik Kruijssen, Andreas Schruha and Paolo Serra for very motivating discussions. The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007–2013/) under grant agreement no. 229517 and the Science and Technology Facilities Council grant number ST/F001166/1. This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1, and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure. VGP acknowledges support from a European Research Council Starting Grant (DEGAS-259586).

REFERENCES

- 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., Baldry I. K., Nichol R., Miller C., Bower R., Glazebrook K., 2004, *ApJ*, 615, L101
- Baugh C. M., 2006, *Rep. Prog. Phys.*, 69, 3101
- Baugh C. M., Cole S., Frenk C. S., 1996, *MNRAS*, 283, 1361
- Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, *MNRAS*, 356, 1191
- Benson A. J., 2005, *MNRAS*, 358, 551
- Benson A. J., 2010, *Phys. Rep.*, 495, 33
- Benson A. J., Džanović D., Frenk C. S., Sharples R., 2007, *MNRAS*, 379, 841
- Bernardi M. et al., 2003, *AJ*, 125, 1817
- Bernardi M., Sheth R. K., Nichol R. C., Schneider D. P., Brinkmann J., 2005, *AJ*, 129, 61
- Bettoni D., Galletta G., García-Burillo S., 2003, *A&A*, 405, 5
- Bigiel F. et al., 2011, *ApJ*, 730, L13
- Blitz L., Rosolowsky E., 2006, *ApJ*, 650, 933 (BR)
- Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Lagos C., Saintonge A., 2014a, *A&A*, 564, A66
- Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Gavazzi G., Lagos C., Saintonge A., 2014b, *A&A*, 564, A67
- Bournaud F. et al., 2011, *ApJ*, 730, 4
- Bournaud F., Elmegreen B. G., Martig M., 2009, *ApJ*, 707, L1
- Bower R. G., Lucey J. R., Ellis R. S., 1992, *MNRAS*, 254, 601
- 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., Vernon I., Goldstein M., Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2010, *MNRAS*, 407, 2017
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Cappellari M. et al., 2011, *MNRAS*, 413, 813
- Cappellari M. et al., 2013, *MNRAS*, 432, 1862
- Catinella B. et al., 2010, *MNRAS*, 403, 683
- Cavaliere A., Fusco-Femiano R., 1976, *A&A*, 49, 137
- Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168
- Cortese L., Catinella B., Boissier S., Boselli A., Heinis S., 2011, *MNRAS*, 415, 1797
- Davis T. A. et al., 2011, *MNRAS*, 414, 968
- Davis T. A. et al., 2014, *MNRAS*, in press ([arXiv:1403.4850](https://arxiv.org/abs/1403.4850))
- De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, *MNRAS*, 366, 499
- de Vaucouleurs G., de Vaucouleurs A., Corwin H. G., Jr, Buta R. J., Paturel G., Fouque P., 1991, *Sky Telesc.*, 82, 621
- Doyle M. T. et al., 2005, *MNRAS*, 361, 34
- Efstathiou G., Lake G., Negroponte J., 1982, *MNRAS*, 199, 1069
- Eke V. R., Cole S., Frenk C. S., Patrick Henry J., 1998, *MNRAS*, 298, 1145
- Elmegreen B. G., 1989, *ApJ*, 338, 178

Elmegreen B. G., Burkert A., 2010, *ApJ*, 712, 294
 Fall S. M., Efstathiou G., 1980, *MNRAS*, 193, 189
 Fanidakis N. et al., 2012, *MNRAS*, 419, 2797
 Font A. S. et al., 2008, *MNRAS*, 389, 1619
 Gammie C. F., 2001, *ApJ*, 553, 174
 Genzel R. et al., 2010, *MNRAS*, 407, 2091
 Giovanelli R. et al., 2005, *AJ*, 130, 2598
 González J. E., Lacey C. G., Baugh C. M., Frenk C. S., Benson A. J., 2009, *MNRAS*, 397, 1254
 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 (Gonzalez-Perez14)
 Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, *ApJ*, 542, 710
 Hambly N. C. et al., 2001, *MNRAS*, 326, 1279
 Hopkins P. F. et al., 2010, *ApJ*, 715, 202
 Huang S., Haynes M. P., Giovanelli R., Brinchmann J., 2012, *ApJ*, 756, 113
 Jarrett T. H., Chester T., Cutri R., Schneider S., Skrutskie M., Huchra J. P., 2000, *AJ*, 119, 2498
 Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, *ApJ*, 675, 1095
 Kauffmann G., 1996, *MNRAS*, 281, 487
 Kauffmann G. et al., 2012, *MNRAS*, 422, 997
 Kaviraj S. et al., 2007, *ApJS*, 173, 619
 Kennicutt R. C., Jr, 1983, *ApJ*, 272, 54
 Keres D., Yun M. S., Young J. S., 2003, *ApJ*, 582, 659
 Khochfar S. et al., 2011, *MNRAS*, 417, 845
 Komatsu E. et al., 2011, *ApJS*, 192, 18
 Krumholz M., Burkert A., 2010, *ApJ*, 724, 895
 Lacey C., Guiderdoni B., Rocca-Volmerange B., Silk J., 1993, *ApJ*, 402, 15
 Lacey C. G., Baugh C. M., Frenk C. S., Silva L., Granato G. L., Bressan A., 2008, *MNRAS*, 385, 1155
 Lagos C. D. P., Lacey C. G., Baugh C. M., Bower R. G., Benson A. J., 2011a, *MNRAS*, 416, 1566
 Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011b, *MNRAS*, 418, 1649 (L11)
 Lagos C. d. P., Bayet E., Baugh C. M., Lacey C. G., Bell T. A., Fanidakis N., Geach J. E., 2012, *MNRAS*, 426, 2142 (Lagos12)
 Lagos C. d. P., Lacey C. G., Baugh C. M., 2013, *MNRAS*, 436, 1787
 Lagos C. D. P., Baugh C. M., Zwaan M. A., Lacey C. G., Gonzalez-Perez V., Power C., Swinbank A. M., van Kampen E., 2014, *MNRAS*, 440, 920
 Lang P. et al., 2014, *ApJ*, 788, 11
 Le Fèvre O. et al., 2000, *MNRAS*, 311, 565
 Lemonias J. J., Schiminovich D., Catinella B., Heckman T. M., Moran S. M., 2013, *ApJ*, 776, 74
 Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, *AJ*, 136, 2782
 Lintott C. J. et al., 2008, *MNRAS*, 389, 1179
 Lisenfeld U. et al., 2011, *A&A*, 534, A102
 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
 Malbon R. K., Baugh C. M., Frenk C. S., Lacey C. G., 2007, *MNRAS*, 382, 1394
 Maraston C., 2005, *MNRAS*, 362, 799
 Marigo P., 2001, *A&A*, 370, 194
 Martig M. et al., 2013, *MNRAS*, 432, 1914
 Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, *ApJ*, 723, 1359
 Meyer M. J., Zwaan M. A., Webster R. L., Schneider S., Staveley-Smith L., 2008, *MNRAS*, 391, 1712
 Meyer M. J. et al., 2004, *MNRAS*, 350, 1195
 Mitchell P. D., Lacey C. G., Baugh C. M., Cole S., 2013, *MNRAS*, 435, 87
 Mo H. J., Mao S., White S. D. M., 1998, *MNRAS*, 295, 319
 Naab T., Burkert A., 2003, *ApJ*, 597, 893
 Naab T. et al., 2013, *MNRAS*, in press ([arXiv:1311.0284](https://arxiv.org/abs/1311.0284))
 Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
 Parry O. H., Eke V. R., Frenk C. S., 2009, *MNRAS*, 396, 1972

Portinari L., Chiosi C., Bressan A., 1998, *A&A*, 334, 505
 Ruiz A. N. et al., 2013, *ApJ*, submitted ([arXiv:1310.7034](https://arxiv.org/abs/1310.7034))
 Saintonge A. et al., 2011, *MNRAS*, 415, 32
 Schiminovich D. et al., 2007, *ApJS*, 173, 315
 Schmidt M., 1968, *ApJ*, 151, 393
 Serra P. et al., 2012, *MNRAS*, 422, 1835
 Serra P. et al., 2014, *MNRAS*, in press ([arXiv:1401.3180](https://arxiv.org/abs/1401.3180))
 Simien F., de Vaucouleurs G., 1986, *ApJ*, 302, 564
 Smith M. W. L. et al., 2012, *ApJ*, 748, 123
 Spergel D. N. et al., 2003, *ApJS*, 148, 175
 Springel V. et al., 2005, *Nature*, 435, 629
 Strateva I. et al., 2001, *AJ*, 122, 1861
 Sutherland R. S., Dopita M. A., 1993, *ApJS*, 88, 253
 Tecce T. E., Cora S. A., Tissera P. B., Abadi M. G., Lagos C. D. P., 2010, *MNRAS*, 408, 2008
 Toomre A., 1977, in Tinsley B. M., Larson D., Campbell R. B. G., eds, *Evolution of Galaxies and Stellar Populations*. Yale Univ. Observatory, New Haven, p. 401
 Wang J. et al., 2014, *MNRAS*, 441, 2159
 Weinzierl T., Jogee S., Khochfar S., Burkert A., Kormendy J., 2009, *ApJ*, 696, 411
 Welch G. A., Sage L. J., Young L. M., 2010, *ApJ*, 725, 100
 White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
 Wuyts S. et al., 2011, *ApJ*, 742, 96
 Young J. S., Knezek P. M., 1989, *ApJ*, 347, L55
 Young L. M. et al., 2011, *MNRAS*, 414, 940
 Young L. M. et al., 2013, *MNRAS*, in press ([arXiv:1312.6318](https://arxiv.org/abs/1312.6318))
 Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, *MNRAS*, 359, L30

APPENDIX A: THE EFFECT OF VARYING THE PARTIAL RAM PRESSURE STRIPPING PARAMETERS

We show in Fig. A1 the $H\text{I}$ gas fractions for the Lacey14+RP model using three different values for ϵ_{strip} , which controls the fraction of the reheated mass from stellar feedback that was driven out from the galaxy after the first passage of the satellite that is affected by stripping. The reheated mass considered here is the one that sits

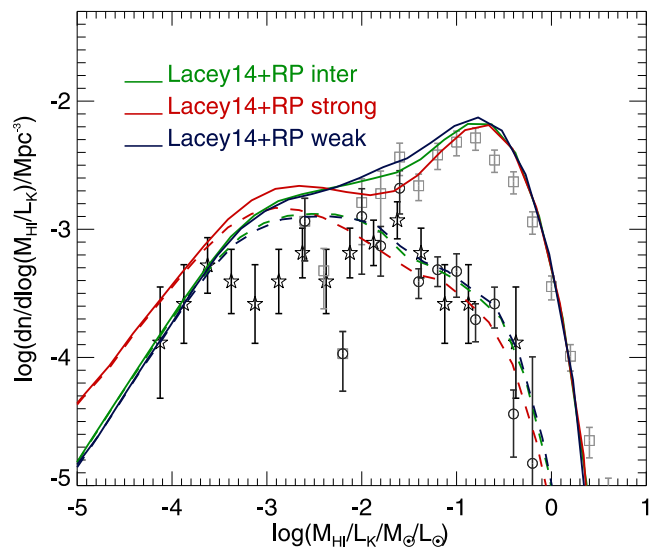


Figure A1. The $H\text{I}$ gas fraction, $M_{H\text{I}}/L_K$, distribution for all galaxies with $L_K > 6 \times 10^9 L_\odot$ (solid lines) and the subsample of ETGs ($B/T > 0.5$; dashed lines) for the Lacey14+RP model. We have adopted three different values for ϵ_{strip} : 0.01 (RP weak), 0.1 (RP inter) and 1 (RP strong). The observations are described in Section 2.

outside the stripping radius. We remind the reader that the latter is calculated at the pericentre of the satellite’s orbit (see Section 3.4). A value of $\epsilon_{\text{strip}} = 1$ implies that the reheated gas gets stripped in subsequent steps at the same rate as the hot gas of the galaxy when the satellite first passed through its pericentre. To adopt this value in the model leads to predictions that are very similar to the Lacey14 model predictions (including strangulation). This shows that $\epsilon_{\text{strip}} = 1$ drives very efficient hot gas stripping close to the fully efficient case. The model using the values $\epsilon_{\text{strip}} = 0.01$ and 0.1 produce very similar gas fractions. Both cases lead to predict number densities of ETGs with H I and H_2 gas fraction > 0.02 higher than both the standard Lacey14 and the version including partial ram pressure stripping with $\epsilon_{\text{strip}} = 1$. This higher number density is due to the higher rates of infalling cold gas, which replenish the ISM with newly cooled gas. We conclude that for values of $\epsilon_{\text{strip}} \lesssim 0.3$ the results presented in the paper do not considerably change, mainly because of a self-regulation of outflows and inflows in satellites: if a very small value of ϵ_{strip} is adopted, it will drive higher accretion rates of gas on to the disc, which will lead to higher SFRs, and therefore higher outflow rates. For values $\epsilon_{\text{strip}} \gtrsim 0.3$ the satellite’s hot gas reservoir is removed too quickly driving very little further gas accretion of newly cooled gas on to the galaxy disc. The latter has the effect of quenching SF in the satellite galaxy quickly, driving low H I and H_2 gas fractions.

APPENDIX B: COMPUTING THE CONTRIBUTION FROM DIFFERENT GAS SOURCES

The equations governing the mass exchange between stars (M_*), gas in the disc (M_g), ejected mass from the disc (M_{eject}) and the hot halo (M_{hot}) are as follows:

$$\dot{M}_* = (1 - R)\psi, \quad (\text{B1})$$

$$\dot{M}_g = \dot{M}_{\text{cool}} - (1 - R)\psi - \dot{M}_{\text{eject}} \quad (\text{B2})$$

$$\dot{M}_{\text{eject}} = \beta \psi \quad (\text{B3})$$

$$\dot{M}_{\text{hot}} = -\dot{M}_{\text{cool}} + \frac{M_{\text{eject}}}{\tau_{\text{rein}}}. \quad (\text{B4})$$

Here, ψ is the instantaneous SFR described in Section 3.1, \dot{M}_{cool} is the cooling rate described in Section 3.3.1, R is the recycled

fraction described in Section 3.2 and β is the efficiency of SNe feedback. The latter depends on the circular velocity as $\beta = (V/V_0)^{-\alpha_{\text{hot}}}$ (see Lagos, Lacey & Baugh 2013 for a discussion of the physical motivation of this parametrization). The parameters adopted in each model are $\alpha_{\text{hot}} = 3.2$ in the Lagos12, Gonzalez-Perez14 and Lacey14 models, $V_0 = 485 \text{ km s}^{-1}$ in Lagos12, $V_0 = 425 \text{ km s}^{-1}$ in Gonzalez-Perez14 and $V_0 = 320 \text{ km s}^{-1}$ in Lacey14. We define the changes in the quantities above in an arbitrary time step as ΔM_{cool} , ΔM_* , ΔM_g and ΔM_{eject} .

In order to follow the three sources of gas in galaxies (galaxy mergers, recycling and gas cooling), we define M_{merg} , M_{recycle} and M_{cooling} , and calculate them as follows. We first add the amount of cooled gas in the cooling component and the one following the total gas in the disc, M_g ,

$$\Delta M_{\text{cooling}} = \Delta M_{\text{cool}}. \quad (\text{B5})$$

We update the quantities M_{cooling} and M_g by adding $\Delta M_{\text{cooling}}$. From M_g , an amount ΔM_* of stars is formed and the amount of gas that is depleted from the ISM is ΔM_* . This mass is subtracted from the quantities M_{merg} , M_{recycle} and M_{cooling} , preserving their fractional contribution to M_g before stars formed. We then update M_g by subtracting ΔM_* .

After stars form, a fraction R is returned to the ISM, and we modify M_{recycle} and M_g by $\Delta M_{\text{recycle}}$ defined as

$$\Delta M_{\text{recycle}} = R \Delta M_*. \quad (\text{B6})$$

From the stars formed, an amount $\Delta M_{\text{eject}} = \beta M_*$ is ejected from the galaxy, and we subtract the amount of gas escaping the disc from M_{merg} , M_{recycle} and M_{cooling} , preserving their fractional contribution to M_g before the ejection of gas. We then update M_g by subtracting ΔM_{eject} . This procedure ensures that $M_{\text{merg}} + M_{\text{recycle}} + M_{\text{cooling}} \equiv M_g$.

During galaxy mergers, we add the amount of gas accreted by the central galaxy from the satellite to M_{merg} before SF takes place, and then we proceed to the set of equations (B1)–(B4) with the formalism described above.

This paper has been typeset from a \LaTeX file prepared by the author.