Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **491,** 5083–5100 (2020)
Advance Access publication 2019 November 22

doi:10.1093/mnras/stz3156

How feedback shapes galaxies: an analytic model

Jaime Salcido [1,2]★ Richard G. Bower [2] and Tom Theuns [2]

[1]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[2]*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham, DH1 3LE, UK*

Accepted 2019 November 7. Received 2019 November 5; in original form 2019 July 31

ABSTRACT

We introduce a simple analytic model of galaxy formation that links the growth of dark matter haloes in a cosmological background to the build-up of stellar mass within them. The model aims to identify the physical processes that drive the galaxy-halo co-evolution through cosmic time. The model restricts the role of baryonic astrophysics to setting the relation between galaxies and their haloes. Using this approach, galaxy properties can be directly predicted from the growth of their host dark matter haloes. We explore models in which the effective star formation efficiency within haloes is a function of mass (or virial temperature) and independent of time. Despite its simplicity, the model reproduces self-consistently the shape and evolution of the cosmic star formation rate density, the specific star formation rate of galaxies, and the galaxy stellar mass function, both at the present time and at high redshifts. By systematically varying the effective star formation efficiency in the model, we explore the emergence of the characteristic shape of the galaxy stellar mass function. The origin of the observed double Schechter function at low redshifts is naturally explained by two efficiency regimes in the stellar to halo mass relation, namely, a stellar feedback regulated stage, and a supermassive black hole regulated stage. By providing a set of analytic differential equations, the model can be easily extended and inverted, allowing the roles and impact of astrophysics and cosmology to be explored and understood.

Key words: galaxies: evolution – galaxies: formation – cosmology: theory.

1 INTRODUCTION

The co-evolution between galaxies and their haloes is perhaps one of the most fundamental aspects of every galaxy formation model. In the current paradigm of galaxy formation, every galaxy forms within a dark matter halo. However, understanding the relationship between a dark matter halo and the galaxies it hosts is not a trivial exercise due to our lack of detailed understanding of the complex baryonic process involved.

In a standard Lambda Cold Dark Matter (ΛCDM) cosmology, gravitationally bound dark matter structures build up hierarchically, by a combination of the smooth accretion of surrounding matter and continuous merging with smaller structures (White & Rees 1978; Qu et al. 2017). The formation and evolution of galaxies within these haloes is thought to be a highly self-regulated process, in which galaxies tend to evolve towards a quasi-equilibrium state where the gas outflow rate balances the difference between the gas inflow rate and the rate at which gas is locked up in stars and black holes (BHs) (e.g. White & Frenk 1991; Finlator & Davé 2008; Bouché et al. 2010; Schaye et al. 2010; Davé, Finlator & Oppenheimer 2012a; Bower et al. 2017). Consequently, galaxy formation is thought to be determined on the one hand by the formation and growth of dark matter haloes, which depends solely on the cosmological background, and on the other hand, by the regulation of the gas content in these haloes, which in turn depends on complex baryonic processes such as radiative cooling, stellar mass-loss, and feedback from stars and accreting BHs. This co-evolution process results in a tight correlation between the properties of galaxies and their dark matter haloes (see e.g. Wechsler & Tinker 2018 for a review).

A fundamental requirement for a successful galaxy formation model is to reproduce the relation between stellar mass and halo mass inferred from observations. However probing the dark matter distribution and its evolution represents an observational challenge. Direct observational probes include galaxy–galaxy lensing (e.g. Brainerd & Specian 2003; Hoekstra, Yee & Gladders 2004; Hudson et al. 2015) and the kinematics of satellite galaxies (e.g. Zaritsky et al. 1993; van den Bosch et al. 2004; Norberg, Frenk & Cole 2008). However, direct observation techniques are limited to low redshifts ($z < 1$), due to the difficulty of resolving individual distant galaxies. Indirect methods include, for example, comparing the abundance and clustering properties of galaxy samples with predictions from phenomenological halo models (e.g. Neyman & Scott 1952; Berlind & Weinberg 2002; Cooray & Sheth 2002; Cowley et al. 2018). This method however, depends heavily on the underlying modelling and assumptions, for

★ E-mail: J.SalcidoNegrete@ljmu.ac.uk

example, the bias with which haloes trace the underlying matter distribution.

From the theoretical point of view, the formation and evolution of dark matter haloes is largely considered a 'solved problem' (see however, van den Bosch et al. 2018). Using extremely accurate measurements of the density perturbations imprinted on to the cosmic microwave background radiation fluctuations as initial conditions (e.g. Planck Collaboration XIII 2016), many different groups have produced convergent results using large cosmological N-body simulations (e.g Springel et al. 2005; Klypin, Trujillo-Gomez & Primack 2011; Trujillo-Gomez et al. 2011; Angulo et al. 2012; Fosalba et al. 2015).

On the other hand, the complex physics of galaxy formation still has many open questions. Different approaches have been used to model the intricate baryonic physics of galaxy formation. The most widely used technique combines the evolution of dark matter with either a *semi-analytical* (e.g. Cole et al. 1994; Somerville et al. 2008; Henriques et al. 2015; Lacey et al. 2016) or *hydrodynamical* (e.g. Vogelsberger et al. 2014; Schaye et al. 2015; Davé, Thompson & Hopkins 2016; Dubois et al. 2016; Pillepich et al. 2018a) treatment of the baryonic processes involved. A key ingredient in both methods that has led us to a better understating of the physics of galaxy formation is the use of physically motivated models for feedback processes (see Somerville & Davé 2015; Naab & Ostriker 2017 for a comprehensive review).

An alternative approach known as *empirical modelling* takes the advantage of the vast number of observational data sets from large galaxy surveys and relate statistical galaxy scaling relations to the evolution of dark matter haloes without assuming strong physical priors (e.g. Behroozi, Wechsler & Conroy 2013; Moster, Naab & White 2013; Rodríguez-Puebla et al. 2016; Moster, Naab & White 2018; Behroozi et al. 2019; Grylls et al. 2019).

While all of these approaches have been very productive, the increasing complexity of the models and simulations make it difficult to pinpoint and understand the fundamental physics driving the results. For instance, in cosmological simulations, 'sub-grid' physics is implemented as micro phenomena that depend only on local gas properties from which macroscopic patterns emerge. However, it is hard to track down the link between what emerges and why (e.g. Bower et al. 2017). In this paper, we examine this issue in detail by adopting the opposite approach. We develop a fully analytic model of galaxy formation derived from a simple relation between the star formation rate (SFR) and halo growth rate that disentangles the role of cosmology from the role of astrophysics in the galaxy formation process. Our model restricts the role of baryonic astrophysics to setting the relation between galaxies and their haloes. With this simple relation, we can use an analytic approximation to the growth of dark matter haloes to predict galaxy properties. By providing a set of analytic equations, the model can be easily 'inverted' and allows for rapid experiments to be conducted, providing a powerful tool to explore the differential effects of baryonic physics, averaged over galaxy scales. Despite its simplicity, the model reproduces self-consistently the shape and evolution of the cosmic SFR density, the specific star formation rate (sSFR) of galaxies, and the galaxy stellar mass function (GSMF), both at the present time and at high redshift.

We validate our results by comparing to numerical hydrodynamic simulations from the EAGLE project. The EAGLE simulation suite[1] (Crain et al. 2015; Schaye et al. 2015) consists of a large number of cosmological hydrodynamical simulations that include different resolutions, simulated volumes, and physical models. These simulations use advanced smoothed particle hydrodynamics (SPH) and state-of-the-art subgrid models to capture the unresolved physics. A complete description of the code and physical parameters used can be found in Schaye et al. (2015). Here we compare to the EAGLE reference simulations that used a flat, ΛCDM cosmology with parameters ($\Omega_m = 0.307$, $\Omega_\Lambda = 0.693$, $h = 0.6777$, $\sigma_8 = 0.8288$, $n_s = 0.9611$) consistent with the Planck Collaboration XVI (2014) results. The calibration strategy of the EAGLE simulations is described in detail by Crain et al. (2015), who also presented additional simulations to demonstrate the effect of parameter variations.

The layout of this paper is as follows: In Section 2 we introduce the analytic model of galaxy formation. We present two models of the effective star formation efficiency: A time-independent efficiency which depends only on halo mass, and an efficiency that depends on the virial temperature of the halo. In Section 3, we explore the effect of the different efficiency parameters in the galaxy formation outputs. Namely, the cosmic SFR density, the sSFR of galaxies, and the GSMF. In Section 4 we compare the results from our model to different observational data sets. We also discuss the need for a time-evolving efficiency in order to reproduce the rapid evolution of the GSMF. We discuss the limitations of our model, and summarize our conclusions in Section 5.

2 AN ANALYTIC MODEL OF GALAXY FORMATION

2.1 The effective star formation efficiency

The formation, evolution, and abundance of dark matter haloes can be predicted accurately when the cosmology and dark matter model (i.e. cold, warm, self-interacting, etc.) is known. Although these processes are highly non-linear, the underlying physics is well understood. However, the gas and stellar content of haloes is much less well understood because of the intrinsic complexity of the baryonic processes, such as cooling, star formation and feedback, that drive it. An empirical approach to populating dark matter haloes with galaxies is to focus on the relation between stellar mass and halo mass inferred from observations. We write this relation as

$$\log_{10}\left(\frac{M_*}{10^{12}\,\mathrm{M}_\odot}\right) = \varepsilon(M_h, t)\log_{10}\left(\frac{M_h}{10^{12}\,\mathrm{M}_\odot}\right) + \log_{10}\mathcal{N}(t), \tag{1}$$

where M_* and M_h are the central galaxy stellar mass and host halo mass respectively, $\mathcal{N}(t)$ is a normalization factor, and $\varepsilon(M_h, t)$ is the logarithmic slope of the stellar to halo mass relation (SHMR). Allowing $\mathcal{N}(t)$ to be a random variable would account for the scatter in the relation, but in this paper we will focus on the mean relation and replace $\mathcal{N}$ by its expectation value. ε is closely related to the galaxy formation efficiency of haloes, and we will explore this connection further below.

Probing the dark matter distribution and its evolution directly represents an observational challenge. Perhaps the simplest, and most commonly used alternative technique is to use (sub-) halo abundance matching to determine the typical SHMR (e.g. Behroozi et al. 2013; Moster et al. 2013). In essence, the SHMR is derived by mapping the theoretical halo mass function and the observed

[1]The galaxy and halo catalogues of the simulation suite, as well as the particle data, are publicly available at http://www.eaglesim.org/database.php (McAlpine et al. 2016).

abundance of galaxies given by the GSMF,

$$\phi(M_*) \equiv \frac{\mathrm{d}n_{gal}}{\mathrm{dlog}_{10}M_*} = \varepsilon^{-1}\,\frac{\mathrm{d}n_{\rm h}}{\mathrm{dlog}_{10}M_h}, \tag{2}$$

where n_{gal} and $n_{\rm h}$ are the co-moving abundances of galaxies and haloes respectively. A subtlety here is that M_* refers to the stellar mass of the central object in the halo. More complex formulations of the abundance matching method allow for the contribution of satellite galaxies to the mass function, but we will keep to the simple approach. This is adequate if the stellar mass function is dominated by central galaxies (e.g. Yang, Mo & van den Bosch 2009; Lan, Ménard & Mo 2016).

Abundance matching studies have consistently shown that a simple picture of the galaxy population is consistent with much of the observational data. The SHMR is a strong function of halo mass but a weak function of cosmic time; it can be approximated well by two power laws that connect at a stellar mass that corresponds roughly to the knee of the stellar mass function (e.g. Moster et al. 2010; Yang et al. 2012; Mitchell et al. 2016). We use this as the basis of a simple model that couples the build-up of dark matter haloes and the build-up of galaxy stellar mass.

We must be careful, however, to distinguish the instantaneous efficiency with which infalling baryons are converted into stars,

$$\epsilon_* = \frac{\dot{M}_*}{f_b \dot{M}_h} \tag{3}$$

(where $f_b = \Omega_b/\Omega_m$ is the cosmic baryon fraction) and the integral of this growth over the history of the halo, equation (1). Note that $\dot{M}_*$ includes both star formation within the central object, and the accretion of infalling stars. We will need to distinguish between the two in order to relate the stellar mass growth to the observed SFR density.

In Appendix A, we show that the build up of dark matter haloes can be described analytically using recent developments of the linear theory originally described by Press & Schechter (1974). This allows the abundance and growth rates of haloes to be derived from the power spectrum of density fluctuations in the early universe. Equation (2) provides a promising approach to connect the growth of haloes to the formation of galaxies, and the observational results suggest that a good starting point is to consider a time-independent SHMR that depends only on halo mass. In addition, we will examine a model in which the efficiency of star formation depends on the halo's virial velocity, as it reflects the evolution of the gravitational potential of the halo (e.g. Sharma & Theuns 2019).

(i) *Halo mass-dependent efficiency (Model I)*

In order to model ϵ_* in equation (3), we begin by assuming that the efficiency of conversion of infalling baryons into stars depends only on halo mass (e.g. Rodríguez-Puebla et al. 2016; Salcido et al. 2018; Tacchella et al. 2018). Motivated by the results from abundance matching techniques (e.g. Behroozi et al. 2013; Rodríguez-Puebla et al. 2016), that estimate a galaxy formation efficiency that peaks at masses similar to Milky-Way sized haloes ($\sim 10^{12}\,\mathrm{M}_\odot$) and falls steeply for higher and lower masses, we model ϵ_* as a double power law, a similar parametrization as that proposed by Moster et al. (2010):

$$\epsilon_*(M_h) = 2\epsilon_{\rm N}\left[\left(\frac{M_h}{M_{\rm crit}}\right)^{-\alpha} + \left(\frac{M_h}{M_{\rm crit}}\right)^{\beta}\right]^{-1}, \tag{4}$$

where $\epsilon_{\rm N}$ is a normalization parameter, and α and $-\beta$ are the power-law slopes at low and high mass respectively. The maximum efficiency occurs at halo mass $M_{\rm crit}$. To agree with observational data, the values of α and β are typically positive, i.e. at low masses, star formation is suppressed because of the efficiency of feedback from star formation, and at higher masses, the cooling of the inflowing gas is suppressed by heating from BHs (e.g. White & Frenk 1991; Bower et al. 2006; Benson 2012; Haas et al. 2013).

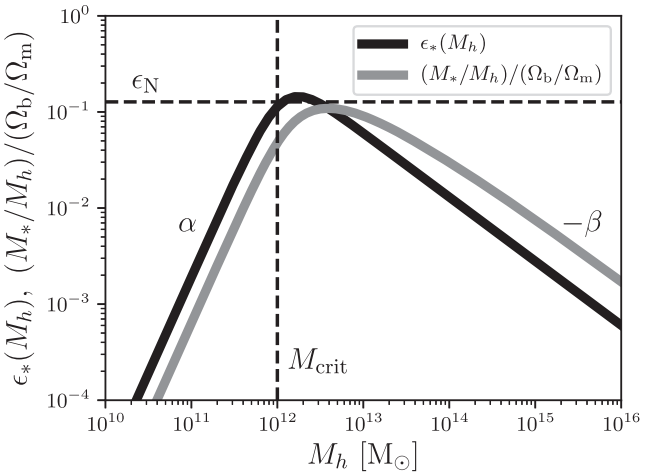

Figure 1. Parametrization of the effective star formation efficiency ϵ_* provided in equation (4). $\epsilon_{\rm N}$ is the normalization parameter, α and β determine the slope of the efficiency at low and high masses, respectively, and $M_{\rm crit}$ locates the transition mass, or peak efficiency. The SHMR is shown for comparison. ϵ_* has the same slopes as M_*/M_h, i.e. α and β, but the normalization of M_*/M_h is different by a factor of $1/(1+\alpha)$, and $1/(1-\beta)$ for low and high mass haloes, respectively.

Because of our assumption that ϵ_* depends only on M_h, we can integrate to determine M_* without needing to know the time evolution of the halo mass.

$$\begin{aligned} M_* &= \int_0^{M_h} \epsilon_* f_b \mathrm{d}M_h' \\ &= \frac{2\epsilon_{\rm N}}{1+\alpha} f_b M_{\rm crit}\left(\frac{M_h}{M_{\rm crit}}\right)^{1+\alpha} F(\eta, \mathit{z}), \end{aligned} \tag{5}$$

where $\eta = (1+\alpha)/(\alpha+\beta)$, $\mathit{z} = (M_h/M_{\rm crit})^{\alpha+\beta}$ and

$$\begin{aligned} F(\eta, \mathit{z}) &= \eta \int_0^1 \frac{x^{\eta-1}}{(1+\mathit{z}x)}\mathrm{d}x \\ &= {}_2F_1(1, \eta; 1+\eta; -\mathit{z}), \end{aligned} \tag{6}$$

where[2] ${}_2F_1(a, b; c; \mathit{z})$, is the Gaussian hypergeometric function. For values of $\alpha > 0$ and $0 < \beta < 1$, in the limit $M_h \ll M_{\rm crit}$, $\lim_{\mathit{z}\to 0} F(\eta, \mathit{z}) = 1$, while for $M_h \gg M_{\rm crit}$, $F(\eta, \mathit{z}) \approx 1/((\eta-1)\,\mathit{z})$. Differentiating equation (5), the logarithmic slope of the stellar mass halo mass relation, ε, can be written in a simple analytic form,

$$\varepsilon = \frac{\mathrm{dlog}M_*}{\mathrm{dlog}M_h} = \frac{1+\alpha}{(1+\mathit{z})F(\eta, \mathit{z})}. \tag{7}$$

Equation (7) describes a smooth transition in slope from $(1+\alpha)$ for $M_h \ll M_{\rm crit}$, to $(1-\beta)$ for $M_h \gg M_{\rm crit}$.

Fig. 1 shows an illustration of the effective star formation efficiency ϵ_* as a function of M_h. The parametrization provides a smooth transition from the α dominated regime (for low halo masses), to the β dominated regime (for high halo masses). The figure shows that ϵ_* has the same slopes as M_*/M_h, i.e. α and β, but the normalization

[2]We have used the symbol z to differentiate from redshift z.

of M_*/M_h is different by a factor of $1/(1+\alpha)$, and $1/(1-\beta)$ for low and high mass haloes respectively. We note that for our chosen parametrization, M_*/M_h is closely approximated by a double power law.

(ii) *Virial temperature-dependent efficiency (Model II)*

As we will discuss in Section 3, an efficiency dependent only on halo mass turns out to be a very good approximation of the stellar mass build up of galaxies because most of the stellar mass builds up when the mass of the halo has roughly its current value. However, a time-independent efficiency model significantly underpredicts the abundance galaxies at high redshifts ($z > 4$), which hints at the need for a time-evolving efficiency model. A purely empirical approach (e.g. Moster et al. 2018; Behroozi et al. 2019) would relax the physical priors and let, in this case, the four efficiency parameters in the model to evolve freely in time. Instead, we consider an alternative model in which ϵ_* depends only on the virial temperature of the halo $T_{\rm vir}$ (and hence the gravitational potential of the halo).

Considering the energetics of galaxy winds suggests that, wind that marginally escape the gravitational binding energy of the galaxy's halo can carry a higher mass loading in lower mass haloes (Dekel & Silk 1986; White & Frenk 1991; Davé, Finlator & Oppenheimer 2012b; Sharma & Theuns 2019). Since the energy required for escape depends on the halo virial temperature, $T_{\rm vir}$, this leads to an inverse scaling of the mass loading, and hence, star formation efficiency. At sufficiently high mass, the energy associated with individual supernovae becomes smaller than the halo binding energy. This may lead to the accumulation of gas around the central BH, and consequently a wind driven by BH accretion, rather than star formation (Dubois et al. 2015). A related argument can also be made based on the buoyancy of gas heated by star formation. Bower et al. (2017) discusses a possible physical origin of a transition from where star formation driven outflows get hotter than the virial temperature of the halo and can escape (i.e. supernovae energy, or entropy, is much greater than the halo binding energy), to where outflows stall inside the halo triggering star formation and BH growth.

In order to explore these effects, we model the effective star formation efficiency as a function of the halo's virial temperature using the same double power-law parametrization as in equation (4),

$$\epsilon_*(T_{\rm vir}) = 2\epsilon_{\rm N}\left[\left(\frac{T_{\rm vir}}{T_{\rm crit}}\right)^{-\alpha} + \left(\frac{T_{\rm vir}}{T_{\rm crit}}\right)^{\beta}\right]^{-1}. \tag{8}$$

Note that the relation between halo mass and virial temperature depends on redshift (see equations 9 and 10), as the density of collapsed haloes decrease as the universe expands. As a result, it is not possible to analytically determine $M_*(M_h, t)$, but the required integrals can easily be evaluated numerically. We discuss this model further in Section 4.1.

2.2 Halo definition

Dark matter haloes are typically identified by growing a sphere outwards from the potential minimum of the dark matter halo out to a radius where the mean interior density equals a fixed multiple of the critical or mean density of the Universe, causing an artificial 'pseudo-evolution' of dark matter haloes by changing the radius of the halo (Diemer, More & Kravtsov 2013). Star formation, however, is governed by the amount of gas that enters these haloes and reaches their central regions. Wetzel & Nagai (2015) show that the growth of dark matter haloes is subject to this 'pseudo-evolution', whereas the accretion of gas is not. Because gas is able to cool radiatively, it decouples from the dark matter, tracking the accretion rate near a radius of $R_{\rm 200m}$, the radius within which the mean density is 200 times the mean density of the universe, $\bar{\rho}$. As we try to connect the mass accretion rate of dark matter haloes to star formation, we define halo masses as the total mass within $R_{\rm 200m}$,

$$M_h = 200\frac{4\pi}{3}R_{\rm 200m}^3\bar{\rho}, \tag{9}$$

where $\bar{\rho}(t) = \rho_0 a(t)^{-3}$.

We assume that during gravitational collapse, the gas experiences strong shocks and thermalizes its kinetic infall energy to the virial temperature of the halo,

$$T_{\rm vir} = \frac{\mu m_p G M_h}{5 k_{\rm B} R_{\rm 200m}}, \tag{10}$$

where we have assumed a uniform cloud of monatomic gas. M_h is the mass of the halo, μ is the mean molecular weight of the gas in the halo, which we have assumed $\mu \approx 0.6$ for a fully ionized plasma of primordial composition, $k_{\rm B}$ is the Boltzmann constant, and m_p the proton mass. Note from equation (9) that for a given halo mass, the radius of the halo, $R_{\rm 200m}$, changes with time as the mean density of the Universe evolves.

2.3 The model

The analytical galaxy formation model developed here is comprised of three main components, which are summarized in the schematic diagram of Fig. 2:

(i) A cosmological model (blue block)
(ii) An astrophysical model that sets the relation between galaxies and their haloes (green block)
(iii) The galaxy formation outputs (orange block)

By using the Taylor expansion solution of the Friedmann equations introduced by Salcido et al. (2018), the formation and evolution of dark matter haloes can be described analytically. This component is shown as the blue block in Fig. 2. The growth rates of haloes depend on the cosmological parameters ρ_0, Λ, H_0, and the shape of the matter density fluctuation power spectrum. We parametrize the variance of the spherically averaged smoothed density field, $S = \sigma^2$, as a power law $S \approx S_0(M_h/10^{12}\,{\rm M}_\odot)^{-\gamma}$ with slope γ. Since we are interested in only a small range of halo mass, this is a sufficiently accurate description. The derivation of these equations are presented in Appendix A. For convenience, we define the cosmology-dependent approximations for the equations that appear below using the function,

$$f_\Lambda(t, A, B) = 1 + A\left(\frac{t}{t_\Lambda}\right)^2 + B\left(\frac{t}{t_\Lambda}\right)^4, \tag{11}$$

where $t_\Lambda = \sqrt{3/\Lambda c^2}$, Λ is the value of the cosmological constant, and the coefficients A and B are obtained by using the Taylor expansion solution of the Friedmann equations in Salcido et al. (2018).

Astrophysical processes (shown in green in Fig. 2), enter into the model through the effective star formation efficiency, which is fully described by the efficiency ϵ_* (equation 3). We consider two models, in which ϵ_* is a function of halo mass or virial temperature. This component of the model is described by four parameters, $M_{\rm crit}$ (or $T_{\rm crit}$), $\epsilon_{\rm N}$, α, and β following equation (4) or equation (8).

In order to simplify the numerical constants in the equations presented in this section, we have substituted the numerical values for the cosmological parameters for a standard ΛCDM universe

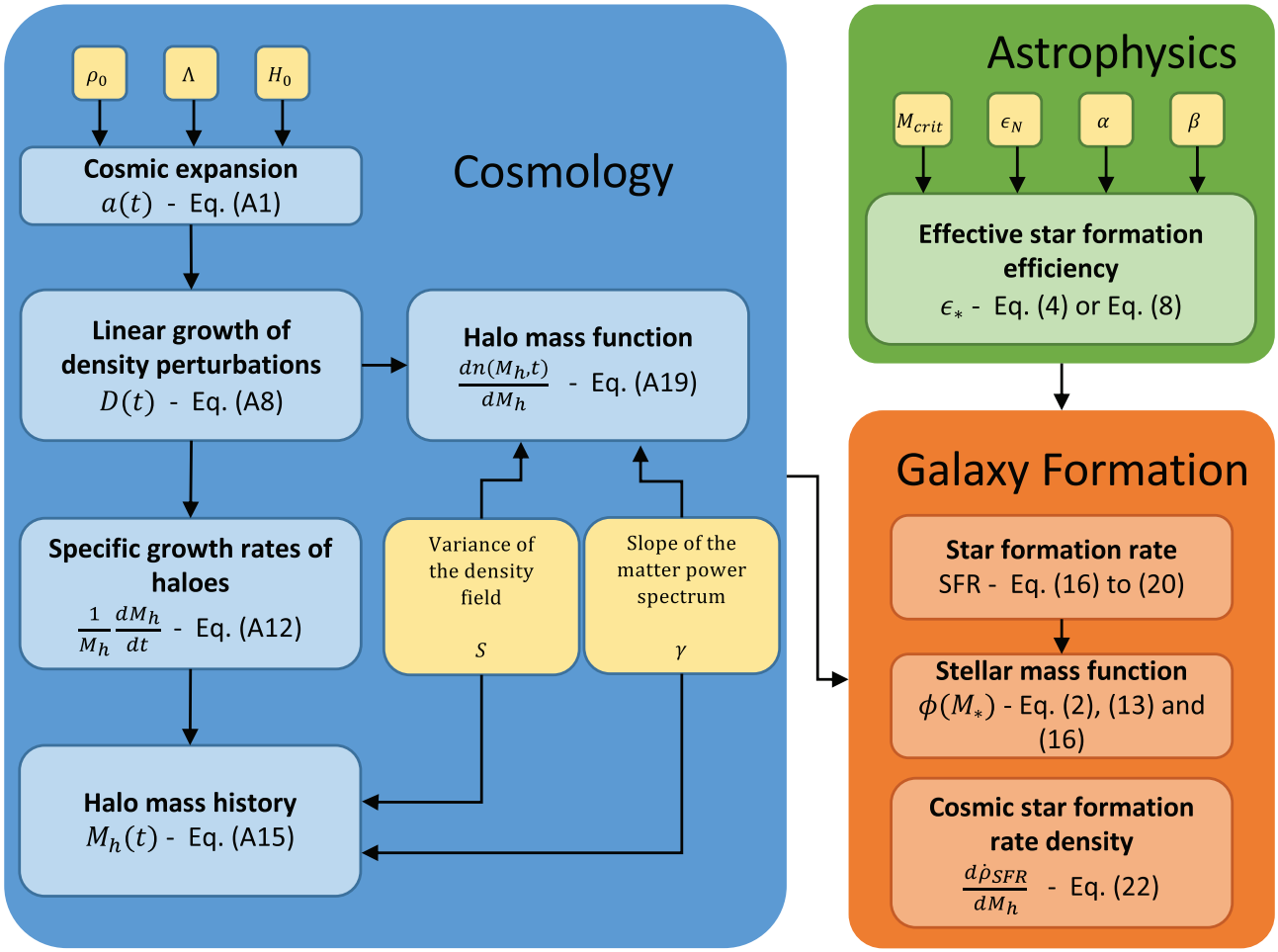

Figure 2. A schematic diagram of the analytic model of galaxy formation. All components in the blue block depend solely on cosmology. By using the Taylor expansion solution to the Friedmann equations in Salcido et al. (2018), all the cosmological components can be calculated analytically for a given cosmology defined by the parameters ρ_0, Λ, H_0, and the shape of the matter power-spectrum parametrized by S and γ. All astrophysical processes (green) enter into the model in terms of the effective star formation efficiency ϵ_*, which is fully described by the four free parameters $M_{\rm crit}$, $\epsilon_{\rm N}$, α and β in equation (4). The galaxy formation outputs are summarized in the orange block.

as inferred by the Planck Collaboration XVI (2014), i.e. $\Omega_{\rm m} = 0.307$, Ω_Λ=0.693, $\Omega_b = 0.04825$, H_0=67.77 km s^{-1} Mpc^{-1}, $\rho_0 = 3.913 \times 10^{10}$ M$_\odot$ Mpc^{-3}, $\gamma = 0.3$, t_0=13.8 Gyr, and $S_0 = 3.98$. The full cosmology dependence of the numerical constants is given in Appendix A, and are highlighted using a coloured superscript (c^*).

(i) *Halo mass history*

The analytic form of the growth rate equations allows us to simply describe the growth of haloes as a function of their present-day mass, M_0:

$$\frac{M_h(t)}{10^{12}\,{\rm M}_\odot} = \left\{ \left(\frac{M_0}{10^{12}\,{\rm M}_\odot} \right)^{-\gamma/2} + 0.31\gamma \left[\left(\frac{t}{t_m} \right)^{-2/3} f_\Lambda(t, 0.16, -0.01) - 1.67 \right] \right\}^{-2/\gamma} \quad (12)$$

,where $t_{\rm m} = \sqrt{3/8\pi G\rho_0}$, and ρ_0 is the mean matter density of the Universe at the present time. For the Planck Collaboration XVI (2014) cosmological parameters, t_m=26.04 Gyr and t_Λ=17.33 Gyr. Fig. 3 shows the individual mass histories for haloes of a given mass M_0 at the present cosmic time (represented by the colour coding).

(i) *Halo mass function*

Using the Press & Schechter formalism (Press & Schechter 1974), the co-moving abundance of haloes of mass M_h at time t is given by,

$$\frac{{\rm d}n(M_h, t)}{{\rm dlog}_{10} M_h} = 5.43 \times 10^{-3}\,{\rm cMpc}^{-3} \left(\frac{M_h}{10^{12}\,{\rm M}_\odot} \right)^{-(1-\frac{\gamma}{2})} \times \left(\frac{t}{t_m} \right)^{-2/3} f_\Lambda(t, 0.16, -0.01) \times \exp\left[-0.13 \left(\frac{M_h}{10^{12}\,{\rm M}_\odot} \right)^{\gamma} \left(\frac{t}{t_m} \right)^{-4/3} f_\Lambda(t, 0.32, 0) \right]. \quad (13)$$

This equation consists of two parts, a low-mass power-law dependence close to M_h^{-1}, and an exponential cut-off at high masses. For a given halo mass, the abundance initially increases as the exponential suppression is reduced, but at late times the halo abundance slowly decreases because of the power-law term.

(i) *The galaxy stellar mass function and the origin of the Schechter function*

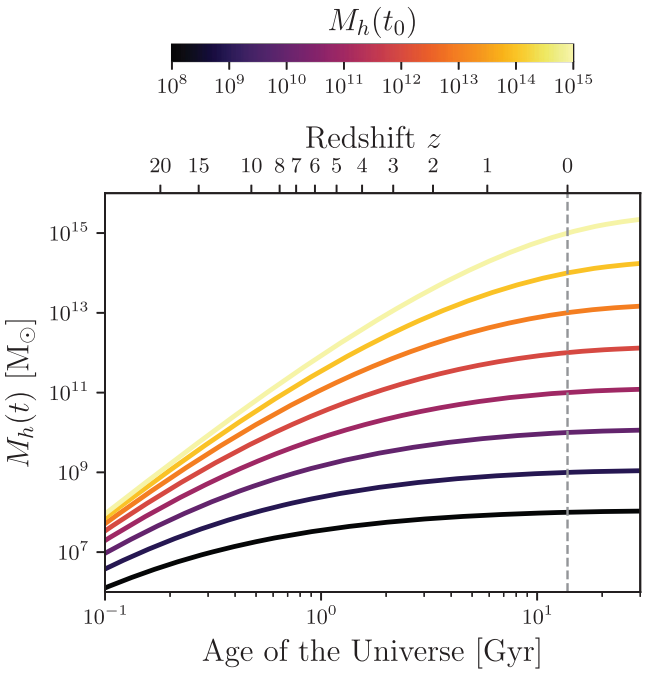

Figure 3. Average halo mass as a function of cosmic time derived in equation (12). A model for the cosmological parameters for a standard ΛCDM universe as inferred by the Planck Collaboration XVI (2014) is shown with solid lines. Colour coding represents different halo masses, M_0, at the present cosmic time t_0, $M_0 = M_h(t_0)$.

The GSMF has been reasonably well measured over much of cosmic time, so that, for a known cosmology, the GSMF provides a good measurement of the efficiency by which haloes convert their baryons into stars. Typically, the GSMF (equation 2) is parametrized by a Schechter function (Schechter 1976),

$$\phi(M) = \phi^* \left(\frac{M}{M^*}\right)^{\alpha} e^{-M/M^*}, \tag{14}$$

where ϕ^* provides the normalization, and M^* is a characteristic galaxy stellar mass where the power-law form of the function cuts off. The form of this function was originally motivated by the halo mass dependence given in equation (13). Importantly, however, the shape GSMF is only indirectly related to the halo mass function (e.g. Benson et al. 2003), with observations showing that the power-law slope is much flatter than that expected for the halo mass function. Moreover, recent measurements of the GSMF at low redshift (e.g. Baldry, Glazebrook & Driver 2008; Li & White 2009; Baldry et al. 2012; Moustakas et al. 2013), have proven that a single Schechter function is insufficient to describe the density of galaxies. Specifically, the low-redshift GSMF shows a clear evidence for a low-mass upturn, or equivalently, a 'pile-up' in the abundance of galaxies around M^*. Typically, a double Schechter function parametrization has been used to better describe observational data,

$$\phi(M) = \left[\phi_1^* \left(\frac{M}{M^*}\right)^{\alpha_1} + \phi_2^* \left(\frac{M}{M^*}\right)^{\alpha_2}\right] e^{-M/M^*}. \tag{15}$$

In the model presented here, the GSMF can be computed as a function of time, by combining the halo mass function (equation 13) and the efficiency of star formation through equations (2) and (7). These equations link the observed shape of the GSMF, to the underlying dark matter halo distribution, and hence to the cosmological background. They also link galaxies to their dark matter haloes, providing valuable information about the efficiency by which haloes convert their baryons into stars.

Further consideration shows that they provide a description of the non-trivial shape of the GSMF and the need for a double Schechter function to describe it. While the underlying distribution of dark matter haloes is theoretically predicted to be a single Schechter function (Press & Schechter 1974), its transformation to the GSMF relies on equation (2). When the halo mass function is multiplied by the inverse of the logarithmic slope of the SHMR, the low-mass end is multiplied by a factor $1/(1+\alpha)$, while the high-mass end is multiplied by a factor of $1/(1-\beta)$. As both α and β are positive, this creates a kink in the gradient, shown with a red dot in the top left-hand panel of Fig. 4. These different changes in the normalization cause a maximum in the logarithmic derivative of (ϕ_h/ε) shown in top right-hand panel of the figure. Hence, there is an inflection point in the distribution, as the second derivative vanishes and changes sign at $\sim M_{\rm crit}$. At this point, the abundance of galaxies decreases slowly, or even rises, as a function of mass, creating a 'bump' at the knee of the GSMF. The sharper the transition (i.e. the larger $\alpha + \beta$), the more pronounced the bump at the knee of the GSM. These effects are illustrated in Fig. 4. Physically, this can be interpreted as galaxies of similar masses 'piling up' at the peak star formation efficiency, i.e. $M_h \approx M_{\rm crit}$, as they rapidly stop forming many more stars. In Section 3, we will systematically vary the four parameter in the efficiency model in equation (4) to investigate their effect on different galaxy formation outputs.

(i) *The galaxy stellar mass growth*

We now have all the necessary ingredients to calculate the stellar mass growth of individual galaxies through cosmic time. Substituting equation (A12) into equation (3), the stellar mass is given by the integral of,

$$\begin{aligned}\frac{{\rm d}M_*}{{\rm d}t} &= \epsilon_* f_b \left[\frac{1}{M_h}\frac{{\rm d}M_h}{{\rm d}t}\right] M_h \\ &= \epsilon_* f_b \; 1.6 \times 10^{10}\,{\rm M}_\odot\,{\rm Gyr}^{-1} \left(\frac{t}{t_m}\right)^{-5/3} \\ &\quad \times f_\Lambda(t, -0.32, 0.06) \left(\frac{M_h}{10^{12}\,{\rm M}_\odot}\right)^{\left(1+\frac{\gamma}{2}\right)},\end{aligned} \tag{16}$$

where ϵ_* is given by either equation (4) or equation (8).

Assuming an instantaneous recycling approximation (Schmidt 1963), the relation between the stellar mass growth due to star formation and the observed galaxy SFR is simply given by,

$${\rm SFR} = \frac{\dot{M}_*}{(1-R)}, \tag{17}$$

where R is the fraction of mass of gas that is instantaneously returned into the interstellar medium by an entire stellar generation. For a universal Chabrier (2003) initial mass function (IMF), $R = 0.41$.

Furthermore, both *in situ* star formation and galaxy mergers contribute to the total stellar build up of galaxies. In low-mass haloes, most of the stellar build up is expected to come from *in situ* star formation, while the most massive galaxies experience almost no internal star formation and grow mainly by mergers with smaller satellite galaxies. Hence, the fractional contribution of accreted stars to the total stellar mass build up of galaxies is a steep function of halo mass (e.g. Rodriguez-Gomez et al. 2016; Qu et al. 2017; Pillepich et al. 2018b). Assuming that all of the stellar growth of haloes of mass $M_{\rm crit}$ and bellow is due to internal star formation, we

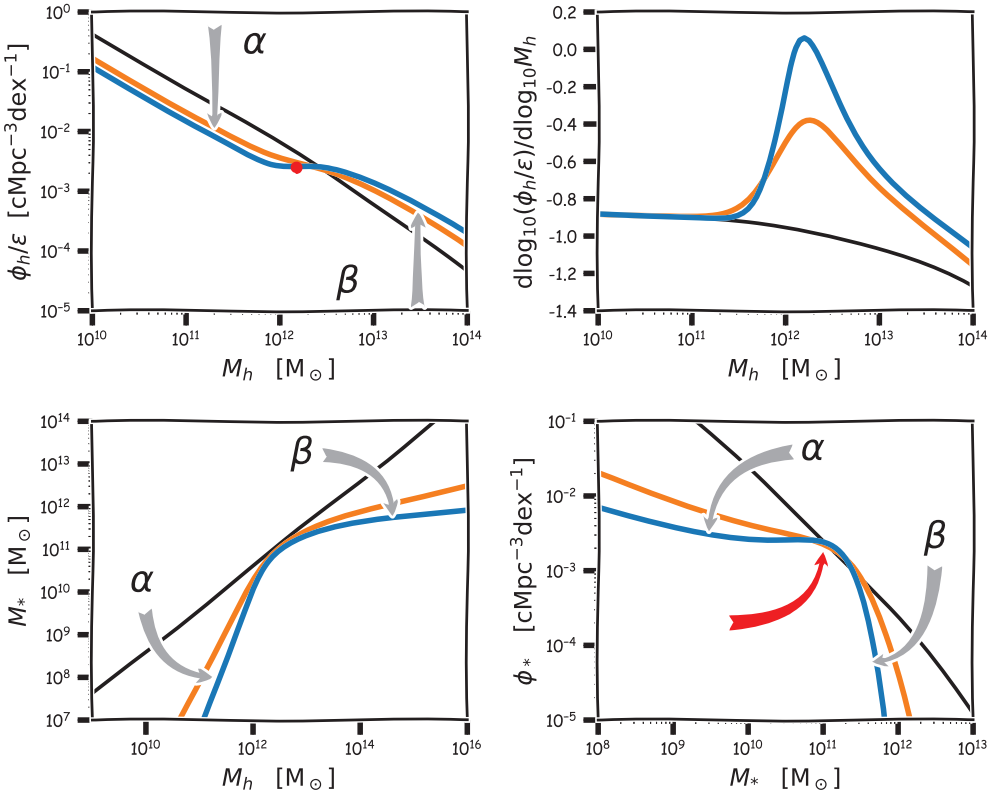

Figure 4. A schematic illustration of the role played by the low-mass and high-mass end slopes of the SHMR in shaping the GSMF (see equation 2). Two arbitrary models are shown. A model with both α and β large is shown in blue. The orange line illustrates a model with smaller α and β. *Top left-hand panel:* Product of the halo mass function and the inverse of the logarithmic slope of the SHMR, ε, given in equation (7). The halo mass function is shown in black. For the blue line, the low-mass end is multiplied by a factor $1/(1+\alpha)$, while the high-mass end is multiplied by a factor of $1/(1-\beta)$. As both α and β are positive, this creates an inflection point (shown as a red dot) in the distribution. *Top right-hand panel:* Logarithmic derivative of (ϕ_h/ε). The different changes in the normalization cause a maximum in the distribution. Hence, there is an inflection point as the second derivative vanishes and changes sign at $\sim M_{\rm crit}$. The black line shows the logarithmic slope of the halo mass function. *Bottom left-hand panel:* The SMHR is shown. The black line shows a relationship of $M_* \propto M_h$. At low masses, SFR is suppressed because of the efficiency of feedback from star formation, yielding a slope of $(1+\alpha)$. At higher masses a slope of $(1-\beta)$ is expected as cooling of the inflowing gas is suppressed by heating from BHs. *Bottom right-hand panel:* The GSMF is shown. The black line shows a relationship of $\phi_* \propto \phi_h$. The low-mass and high-mass end slopes of the SHMR suppress the abundance of low and high mass galaxies, respectively, but also create a characteristic 'bump' at the knee of the GSMF.

parametrize the fraction of stellar mass growth from *in situ* SFR by a broken power law as,

$$f_{\rm SFR} = \begin{cases} 1 & \text{for } M_h \le M_{\rm crit} \\ \left(\frac{M_h}{M_{\rm crit}}\right)^{\eta} & \text{for } M_h > M_{\rm crit}, \end{cases} \tag{18}$$

where $M_{\rm crit}$ is the effective star formation peak efficiency defined in equation (4). For the virial temperature-dependent efficiency model (Model II), $M_{\rm crit}$ also varies with time, and can be calculated using the critical virial temperature in equation (10). We fix the value of η by assuming that at redshift $z = 0$, where $M_{\rm crit} \approx 10^{12}\,{\rm M}_\odot$, $f_{\rm SFR}(10^{13}\,{\rm M}_\odot)$ is ~ 50 per cent (Pillepich et al. 2018b), hence,

$$\eta = \log_{10}(0.5)/(13-12) = -0.3. \tag{19}$$

Putting equations (17) and (18) together, the fraction of stellar mass growth of central galaxies due to *in situ* formation is given by,

$$\text{SFR} = \frac{\dot{M}_*}{(1-R)} \times f_{\rm SFR}. \tag{20}$$

(i) *The cosmic SFR density*

The total cosmic SFR density is given by the integral of all star formation in all haloes,

$$\begin{aligned} \dot{\rho}_{\rm SFR}(t) &= \int \dot{M}_* \frac{f_{\rm SFR}}{(1-R)} \frac{{\rm d}n(M_h,t)}{{\rm dlog}_{10}M_h}\, {\rm dlog}_{10}M_h \\ &= \int \epsilon_* f_b \dot{M}_h \frac{f_{\rm SFR}}{(1-R)} \frac{{\rm d}n(M_h,t)}{{\rm dlog}_{10}M_h}\, {\rm dlog}_{10}M_h. \end{aligned} \tag{21}$$

Using the stellar mass growth rate from equation (16), the halo mass function from equation (13), together with the effective star formation efficiency from equation (3), the contribution to the cosmic SFR density from haloes of mass M_h (the integrand of equation 21) is given by,

$$\begin{aligned} \frac{{\rm d}\dot{\rho}_{\rm SFR}}{{\rm dlog}_{10}M_h} &= \epsilon_* f_b \frac{f_{\rm SFR}}{(1-R)}\, 8.7\times10^7\,{\rm M}_\odot\,{\rm Gyr}^{-1}\,{\rm cMpc}^{-3} \\ &\times \left(\frac{M_h}{10^{12}\,{\rm M}_\odot}\right)^{\gamma} \left(\frac{t}{t_m}\right)^{-7/3} f_\Lambda(t,-0.16,0) \\ &\times \exp\left[-0.13\left(\frac{M_h}{10^{12}\,{\rm M}_\odot}\right)^{\gamma}\left(\frac{t}{t_m}\right)^{-4/3} f_\Lambda(t,0.32,0)\right], \end{aligned} \tag{22}$$

where ϵ_* is modelled using equation (4) or equation (8). The differential form of equation (22) explicitly shows the contribution from haloes of different masses M_h, to the total cosmic SFR density.

Equations (2), (12), (13), (16), (20), and (22), together with a model of the effective star formation efficiency, equation (4) or equation (8), provide a full mathematical framework to explore the effects of cosmology and baryonic physics on galaxy formation. In the next section, we will explore the effect of the different efficiency parameters on the galaxy SFR, GSMF, and the cosmic SFR density.

3 THE IMPACT OF THE EFFECTIVE STAR FORMATION EFFICIENCY

We now use our model to explore the effect of the different efficiency parameters in the galaxy formation outputs in the orange block of Fig. 2. It is common to characterize galaxy properties over halo masses, and for simplicity, in this section we will only use a halo mass-dependent star formation efficiency model (i.e. ϵ_* is constant across cosmic time).

It has been estimated that the SHMR peaks at masses similar to Milky Way sized haloes ($\sim 10^{12}\,\mathrm{M}_\odot$). Typically, at low masses, the SFR is suppressed because of the efficiency of stellar feedback. On the other hand, at higher masses the cooling of the inflowing gas is suppressed by heating from supermassive BHs (e.g. White & Frenk 1991; Bower et al. 2006; Benson 2012). The 'Fiducial' model captures this behaviour with both α and β being positive and equal to 0.75.

We consider five alternative models varying the efficiency parameters systematically to explore the physics of galaxy formation. An extreme idealized case label as 'Constant', describes a model where a fixed fraction of the baryon budget is turned into stars, regardless of the halo mass. The 'No AGN' model describes a scenario where the efficiency of feedback processes is weak for massive objects. Physically, this could be thought as a model where feedback from active galactic nuclei is inefficient. The 'No SN' model describes a scenario where the efficiency of feedback processes is weak in small haloes. Physically, this could be thought as a model where feedback from supernovae is inefficient. While much more computationally expensive, a similar behaviour to these models is reproduced in full hydrodynamical simulations (see Appendix B for a couple of examples). In Fig. 5 we show the median stellar-halo mass ratio for three variations of the EAGLE simulations where the subgrid prescription for stellar and AGN feedback have been removed. Indeed, our models capture the overall behaviour attained in simulations. An additional model labelled '$M_{\rm crit} = 10^{10}$' explores the effect of changing the critical, or transition, halo mass. A final model labelled 'High efficiency' has the same slopes as the fiducial model, but with a different normalization. We show in Fig. 6 the effective star formation efficiency for the six models, and their parameters are summarized in Table 1.

3.1 The build up of stellar mass

First, we explore the effect of the star formation efficiency in the build up of stellar mass in individual haloes for the six models, which can be calculated by integrating equation (16).

Fig. 7 shows an example of the evolution of the stellar mass in a halo of mass $M_0 = 10^{13}\,\mathrm{M}_\odot$ at the present time for the six efficiency models. For the constant efficiency model, the stellar mass grows steadily with time, and starts to slow down only at late times due to cosmic expansion, tracking the mass assembly of the dark matter halo. For the fiducial model, the build up of

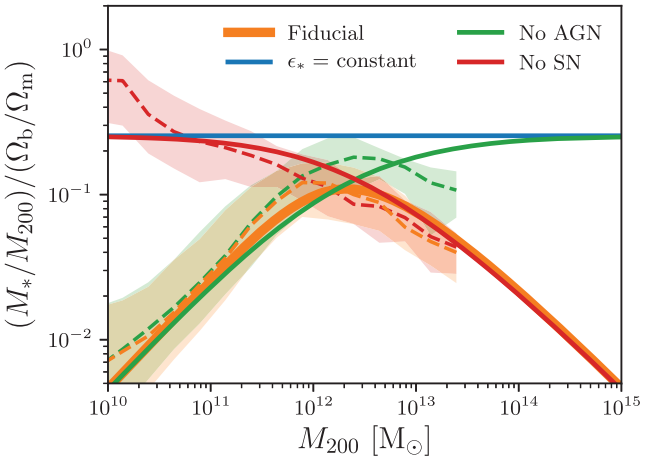

Figure 5. Median stellar-halo mass ratio for central galaxies for three variations of the EAGLE $(50\,\mathrm{cMpc})^3$ simulations at redshift $z = 0$ (dashed lines), compared to their equivalent analytic effective star formation efficiency model (solid lines). The orange line shows the Ref-L050N0752 EAGLE model (Crain et al. 2015; Schaye et al. 2015), which uses the same calibrated sub-grid parameters as the reference model $(100\,\mathrm{cMpc})^3$, ran with the same resolution, but in a smaller volume. The 'No AGN' run (green) uses the same calibrated sub-grid parameters as the reference model but removing feedback from BHs. For the 'No SN' model (red), feedback from star formation has been removed. We note that there is no EAGLE equivalent to the 'constant' (or 'no feedback') model. The faint shaded regions enclose the 10th–90th percentiles. While much more computationally expensive, the behaviour of the full hydrodynamical simulations is well approximated by the analytic models introduced here.

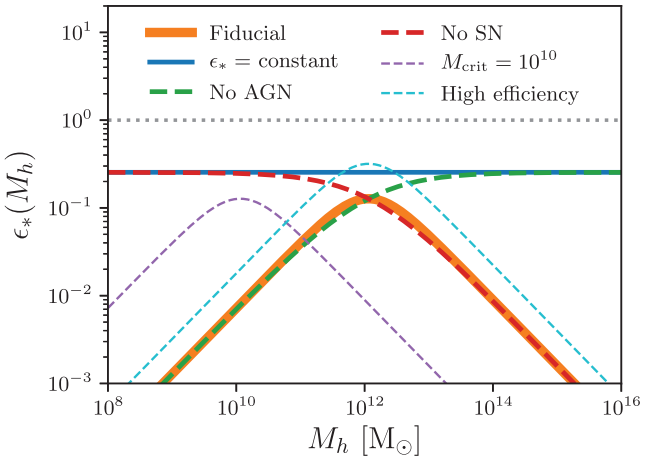

Figure 6. The effective star formation efficiency ϵ_* as a function of halo mass for the six models described in Table 1. For the 'Fiducial' model, the efficiency peaks at masses similar to Milky Way sized haloes ($10^{12}\,\mathrm{M}_\odot$) and fall steeply for higher and lower masses with $\alpha = 0.75$ and $\beta = 0.75$. For the 'Constant' model, a fixed fraction of the baryon budget is turned into stars, regardless of the halo mass. The 'No AGN' model describes a scenario where the efficiency of feedback process is weak for massive objects. The 'No SN' model describes a scenario where the efficiency of feedback process is weak in small haloes. The '$M_{\rm crit} = 10^{10}$' model explore the effect of changing the critical halo mass. The 'High efficiency' model has the same slopes as the fiducial model, but with a higher normalization. A 100 per cent efficiency is shown with a grey dotted line.

Table 1. Effective star formation efficiency parameters for the six idealized models. To agree with observational data, the values of α and β are typically positive. The 'Fiducial' model captures this behaviour while the five variations systematically explore the effect of the effective star formation efficiency on the physics of galaxy formation.

	ϵ_N	M_{crit}	α	β
Fiducial	0.125	10^{12}	0.75	0.75
Constant	0.250	10^{12}	0	0
No AGN	0.125	10^{12}	0.75	0
No SN	0.125	10^{12}	0	0.75
$\boldsymbol{M_{crit} = 10^{10}}$	0.125	10^{10}	0.75	0.75
High efficiency	0.320	10^{12}	0.75	0.75

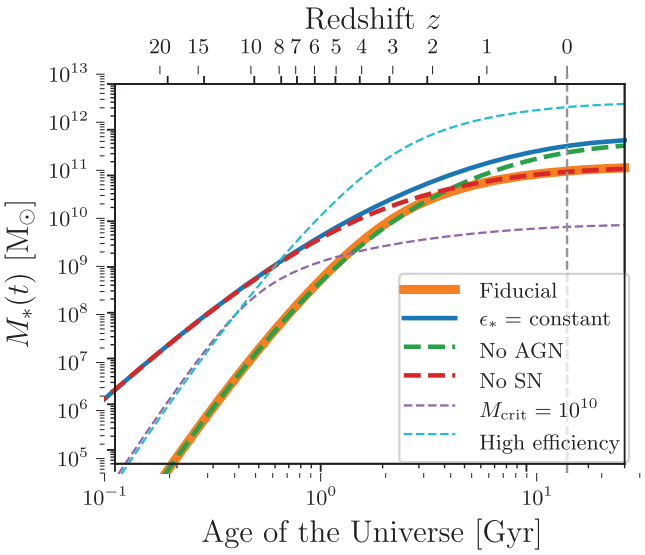

Figure 7. An example of the evolution of the stellar mass in a halo of mass $M_0 = 10^{13}$ M$_\odot$ at the present time calculated by integrating equation (16). The different colours represent the different efficiency models. For the constant efficiency model, the stellar mass grows steadily with time tracking the mass assembly of the dark matter halo. For the fiducial, $M_{crit} = 10^{10}$ M$_\odot$ and high efficiency models, the build up of stellar mass is faster (steeper slope), but once the corresponding critical halo mass is reached, the stellar mass plateaus and the halo hardly produces any additional stellar mass. The high efficiency model has the same shape as the fiducial model, but a higher normalization. As expected, the No SN and No AGN models build up more stellar mass before and after the halo reaches critical halo mass, respectively.

stellar mass is faster (steeper slope). Once the critical halo mass is reached ($M_{crit} = 10^{12}$ M$_\odot$, corresponding to $M_* \approx 10^{10}$ M$_\odot$ for this model), the stellar mass plateaus. The No AGN model has a similar behaviour at early times, but once the critical halo mass is reached, star formation does not slow down and the halo reaches a higher stellar mass at the present time. On the other hand, the No SN model produces much more stellar mass at early times, but once the critical halo mass is reached, star formation slows down, and the halo reaches a similar final stellar mass as the fiducial model. The $M_{crit} = 10^{10}$ M$_\odot$ model presents a similar behaviour to the fiducial model, i.e. once the halo reaches the critical mass it hardly produces any additional stellar mass. However, as the critical mass is lower for this model, the transition happens at earlier times. The high efficiency model has the same shape as the fiducial model, but as expected, a higher normalization.

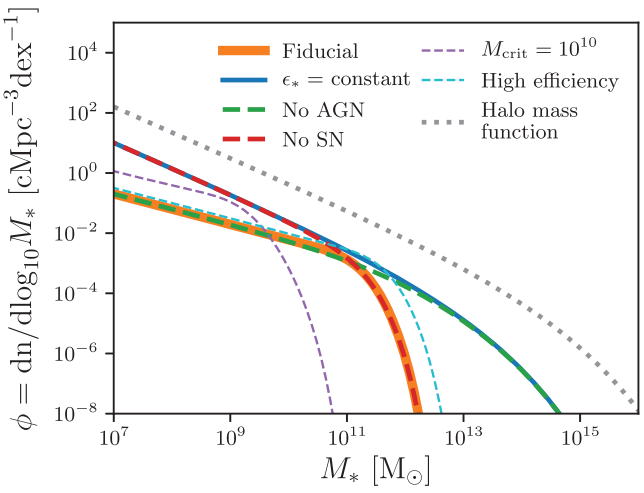

Figure 8. The GSMF at the present time for the six efficiency models described in Table 1. All models with $\beta = 0.75$, i.e. at high masses cooling is suppressed by AGN feedback, exhibit a sharp cut-off at the critical halo mass (Fiducial, No SN, $M_{crit} = 10^{10}$ M$_\odot$ and high efficiency). This shows that AGN feedback is mainly responsible for the characteristic knee of the GSMF. The location of the knee is determined both by the critical halo mass in the star formation efficiency, and the normalization ϵ_N, as this causes a horizontal shift of the whole distribution. For the No SN model, the slope of the faint end of the GSMF is much steeper. If feedback process are inefficient both at the low mass and high mass end (a constant fraction of the baryon budget is turned into stars), the GSMF is identical to the halo mass function (dotted grey line) but shifted in mass by a constant value.

3.2 The stellar mass function

We use equations (2), (5), and (13) to calculate the GSMF at redshift $z = 0$ for the six efficiency models. We note that for a time evolving efficiency, equation (16) should be used to calculate the stellar mass of any halo as a function of time. These equations allow us to obtain the SHMR and convolve it with the halo mass function to calculate the GSMF.

Fig. 8 shows the GSMF at the present time for the six efficiency models. For the constant model, as it has been pointed out before (e.g Benson et al. 2003), if feedback process are inefficient both at the low mass and high mass end, i.e. a constant fraction of the baryon budget is turned into stars in every halo, the GSMF does not exhibit the characteristic knee obtained in observations and is identical to the halo mass function (dotted grey line) but shifted in mass by a constant value. Once feedback processes are implemented, the location of the knee of the GSMF is determined by the critical mass in the star formation efficiency (fiducial, $M_{crit} = 10^{10}$). Changing the critical mass also changes the normalization of the distribution. All models with $\beta = 0.75$, i.e. at high masses cooling is suppressed by AGN feedback, exhibit a sharp cut-off at the transition mass. Hence, AGN feedback is mainly responsible for the knee of the GSMF. The No AGN model has the same shallow slope at the faint end of the GSM function as the fiducial model, with a slight bend at high masses driven only by the exponential cut-off of the halo mass function. The No SN model presents the same knee as the fiducial model, but the slope of the faint end of the GSM function is much steeper. As discussed in Section 2, the low-mass and high-mass end slopes of the SHMR produce a 'bump' at the knee of the GSMF. Finally, the high efficiency model, as α and β are the same as for the fiducial model, the shape of the GSMF is the same. i.e. the relative abundance of galaxies to their haloes, and hence the shape, is independent of the normalization (as ε does not depend on ϵ_N in

equation 7). For a given halo mass, changing the normalization maps that halo mass to a different galaxy mass. Hence, a change in the normalization, ϵ_N, shifts the whole distribution only horizontally. In this case, the high efficiency model shifts the GSMF to the right compared to the fiducial model.

3.3 The cosmic SFR density

The cosmic history of star formation is perhaps one of the most fundamental observables of our Universe. It has been observed to peak approximately 3.5 Gyr after the big bang ($z \approx 2$), and decline exponentially thereafter (for a review see Madau & Dickinson 2014). Different groups have tried to model the complex physics driving the cosmic SFR by using, for example, full hydrodynamical simulations (e.g. Schaye et al. 2015; Davé et al. 2016; Dubois et al. 2016; Pillepich et al. 2018a). Our analytic model disentangles the role of cosmology from the role of astrophysics, which in turn, allows us to examine the effect of the different efficiency parameters on the cosmic SFR density.

We begin by noting that the behaviour of equation (22) is governed by two main factors. First, a multiplier term that originates from both, the halo accretion rate, and the halo mass function, and is $\propto t^{-7/3}$. This, comes from the dynamical time-scale of the universe getting larger. Secondly, an exponential term contribution due to the build up of haloes in the halo mass functions that is $\propto e^{-t^{-4/3}}$. For a given halo mass then, the exponential term dominates at early times, and the contribution to the cosmic SFR density is driven by the exponential build up of haloes. At late times, the exponential term asymptotically tends to a constant value, and the further evolution of the cosmic SFR is dominated by the multiplier term, i.e. it behaves as a power law. As discussed in Salcido et al. (2018), the contribution of dark energy is only relevant at late times, and at its observed value, it has a negligible impact on the star formation in the Universe.

Fig. 9 shows the integrated cosmic SFR density for the six efficiency models computed using equation (22). For the fiducial model, while smaller haloes are more abundant than large objects, a smaller fraction of the inflowing material is converted into stars. As a result, the SFR density is dominated by the largest haloes in which star formation is able to proceed without generating efficient feedback. The smaller haloes only contribute significantly at very early times, when the abundance of larger objects is strongly suppressed by the exponential term in the mass function. We see therefore that the contribution of haloes of mass $\approx M_{\rm crit} = 10^{12}\,{\rm M}_\odot$, is representative of most of the SFR in the model.

If star formation is efficient at all halo masses (constant model), then the cosmic SFR behaves like a power law with time, which only deviates from this behaviour at late times due to the suppression due to the cosmological constant.

Examining the No SN model reveals the origin of the peak in the cosmic history of star formation is the efficient feedback in low mass galaxies. Without a mechanism to suppress star formation in small haloes, the history of the cosmic SFR density would not have its characteristic peak. Supernovae feedback is then mainly responsible for shaping the cosmic SFR density of the Universe. On the other hand, examining the No AGN model reveals that efficient feedback in high mass haloes only has a moderate effect on shaping the cosmic star formation. Without a mechanism to prevent star formation in massive galaxies, the cosmic SFR density would still exhibit a peak, only changing mildly its amplitude and localization. However, the slope of the decline would be similar (orange versus green dashed lines).

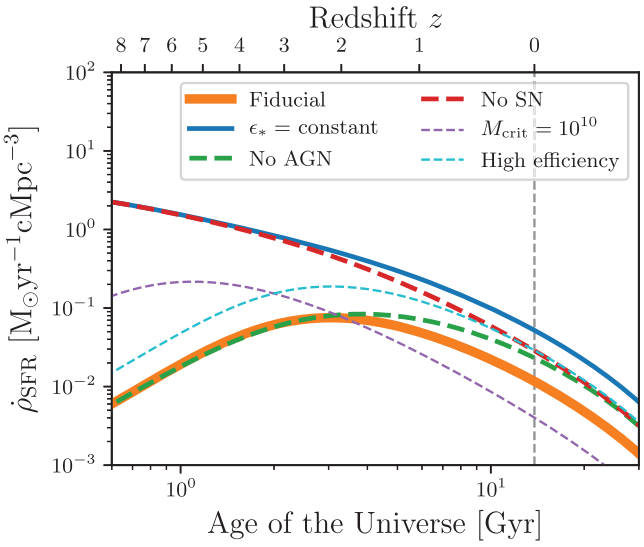

Figure 9. The cosmic SFR density for the six efficiency models. As small haloes only contribute significantly at very early times, when the abundance of larger objects is strongly suppressed by the exponential term in the mass function, all models with $\alpha = 0.75$, i.e. at low masses the SFR is suppressed because of the efficiency of stellar feedback, exhibit the characteristic peak in the observed cosmic history of star formation (Fiducial, No AGN, $M_{\rm crit} = 10^{10}\,{\rm M}_\odot$ and high efficiency). On the other hand, both models with $\alpha = 0$ (constant and No SN), do not exhibit the peak. This shows that supernovae feedback is mainly responsible for shaping the cosmic SFR density of the Universe. The figure shows that changing the transition mass $M_{\rm crit}$ has a great impact on the localization and normalization of the SFR peak, i.e. the SFR density is dominated by the largest haloes in which star formation is able to proceed without generating efficient feedback. AGN feedback only has a moderate effect on shaping the cosmic star formation, changing mildly its amplitude and localization (No AGN model). The high efficiency model the same shape as the fiducial model, but with a higher normalization.

Changing the transition mass $M_{\rm crit}$ has a great impact on the localization of the SFR peak. As in the fiducial model, the contribution of haloes of mass $\approx M_{\rm crit}$ is representative of most of the SFR in the $M_{\rm crit} = 10^{10}$. Hence, the peak happens at earlier times, but also has a higher normalization, as $10^{10}\,{\rm M}_\odot$ haloes are more abundant than $10^{12}\,{\rm M}_\odot$ haloes.

Finally for the high efficiency model, the shape of the SFR of the Universe is identical to the fiducial model, but with a higher normalization.

4 FITTING OBSERVATIONS

In this section we compare the galaxy formation outputs from our analytic model with different observational data sets. We begin by calibrating our model to reproduce the GSMF at $z \sim 0$ using observations from the Galaxy And Mass Assembly (GAMA) survey (Baldry et al. 2012) and the Sloan Digital Sky Survey (SDSS) (Moustakas et al. 2013).[3] We use the reduced chi-squared statistic to derive the best-fitting effective star formation efficiency $\epsilon_*(M_h)$. Because the model is fully analytic, this calibration process is fast and easy to perform. Fig. 10 shows the best-fitting model in orange.

[3]In this paper we used the standardized GSMF data from Behroozi et al. (2019), which assumes a Chabrier (2003) IMF, a Bruzual & Charlot (2003) stellar population synthesis model, dust corrections from Calzetti et al. (2000), and UV-stellar mass corrections.

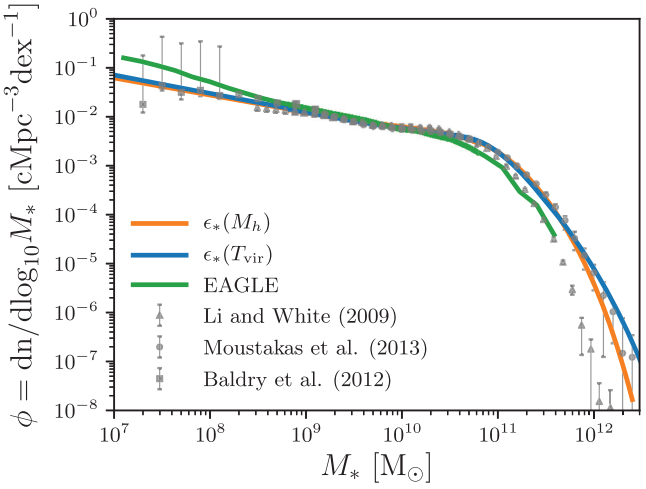

Figure 10. Redshift $z = 0.1$ GSMF for the best-fitting parameters for the halo mass-dependent model ($\epsilon_*(M_h)$) and the virial temperature-dependent model ($\epsilon_*(T_{\rm vir})$). Observational data with their associated uncertainties from Li & White (2009), Baldry et al. (2012), Moustakas et al. (2013) are shown with symbols. Both efficiency models provide a good fit to the present-day GSMF (see also the reduced χ^2 statistics in Table 2). Results from the EAGLE reference simulation are shown in green for reference.

Table 2. Best-fitting parameters for the halo mass-dependent (*Model I*, $\epsilon_*(M_h)$), and virial temperature-dependent (*Model II*, $\epsilon_*(T_{\rm vir})$) star formation efficiency models. For the virial temperature-dependent model, $M_{\rm crit}$ is given at redshift $z = 0$, which corresponds to a critical virial temperature $T_{\rm crit} = 10^{5.3}$ K. As $T_{\rm crit}$ is kept constant, $M_{\rm crit} \propto a(t)^{3/2}$ (see Section 2.1). χ^2_ν is the reduced chi-squared statistic used for goodness of fit testing.

Model	$\epsilon_{\rm N}$	$M_{\rm crit}$ ($\mathrm{M_\odot}$)	α	β	χ^2_ν
Mass-dependent	0.178	$10^{11.68}$	1.537	0.656	1.5
Temperature-dependent	0.140	10^{12}	2.377	0.834	1.6

Results from the EAGLE reference simulation are shown in green for reference. The figure shows that a constant halo mass-dependent efficiency model provides an excellent fit to the present-day GSMF (with reduced $\chi^2_\nu = 1.5$). The best best-fitting efficiency parameters are shown in Table 2.

4.1 Contrasting halo mass and virial temperature efficiency models

Having established the best-fitting efficiency parameters for the model, we can study the evolution of the model outputs. By construction, $\epsilon_*(M_h)$ is only a function of halo mass and is fixed in time. Hence, the evolution of the GSFM depends only on the evolution of the abundance of haloes of mass M_h as a function of time, as described by the halo mass function. Fig. 11 shows the evolution of the predicted GSMF for the halo mass-dependent star formation efficiency model in dashed lines. Different panels and colours represent different redshifts. Observational data from Baldry et al. (2012), Moustakas et al. (2013), Tomczak et al. (2014), Ilbert et al. (2013), Muzzin et al. (2013), and Song et al. (2016)[4] are shown with coloured symbols.

[4]In this paper we used the standardized GSMF data from Behroozi et al. (2019), which assumes a Chabrier (2003) IMF, a Bruzual & Charlot (2003) stellar population synthesis model, dust corrections from Calzetti et al. (2000), and UV-stellar mass corrections.

Remarkably, a simple halo mass-dependent efficiency model reproduces very well the evolution of the GSMF up to redshift $z \approx 4$. While the observed data at higher redshifts are highly uncertain, the halo mass-dependent model significantly underpredicts the abundance of distant galaxies. This may hint to the need of a time-evolving efficiency model. As discussed in Section 2.1, it is reasonable to assume that the efficiency of star formation should be a function of the halo's virial temperature, which naturally evolves with cosmic time. In this section, we investigate an effective star formation efficiency model that depends on virial temperate.

4.1.1 *The three stages of galaxy formation*

An effective star formation efficiency model characterized by a time-independent critical virial temperature $T_{\rm crit}$ assumes that there exists a critical halo virial temperature at which there is a transition from where star formation driven outflows can escape, to where outflows stall inside the halo. Using the viral temperature of the halo to parametrize this tipping point, provides a natural evolution of the star formation efficiency. For a fixed halo mass, early collapsed haloes are more compact (denser), and one might expect a higher efficiency (for haloes with $T_{\rm vir} < T_{\rm crit}$).

In this simple picture, we can distinguish three stages of parametrize formation,[5] characterized by the virial temperature of the halo:

(i) **Stellar feedback regulated stage:** Star formation driven outflows effectively regulate the gas content of galaxies residing in haloes with virial temperature $T_{\rm vir} < T_{\rm crit}$. In this stage, efficient outflows prevent the density of central star forming gas building up, suppressing the growth of the central BH.

(ii) **Efficient star forming/rapid growing black hole stage:** As haloes grow, the virial temperature increases to the point that the stellar outflows are no longer buoyant relative to their surroundings, and therefore stall (i.e $T_{\rm vir} \approx T_{\rm crit}$). The density of gas builds up within the halo triggering high SFRs and rapid BH growth.

(iii) **Black hole feedback regulated stage:** In haloes with $T_{\rm vir} > T_{\rm crit}$, the central BH is massive enough to produce efficient AGN feedback, in turn, regulating the gas content of the halo and preventing further star formation.

An additional advantage of using the virial temperature to characterize the star formation efficiency is that we can add a proxy for the effect of cosmic reionization. Ultraviolet radiation from the first stars formed reionized neutral hydrogen, raising its entropy to a temperature of $\approx 10^4$ K. This process prevented further cooling, hence preventing star formation in haloes with $T_{\rm vir} < 10^4$ K (Doroshkevich, Zel'dovich & Novikov 1967; Couchman & Rees 1986; Rees 1986; Efstathiou 1992; Loeb & Barkana 2001). As a result of this suppression of star formation, only a fraction of the haloes with present-day mass $\approx 10^{10}\,\mathrm{M_\odot}$ form a galaxy, and no galaxies form below a halo mass of $\approx 10^7\,\mathrm{M_\odot}$ (Sawala et al. 2013, 2016; Fitts et al. 2017; Bose, Deason & Frenk 2018). We therefore include the effect of reionization by setting $\epsilon_*(T_{\rm vir} < 10^4\,\mathrm{K}) = 0$.

[5]While perhaps closely related, we distinguish these three *stages* of galaxy formation from the three *phases* in Clauwens et al. (2018), as the latter refer mainly to a morphological evolution, rather than the entropy state and buoyancy of the gas.

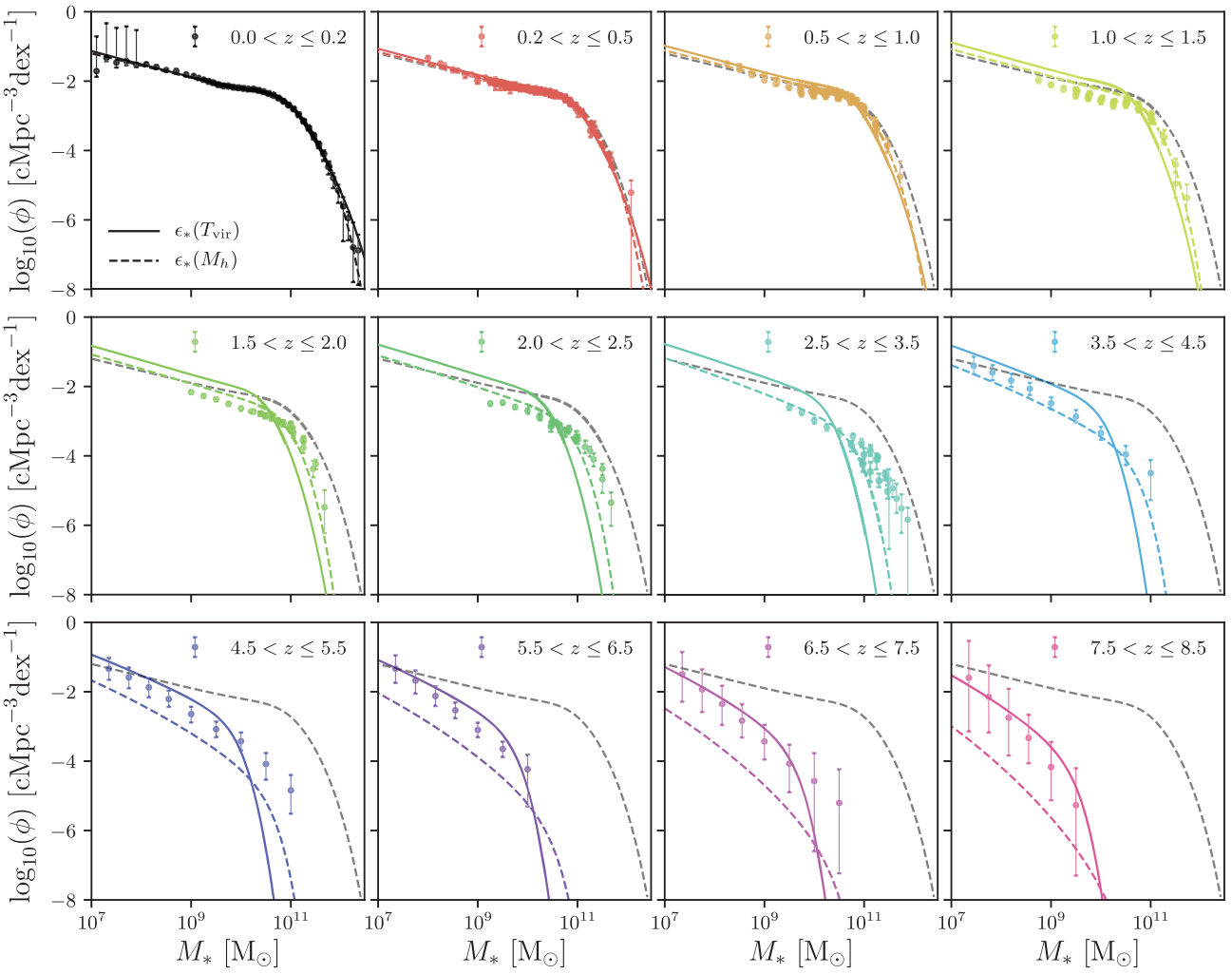

Figure 11. Evolution of the predicted GSMF for the halo mass-dependent, and the virial temperature-dependent star formation efficiency models. Different panels and colours represent different redshifts. Observational data with their associated uncertainties from Baldry et al. (2012), Moustakas et al. (2013), Tomczak et al. (2014), Ilbert et al. (2013), Muzzin et al. (2013), Song et al. (2016) are shown by coloured symbols. The halo mass-dependent model is shown in dashed lines ($\epsilon_*(M_h)$). The virial temperature-dependent model is shown in solid lines ($\epsilon_*(T_{\rm vir})$). The redshift $z \sim 0$ halo mass-dependent model is reproduced in each panel as a grey dashed curve to highlight the evolution. While both models have been calibrated to reproduce the GSMF at $z \sim 0$, the halo mass-dependent model reproduces very well the evolution of the GSMF up to redshift $z = 4$, but significantly underpredicts the abundance of galaxies at higher redshift. On the other hand, the virial temperature model provides a good fit both at low and high redshift, but the evolution is too rapid at intermediate redshift ($z = 1$ to $z = 4$).

Of course, one could think of more complex ways in which the expected star formation efficiency might evolve with cosmic time. For instance, the evolution of cooling versus free-fall time of a cloud of gas, the evolution of metallicity, and the UV background radiation, might all result in a more complex evolution. However, the aim here is to describe the main features of the universe as simply as possible, and so we leave exploration of more complex models for future work.

We calibrate the $\epsilon_*(T_{\rm vir})$ model to the GSMF at $z = 0$ using the reduced chi-squared statistic to derive the best-fitting parameters. The best fit efficiency parameters are shown in Table 2.

It is important to highlight that the models were calibrated to reproduce only the observed GSMF at redshift $z \sim 0$. Fig. 10 shows the best-fitting virial temperature model in blue. The figure shows that both the halo mass-dependent efficiency ϵ_*, and the virial temperature-dependent efficiency $\epsilon_*(T_{\rm vir})$ models, provide a good fit to the present-day GSMF, both at the faint end and at the knee. Fig. 11 shows that a star formation efficiency as a function of the virial temperature of the halo provides a good fit to the abundance of galaxies both at low and high redshift, but the evolution is too rapid at intermediate redshift ($z = 1$ to $z = 4$).

Fig. 12 shows the build up of the stellar mass within haloes using both models (calculated integrating equation 16). Colour coding represents different present-day halo mass $M_h(t_0)$. The transition of the star formation efficiency at $M_{\rm crit}$ can be clearly seen in very massive haloes, where there is a rapid rise of stellar mass, then, when the halo reaches the critical mass (or virial temperature), the build up of stellar mass slows down significantly. The change in slope is due to AGN feedback becoming efficient in those haloes, preventing any further star formation.

Fig. 13 shows the predicted cosmic SFR for the two efficiency models. Using the analytic model, we can clearly see the contribution to the integrated SFR density from dark matter haloes of different masses (per dex) shown as coloured dashed lines (only

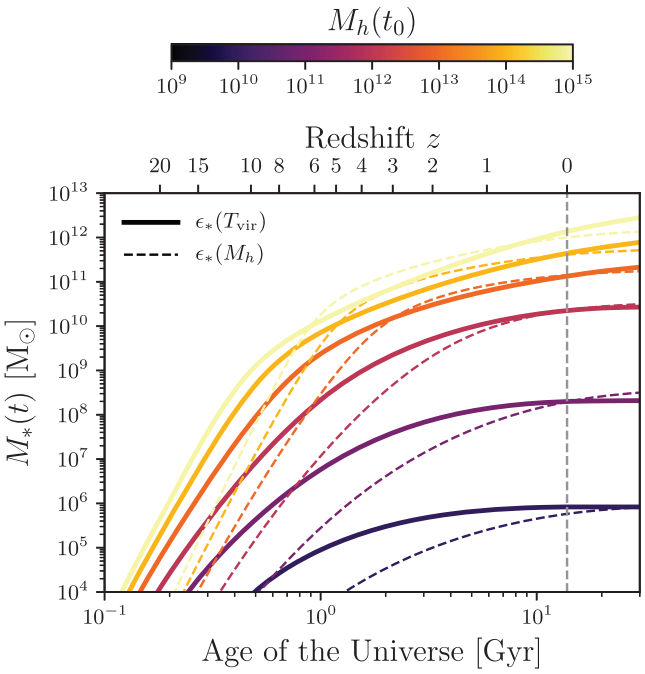

Figure 12. Evolution of the stellar mass within haloes using the best-fitting parameters for both models (Table 2). The halo mass-dependent model is shown in dashed lines, while the virial temperature efficiency model is shown in solid lines. Colour coding represents different present-day halo masses $M_h(t_0)$. Once massive haloes reach the critical mass, the build up of stellar mass slows down significantly, i.e. the change in slope of the curves is due to AGN feedback becoming effective in those haloes.

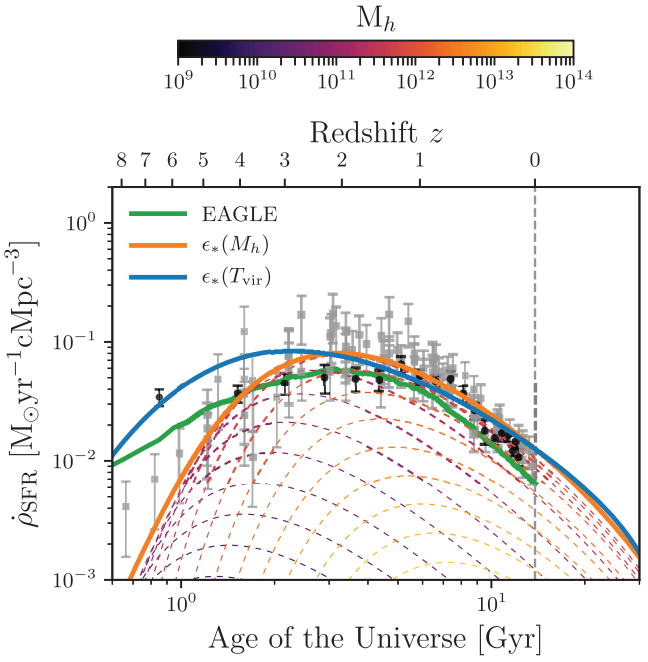

Figure 13. The predicted SFR history of the Universe for the two efficiency models presented in this paper. Coloured lines show the contributions from dark matter haloes of different masses (per dex), using the star formation efficiency described by equation (22), and using the virial temperature efficiency model. The total SFR for the virial temperature efficiency model is shown in blue. The time-independent efficiency model is shown in orange. Results from the EAGLE simulation are shown in green for reference. Observational data compiled by Behroozi et al. (2013) are shown as grey symbols. Observational data from Driver et al. (2018) are shown as black symbols. The analytic model using a halo mass-dependent star formation efficiency reproduces the amplitude and shape of the cosmic SFR density remarkably well.

shown for the halo mass-dependent efficiency model). The total SFR for the virial temperature efficiency model is shown in blue. The halo mass-dependent efficiency model is shown in orange. Results from the EAGLE simulation are shown in green for reference. Observational data compiled by Behroozi et al. (2013) are shown as grey symbols. The latest observational results from the GAMA survey from Driver et al. (2018) are shown as black symbols. The model using a halo mass efficiency reproduces the amplitude and shape of the observed SFR density remarkably well, while the virial temperature-dependent efficiency model produces a higher SFR at high redshift.

Fig. 14 shows the predicted SHMR from both efficiency models. Colour coding represents different redshifts. The virial temperature efficiency model is shown in solid lines. The halo mass-dependent efficiency model is shown with a dashed line (only shown for $z = 0$ as the halo mass-dependent efficiency model is constant in time). The critical halo mass predicted in Bower et al. (2017) is shown in vertical dotted lines, which roughly coincide with the peak efficiency for a viral temperature efficiency model. Recently, McAlpine et al. (2018) showed that the critical halo mass predicted in Bower et al. (2017) agrees remarkably well with the triggering of a rapid black hole growth stage in the EAGLE simulations.

The model using a virial temperature efficiency predicts an SHMR relation that differs from observational contains using abundance and clustering properties of galaxy samples with predictions from a phenomenological halo models. For example, recently Cowley et al. (2018) calculated that the peak of the SHMR shifts to higher masses at earlier times. These methods however, depend heavily on the underlying modelling and assumptions. More sophisticated

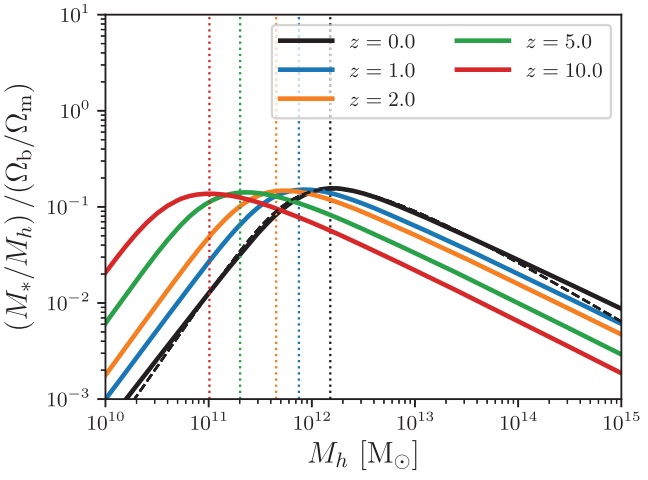

Figure 14. Predicted SHMR for the time-independent, and evolving star formation efficiency models. Colour coding represents different observed redshifts. The halo mass-dependent model is shown as a dashed black line. The vertical dotted lines indicate the critical mass derived in Bower et al. (2017), which tracks the triggering of a rapid black hole growth stage in the EAGLE simulations (McAlpine et al. 2018).

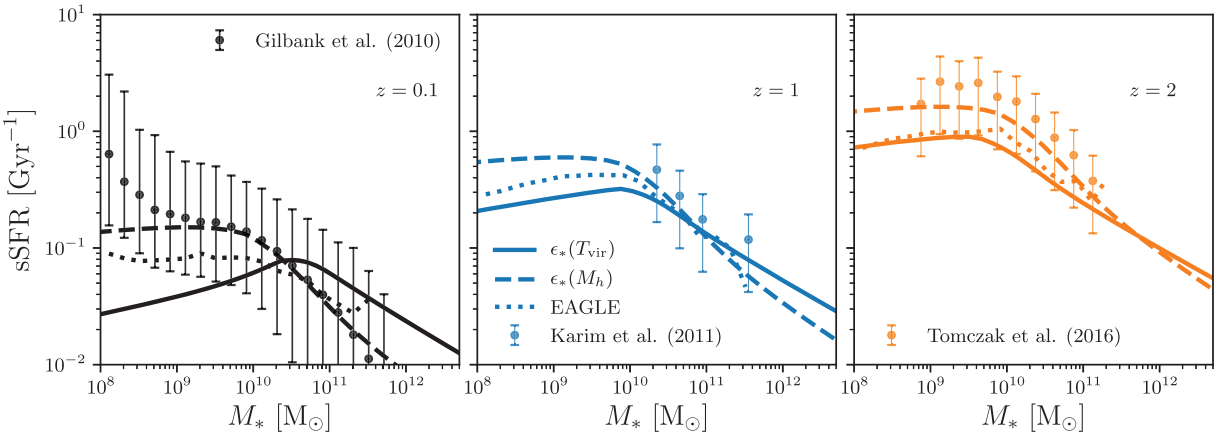

Figure 15. The sSFR of galaxies at different redshifts. The model using a virial temperature efficiency is shown in solid lines. The halo mass-dependent model is shown in dashed lines. Results from the EAGLE simulations are shown in dotted lines for reference. Observational data from Gilbank et al. (2010), Karim et al. (2011), Tomczak et al. (2016) are shown as symbols. The halo mass-dependent model is in remarkable agreement with observational data sets.

empirical models (e.g. Moster et al. 2018; Behroozi et al. 2019) find that the peak in the SHMR moves first to higher masses for low redshifts, and then to lower masses at high redshifts.

Finally, in Fig. 15 we show the sSFR of galaxies for different redshifts. The halo mass-dependent model is shown in dashed lines. The model using a virial temperature efficiency is shown in solid lines. Results for central galaxies from the EAGLE simulations are shown in dotted lines for reference. Observational data from Gilbank et al. (2010), Karim et al. (2011), and Tomczak et al. (2016) are shown as symbols. While not calibrated to reproduce the sSFR of galaxies, the agreement of the halo mass-dependent model with the observational data is remarkable.

5 DISCUSSION AND CONCLUSIONS

In our current paradigm of galaxy formation, every galaxy forms within a dark matter halo. Due to the tight correlation observed between the properties of galaxies and their host haloes, it is natural to expect that individual galaxy assembly could be correlated with halo assembly (see Wechsler & Tinker 2018 for a review).

In this paper we developed a fully analytic model of galaxy formation that connects the growth of dark matter haloes in a cosmological background, with the build up of stellar mass within these haloes. The model restricts the role of baryonic astrophysics to setting the relation between galaxies and their haloes. We assume an effective star formation efficiency which captures all the physical processes involved in the conversions of gas into stars, i.e. cooling, star formation law, feedback mechanisms, etc.

We show that galaxy formation is revealed as a simple process where the effective star formation efficiency within haloes is only a function of their mass. We show that all the complex physics of galaxy formation, the interplay between cosmology and baryonic process can be understood as a simple set of equations. Despite its simplicity, the model reproduces self-consistently the shape and evolution of the cosmic SFR density, the sSFR of galaxies, and the GSMF, both at the present time and at high redshift.

We use our model to investigate the origin of the characteristic shape of the GSMF and the need for a double Schechter function to describe it. Using the logarithmic slope of the SHMR, the model naturally explains an inflection point in the distribution causing the characteristic 'bump' observed at the knee of the GSMF.

To demonstrate the flexibility and power of our mathematical framework, we introduced a physically motived model for the effective star formation efficiency, characterized by a time-independent critical virial temperature, $T_{\rm crit}$. The model assumes that there exists a critical halo virial temperature at which there is a transition from where star formation driven outflows can escape, to where outflows stall, triggering high SFRs and rapid BH growth. We demonstrate that this model can reproduce the GSMF at high redshift ($z > 4$) better than a simple halo mass-dependent model, but the evolution at intermediate redshifts is to rapid to reproduce observations.

While the aim of this paper is not to present a 'perfect' model fitted to reproduce a large set of observational constraints, the two variations of an effective star formation efficiency presented here, already provide very valuable information about the average evolution of the galaxy population within a cosmological background. Furthermore, the model can be easily extended to include further modelling (such as time evolution of the model parameters, or a prescription for satellite galaxies, e.g. Grylls et al. 2019) or the use of advanced gradient-based minimization and Markov Chain Monte Carlo algorithms to fit to a larger number of data sets. Additionally, the model can be easily adapted to combine the equations developed here, with for example, halo merger trees from a dark matter simulation.

Our model is limited to the connection between haloes and central galaxies only. Sub-haloes and satellite galaxies are subject to complex processes, such as tidal and ram pressure stripping, which are not included.

Finally, one of the main advantages of the model is that by providing a set of analytic equations, the model can be easily 'inverted' and allows for rapid experiments to be conducted, providing a great tool to explore the differential effects of baryonic physics, averaged over galaxy scales. We conclude therefore that there is a clear opportunity to use the analytic model developed in this paper to improve theoretical galaxy formation models.

ACKNOWLEDGEMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 769130). This work was supported by the Science and Technology Facilities

Council [ST/P000541/1]. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1 and ST/R002371/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

REFERENCES

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945
Baldry I. K. et al., 2012, MNRAS, 421, 621
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Benson A. J., 2012, New Astron., 17, 175
Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bose S., Deason A. J., Frenk C. S., 2018, ApJ, 863, 123
Bouché N. et al., 2010, ApJ, 718, 1001
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Bower R. G., Schaye J., Frenk C. S., Theuns T., Schaller M., Crain R. A., McAlpine S., 2017, MNRAS, 465, 32
Brainerd T. G., Specian M. A., 2003, ApJ, 593, L7
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Chabrier G., 2003, PASP, 115, 763
Clauwens B., Schaye J., Franx M., Bower R. G., 2018, MNRAS, 478, 3994
Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015, MNRAS, 450, 1514
Couchman H. M. P., Rees M. J., 1986, MNRAS, 221, 53
Cowley W. I. et al., 2018, ApJ, 853, 69
Crain R. A. et al., 2015, MNRAS, 450, 1937
Davé R., Finlator K., Oppenheimer B. D., 2012a, MNRAS, 421, 98
Davé R., Finlator K., Oppenheimer B. D., 2012b, MNRAS, 421, 98
Davé R., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265
Dekel A., Silk J., 1986, ApJ, 303, 39
Diemer B., More S., Kravtsov A. V., 2013, ApJ, 766, 25
Doroshkevich A. G., Zel'dovich Y. B., Novikov I. D., 1967, Sov. Astron., 11, 233
Driver S. P. et al., 2018, MNRAS, 475, 2891
Dubois Y., Volonteri M., Silk J., Devriendt J., Slyz A., Teyssier R., 2015, MNRAS, 452, 1502
Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, MNRAS, 463, 3948
Efstathiou G., 1992, MNRAS, 256, 43P
Finlator K., Davé R., 2008, MNRAS, 385, 2181
Fitts A. et al., 2017, MNRAS, 471, 3547
Fosalba P., Crocce M., Gaztañaga E., Castander F. J., 2015, MNRAS, 448, 2987
Gilbank D. G., Baldry I. K., Balogh M. L., Glazebrook K., Bower R. G., 2010, MNRAS, 405, 2594
Grylls P. J., Shankar F., Zanisi L., Bernardi M., 2019, MNRAS, 483, 2506
Haas M. R., Schaye J., Booth C. M., Dalla Vecchia C., Springel V., Theuns T., Wiersma R. P. C., 2013, MNRAS, 435, 2931
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Hoekstra H., Yee H. K. C., Gladders M. D., 2004, ApJ, 606, 67
Hudson M. J. et al., 2015, MNRAS, 447, 298
Ilbert O. et al., 2013, A&A, 556, A55
Karim A. et al., 2011, ApJ, 730, 61
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Lacey C. G. et al., 2016, MNRAS, 462, 3854
Lan T.-W., Ménard B., Mo H., 2016, MNRAS, 459, 3998
Li C., White S. D. M., 2009, MNRAS, 398, 2177
Loeb A., Barkana R., 2001, ARA&A, 39, 19
Madau P., Dickinson M., 2014, ARA&A, 52, 415
McAlpine S. et al., 2016, Astron. Comput., 15, 72
McAlpine S., Bower R. G., Rosario D. J., Crain R. A., Schaye J., Theuns T., 2018, MNRAS, 481, 3118
Mitchell P. D., Lacey C. G., Baugh C. M., Cole S., 2016, MNRAS, 456, 1459
Moster B. P., Somerville R. S., Maulbetsch C., Bosch F. C. v. d., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Moster B. P., Naab T., White S. D. M., 2018, MNRAS, 477, 1822
Moustakas J. et al., 2013, ApJ, 767, 50
Muzzin A. et al., 2013, ApJ, 777, 18
Naab T., Ostriker J. P., 2017, ARA&A, 55, 59
Neistein E., van den Bosch F. C., Dekel A., 2006, MNRAS, 372, 933
Neyman J., Scott E. L., 1952, ApJ, 116, 144
Norberg P., Frenk C. S., Cole S., 2008, MNRAS, 383, 646
Pillepich A. et al., 2018a, MNRAS, 473, 4077
Pillepich A. et al., 2018b, MNRAS, 475, 648
Planck Collaboration XVI, 2014, A&A, 571, A16
Planck Collaboration XIII, 2016, A&A, 594, A13
Press W. H., Schechter P., 1974, ApJ, 187, 425
Qu Y. et al., 2017, MNRAS, 464, 1659
Rees M. J., 1986, MNRAS, 218, 25P
Rodriguez-Gomez V. et al., 2016, MNRAS, 458, 2371
Rodríguez-Puebla A., Primack J. R., Behroozi P., Faber S. M., 2016, MNRAS, 455, 2592
Salcido J. et al., 2018, MNRAS, 477, 3744
Sawala T., Frenk C. S., Crain R. A., Jenkins A., Schaye J., Theuns T., Zavala J., 2013, MNRAS, 431, 1366
Sawala T. et al., 2016, MNRAS, 456, 85
Schaye J. et al., 2010, MNRAS, 402, 1536
Schaye J. et al., 2015, MNRAS, 446, 521
Schechter P., 1976, ApJ, 203, 297
Schmidt M., 1963, ApJ, 137, 758
Sharma M., Theuns T., 2019, preprint (arXiv:1906.10135)
Somerville R. S., Davé R., 2015, ARA&A, 53, 51
Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481
Song M. et al., 2016, ApJ, 825, 5
Springel V. et al., 2005, Nature, 435, 629
Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, ApJ, 868, 92
Tomczak A. R. et al., 2014, ApJ, 783, 85
Tomczak A. R. et al., 2016, ApJ, 817, 118
Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, ApJ, 742, 16
van den Bosch F. C., Norberg P., Mo H. J., Yang X., 2004, MNRAS, 352, 1302
van den Bosch F. C., Ogiya G., Hahn O., Burkert A., 2018, MNRAS, 474, 3043
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435
Wetzel A. R., Nagai D., 2015, ApJ, 808, 40
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
Yang X., Mo H. J., van den Bosch F. C., 2009, ApJ, 695, 900
Yang X., Mo H. J., van den Bosch F. C., Zhang Y., Han J., 2012, ApJ, 752, 41
Zaritsky D., Smith R., Frenk C., White S. D. M., 1993, ApJ, 405, 464

APPENDIX A: DERIVATION OF THE MODEL

A1 Cosmological expansion

Here we provide a brief summary of the analytic solution of the Friedmann equation developed in Salcido et al. (2018). Equations with full cosmology dependence of the numerical constants are highlighted using a coloured superscript ($^{c^*}$), and can be used for arbitrary flat ΛCDM cosmologies.

Using a Taylor expansion, the expansion factor of the Universe can be written as,

$$a^{c^*}(t) \propto \left[\frac{3}{2}\frac{t}{t_m}\right]^{2/3}\left(1+\frac{1}{4}\left(\frac{t}{t_\Lambda}\right)^2+\frac{1}{80}\left(\frac{t}{t_\Lambda}\right)^4+...\right), \quad \text{(A1)}$$

where the *matter time-scale* is given by,

$$t_{\rm m}^{c^*}=\sqrt{\frac{3}{8\pi G\rho_0}}=\frac{1}{H_0\sqrt{\Omega_{\rm m,0}}}, \quad \text{(A2)}$$

and the *dark energy time-scale* is given by,

$$t_\Lambda^{c^*}=\sqrt{\frac{3}{\Lambda c^2}}=\frac{1}{H_0\sqrt{\Omega_{\Lambda,0}}}. \quad \text{(A3)}$$

For the cosmological parameters given at the end of Section 1, $t_{\rm m}=26.04\,{\rm Gyr}$ and $t_\Lambda=17.33\,{\rm Gyr}$. At the present day, $t\equiv t_0=13.82\,{\rm Gyr}$, so that $\frac{t_0}{t_m}=0.53$ and $\frac{t_0}{t_\Lambda}=0.8$. By convention, equation (A1) is normalized so that $a(t_0)=1$.

A2 The growth of density perturbations and the halo accretion rates

Dark matter structures are assumed to have grown from small initial density perturbations. Expressing the density, ρ, in terms of the density perturbation contrast against a density background,

$$\rho(\mathbf{x},t)=\bar\rho(t)[1+\delta(\mathbf{x},t)], \quad \text{(A4)}$$

the differential equation that governs the time dependence of the growth of linear perturbations in a pressureless fluid, such as e.g. dark matter, can be written as

$$\frac{{\rm d}^2\delta}{{\rm d}t^2}+2\frac{\dot a}{a}\frac{{\rm d}\delta}{{\rm d}t}-4\pi G\bar\rho\delta=0. \quad \text{(A5)}$$

The growing mode of equation (A5) can be written as,

$$\delta(t)=D(t)\delta(t_0), \quad \text{(A6)}$$

where $D(t)$ is the linear growth factor, which determines the normalization of the linear matter power spectrum relative to the initial density perturbation power spectrum, and is computed by the integral

$$D^{c^*}(t)\propto\frac{\dot a}{a}\int_0^t\frac{{\rm d}t'}{\dot a^2(t')}. \quad \text{(A7)}$$

Using the power-series approximation for $a(t)$ from equation (A1), keeping the leading order terms and using the definition of f_Λ in equation (11), we can obtain an analytic solution of equation (A7),

$$D^{c^*}(t)=\left[\frac{3}{2}\frac{t}{t_m}\right]^{2/3}\frac{2}{5}t_m^2K_D\,f_\Lambda(t,-0.16,0.04), \quad \text{(A8)}$$

where K_D is a normalization constant with units of time^{-2}. By convention, K_D is chosen so that $D(t_0)=1$. For the cosmological parameters inferred by the Planck Collaboration XVI (2014), $K_D=4.7\times10^{-3}\,{\rm Gyr}^{-2}$. Collecting the numerical and cosmology-dependent constants together gives,

$$D(t)\approx1.671\left[\frac{t}{t_m}\right]^{2/3}f_\Lambda(t,-0.16,0.04). \quad \text{(A9)}$$

The growth rates of linear perturbations do not directly predict the growth rates of haloes; however, we can directly connect the two through the approach developed by Press & Schechter (1974). Correa et al. (2015) showed that the accretion rates of haloes can be written as (see also Neistein, van den Bosch & Dekel 2006),

$$\left(\frac{1}{M_h}\frac{{\rm d}M_h}{{\rm d}t}\right)^{c^*}=\sqrt{\frac{2}{\pi}}\frac{(\delta_c/D)}{S^{1/2}(q^\gamma-1)^{1/2}}\frac{1}{D}\frac{{\rm d}D}{{\rm d}t}, \quad \text{(A10)}$$

where M_h is the halo mass and S is the variance of the density field on the length scale corresponding the halo mass. δ_c is a parameter that represents a threshold in the linearly extrapolated density field for halo collapse. We assume $\delta_c=1.68$ (Press & Schechter 1974). The parameters, q and γ, are related to the shape of the power-spectrum around the halo mass M_h. The scale dependence of the density field is approximated as a power-law around $10^{12}\,{\rm M}_\odot$ haloes as $S=S_0(M_h/10^{12}\,{\rm M}_\odot)^{-\gamma}$. Correa et al. (2015) find that this prescription works for different cosmologies because the halo mass histories are mainly driven by changes in σ_8 and Ω_m. For the cosmological parameters inferred by the Planck Collaboration XVI (2014), $S_0\approx3.98$, $\gamma\approx0.3$, and $q\approx3.16$. Collecting the numerical and cosmology-dependent constants together yields,

$$\frac{1}{M_h}\frac{{\rm d}M_h}{{\rm d}t}=1.05\left(\frac{M_h}{10^{12}\,{\rm M}_\odot}\right)^{-\gamma/2}\frac{1}{D^2}\frac{{\rm d}D}{{\rm d}t}. \quad \text{(A11)}$$

Using the series approximation equation (A9), the specific growth rate of haloes can be written as,

$$\left(\frac{1}{M_h}\frac{{\rm d}M_h}{{\rm d}t}\right)^{c^*}=\frac{2.66}{\sqrt{S_0}K_D\,t_m^3}\left(\frac{t}{t_m}\right)^{-5/3}\times\left(\frac{M_h}{10^{12}\,{\rm M}_\odot}\right)^{\gamma/2}f_\Lambda(t,-0.32,0.06). \quad \text{(A12)}$$

This differential equation can be solved by separation of variables to obtain the average mass history of dark matter haloes,

$$\int_M^{M_0}\left(\frac{M_h'}{10^{12}\,{\rm M}_\odot}\right)^{-\left(\frac{\gamma}{2}+1\right)}\frac{{\rm d}M_h}{10^{12}\,{\rm M}_\odot}=\frac{2.66}{\sqrt{S_0}K_D\,t_m^3}\int_t^{t_0}\left(\frac{t'}{t_m}\right)^{-5/3}f_\Lambda(t,-0.32,0.06)\,{\rm d}t' \quad \text{(A13)}$$

where M_0 is the mass of a halo today. Integrating both sides and solving for $M(t)$ yields,

$$\frac{2}{\gamma}\left[\left(\frac{M_h}{10^{12}\,{\rm M}_\odot}\right)^{-\gamma/2}-\left(\frac{M_0}{10^{12}\,{\rm M}_\odot}\right)^{-\gamma/2}\right]=\frac{4}{\sqrt{S_0}K_Dt_m^2}\left[\left(\frac{t}{t_m}\right)^{-2/3}f_\Lambda(t,0.16,-0.01)-\left(\frac{t_0}{t_m}\right)^{-2/3}f_\Lambda(t_0,0.16,-0.01)\right]. \quad \text{(A14)}$$

Note that in the case $\gamma\to0$ the LHS becomes the logarithm of the mass ratio M_h/M_0, and all haloes grow by the same factor in a given time interval. For realistic power spectra, however, the relative growth rate increases with mass because massive haloes arise from increasingly rare fluctuations in the initial density perturbation field.

We can re-write equation (A14) as an explicit equation for the halo mass as a function of time. This form is useful for symbolic substitution into calculations that are driven by the halo mass.

$$\frac{M_h^{c^*}(t)}{10^{12}\,\mathrm{M}_\odot} = \left\{ \left(\frac{M_0}{10^{12}\,\mathrm{M}_\odot}\right)^{-\gamma/2} + \frac{2\gamma}{\sqrt{S_0}K_D t_m^2}\left[\left(\frac{t}{t_m}\right)^{-2/3} \times f_\Lambda(t, 0.16, -0.01) - \left(\frac{t_0}{t_m}\right)^{-2/3} \times f_\Lambda(t_0, 0.16, -0.01)\right]\right\}^{-2/\gamma}. \quad \text{(A15)}$$

As $t \to 0$, the mass of the halo becomes small compared to the final mass so that we can write,

$$\frac{M_h(t)}{10^{12}\,\mathrm{M}_\odot} \approx \left\{ \left(\frac{M_0}{10^{12}\,\mathrm{M}_\odot}\right)^{-\gamma/2} + \frac{2\gamma}{\sqrt{S_0}K_D t_m^2}\left(\frac{t}{t_m}\right)^{-2/3}\right\}^{-2/\gamma}, \quad \text{(A16)}$$

where the first term in the RHS is much smaller than the second term. This shows that masses of early haloes depend very weakly on their average final mass, and that the halo mass initially grows roughly $\propto t^4$, (since $\gamma \approx 1/3$).

Finally, collecting the numerical and cosmology-dependent constants together, we can write equation (A15) as,

$$\frac{M_h(t)}{10^{12}\,\mathrm{M}_\odot} = \left\{ \left(\frac{M_0}{10^{12}\,\mathrm{M}_\odot}\right)^{-\gamma/2} + 0.31\gamma\left[\left(\frac{t}{t_m}\right)^{-2/3} f_\Lambda(t, 0.16, -0.01) - 1.67\right]\right\}^{-2/\gamma}. \quad \text{(A17)}$$

A3 The halo mass function

In the Press & Schechter analysis, the co-moving abundance of haloes of mass M_h at time t is given by (Press & Schechter 1974),

$$\frac{\mathrm{d}n^{c^*}(M_h, t)}{\mathrm{d}M_h} = \frac{\rho_0}{M_h^2}\frac{\delta_c\gamma}{\sqrt{2\pi}S^{1/2}}\frac{1}{D}\exp\left(-\frac{\delta_c^2}{2SD^2}\right), \quad \text{(A18)}$$

where we have assumed that the density power spectrum is a power law with exponent γ and written the co-moving matter density of the Universe as ρ_0 following our convention. Using the evolution of the growth factor given by equation (A9) and keeping the leading order terms we obtain,

$$\frac{\mathrm{d}n^{c^*}(M_h, t)}{\mathrm{d}\log_{10}M_h} = \frac{2.94\times10^{-12}\,\mathrm{M}_\odot^{-1}\rho_0\gamma}{\sqrt{S_0}K_D t_m^2}\left(\frac{M_h}{10^{12}\,\mathrm{M}_\odot}\right)^{-(1-\frac{\gamma}{2})} \times \left(\frac{t}{t_m}\right)^{-2/3} f_\Lambda(t, 0.16, -0.01) \times \exp\left[\frac{-5.14}{S_0K_D^2t_m^4}\left(\frac{M_h}{10^{12}\,\mathrm{M}_\odot}\right)^{\gamma}\left(\frac{t}{t_m}\right)^{-4/3} f_\Lambda(t, 0.32, 0)\right]. \quad \text{(A19)}$$

For the cosmological parameters adopted in this paper, $\rho_0 = 3.913\times10^{10}\,\mathrm{M}_\odot\,\mathrm{Mpc}^{-3}$. Substituting for values of the constants and cosmological parameters, we can write equation (A19) as,

$$\frac{\mathrm{d}n(M_h, t)}{\mathrm{d}\log_{10}M_h} = 5.43\times10^{-3}\,\mathrm{cMpc}^{-3}\left(\frac{M_h}{10^{12}\,\mathrm{M}_\odot}\right)^{-(1-\frac{\gamma}{2})} \times \left(\frac{t}{t_m}\right)^{-2/3} f_\Lambda(t, 0.16, -0.01) \times \exp\left[-0.13\left(\frac{M_h}{10^{12}\,\mathrm{M}_\odot}\right)^{\gamma}\left(\frac{t}{t_m}\right)^{-4/3} f_\Lambda(t, 0.32, 0)\right]. \quad \text{(A20)}$$

APPENDIX B: COMPARISON WITH HYDRODYNAMICAL SIMULATIONS

In this section, we compare three variations of the EAGLE $(50\,\mathrm{cMpc})^3$ simulations to their equivalent analytic effective star formation efficiency model. The Ref-L050N0752 EAGLE model (Crain et al. 2015; Schaye et al. 2015; McAlpine et al. 2016), uses the same calibrated sub-grid parameters as the reference model $(100\,\mathrm{cMpc})^3$, ran with the same resolution, but in a smaller volume. The 'No AGN' run uses the same calibrated sub-grid parameters as the reference model but removing feedback from BHs. For the 'No SN' model (red, introduced in Bower et al. 2017), feedback from star formation has been removed. We note that, while the EAGLE 'No SN' simulation removes the effect of star formation feedback, it still includes the effect of cosmic reionization. Hence, there is a suppression of star formation is small haloes. In order to compare with the simulations, we have included the effect of cosmic reionization, $\epsilon_*(T_{\rm vir} < 10^4\,\mathrm{K}) = 0$ in both the 'No SN' and 'constant' star formation efficiency models. Finally, there is no EAGLE equivalent to the 'constant' (or 'no feedback') model. In Fig. B1 we compare the GSMF at $z = 0.1$, and in Fig. B2, we compare the SFR history of the Universe.

While much more computationally expensive, the behaviour of the full hydrodynamical simulations is well approximated by the analytic models introduced here.

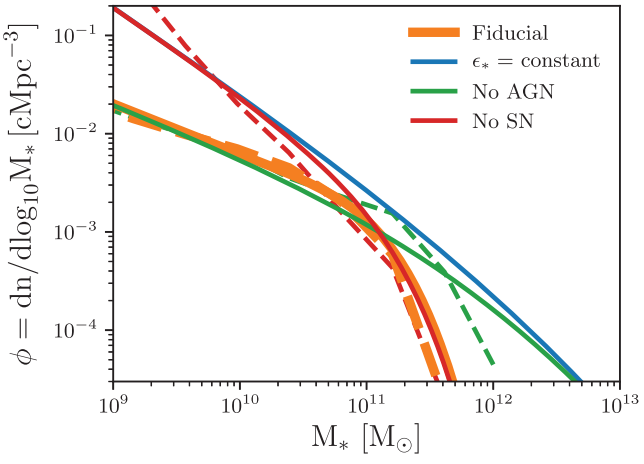

Figure B1. Redshift $z = 0.1$ GSMF for three variations of the EAGLE $(50\,\mathrm{cMpc})^3$ simulations at redshift $z = 0$ (dashed lines), compared their equivalent analytic effective star formation efficiency model (solid lines). The orange line shows the Ref-L050N0752 EAGLE model. The 'No AGN' and 'No SN' models are shown in green and red, respectively. While the models were not calibrated to reproduce their hydrosimulation equivalent, they capture their overall behaviour reasonably well. The small differences are consistent with the differences in the efficiency parameters (see Fig. 5).

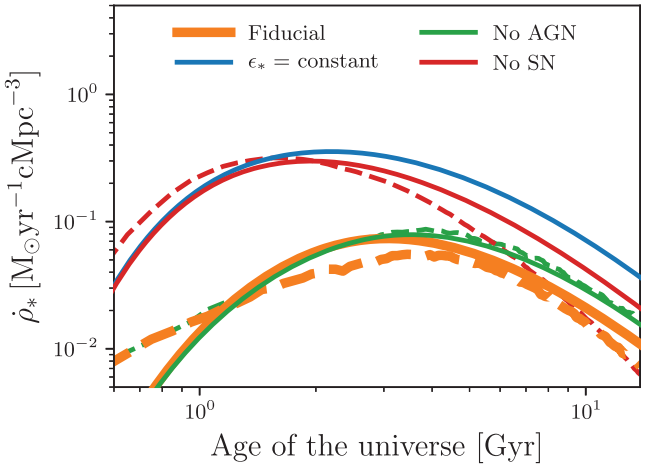

Figure B2. SFR history of the Universe for three variations of the EAGLE (50 cMpc)^3 simulations (dashed lines), compared their equivalent analytic effective star formation efficiency model (solid lines). The orange line shows the Ref-L050N0752 EAGLE model. The 'No AGN' and 'No SN' models are shown in green and red respectively. In order to compare with the simulations, we have included the effect of cosmic reionization, $\epsilon_*(T_{\rm vir} < 10^4\,{\rm K}) = 0$ in both the 'No SN' and 'constant' star formation efficiency models. While the models were not calibrated to reproduce their hydrosimulation equivalent, they capture their overall behaviour reasonably well. The small differences are consistent with the differences in the efficiency parameters (see Fig. 5).

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