On the puzzling plateau in the specific star formation rate at $z = 2–7$

Simone M. Weinmann,1* Eyal Neistein2,3 and Avishai Dekel4

1Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, the Netherlands
2Max-Planck-Institute for Extraterrestrial Physics, Giessenbachstrasse 1, 85748 Garching, Germany
3Max-Planck Institut für Astrophysik, Karl-Schwarzschild-Str. 1, Postfach 1317, 85741 Garching, Germany
4Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel

Accepted 2011 July 12. Received 2011 July 4; in original form 2011 March 15

ABSTRACT

The observational indications for a constant specific star formation rate (sSFR) in the redshift range $z = 2–7$ are puzzling in the context of current galaxy-formation models. Despite the tentative nature of the data, their marked conflict with theory motivates a study of the possible implications. The plateau at $sSFR \sim 2 \, \text{Gyr}^{-1}$ is hard to reproduce because (a) its level is low compared to the cosmological specific accretion rate at $z \gtrsim 6$, (b) it is higher than the latter at $z \sim 2$, (c) the natural correlation between SFR and stellar mass makes it difficult to manipulate their ratio, and (d) a low SFR at a high $z$ makes it hard to produce enough massive galaxies by $z \sim 2$. Using a flexible semi-analytic model, we explore ad hoc modifications to the standard physical recipes trying to obey the puzzling observational constraints. Successful models involve non-trivial modifications, such as (a) a suppressed SFR at $z \gtrsim 4$ in galaxies of all masses, by enhanced feedback or reduced SFR efficiency, following an initial active phase at $z > 7$; (b) a delayed gas consumption into stars, allowing the gas that was prohibited from forming stars or ejected at high $z$ to form stars later in more massive galaxies; and (c) enhanced growth of massive galaxies, in terms of either faster assembly or more efficient starbursts in mergers, or by efficient star formation in massive haloes.

Key words: galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: statistics.

1 INTRODUCTION

The bulks of the stellar mass observed today in galaxies are built up at a high redshift, where star formation and mass assembly are very efficient. This makes observations at high redshifts extremely important in understanding galaxy evolution in general. For example, the maximum value of the mean cosmological star formation rate (SFR) density is achieved at redshift 1–3 (e.g. Hopkins & Beacom 2006).

Although high-redshift observations are naturally plagued by larger uncertainties than local data, it is clear that a successful model of galaxy formation and evolution should match them within realistic error margins. Current models are often mainly tuned to reproduce the low-redshift Universe, and it is therefore very instructive to test them against high-redshift observations. For example, Fontanot et al. (2009) and Guo et al. (2010) point out that the low-mass end of the stellar mass function is built too quickly at high redshifts in the current semi-analytic models (SAMs hereafter), while Khochar et al. (2007) claim that their SAM can reproduce the evolution of the faint end of the luminosity function. It seems, however, likely that the evolution of the stellar mass function over time alone is insufficient to fully constrain the models, as was argued by Neistein & Weinmann (2010). Consequently, it is a crucial next step to compare the observed SFR of high-redshift galaxies to models in more detail than previously done, where the main focus was in trying to match the global star formation rate density.

Pioneering observational estimates of the SFR ($\dot{m}_{\text{star}}$) and stellar mass ($m_{\text{star}}$) indicate that the specific SFR ($sSFR = \dot{m}_{\text{star}}/m_{\text{star}}$) is roughly constant in time throughout the redshift range $z = 2–7$, for galaxies with roughly the same mass $m_{\text{star}} \sim (0.2–1) \times 10^{10} \, M_\odot$, at a level $sSFR \sim 1–2 \, \text{Gyr}^{-1}$, (e.g. Stark et al. 2009a; González et al. 2010; Labbé et al. 2010a,b). There are indications that this sSFR plateau is associated with a rather constant sSFR within each galaxy as it grows (Stark et al. 2009a; Papovich et al. 2011). The plateau is hard to reconcile with the current theoretical wisdom for several reasons.

In current models of galaxy formation the high-$z$ SFR is assumed to a large extent to be driven by the fresh gas supply [see Kereš et al. 2005; Dekel, Sari & Ceverino 2009; Bouché et al. 2010; and basically all the SAM; but see also Narayanan et al. (2010) who discuss a potential merger origin for the population of submillimetre galaxies at $z \sim 2$]. The specific cosmological accretion rate of baryons is steeply declining with time, $M/M_\odot \propto (1+z)^{-3}$ (Neistein

*E-mail: weinmann@strw.leidenuniv.nl

© 2011 The Authors

Monthly Notices of the Royal Astronomical Society © 2011 RAS

Downloaded by guest on 29 July 2018
2 THE PROBLEM

2.1 Observations

A correlation between stellar mass and SFR has been observed at various redshifts up to $z \sim 4$ (e.g. Daddi et al. 2007; Stark et al. 2009; Karim et al. 2011). Fig. 1 shows a compilation of observational estimates of sSFR as a function of redshift for star-forming galaxies\(^1\) of a similar stellar mass $\sim (0.2 - 1) \times 10^{10} M_\odot$. The data reveal a rather constant sSFR $\sim 2$ Gyr\(^{-1}\) in the redshift range $z = 2 - 7$ (Feulner et al. 2005; Yan et al. 2006; Daddi et al. 2008; Dekel et al. 2009). In particular, at $z \sim 7$ the specific accretion rate is higher than the observed sSFR by a factor of a few, and at $z \sim 2$ it is lower than the sSFR by a similar factor. A constant sSFR during the evolution of each galaxy (main progenitor) would require either an exponential growth in time of $m_{\text{gas}}$ and $m_{\text{star}}$ (this is if most stars are formed in situ to the main progenitor), or a non-trivial combination of effective SFR and stellar assembly rate as a function of time and mass. An obvious related difficulty is introduced by the fact that a low SFR at high $z$ could make it hard to produce enough massive galaxies by $z \sim 2$ to match the bright end of the observed galaxy mass function at that epoch.

Given the marked contrast between the indicated observation and the current models of galaxy formation, we employ a special semi-analytic tool. Traditional SAMs (e.g. Kauffmann, White & Guideroni 1993; Cole et al. 2000; De Lucia & Blaizot 2007) describe the processes that are responsible for galaxy evolution by physically motivated recipes that are fixed a priori. They are thus geared to solve the ‘forward problem’, i.e. test to what extent the assumed set of physical recipes provides a match to the observed properties of the galaxy population. This methodology is not ideal for exploring a large variety of physical recipes, some of which may need to deviate significantly from the standard assumptions. Our approach here is to solve the ‘inverse problem’, where we investigate how the observations at high $z$ constrain the basic recipes, especially those associated with the processes of mergers, star formation and feedback. This is achievable using the method of Neistein & Weinmann (2010, NW10 hereafter), which allows us freedom in choosing the recipes of interest, and a very efficient exploration of a broad parameter space.

In order to reproduce the sSFR plateau, we will try to either lower the SFR efficiency after a very short period of high efficiency at $z \geq 7$, or to enhance the suppression of SFR by feedback at $z \geq 4$, especially in high-mass haloes, to be followed by an enhanced SFR in the retained or reincorporated gas at $z \sim 2 - 3$. In both cases, the low sSFR at high $z$ makes it difficult to form enough massive galaxies at $z \sim 1 - 3$, unless the rate of mass assembly due to mergers and the associated starbursts are pushed to their limits.

The models presented in this paper are not a priori physically motivated. However, we have tried to keep them simple, and to minimize deviations both from the standard model and from a monotonically dependence on halo mass and time. We also try to keep the models as physically plausible as possible. Our aim in this paper is not to find the ‘right’ model. The main goal is to investigate the approximate nature of the needed changes, and how they impact on the properties of the galaxy population other than the sSFR.

The measurements of SFR and stellar mass at high $z$ are still at their infancy, and therefore their interpretation, in particular the sSFR plateau, is uncertain and highly controversial. The main source of uncertainty is the obscuration by dust, where several authors bring convincing arguments for little or no dust at $z \geq 4$ (e.g. Bouwens et al. 2009; Finkelstein et al. 2010), while others do apply dust corrections and obtain higher values of sSFR. Stellar masses may also be affected by systematic errors. For example, uncertainties in the treatment of the TP AGB-phase may lead to errors in the stellar mass estimates at $z = 2 - 3$ by factors of $\sim 2$ (e.g. Maraston et al. 2006; Magdis et al. 2010a). The disagreement between a constant sSFR and our current theoretical wisdom is so pronounced that it motivates a study of the theoretical implications despite the observational uncertainties, adopting the validity of the sSFR plateau as a working assumption. An alternative way out from the puzzle might be to assume a time-dependent stellar initial mass function (IMF)\(^1\) typically selected at high redshift as Lyman-break galaxies (LBGs, Steidel et al. 1999), which possibly excludes a population of low-SFR galaxies (e.g. Richard et al. 2011). This will, however, only increase differences between model and basic theoretical predictions.

\(^{1}\)Typically selected at high redshift as Lyman-break galaxies (LBGs, Steidel et al. 1999), which possibly excludes a population of low-SFR galaxies (e.g. Richard et al. 2011). This will, however, only increase differences between model and basic theoretical predictions.

\(^{1}\)
2007; Eyles et al. 2007; Stark et al. 2009; González et al. 2010; Labbé et al. 2010a,b; Magdis et al. 2010a,b; Rodighiero et al. 2010; McLure et al. 2011) except for three higher estimates (Yabe et al. 2009; Schaerer & de Barros 2010; Shim et al. 2011).^2 The main reason for these higher estimates is the larger correction for dust extinction assumed by these authors (to be discussed in Section 5.3), as well as different treatments of nebular emission lines and different assumptions concerning star-formation histories (to be discussed in Section 4.6). Additionally, Shim et al. (2011) only include galaxies with indications of Hα emission, which will bias the estimate of the sSFR high. We note that all the estimates above do not include submillimetre galaxies, which are outliers to the relation between stellar mass and SFR, simply because these tend to have stellar masses above the limit we consider here (e.g. Daddi et al. 2007). At \( z < 2 \), the sSFR declines steeply (Elbaz et al. 2007; Noeske et al. 2007; Dunne et al. 2009; Oliver et al. 2010; Rodighiero et al. 2010). The grey belt in Fig. 1 tries to capture the overall trend, reflecting an uncertainty of \( \pm 0.3 \) dex as estimated by González et al. (2010), and ignoring the two high estimates. As will be discussed below, this observed sSFR plateau is puzzling – its level is surprisingly high at \( z \sim 2 \) and surprisingly low at \( z > 4 \). For the purpose of the theoretical analysis of the current paper, we adopt the sSFR plateau as marked by the grey belt.

2.2 Tension with theory

Here, we outline the main potential points of tension between the observed sSFR and theoretical predictions both from relatively detailed SAMs and simple analytical arguments.

2.2.1 Tension with SAMs

In Fig. 2, we show the sSFR as a function of redshift at a fixed mass for four of the models presented in NW10, in comparison with the observed sSFR plateau. To account for the observational completeness limit as indicated by Stark et al. (2009), we only take into account model galaxies with log (SFR) > 0.25, 0.45 and 0.5 yr\(^{-1}\) at \( z = 4, 5 \) and 6, respectively. The four models are described in detail in NW10 with the same numbering used here and can be summarized as follows: (i) model without supernova (SN) feedback; (ii) model without ejective SN feedback; (iii) model including cold accretion (which is the model most similar to other standard SAMs); and (v) model in which cooling and star formation shut down after major mergers. All of those models have been tuned to reproduce key observables like the stellar mass functions at different redshifts, and star formation rate at \( z = 0 \). Remarkably, all models show a closely similar behaviour despite their fundamental differences, all in disagreement with observations. They overestimate the sSFR by about an order of magnitude at \( z \sim 6 \), and underestimate it by around 0.3 dex at \( z \sim 2 \).

Other current SAMs show a similar behaviour. For example, Lacey et al. (2011) show that the sSFR of galaxies in the Baugh et al. (2005) model have sSFR > 10 Gyr\(^{-1}\) at \( z \sim 6 \), an order of magnitude higher than observational results. Daddi et al. (2007) indicate that their observed SFR at \( z \sim 2 \) is significantly higher than the values predicted by the model of Kitzbichler & White (2007). Finally, we have confirmed ourselves that the model of De Lucia & Blaizot (2007) predicts results very similar to our model predictions shown in Fig. 2 (see also Guo & White 2008, their fig. 3).

2.2.2 Tension with basic theoretical considerations

The observed sSFR plateau is in disagreement with the standard wisdom concerning galaxy evolution. First, the average specific accretion rate into dark matter haloes of a given mass is rapidly increasing with redshift, roughly in proportion to \((1 + z)^{2.5}\) (Neistein & Dekel 2008). Secondly, galaxies are denser and more gas-rich at high redshifts, which is expected to lead to a higher SFR (e.g. Dutton, van den Bosch & Dekel 2010). In agreement with these studies, Bouché et al. (2010) find in their idealized model that the sSFR of individual galaxies is indeed monotonically decreasing with time. To illustrate the first point, we show in Fig. 2 the approximation for the average specific dark matter accretion rate \( \dot{M}/M \) for haloes with log \( (M_{\text{halo}}) \sim 10^{12} \) \( M_\odot \) (Neistein & Dekel 2008), and compare it to the standard model, described in Section 4.1, Fig. 3. Remarkably, when we remove feedback from the standard model, it predicts an sSFR evolution in excellent agreement with twice the specific dark matter accretion rate. As shown in Fig. 3, the remaining factor of 2 difference is fully explained by instantaneous stellar mass-loss. Given that the ratio between stellar mass and halo mass is only about 5 per cent for the galaxies considered here, this agreement is noteworthy. We see in the figure that the feedback as implemented in the standard model does not have a significant effect on the sSFR at \( z > 4 \), while it gradually reduces the sSFR at lower redshifts. Compared to the models of NW10 shown in Fig. 2, the standard model that we use in this paper has a low sSFR at \( z < 2 \), due to efficient feedback at late times. This is not relevant to studying the plateau.

\(^2\) We obtained part of the estimates by dividing the median SFR by the median stellar mass (see Yan et al. 2006; Eyles et al. 2007; Yabe et al. 2009). In some other cases, the estimates are based on an extrapolation of the sSFR–stellar mass relation to a stellar mass of \( \sim 0.5 \times 10^{10} \) \( M_\odot \) (Daddi et al. 2007; Rodighiero et al. 2010).
3 THE FORMALISM

In this section we describe the formalism we use for modelling the evolution of galaxies. For more details about the methodology the reader is referred to NW10. It is shown there that the results of our model are very similar to those given by a standard SAM, although the recipes are simplified and schematic. In the context of this work, the simplicity of the model allows us to tune it easily, without losing the complex interplay between different processes such as dark matter growth, cooling, star formation (SF), feedback and merging. The code is available for public usage through the Internet (see http://www.mpa-garching.mpg.de/galform/semag).

3.1 Merger trees

We use merger trees extracted from the Millennium N-body simulation (Springel et al. 2005). This simulation was run using the cosmological parameters \((\Omega_m, \Omega_\Lambda, h, \sigma_8) = (0.25, 0.75, 0.73, 0.9)\), with a particle mass of \(8.6 \times 10^9 \ h^{-1} \ M_\odot\) and a box size of 500 \(h^{-1} \) Mpc. The merger trees used here are based on subhaloes identified using the SUBFIND algorithm (Springel et al. 2001). They are defined as the bound density peaks inside friends-of-friends (FoF) groups (Davis et al. 1985). More details on the simulation and the subhalo merger-trees can be found in Springel et al. (2005) and Croton et al. (2006). The mass of each subhalo (referred to as \(M_h\) in what follows) is determined according to the number of particles it contains. Within each FoF group the most massive subhalo is termed the central subhalo of this group. Throughout this paper we will use the term ‘haloes’ for both subhaloes and the central (sub)halo of FoF groups.

3.2 Quiescent evolution

Each galaxy is modelled by a four-component vector,

\[
\mathbf{m} = \begin{pmatrix} m_{\text{star}} \\ m_{\text{cold}} \\ m_{\text{fil}} \\ m_{\text{fb}} \end{pmatrix},
\]

where \(m_{\text{star}}\) is the mass of stars, \(m_{\text{cold}}\) is the mass of cold gas within the disc, \(m_{\text{fil}}\) is the mass within cold filaments streaming within the host halo into the central galaxy and \(m_{\text{fb}}\) is the mass currently made unavailable for star formation by stellar feedback. We use the term ‘quiescent evolution’ to mark all the evolutionary processes of a galaxy, except those related to mergers.

All the models in this work assume that fresh gas is added to a galaxy only by cold filaments, increasing the mass of \(m_{\text{fil}}\). The infall rate into filaments is assumed to be proportional to the dark matter growth rate:

\[
[m_{\text{fil}}]_{\text{accretion}} = 0.17 \dot{M}_h.
\]

Here 0.17 is the cosmic baryonic fraction and \(\dot{M}_h\) is the rate of dark matter smooth accretion which does not include mergers with resolved progenitors (if \(\dot{M}_h < 0\) we use a gas accretion rate of zero).

The mass of cold gas within the disc is increased due to the free infall of cold filaments from the outer parts of the host halo. We mimic this effect by assuming that gas joins the disc with a specific rate \(f_c:\n
\[
[m_{\text{cold}}]_{\text{fil}} = -[m_{\text{fil}}]_{\text{fil}} = f_cm_{\text{fil}}.
\]

The efficiency \(f_c = f_c(M_h, t)\) is a function of the host halo mass \(M_h\) and the cosmic time \(t\) only, and is given in units of \(\text{Gyr}^{-1}\).

We assume that the SF rate is proportional to the amount of cold gas:

\[
[m_{\text{star}}]_{\text{SF}} = -[m_{\text{cold}}]_{\text{SF}} = f_c m_{\text{cold}},
\]

where \(f_c = f_c(M_h, t)\) is a function of the halo mass and time, in units of \(\text{Gyr}^{-1}\). For each SF episode we assume that a constant fraction of the mass is returned to the cold gas component due to SN events and stellar winds. This recycling is assumed to be instantaneous, and contributes

\[
[m_{\text{cold}}]_{\text{recycling}} = -[m_{\text{star}}]_{\text{recycling}} = R [m_{\text{star}}]_{\text{SF}}.
\]

Following NW10, we use \(R = 0.5\) for all models. This is the recycled fraction for a Chabrier (2003) IMF at 13.5 Gyr after a starburst according to the Bruzual & Charlot (2003) stellar population models. We note that this is the only point where the assumption on the IMF enters our model.

Cold gas can be affected by feedback, which means that it becomes unavailable for star formation and moves from the cold phase to the feedback phase. Assuming that stellar feedback immediately follows star formation, this feedback should be in proportion to the SF rate:

\[
[m_{\text{fb}}]_{\text{feedback}} = -[m_{\text{cold}}]_{\text{feedback}} = f_d [m_{\text{star}}]_{\text{SF}} = f_d f_c m_{\text{cold}}.
\]

We model feedback by a function of halo mass and time, \(f_d = f_d(M_h, t)\).

Once gas has been made unavailable for star formation due to feedback, we allow it to return to the cold phase, with a reincorporation efficiency:

\[
[m_{\text{cold}}]_{\text{rec}} = -[m_{\text{fb}}]_{\text{rec}} = f_r m_{\text{fb}}.
\]

\[\text{© 2011 The Authors, MNRAS 417, 2737–2751}\]

Monthly Notices of the Royal Astronomical Society © 2011 RAS
Note that the feedback mechanism we use is moving gas from the cold phase to the feedback phase. The mass within filaments is not participating in the feedback and reincorporation processes. If the cold gas mass becomes negative in a given time-step due to strong feedback, the star formation rates are adjusted such that a cold gas mass of zero is produced.

To conclude, each process is described by one function which depends on the host halo mass and time only. All processes discussed in this section can be written in a compact form by using the following differential equations:

\[ m = Am + BM_h, \]  

where

\[
A = \begin{pmatrix}
0 & (1 - R)f_s & 0 & 0 \\
0 & -(1 - R)f_s - f_a f_c & f_c & f_c \\
0 & 0 & -f_c & 0 \\
0 & f_a f_c & 0 & -f_c
\end{pmatrix},
\]

\[
B = \begin{pmatrix}
0 \\
0 \\
0.17 \\
0
\end{pmatrix}.
\]

Photoionization heating of the intergalactic medium is assumed to suppress the amount of cold gas available for SF within low-mass haloes. This effect is critical for modelling the formation of dwarf galaxies. The minimum halo mass of \( \sim 2 \times 10^{10} h^{-1} M_\odot \) in the Millennium Simulation, which we use here, does however not allow a detailed modelling of small mass galaxies. Thus, instead of implementing a detailed treatment of reionization, we simply assume that all the gas is kept hot until redshift 9, where cooling and SF are allowed to start. This is a higher redshift than the one in NW10, which we found is needed to produce a high enough number of galaxies at very high redshifts.

### 3.3 Mergers and satellite galaxies

Satellite galaxies are defined as all galaxies inside an FoF group except the main galaxy inside the central (most massive) subhalo. Once the subhalo corresponding to a given galaxy cannot be resolved anymore, it is considered as having merged with the central halo. Due to the effect of dynamical friction, the galaxy is then assumed to spiral towards the centre of the FoF group and merge with the galaxy in the central halo after a significant delay time.

At the last time the dark matter subhalo of a satellite galaxy is resolved we compute its distance from the central halo \( r_{sat} \), and estimate the dynamical friction time using the formula of Binney & Tremaine (1987):

\[ t_{df} = \alpha_{df} \times \frac{1.17 V_{c,s}^2}{Gm_{sat} \ln(1 + M_h/m_{sat})}. \]  

For \( m_{sat} \) we use the baryonic (stars + cold gas) mass of the satellite galaxy plus the minimum subhalo mass which can be resolved by the Millennium Simulation. \( V_c \) and \( M_h \) are the virial velocity and mass of the central subhalo. If a satellite falls into a larger halo together with its central galaxy, we update \( t_{df} \) for both the objects according to the new central galaxy.

While satellite galaxies move within their FoF group, they suffer from loss of their extended gas reservoir due to tidal stripping. We assume that all satellite galaxies are losing their reservoir of filament gas exponentially, on a time-scale of a few Gyr. In order to properly model this stripping we modify \( A \) by subtracting a constant \( \alpha_b \) from one of its elements:

\[
A_{sat}(3, 3) = -f_c - \alpha_b.
\]

Note that a constant in the diagonal of \( A \) gives an exponential time dependence. However, the actual dependence of \( m_{c,s} \) on time for satellite galaxies is more complicated due to contributions from accretion. In general the parameter \( \alpha_b \) should depend on the dynamical time of the host halo. For simplicity we consider it to be a constant here. We assume that the gas which is in the feedback phase is not stripped.

When galaxies finally merge we assume that an SF burst is triggered. We follow Mihos & Hernquist (1994), Somerville, Primack & Faber (2001) and Cox et al. (2008) and model the amount of stars produced by

\[
\Delta m_{star} = f_{burst}(m_{c,s} + m_{c,s}),
\]

where

\[ f_{burst} = \sigma_b \left( \frac{m_1}{m_2} \right)^{\frac{4}{3}}. \]

Here \( m_i \) are the baryonic masses of the progenitor galaxies (cold gas plus stars), \( m_{c,s} \) is their cold gas mass, and \( \sigma_b, \alpha_b \) are constants.

The burst duration has been shown to vary in hydrodynamical simulations between tens of Myr to a few Gyr depending on the merger mass ratio, and whether multiple bursts are considered or just the main peak (Cox et al. 2008). We use a time-scale of 10 Myr in all our models, following De Lucia & Blaizot (2007). Merger-induced bursts cause feedback in the same way as quiescent star formation.

### 4 MODIFYING A SIMPLE MODEL

Below, we present a simple standard SAM, which we use as a starting point for our tuning procedure. This simple model includes features common to many current models of galaxy formation. It is kept as simple as possible in order to facilitate tuning and to simplify interpretation of changes to the model. In Section 4.2, we present six alternative models which reproduce the sSFR plateau.

#### 4.1 The standard model

The ingredients of this model were chosen based on the simplicity of the physical processes involved and the consistency with observational constraints.

(i) We assume that all accretion is cold, with \( f_c = 1/t_{\text{dyn}} \), and that accretion is quenched at \( z < 3 \) above \( M_h = 1.6 \times 10^{12} M_\odot \) (e.g. Cattaneo et al. 2006; Dekel & Birnboim 2006; Ocvirk, Pichon & Teyssier 2008; Dekel et al. 2009). To avoid sharp breaks in the stellar mass function, we smoothed the transition both in redshift and in mass by hand, as shown in Fig. 4.

(ii) The rate at which cold gas is turned into stars is

\[ f_s = \epsilon_s/t_{\text{dyn}}, \]

where \( \epsilon_s \) is the star formation efficiency. For \( t_{\text{dyn}} \), we assume

\[ t_{\text{dyn}} = \frac{3\lambda}{V_{\text{vir}}^2} \sim 0.0072 t_{\text{stable}}. \]

where the halo spin parameter \( \lambda = 0.03 \) [according to the mean value found by Muñoz-Cuartas et al. (2011) in N-body simulations]. We use a constant star formation efficiency \( \epsilon_s = 0.01 \), as shown in Fig. 5 as blue line. This is comparable to the estimates of Krumholz & Tan
is the usual way to reproduce the low-mass end of the stellar mass function (Dekel & Silk 1986). The formation of stars affects the surrounding cold gas, making part of it unavailable for star formation for a certain amount of time. The rate at which gas is made unavailable for star formation is equal to the SFR times an efficiency $f_d$. As in most other SAMs, $f_d$ is a combination of a constant term (which is called the ‘reheating’ part of feedback in De Lucia & Blaizot 2007) and a term that is inversely proportional to the virial velocity to some power, which should roughly mimic ejection of material due to SN explosions (e.g. Bower et al. 2006; De Lucia & Blaizot 2007; Guo et al. 2010). The feedback efficiency is thus given by

$$f_d = \delta + \left(\frac{V_{\text{vir}}}{\gamma}\right)^{-\alpha}.$$  

(17)

We use $\alpha = 3.5, \gamma = 161 \text{ km s}^{-1}, \delta = 1$.³ We do not let log($f_d$) be higher than 2.5, and we enforce a constant value for $f_d$ at $z \gtrsim 5.7$. The value of $f_d$ as a function of halo mass and time is shown in Fig. 6, top panel.

(iv) There is no reincorporation, i.e. we assume that cold gas that has been made unavailable for star formation by feedback once never returns to the cold gas reservoir.

(v) We treat dynamical friction as in NW10, with a pre-factor of $a_{\text{df}} = 3$.

(vi) No merger-induced starbursts are included.

(vii) The gas residing in the filaments of satellites is stripped with an exponential time-scale of $4 \text{ Gyr}$.

Overall our model has similar basic scalings as other SAMs, although the parameter values are slightly different, and the recipes are simplified.

We ran the standard model (as well as all the following models) on the full volume of the Millennium Simulation (Springel et al. 2005). The resulting stellar mass functions (SMFs) at different redshifts and the sSFR–$z$ relation (defined in what follows as the sSFR as a function of redshift at stellar mass $2 \times 10^{9}$–$10^{10} \text{ M}_\odot$) are shown in Figs 7 and 8 as thin blue lines. To account for the observational completeness limit as indicated by Stark et al. (2009), we only take into account galaxies with log(SFR) $> 0.25, 0.45$ and $0.5 \text{ yr}^{-1}$ at $z = 4, 5$ and 6, respectively, in Fig. 8. Clearly, the observed relation in the mass bin $2 \times 10^{9}$–$10^{10} \text{ M}_\odot$ is not reproduced, very similar to the models from NW10 shown in Fig. 2. On the other hand, note that the evolution of the SMF is reproduced well (Fig. 7), which is important for the following discussion of non-standard models. We show the relation between sSFR and stellar mass at $z = 4$ and $6$ in Fig. 9, top-left panels. The evolution of the mean cosmological SFR density is shown as a thin blue line in Fig. 10. We have checked that including merger-induced bursts according to Croton et al. (2006) does not have a significant impact on any of the results shown here.

### 4.2 Tuning

We next gradually modify the selected ingredients of the standard model. Our first goal is to bring the median sSFR at $z = 2$–$6$ and in the mass bin $(0.2$–$1) \times 10^{10} \text{ M}_\odot$ into a better agreement with the observed plateau, the grey band in Fig. 1. We also consider the relation between sSFR and stellar mass at $z = 4$ and $6$, shown in Fig. 9, in order to verify that the median sSFR tends to decrease with an increasing stellar mass. We do not attempt a fit with the observed sSFR–mass relation outside the mass bin $(0.2$–$1) \times 10^{10} \text{ M}_\odot$, because the lower mass bin is strongly affected by incompleteness (e.g. Stringer et al. 2011), while results for the higher mass bin seem controversial (compare Stark et al. 2009 to Lee et al. 2010a).

Our secondary goal is to avoid severe deviations from the observed

³ A non-zero delta is necessary in order to have some feedback even in massive galaxies – otherwise, even in the absence of cooling, the gas coming from stellar recycling and mergers is enough to keep up relatively high star formation rates in massive galaxies.
The sSFR plateau at $z = 2-7$

Figure 6. Feedback as a function of halo mass and redshift in our models. Top left: feedback efficiency $f_d$ in the standard model and all SF models. Top right: $f_d$ in model FB1. Bottom left: $f_d$ in model FB2. Contours mark $\log(f_d) = -0.5, 0, 0.5, 1, 1.5$ and 2. Bottom right: reincorporation efficiency $f_{rc}$ in the unit of $1/t_{dyn}$ in model FB2. The main new feature in the FB models is strong feedback at $z \geq 4$, and especially so in massive galaxies. Reincorporation is effective at $z < 3$ and in massive galaxies.

SMF at intermediate redshifts, and from the observed evolution of mean SFR density.

We do not attempt to fit the observed sSFR at $z < 2$, as we believe that the mismatch in the sSFR between observations and models at $z < 2$, where the SFR in the universe declines rapidly, is a separate problem that deserves a dedicated study. It is discussed in more detail elsewhere (e.g. Fontanot et al. 2009; Karim et al. 2011). Additionally, we find fitting the $z > 2$ sSFR challenging enough even prior to adding the additional constraints imposed by the low-redshift data. On a separate note, we point out that observations of sSFR in galaxies with masses $(0.2-1) \times 10^{10} M_\odot$ at $z = 1-2$ are rather uncertain, probably more so than at higher redshifts, as they tend to fall below the current stellar mass completeness limits at these redshifts (e.g. Dunne et al. 2009; Rodighiero et al. 2010). This is due to the decline of the global star formation rate density and the increasing importance of dust, which decrease the intrinsic luminosity of galaxies at $z < 2$.

For tuning the models, we focus on one model ingredient at the time, i.e. the star formation efficiency or the feedback efficiency. Using a table with discrete values of the efficiency in eight bins in time, and 10 bins in halo mass, we start by fitting the plateau at the highest redshift and then subsequently continue to lower redshifts. We try to keep dependencies on halo mass and time as monotonic as possible, to limit the number of possible models, and also because monotonic dependencies are easier to motivate physically. Once we have found a solution for the plateau from $z = 2-7$, we compare the model results with the other observational constraints. Depending on the outcome, we discard the model or improve the fit to the SMFs and the sSFR–stellar mass relation by making additional changes to the efficiencies. Part of our models (SF2, SF3, FB2) are then further improved by simple changes to one or two other model ingredients. We point out that we have not carried out a systematic study covering all the parameter spaces, but a process of trial and error that is geared towards finding successful solutions which fit the constraints in question. Fitting the sSFR plateau and the evolution of the SMF together is not trivial, and a substantial number of iterations were needed for each model to arrive at the solutions presented below.

The different models are summarized in Table 1.

4.3 Reproducing the sSFR plateau

In this section, we explore two alternative modifications to the standard model that aim at reproducing the sSFR plateau: (i) models ‘SF’, in which $\epsilon_s$ is reduced at high redshifts; and (ii) models ‘FB’, where the feedback efficiency is enhanced at high redshifts. Results for all the models are shown in Figs 7–10.

4.3.1 Model SF1 – tune $\epsilon_s$

In model SF1, we tune the star formation efficiency (SFE) parameter, $\epsilon_s$, while all other model parameters are kept fixed as in the standard model. With $\epsilon_s$ varying in time as shown in Fig. 5 (top...
Figure 7. Stellar mass function (SMF) of galaxies at different redshift bins, for our different models as indicated and as summarized in Table 1. Observational data are marked by grey symbols. At \( z > 0 \), the model stellar mass is convolved with a Gaussian error distribution of standard deviation 0.25 dex, which includes the differences in the IMFs assumed by the different observers. Data at \( z = 0 \) are by Li & White (2009, circles), Baldry, Glazebrook & Driver (2008, crosses) and Panter et al. (2007, pluses). Data at \( z > 0 \) are from Bundy et al. (2006, \( z = 0.75-1 \), circles); Borch et al. (2006, \( z = 0.8-1 \), crosses); Pérez-González et al. (2008, \( z = 0.8-1, 1.6-2 \) and 2.5-3, the plus signs); Fontana et al. (2006, \( z = 0.8-1, 1.6-2 \) and 2-3, stars); Drory et al. (2004, \( z = 0.8-1 \), upward-pointing triangles); Drory et al. (2005, \( z = 0.75-1.25, 1.75-2.25 \) and 2.25-3, diamonds and squares); and Marchesini et al. (2009, \( z = 1.3-2 \) and 2-3, right-pointing triangles). Model SMFs are plotted at \( z = 0 \) (top-left panel); \( z = 0.8, 1, 1.2 \) (top-right panel); \( z = 1.2, 1.5, 2 \) (bottom-left panel); and \( z = 2, 2.5, 3 \) (bottom-right panel).

Figure 8. Evolution of sSFR for galaxies with stellar mass in the fixed bin \( 2 \times 10^9-10^{10} \, M_\odot \) for the models indicated and listed in Table 1. At \( z \geq 4 \), only galaxies with SFR above the completeness limits given by Stark et al. (2009) are included. Error bars denote \( \pm 1 \sigma \) at \( z > 2 \). Each of the redshift bins contains more than 50 galaxies. All the models shown reproduce the sSFR plateau at \( \sim 0.01 \).

Panel), the model reproduces the sSFR plateau at \( z = 2-6 \) as shown in Fig. 8. While \( \epsilon_s \) does not need to depend on halo mass, it is not monotonic with time. In order to have an sSFR plateau starting from redshift \( z_p \) (chosen here to be \( z_p \sim 7 \)), this model requires an earlier epoch where \( \epsilon_s \) is well above its fiducial value 0.01. This early star formation is needed in order to produce enough galaxies of large enough stellar mass by \( z_p \), after which the much lower SFR adds only slowly to the stellar mass. At the onset of the plateau, \( \epsilon_s \) has to drop to values well below 0.01, and it should continue to gradually decline till \( z \sim 4 \), in order to match the high observed sSFR in this regime. It catches up with the fiducial value \( \epsilon_s \sim 0.01 \) only at low redshifts. We note that model SF1 predicts that the sSFR plateau does not extend all the way to the epoch of the emergence of the first galaxies; it is preceded by a starburst epoch.

Unfortunately, Fig. 7 shows that model SF1 does not produce enough massive galaxies to match the observed mass function at intermediate redshifts, despite the high initial \( \epsilon_s \). Model SF1 thus reveals a non-trivial tension between two sets of data, namely the low values of sSFR at high redshifts versus the high-mass end of the SMF at intermediate redshifts. This indicates that the high values of sSFR obtained in SAMs at high redshifts are needed there for the purpose of building up massive enough galaxies by \( z \sim 2 \), and are therefore not easy to avoid.

4.3.2 Model FB1 - tune feedback

In model FB1, we only tune the stellar feedback parameter \( f_d \), while all other model parameters are kept fixed as in the standard model. The values of \( f_d \) as a function of mass and redshift are shown in Fig. 6 (top-right panel), and the resultant sSFR evolution and SMF are shown in Figs 7 and 8. We find that fitting the sSFR plateau requires very high stellar feedback efficiencies at early times.
Figure 9. sSFR as a function of stellar mass, at $z = 4$ (top panels) and $z = 6$ (bottom panels). The results of model FB1 are very similar to model FB2. The grey-scale refers to log number density of galaxies in units of $\text{Mpc}^{-3} \text{dex}^{-2}$.

$(z \gtrsim 3)$, and especially so for massive haloes. This is needed in order to balance the high accretion rates of these haloes. A high feedback efficiency is not needed in the massive haloes at later times because the cooling in them is set to zero at $z \gtrsim 3$. At lower halo masses $M_h \lesssim 10^{11} \text{M}_\odot$, moderately strong feedback is needed at all redshifts. Note that stellar feedback is not entirely monotonic here with either mass or time; we improve this in the following model FB2 by adding reincorporation as an additional model ingredient.

As in model SF1, we do not reproduce the high-mass end of the SMF with this model, again indicating that suppressing the efficiency of star formation at a high redshift leads to an underproduction of massive galaxies, if no other changes to the model are made. As can be seen in Fig. 7, this shortage is more severe than in the SF models, and more so at higher redshifts. We did not manage to improve this aspect of the FB1 model so far, since tuning the FB models is more difficult than tuning the SF models. The reason for this is that the star formation rate of galaxies is directly proportional to $\epsilon_s$, while its dependence on the feedback efficiency is non-linear.

All our models that reproduce the sSFR plateau at $z < 7$ require an earlier phase of high SFR, with an sSFR higher than the plateau level by a factor of a few. This provides a high enough stellar mass at $z \sim 6$, which allows the desired low sSFR at the plateau level while the SFR is driven to high values by the high accretion rate. For example, in order to reach by $z \sim 7$ a mass of $2 \times 10^9 \text{M}_\odot$, the minimum mass of galaxies on the plateau, a galaxy needs an average SFR $\sim 10 \text{M}_\odot \text{yr}^{-1}$ between the onset of star formation at $z \sim 9$ and 7, say. This implies an sSFR $\sim 5 \text{Gyr}^{-1}$ at $z \sim 7$, and even higher values at earlier times, when the stellar mass is smaller.
Such an early phase of high SFR has to be introduced by hand when $\epsilon_s$ is set to low values at $z < 7$, as in model SF1. In model FB1, on the other hand, the high SFR at $z > 7$ occurs naturally, because the gas available for SFR in a growing galaxy is proportional to the instantaneous accretion rate, which is only slowly increasing with time. This is in contrast to the SF models, where the SFR is proportional to the accumulated gas in the galaxy and therefore tends to increase faster with time.

4.3.3 Model FB2 – tune feedback and reincorporation

In model FB2, we improve on model FB1 by adding reincorporation from the feedback phase back to the cold phase as an additional model ingredient. This makes it possible to keep the variation in the feedback efficiency with time monotonic (although the trend is opposite at low and high halo masses). With model FB2, we can reproduce the sSFR plateau very well down to $z \sim 2$ (Fig. 8, black thick solid line), but again fail to maintain the high-mass end of the SMF at $z \gtrsim 2$. Feedback and reincorporation efficiencies in model FB2 are shown in Fig. 6, left- and right-bottom panel, respectively. Reincorporation is needed in order to keep up high sSFR at $z < 3$ for haloes with masses $M_h \gtrsim 10^{12} M_\odot$. We set it to very low values outside this range. The efficiency of reincorporation in the range of halo mass and time when it is needed is then quite similar at all redshifts when expressed in units of $1/t_{\text{dyn}}$, with a slight peak at around $z \sim 2$.

4.4 Reproducing the sSFR and the evolution of the SMF

While we have found in the previous section that it is possible to reproduce the sSFR plateau once the star formation or feedback efficiency is allowed to vary in time and with mass, all the models discussed so far underpredict the high-mass end of the SMF at $z \gtrsim 2$. In this section, we attempt to improve the fit to the SMF at intermediate redshifts by an additional modification to the model. For simplicity, we focus on modifications to model SF1. Fig. 7 shows that these models, SF2–SF4, reproduce the high-mass end of the intermediate-redshift SMF better than the previous models. Fig. 9 demonstrates that this success is associated with a population of galaxies more massive than $10^9 M_\odot$ with high SFR at $z \sim 4$, which were missing in the previous models. We explore here three different ways for producing this population. In model SF2, we boost the efficiency of merger-induced starbursts. In model SF3, we speed up the merger rate. In model SF4, we introduce a rather involved variation of $\epsilon_s$ both with halo mass and time, making star formation more efficient in relatively high mass haloes while keeping it inefficient in lower mass haloes.

4.4.1 Model SF2 – tune $\epsilon_s$ and mergers

In model SF2, we reproduce the required massive galaxies at $z \sim 2$ by boosting merger-induced starbursts. We first tried the starburst prescription of Croton et al. (2006) where $f_{\text{burst}} = 0.56(m_1/m_2)^{3/2}$, but this had only a little effect on the SMF. However, with $f_{\text{burst}} = (m_1/m_2)^{3/2}$ at $z > 1$, followed by the Croton et al. (2006) prescription at $z < 1$, model SF2 produces a higher abundance of high-mass galaxies in a better agreement with the SMF at an intermediate redshift (Fig. 7), while still reproducing the sSFR plateau (Fig. 8).

An additional parameter governing merger-induced starbursts is the burst duration, which is set to 10 Myr. We find that smearing the bursts over a much longer duration (e.g. 500 Myr) makes very little difference to the results. As we only consider the median and the 68 per cent range, a small population of starbursting galaxies can be present without ruining the sSFR plateau. Strongly starbursting galaxies might also not make it into the Stark et al. (2009) or González et al. (2010) samples, as the highly star forming regions in galaxies tend to be strongly dust attenuated (e.g. Charlot & Fall 2000), and are therefore a good way of building up stellar mass in a hidden mode.

4.4.2 Model SF3 – tune $\epsilon_s$ and mergers

Model SF3 is a variation of model SF2, where the high starburst efficiency at $z > 1$ is replaced by a shorter characteristic time for dynamical friction. The fiducial dynamical friction pre-factor of $\alpha_{\text{df}} = 3$ is replaced by $\alpha_{\text{df}} = 0.1$ at $z > 1$ and $\alpha_{\text{df}} = 5$ at $z < 1$, while the starbursts are moderate, $f_{\text{burst}} = 0.2(m_1/m_2)^{3/2}$. The rapid stellar assembly in this model boosts the build-up of stellar mass, and we need to lower the efficiency of star formation accordingly (Fig. 5, top panel). This model provides a sensible fit to the SMF (Fig. 7) and it reproduces the sSFR plateau (Fig. 8).
4.4.3 Model SF4 – tune $\epsilon_s$

Model SF4 demonstrates a third way to build up high-mass galaxies by moderate redshifts without enhancing the contribution of mergers. In this model, the SFE, $\epsilon_s$, is varied as a function of both halo mass and time as shown in Fig. 5. This allows for a simultaneous fit to the SMF (Fig. 7) and the sSFR plateau (Fig. 8). We note that in this model, the sSFR does not in general decrease with increasing stellar mass, as shown in Fig. 9, for the $z = 6$ results, bottom-right panel. This is contrary to all observations of this relation we are aware of, which tend to find that the median sSFR of galaxies decreases with increasing stellar mass.

4.5 Summary of model results

In this section, we summarize the successes and failures of the models, guided by Figs 7, 8 and 10. From Fig. 7, which shows the SMFs up to $z \sim 3$, it becomes clear that only models SF3 and SF4 manage to reproduce the SMF at $z > 1.2$.

In Fig. 8, we show that all our models except the standard models provide a reasonable fit to the sSFR plateau from $z \sim 2$–$6$. At $z < 2$, the sSFR is lower than the observational estimates. This may be partially due to the incompleteness of the observations, which misses galaxies with low SFR especially at low redshifts, where the global SF in the Universe has declined. The mismatch at $z \sim 0$, however, is probably real and already apparent in the standard model. It is likely connected with a similar underproduction of the sSFR present in other recent SAMs (e.g. Fontanot et al. 2009; Guo et al. 2011). We note that our efforts to improve the match to the sSFR plateau lead to an increase in the sSFR at $z < 2$ and thus to a better agreement with observations.

In Fig. 10 we show the global SFR density as a function of redshift in our models. The upper grey belt represents the directly observed SFR density from Hopkins & Beacom (2006), while the lower belt is derived from the growth curve of the stellar mass (Wilkins et al. 2008). More recent estimates, such as those from Bouwens et al. (2009) and Kistler et al. (2009), are in agreement with the upper belt. The observational results that are directly measured from star formation, and those that are obtained indirectly from the evolution of the stellar mass density are thus clearly inconsistent, which is the reason that SAMs (like, e.g. Guo et al. 2011) usually have an SF density somewhere in between these two constraints. Results of our various models start to deviate considerably at $z > 2$, but the discrepancy between the observational results makes it impossible to use them to constrain models. The highest global SFR density at $z > 3$ is reached by the standard model, followed by models that have the best fit to the SMF at a high redshift, namely model SF3 and SF4.

4.6 The sSFR of individual galaxies

In Fig. 11, we follow the growth of one individual model galaxy following the main branch of its merger history. Shown as a function of redshift is the mass in the dark matter, cold gas and stars, the SFR and sSFR, for the standard model, SF2 and FB2. The resulting stellar mass at $z = 0$ is similar in the three models. However, the star-formation history and the growth of stellar mass are different—in both model SF2 and FB2, the build-up of stellar mass is delayed compared to the standard model.

We note that at $z > 2$ the SFR in individual galaxies is rising with time for all models. This is a common feature to all our model galaxies, and is consistent with the predictions of Finlator et al. (2006) and Finlator, Oppenheimer & Davé (2011), as well as with the observational finding of Papovich et al. (2011), Maraston et al. (2010) and Lee et al. (2010a,b). It is emerging naturally from the fact that in an individual galaxy, the accretion rate is expected to be growing in time in its rapid growth phase at $z \geq 2$ (Neistat & Dekel 2008).

We see in Fig. 11 that as the galaxy evolves from $z = 6$ to $z = 2$, the SFR declines at a rather slow pace under models SF2 and FB2 compared to the steeper decline in the standard model. The sSFR plateau discussed in the previous sections, which refers to the median value for galaxies of a fixed mass at different redshifts, does not necessarily imply a constant sSFR in individual galaxies as they grow. However, the observations of Stark et al. (2009) and Papovich et al. (2011, figs. 2 and 3) do argue for a constant sSFR in individual galaxies as well. It is thus encouraging that our models that reproduce the sSFR plateau also come close to reproducing the constancy of the sSFR in individual galaxies.

5 DISCUSSION

In Sections 5.1 and 5.2, we comment on the physical plausibility of the ad hoc modifications suggested by our models. This discussion is not complete or conclusive – it is only meant to raise some of the relevant issues and to trigger further discussion. The limited goal of the current paper remains to point out the possible nature of the modifications to the standard models that are required for reproducing the sSFR plateau together with the SMF at a moderate redshift.

5.1 Changes to the quiescent evolution of galaxies

(i) SFE: in the SF models, $\epsilon_s$ has to vary in time, with a very high efficiency at $z \geq 7$, and a low efficiency at $z = 3$–$6$. The latter is consistent with the prediction of Gnedin & Kravtsov (2010) that the low metallicity and high UV radiation at $z \sim 3$ should significantly lower the normalization of the Kennicutt–Schmidt relation. This effect is particularly strong for low gas surface densities. An effective suppression of star formation in high-redshift galaxies and at low masses turns out to be a natural outcome of the low metallicity in these galaxies (Krumholz & Dekel 2011). The low dust content enables the UV radiation from stars to heat the gas (while dissociating H$_2$ molecules) and to prevent further star formation, until enough metallicity is built up and the SFR regions are shielded. Observational studies at $z \sim 3$ seem to agree with this prediction (e.g. Wolfe & Chen 2006; Rafelski, Wolfe & Chen 2011). Furthermore, both SAMs and hydrodynamical simulations have suggested that a low $\epsilon_s$ at a high redshift improves the match with observations in a number of ways (Baugh et al. 2005; Lacey et al. 2011; Agertz, Teyssier & Moore 2011). It is more difficult to explain the high $\epsilon_s$ at $z \geq 7$, necessary for producing enough galaxies with a high stellar mass by $z \sim 6$. What might help is a feedback mechanism that causes an abrupt change in the mode of star formation after a very active initial phase. Model SF4 suggests in addition an increase in $\epsilon_s$ with halo mass. Such a variation might result from faster metal enrichment or less efficient feedback in more massive haloes.

(ii) Stellar feedback: the FB models suggest a strong early stellar feedback that is not limited to small galaxies. This is in potential disagreement with conventional models of SN feedback, in which the effect of feedback at a fixed halo mass is expected to be weaker at a high redshift, where the deeper potential well makes it harder to eject gas from the galaxy (Dekel & Silk 1986; Dekel
& Birnboim 2006). One should therefore consider additional feedback mechanisms that may vary in the required way with time and mass. For example, a metallicity-dependent stellar feedback (Nishi & Tashiro 2000; Krumholz & Dekel 2011) will be especially effective at high redshifts. The required strong feedback at high halo masses $[\log (M_h) > 11.5$ at $t_{\text{Hubble}} < 2]$ may be achieved by active galactic nucleus (AGN) feedback from massive black holes rather than stellar feedback.

(iii) Reincorporation: model FB2 requires a high rate of reincorporation only at $z \lesssim 3$. In other SAMs, the reincorporation rates scale with $1/t_{\text{dyn}}$, and therefore gradually increase towards a higher redshift, making the effective stellar feedback less efficient at a high $z$. A weaker time dependence, more consistent with FB2, is found in simulations that incorporate strong momentum-driven winds (Oppenheimer et al. 2010). Another option is ‘reincorporation’ that utilizes a gas reservoir in the galaxy rather than gas that has been ejected earlier. For example, this occurs naturally for a metallicity-dependent feedback. As long as the metal content in a galaxy is low, gas is not well shielded from ultraviolet (UV) radiation by dust, leading to efficient heating and $H_2$ dissociation. Once the metal content reaches a certain threshold value, the gas is shielded and star formation becomes efficient. The abrupt appearance at $z \sim 3$ of reincorporation in FB2 is an oversimplification, but the general increase in reincorporation with halo mass is plausible both for ejective and non-ejective feedback.

5.2 Changes to the recipes for mergers

(i) Enhanced starburst efficiency at high redshift: in model SF2, the starburst efficiency, defined according to equation 13, is higher at a higher $z$. This is not to be confused with the SF efficiency $\epsilon_s$, which is relevant for quiescent star formation. An increased starburst efficiency at a high $z$ is plausible due to the shorter dynamical times. The gas-rich mergers at a high redshift may in principle be very different from today’s more familiar gas-poor mergers. High-resolution gas-rich merger simulations indeed indicate an effective SFE that is higher than in non-merger situations (Teyssier, Chapon & Bournaud 2010).

(ii) Enhanced merger rate at a high redshift: in model SF3, the dynamical friction pre-factor $\alpha_{df}$ is reduced to 0.1 at $z > 1$, which speeds up the merger process by a factor of $\sim 20$–30 compared to the standard model and other SAMs (Croton et al. 2006; De Lucia & Blaizot 2007). At a high redshift, the merging galaxies flow from the virial radius to the central galaxy along narrow radial streams associated with the cosmic-web filaments in about one-half of a halo-crossing time, $R_{\text{vir}}/V_{\text{vir}}$, which is $\sim 0.1t_{\text{Hubble}}$ (Dekel & Birnboim 2006). One should therefore consider additional feedback mechanisms that may vary in the required way with time and mass. For example, a metallicity-dependent stellar feedback (Nishi & Tashiro 2000; Krumholz & Dekel 2011) will be especially effective at high redshifts. The required strong feedback at high halo masses $[\log (M_h) > 11.5$ at $t_{\text{Hubble}} < 2]$ may be achieved by active galactic nucleus (AGN) feedback from massive black holes rather than stellar feedback.

(iii) Reincorporation: model FB2 requires a high rate of reincorporation only at $z \lesssim 3$. In other SAMs, the reincorporation rates scale with $1/t_{\text{dyn}}$, and therefore gradually increase towards a higher redshift, making the effective stellar feedback less efficient at a high $z$. A weaker time dependence, more consistent with FB2, is found in simulations that incorporate strong momentum-driven winds (Oppenheimer et al. 2010). Another option is ‘reincorporation’ that utilizes a gas reservoir in the galaxy rather than gas that has been ejected earlier. For example, this occurs naturally for a metallicity-dependent feedback. As long as the metal content in a galaxy is low, gas is not well shielded from ultraviolet (UV) radiation by dust, leading to efficient heating and $H_2$ dissociation. Once the metal content reaches a certain threshold value, the gas is shielded and star formation becomes efficient. The abrupt appearance at $z \sim 3$ of reincorporation in FB2 is an oversimplification, but the general increase in reincorporation with halo mass is plausible both for ejective and non-ejective feedback.

5.2 Changes to the recipes for mergers

(i) Enhanced starburst efficiency at high redshift: in model SF2, the starburst efficiency, defined according to equation 13, is higher at a higher $z$. This is not to be confused with the SF efficiency $\epsilon_s$, which is relevant for quiescent star formation. An increased starburst efficiency at a high $z$ is plausible due to the shorter dynamical times. The gas-rich mergers at a high redshift may in principle be very different from today’s more familiar gas-poor mergers. High-resolution gas-rich merger simulations indeed indicate an effective SFE that is higher than in non-merger situations (Teyssier, Chapon & Bournaud 2010).

(ii) Enhanced merger rate at a high redshift: in model SF3, the dynamical friction pre-factor $\alpha_{df}$ is reduced to 0.1 at $z > 1$, which speeds up the merger process by a factor of $\sim 20$–30 compared to the standard model and other SAMs (Croton et al. 2006; De Lucia & Blaizot 2007). At a high redshift, the merging galaxies flow from the virial radius to the central galaxy along narrow radial streams associated with the cosmic-web filaments in about one-half of a halo-crossing time, $R_{\text{vir}}/V_{\text{vir}}$, which is $\sim 0.1t_{\text{Hubble}}$ (Dekel & Birnboim 2006). One should therefore consider additional feedback mechanisms that may vary in the required way with time and mass. For example, a metallicity-dependent stellar feedback (Nishi & Tashiro 2000; Krumholz & Dekel 2011) will be especially effective at high redshifts. The required strong feedback at high halo masses $[\log (M_h) > 11.5$ at $t_{\text{Hubble}} < 2]$ may be achieved by active galactic nucleus (AGN) feedback from massive black holes rather than stellar feedback.

(iii) Reincorporation: model FB2 requires a high rate of reincorporation only at $z \lesssim 3$. In other SAMs, the reincorporation rates scale with $1/t_{\text{dyn}}$, and therefore gradually increase towards a higher redshift, making the effective stellar feedback less efficient at a high $z$. A weaker time dependence, more consistent with FB2, is found in simulations that incorporate strong momentum-driven winds (Oppenheimer et al. 2010). Another option is ‘reincorporation’ that utilizes a gas reservoir in the galaxy rather than gas that has been ejected earlier. For example, this occurs naturally for a metallicity-dependent feedback. As long as the metal content in a galaxy is low, gas is not well shielded from ultraviolet (UV) radiation by dust, leading to efficient heating and $H_2$ dissociation. Once the metal content reaches a certain threshold value, the gas is shielded and star formation becomes efficient. The abrupt appearance at $z \sim 3$ of reincorporation in FB2 is an oversimplification, but the general increase in reincorporation with halo mass is plausible both for ejective and non-ejective feedback.
et al. 2009). This implies a merging time substantially shorter than the dynamical friction time estimated for the gradual spiralling in. Hopkins et al. (2010) point out that the standard estimate of merger time based on dynamical friction is indeed an overestimate when the approach of the satellite is along a radial orbit, or when the dynamical friction estimate starts at a distance larger than 0.1–0.2 $R_{\text{vir}}$, where the satellite is no longer properly resolved in the simulation. Both of these conditions tend to be valid at a high redshift (Hopkins et al. 2010; Wetzel 2011), and together they may lead to an order of magnitude overestimate of the merger time in the standard model. We used the Millemium Simulation to verify that with $\alpha_{\text{sf}} = 0.1$, the merging time is comparable to the halo-crossing time.

5.3 Dust at a high redshift

The validity of the sSFR plateau at high redshifts crucially depends on the correction adopted for dust extinction by different authors. At $z \sim 2$, the estimates for dust extinction in LBGs from the UV slopes are confirmed by radio estimates (Pannella et al. 2009) and from comparison with a local sample of Lyman break analogues (Overzier et al. 2011). At higher redshifts, the estimates are more uncertain. Stark et al. (2009), González et al. (2010) and Labbé et al. (2010a,b) all assume practically no dust extinction at $z > 4$. This assumption is supported by simple theoretical considerations and several observational studies. It is expected that high-redshift galaxies should have a lower dust content than today’s galaxies simply because they had less time to produce metals and dust. Indeed, theoretical models (e.g. Guo & White 2009) often assume a lower dust extinction at a higher redshift. Bouwens et al. (2009) estimated the dust extinction at a high redshift based on the observed UV continuum slope, and found low values at $z \gtrsim 4$, especially for the dominant population of galaxies with relatively low UV luminosities. Labbé et al. (2010b), Finkelstein et al. (2010) and Bouwens et al. (2010) also report on low or zero dust extinction at $z \sim 6$–8. Bramer & van Dokkum (2007) tested whether the usual LBG selection criteria miss a significant number of dusty galaxies at $z \sim 4$ by selecting according to Balmer breaks instead, and found that this is not the case, again indicating that a strong dust obscuration at a high redshift is rare. This may, however, not be the case for massive galaxies, which potentially already have a high dust content at $z \sim 3$–4 (Mancini et al. 2009; Marchesini et al. 2010) and thus may be missed by the usual LBG selection criteria. Also, Schaerer & de Barros (2010) and Yabe et al. (2009) argue for high correction factors of $\sim 9$ when translating the UV flux to SFR at $z = 5$–7. We note that in order to reconcile the plateau in the sSFR with current SAMs by the effect of dust-extinction alone, the dust extinction would have to increase with redshift, a trend which is opposite to basic theoretical expectations. While the ‘dust’ has not settled yet on this debate, it seems that the evidence for an sSFR plateau is intriguing enough to justify a serious theoretical consideration; it is left to future observations to tell whether this is indeed a valid strong constraint or a fluke.

5.4 Comparison to previous work

Previous efforts to understand the properties of star-forming galaxies at $z > 4$ include Finlator et al. (2006, 2011), Night et al. (2006), Mao et al. (2007), Nagamine et al. (2008), Stark et al. (2009), Lee et al. (2009), Khochfar & Silk (2011), Lacey et al. (2011), Lo Faro et al. (2009) and Stringer et al. (2011). Here we compare our findings to some of those studies.

Lee et al. (2009) concluded based on clustering analysis that the duty cycle of star formation in high-redshift galaxies must be short, consisting of episodes of $<0.35$ Gyr in each galaxy. Stark et al. (2009) came to similar conclusions. This would be in apparent conflict with the need to build up the massive end of the SMF by $z \sim 2$–3. Our model SF1, which reproduces the sSFR plateau, assumes a continuous mode of star formation, and even then, it underpredicts the massive end of the SMF. If the low sSFR were associated instead with short-lived bursts of star formation, it would have been even harder to build up the stellar mass quickly enough. A way to reconcile a short duty cycle with the required fast build-up of stellar mass might be that the short-lived episodes of SF are only the final phase of much more active and dust-obscured starbursts, enough for building up the high-mass end of the SMF but undetectable in their violent phase. Finlator et al. (2011) suggest that the clustering data by Lee et al. (2009) can also be explained without a bursty star formation history, if there are strong outflows that lead to an increased scatter in the relation between baryonic mass and host halo mass.

Khochfar & Silk (2011) present a model that reproduces the sSFR plateau at $z > 4$ with (i) a very high frequency of starbursts at a high redshift and (ii) a burst efficiency that scales with the inverse of the halo circular velocity to the third power. In this way, star formation in galaxies in the mass range relevant for the plateau is inefficient at a high redshift, but stellar mass grows efficiently during mergers in low-mass haloes. Their model thus resembles a combination of our models SF2 and SF3.

The simultaneous need for a low sSFR at a high redshift and a fast build-up of stellar mass seems to be in odds with the notion that the observed growth rate in stellar mass density is low compared to that predicted by integrating the observed SFR density (Wilkins et al. 2008). These seemingly inconsistent problems may reflect a difference between the overall evolution of stellar mass density and the evolution of the population dealt with here, involving relatively massive galaxies in a fixed mass bin.

Lo Faro et al. (2009), comparing a SAM for LBGs at $z \sim 4$–6 to observations, found that their model produces an excess of star-forming galaxies with stellar masses $\sim 10^{9}$–$10^{10}$ $M_{\odot}$ at $z \sim 4$, within the range relevant for our study of the sSFR plateau. They conclude that some feedback mechanism must suppress star formation at early times in the corresponding haloes of $\sim 10^{11}–10^{12} M_{\odot}$, but it should become ineffective at later times in haloes of similar mass. As Lo Faro et al. (2009) point out, such a behaviour of the feedback efficiency is not a part of the standard feedback models. Their conclusions are thus similar to ours, although their conclusions originate from a different observational constraint.

6 CONCLUSIONS

We explored possible ad hoc modifications to standard SAMs that reproduce a constant sSFR at $z > 2$ while growing enough massive galaxies by $z \sim 2$, and came to the following conclusions.

(i) We confirm that an sSFR plateau at $z = 2$–6 is in a robust disagreement with current models, as claimed in Bouché et al. (2010). We have demonstrated that, in a fixed stellar mass bin of $2 \times 10^{9}$–$10^{10} M_{\odot}$, the common feature of a variety of standard SAMs is a gradual decline of the sSFR with time, associated with the decline of the total specific accretion rate.

(ii) We find that it is possible to reproduce the sSFR plateau together with the SMF at $z \sim 2$ via non-trivial modifications to the standard SAMs. Three different modifications seem necessary,
related to three different observational features: (1) the low sSFR at \( z > 4 \); (2) the high sSFR at \( z \sim 2-3 \); and (3) the abundance of high-mass galaxies at \( z \sim 2-3 \). The low sSFR at \( z > 4 \) is reproduced either by strong stellar feedback at a high redshift in all masses, or by inefficient star formation at a high redshift following a phase of very efficient star formation at a very high redshift. The high sSFR at \( z = 2-3 \) could be caused either by a drop in the feedback \( \sim > 20 \), or by introducing a non-trivial dependence of SFE on time and mass.

We have demonstrated that the simple SAM of NW10 is useful for exploring how current SAMs should be modified in order to match new observational constraints, and for pointing out apparently conflicting observational constraints. A similar method may be used in addressing other puzzling observations such as the tilt in the relation of sSFR and stellar mass (e.g. Somerville et al. 2008), the fraction of passive galaxies as a function of stellar mass at \( z = 0 \) (e.g. Weinmann et al. 2010) or the fraction of AGN as a function of stellar mass (Fontanot et al. 2011).

We learn that the observed sSFR at a high redshift has the potential for posing powerful constraints on the physical processes of star formation and feedback. If the sSFR plateau, as observed by Stark et al. (2009), Labbé et al. (2010a,b) and González et al. (2010), is confirmed, it will provide invaluable information on the baryonic physical processes at a high \( z \), indicating that it could be different from those at low redshifts. One should be eagerly waiting for new developments in the observations of SFR and stellar mass at a high redshift.

While we started a discussion of the physical plausibility of the required modifications to the standard recipes of galaxy-formation models, the modifications suggested above are primarily of an ad hoc nature. They deserve a thorough theoretical study of the physical mechanisms that could be responsible for the required variation with time and mass.

ACKNOWLEDGMENTS

We thank Rychard Bouwens, Sadegh Khochfar and the anonymous referee for helpful comments on the draft, and Rachel Somerville, Ben Oppenheimer, Ivo Labbé, Niv Drory, Emmanuelle Daddi, Marcel Haas, Raanan Nordon and Giulia Rodighiero for useful discussion. EN was partially supported by the Minerva fellowship during this project. AD was partially supported by ISF grant 60/08, by GIF grant G-1052-104.7/2009, by a DIP grant, and by NSF grant AST-1010033. SQL data bases containing the Millennium Simulations are publicly released at http://www.mpa-garching.mpg.de/millennium. The code used to generate semi-analytic models based on the NW10 method is publicly available at http://www.mpa-garching.mpg.de/galform/sesam. The Millennium site was created as a part of the activities of the German Astrophysical Virtual Observatory.

REFERENCES

Agerz O., Teyssier R., Moore B., 2011, MNRAS, 410, 1391
Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945
Baug C. M., Lacey C. G., Cen R. C., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191
Binney J., Tremaine S., 1987, Galactic Dynamics. Princeton Univ. Press, Princeton, NJ
Borch A. et al., 2006, A&A, 453, 869
Bouché et al., 2010, ApJ, 718, 1001
Bouwens R. J. et al., 2009, ApJ, 705, 936
Bouwens R. J. et al., 2010, ApJ, 708, 69
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frank C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 465
Brammer G. B., van Dokkum P. G., 2007, ApJ, 654, L107
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Bundy K. et al., 2006, ApJ, 651, 120
Cattaneo A., Dekel A., Devriendt J., Guideroni B., Blaizot J., 2006, MNRAS, 370, 1651
Chabrier G., 2003, PASP, 115, 763
Charlot S., Fall S. M., 2000, ApJ, 718, 731
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Cox T. J., Jonsson P., Somerville R. S., Primack J. R., Dekel A., 2008, ApJ, 384, 386
Croton D. J. et al., 2006, MNRAS, 365, 11
Daddi E. et al., 2007, ApJ, 670, 156
Davé R., 2011, in Treyer M., Lee J. C., Seibert M. H., Wyder T., Neill J., eds, ASP Conf. Ser. Vol. 440, UP2010: Have Observations Revealed a Variable Upper End of the Initial Mass Function? Astron. Soc. Pac., San Francisco, p. 353
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dekel A., Silk J., 1986, ApJ, 303, 39
Dekel A., Sari R., Ceverino D., 2009, ApJ, 703, 785
Drory N., Bender R., Feulner G., Hopp U., Maraston C., Snigula J., Hill G. J., 2004, ApJ, 608, 742
Drory N., Salvato M., Gabasch A., Bender R., Hopp U., Feulner G., Pannella M., 2005, ApJ, 608, 742
Dunne L. et al., 2009, MNRAS, 394, 3
Dutton A. A., van den Bosch F. C., Dekel A., 2010, MNRAS, 405, 1690
Elbaz D. et al., 2007, A&A, 468, 33
Eyles L. P., Bunker A. J., Ellis R. S., M., 2005, ApJ, 608, 742
Finkelstein S. L., Papovich C., Giavalisco M., Reddy N. A., Ferguson H. C., Koekemoer A. M., Dickinson M., 2010, ApJ, 719, 1250
Finlator K., Davé R., Papovich C., Hernquist L., 2006, ApJ, 639, 672
Finlator K., Oppenheimer B. D., 2011, MNRAS, 410, 1703
Fontana A. et al., 2006, A&A, 459, 745
Fontanot F., De Lucia G., Monaco P., Somerville R. S., Santini P., 2009, MNRAS, 397, 1776
Fontanot F., Pasquali A., De Lucia G., van den Bosch F. C., Somerville R., Kang X., 2011, MNRAS, 413, 957
Fuehner G., Gabasch A., Salvato M., Drory N., Hopp U., Bender R., 2005, ApJ, 633, L9
Gnedin N. Y., Kravtsov A. V., 2010, ApJ, 714, 287
González V., Labbé I., Bouwens R. J., Illingworth G., Franx M., Kriek M., Brammer G. B., 2010, ApJ, 713, 115

© 2011 The Authors, MNRAS 417, 2737–2751
Monthly Notices of the Royal Astronomical Society © 2011 RAS
