MNRAS **472,** 1576–1592 (2017)
Advance Access publication 2017 August 18

doi:10.1093/mnras/stx2132

A minimalist feedback-regulated model for galaxy formation during the epoch of reionization

Steven R. Furlanetto,[1]★ Jordan Mirocha,[1] Richard H. Mebane[1] and Guochao Sun[2]

[1]*Department of Physics & Astronomy, University of California, Los Angeles, Los Angeles, CA 90095, USA*
[2]*Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA*

Accepted 2017 August 15. Received 2017 August 15; in original form 2016 October 30

ABSTRACT

Near-infrared surveys have now determined the luminosity functions of galaxies at $6 \lesssim z \lesssim 8$ to impressive precision and identified a number of candidates at even earlier times. Here, we develop a simple analytic model to describe these populations that allows physically motivated extrapolation to earlier times and fainter luminosities. We assume that galaxies grow through accretion on to dark matter haloes, which we model by matching haloes at fixed number density across redshift, and that stellar feedback limits the star formation rate. We allow for a variety of feedback mechanisms, including regulation through supernova energy and momentum from radiation pressure. We show that reasonable choices for the feedback parameters can fit the available galaxy data, which in turn substantially limits the range of plausible extrapolations of the luminosity function to earlier times and fainter luminosities: for example, the global star formation rate declines rapidly (by a factor of $\sim$20 from $z = 6$ to 15 in our fiducial model), but the bright galaxies accessible to observations decline even faster (by a factor $\gtrsim$400 over the same range). Our framework helps us develop intuition for the range of expectations permitted by simple models of high-z galaxies that build on our understanding of 'normal' galaxy evolution. We also provide predictions for galaxy measurements by future facilities, including *James Webb Space Telescope* and *Wide-Field Infrared Survey Telescope*.

Key words: galaxies: high-redshift – dark ages, reionization, first stars – cosmology: theory.

1 INTRODUCTION

One of the major goals of extragalactic astrophysics is to map the formation of the first galaxies and their evolution into the mature objects we observe at $z \lesssim 1$. Much of that story has been revealed over the last two decades, with observational astronomers pushing the frontiers past the peak era of star formation, where we are now developing a sophisticated understanding of galaxy evolution. Over the past several years, the Wide Field Camera 3 on the *Hubble Space Telescope* has enabled a census of star formation in the first billion years of the Universe's history at $z \sim$ 6–8 (McLure et al. 2013; Schenker et al. 2013; Atek et al. 2015a; Bouwens et al. 2015a; Finkelstein et al. 2015; Bowler et al. 2017), with a few sources now detected at $z \sim$ 9–11 (Oesch et al. 2013, 2015; Ishigaki et al. 2015; McLeod et al. 2015; Bouwens et al. 2016; McLeod, McLure & Dunlop 2016).

While we know little more than the abundance of bright galaxies at this time, enough progress has been made to begin modelling the contents, formation and evolution of these early generations of galaxies. There is (as yet) no evidence that these galaxies have particularly unusual stars (Dunlop et al. 2012; Bouwens et al. 2014), though there are hints that the detailed processes of star formation differ (Muñoz & Furlanetto 2014; Capak et al. 2015). Qualitatively, their behaviour is consistent with galaxies at lower redshifts (Behroozi, Wechsler & Conroy 2013; Sun & Furlanetto 2016), with a peak in the star formation efficiency at a halo mass $m_{\rm h} \sim 10^{11}$–10^{12} M$_\odot$ and $\lesssim$10 per cent of the baryons converted into stars.

With this basic picture in hand, extensive theoretical work is underway to understand the processes driving these sources. Much of this has been performed in the context of detailed numerical simulations (e.g. Jaacks et al. 2012; Salvaterra et al. 2013; Genel et al. 2014; Ocvirk et al. 2016; Trac, Cen & Mansfield 2015; Feng et al. 2016; Gnedin 2016; Waters et al. 2016) and semi-analytic models (e.g. Trenti et al. 2010; Tacchella, Trenti & Carollo 2013; Cai et al. 2014; Dayal et al. 2014; Liu et al. 2016; Mutch et al. 2016b; Yue, Ferrara & Xu 2016). Here we take a different approach, developing a flexible, analytic model of galaxy formation from a set of simple, transparent physical assumptions. While such a model cannot be used to describe the detailed properties of the galaxies, it provides a basis for understanding the qualitative features of the galaxy populations (which are, so far, the limits of our observations). Importantly, it also allows *physically motivated* extrapolation

★E-mail: sfurlane@astro.ucla.edu

to both early times and faint luminosities, guiding our understanding of empirical extrapolations that are now widely used (Robertson et al. 2013, 2015; Mason, Trenti & Treu 2015b; Visbal, Haiman & Bryan 2015; Mashian, Oesch & Loeb 2016; Sun & Furlanetto 2016; Mirocha, Furlanetto & Sun 2017). Such extrapolations are crucial not only to observations of galaxies at earlier times (with, e.g. the future *James Webb Telescope*, or *JWST*, and the *Wide-Field Infrared Survey Telescope*, or *WFIRST*) but also to measurements of the reionization process, which depends on the *cumulative* number of photons emitted by the entire galaxy population.

The model presented below is similar in spirit to those of Dayal et al. (2014) and Trenti et al. (2010), though simpler and more flexible than both, and is inspired by the basic physics driving galaxy formation models at lower redshifts. We assume galaxies grow inside dark matter haloes and form stars as those haloes accrete more material. We set the star formation rate by assuming that stellar feedback, through radiation or supernovae, ejects the remaining inflowing material. We will show that such a framework can fit all available observations with reasonable parameter choices and that calibrating to those observations substantially limits the expected behaviour at faint luminosities and earlier times.

This paper is organized as follows. In Section 2, we describe our treatment of dark matter haloes, including their abundance and growth, which forms the basis for our calculations. In Section 3, we describe the feedback model that specifies the star formation rates in our model. In Section 4, we compare our model to observations of distant galaxies and show predictions for the overall star formation rate. In Section 5, we derive some simple properties of the galaxy population in the context of our model. Finally, in Section 6 we consider some of the model's implications for reionization and metal production, and we conclude in Section 7.

The numerical calculations here assume $\Omega_{\rm m} = 0.308$, $\Omega_\Lambda = 0.692$, $\Omega_{\rm b} = 0.0484$, $h = 0.678$, $\sigma_8 = 0.815$ and $n_{\rm s} = 0.968$, consistent with the recent results of Planck Collaboration XIII (2016). Unless specified otherwise, all distances quoted herein are in comoving units.

2 DARK MATTER HALOES

Our model for the galaxy population will use a few simple ingredients for the properties of galaxies, intentionally choosing the simplest options permitted by detailed observations and simulations. In this section, we will describe their cosmological context, the dark matter haloes.

2.1 The halo mass function

We assume that galaxies inhabit dark matter haloes and let $n_{\rm h}(m, z)$ be the comoving number density of dark matter haloes with masses in the range $(m, m + {\rm d}m)$ at a redshift z. Following convention, we write this as

$$n_{\rm h}(m, z) = f(\sigma)\frac{\bar{\rho}}{m}{\rm d}\ln\frac{(1/\sigma)}{{\rm d}m}, \quad (1)$$

where $\bar{\rho}$ is the comoving matter density, $\sigma(m, z)$ is the rms fluctuation of the linear density field, smoothed on a scale m and $f(\sigma)$ is a dimensionless function that parametrizes the barrier-crossing distribution of the linear density field. We assume for simplicity that this function is 'universal,' though simulations show there may be small deviations (Tinker et al. 2008). For our default calculations, we take $f(\sigma)$ from a fit to recent high-z cosmological simulations (Trac et al. 2015):

$$f(\sigma) = 0.150\left[1 + \left(\frac{\sigma}{2.54}\right)^{-1.36}\right]{\rm e}^{-1.14/\sigma^2}. \quad (2)$$

This fit matches the Tinker et al. (2008) predictions (which were based on simulations at lower redshifts) very well for low-mass haloes, but it has ∼20 per cent more haloes at $\gtrsim 10^{10}$ M$_\odot$ (see Trac et al. 2015 for a more detailed comparison). The older Sheth & Tormen (2002) fit, motivated by ellipsoidal collapse, predicts even more high-mass haloes but again agrees at small and moderate masses. We have also recomputed our results for the Sheth & Tormen (2002) mass function in order to test their sensitivity to uncertainties in the halo mass function. Differences in the normalization between model mass functions can simply be absorbed into the uncertain efficiency factors below. Thus only the shape matters significantly, and such differences usually only occur on the very massive end.

We will further assume that only haloes above a specified minimum mass $m_{\rm min}$ can form stars. There are several potential physical reasons for such a threshold. The first stars likely form through the cooling of molecular hydrogen, forming Pop III stars. However, molecular hydrogen is destroyed by a weak UV background; the observed galaxy population is very unlikely to harbour such stars. In the spirit of extrapolating only the known galaxy population, we therefore ignore this possibility in the following. Thus, only if the halo's virial temperature $T_{\rm vir} \gtrsim 10^4$ K can atomic line cooling in primordial gas allow efficient gas cooling and clumping to the densities required for star formation. Second, if the intergalactic medium (IGM) has been photoheated through reionization, accretion is suppressed on to galaxies with $m \lesssim 10^9$ M$_\odot$ (Noh & McQuinn 2014). We will use the atomic cooling threshold, which corresponds to a mass $m \sim 10^8$ M$_\odot$. Feedback from reionization would then be reflected in the shape of the star formation efficiency (see below).

2.2 Accretion rates

The next ingredient of our model is how, on average, haloes accrete matter. We will assume that the rate at which galaxies form stars depends on this overall accretion rate, as described in the next section. For simplicity, we will ignore scatter in this relation: though mergers are an important part of halo growth, even at moderate redshifts the majority of matter is acquired through smooth ongoing accretion (Goerdt et al. 2015).

For a wide range of redshifts, and over moderately large halo masses, simulations have measured relations similar to (McBride, Fakhouri & Ma 2009; Goerdt et al. 2015; Trac, Cen & Mansfield 2015)

$$\dot{m} = Am^\mu(1 + z)^\beta, \quad (3)$$

where A is a normalization constant, m is the halo mass, $\mu \gtrsim 1$ in the simulations,[1] and β converges to 5/2 at large redshifts. Dekel et al. (2013) (see also Neistein, van den Bosch & Dekel 2006; Neistein & Dekel 2008) argued that this form can be generically understood through the extended Press-Schechter algorithm (Lacey & Cole 1993): μ follows from the shape of the matter power spectrum, while $\beta \approx 5/2$ follows from the way the halo mass function depends on redshift (through the critical overdensity for collapse).

[1] Specifically, for example, McBride et al. (2009) find $\mu = 1.127$ at moderate redshifts, while Trac et al. (2015) find $\mu = 1.06$ at $z \sim 6$–10 and $10^8 \lesssim (m/{\rm M}_\odot) \lesssim 10^{13}$.

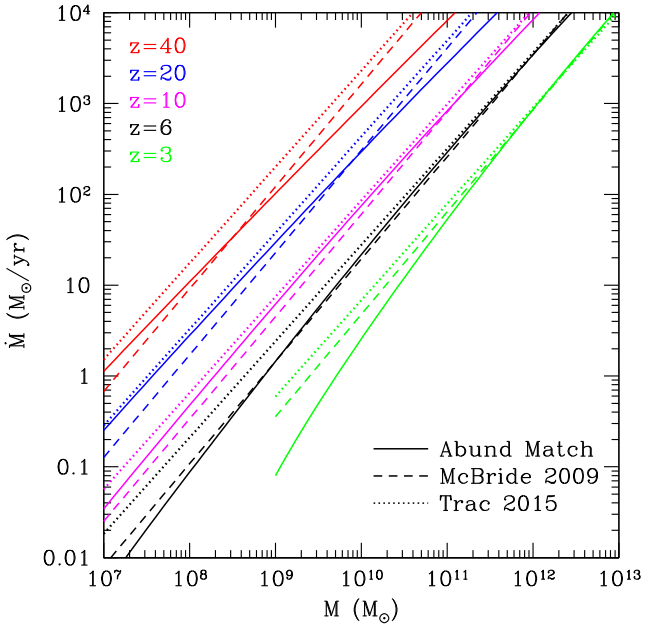

Figure 1. Total accretion rate as a function of halo mass, for $z = 3, 6, 10, 20$ and 40 (from bottom to top). The solid lines take our fiducial abundance-matched accretion rates (from equation 4). We compare these to the fits from McBride et al. (2009) (dashed lines) and from Trac et al. (2015) (dotted lines). The former is based on numerical simulations of moderate to large haloes at $z \lesssim 6$ (note the close agreement between the two approaches at $z = 3$ for $M \gtrsim 10^{11}$ M$_\odot$), while the latter is based on numerical simulations at $z \sim 6$–10.

Unfortunately, the validity of this relation has not been tested at very high redshifts or at very small masses, ranges that are of interest to us. Some of our results are quite sensitive to such uncertainties, as we will be following halo growth over many e-foldings in mass. We therefore use a slightly more sophisticated model for our main results. We make the ansatz that haloes maintain their overall number density as they evolve according to the mass function. The idea is similar to abundance matching (Vale & Ostriker 2004), which populates the galaxy luminosity function with haloes by matching number densities, and to studies that interpret the growth of the galaxy population by comparing objects at fixed number density across many redshifts (van Dokkum et al. 2010). Our approach is the direct analogue of the latter, as we match the halo mass function at different redshifts in order to determine the accretion rate. That is, we demand that at any given pair of redshifts z_1 and z_2 a halo has masses m_1 and m_2 such that

$$\int_{m_1}^{\infty} \mathrm{d}m\; n_\mathrm{h}(m|z_1) = \int_{m_2}^{\infty} \mathrm{d}m\; n_\mathrm{h}(m|z_2). \tag{4}$$

We then assign an accretion rate $\dot{m}(m, z)$ by demanding that this is true for all redshifts and all halo masses. We note that the analogous scatter-free mapping between galaxies across redshifts is too simple to explain detailed observations (Wellons & Torrey 2017), but given the limits of the current observations and the lack of better theoretical models at the masses and redshifts of interest, we choose the simplest possible approach that still guarantees self-consistency between the halo mass assembly histories and the mass function across many redshifts.

Fig. 1 compares these two approaches across a broad range of redshifts and halo masses. Although none of these models is likely to be correct in detail, given the inevitable stochastic growth rates of haloes, it is reassuring that they match so well without any tuning required. The approaches disagree at small masses at $z \sim 3$, where the simulation fit has not been tested and where abundance matching is complicated by the behaviour of the mass function,[2] but they match very well at $M \gtrsim 10^{10.5}$ M$_\odot$. Our model also diverges from the simulation fits at higher redshifts. This is not surprising: the Trac et al. (2015) result is fit to haloes at $z \sim 6$–10 over the range $\sim 10^8$–10^{12} M$_\odot$. Except for the smallest haloes at $z \sim 6$, our model provides an excellent match to the accretion rates over the range in which it can be compared directly to simulations.

It is important to note that, although these methods match rather well, the deviations nevertheless cannot be neglected for many applications. Because the specific mass accretion rate (i.e. $\dot{m}_\mathrm{h}/m_\mathrm{h}$) is nearly independent of mass, any given halo grows (almost) exponentially with redshift. Dekel et al. (2013) show that, neglecting the weak mass dependence in the specific accretion rate and taking the $\mu = 5/2$,

$$m_\mathrm{h}(z) \approx m_0 \mathrm{e}^{-\alpha(z-z_0)}, \tag{5}$$

where $\alpha = 0.79$, and m_0 is the halo's mass at z_0. The parameter α is proportional to the normalization of the accretion rate, so small deviations in it cause large deviations in the history of individual haloes when extrapolated over long time intervals.

We note here that ongoing accretion on to existing galaxies is not the only route through which stars can form in our model: as haloes cross the star formation threshold m_min, they convert a fraction of their total baryonic mass to stars. Because the mass function is so steep at high redshifts, this channel can in principle account for a fair fraction of the total star formation, as we show explicitly in Appendix A. Our model ignores the luminosity generated by this phase, because the physics of this process is so uncertain (as is the gas content of the haloes, if they were subject to earlier bursts of, e.g. Pop III star formation). Assuming that the star formation occurs over a fraction of the Hubble time (corresponding to the gas cooling time and/or the dynamical time of the halo), this will not impact our predictions within the observable range, as the star formation efficiency within these systems is very small.

2.3 Quenching of accretion

At large halo masses, semi-analytic models of galaxy formation often appeal to quenching from high virial temperatures and/or AGN feedback to decrease the star formation efficiency in accord with observations. We will include the effects of virial shock heating as an example here, though it will not have a large effect on most of our results. Faucher-Giguère, Kereš & Ma (2011) found that the fraction of gas able to cool on to the central galaxy in the presence of the virial shock is[3]

$$f_\mathrm{shock} \approx 0.47 \left(\frac{1+z}{4}\right)^{0.38} \left(\frac{m_\mathrm{h}}{10^{12}\ \mathrm{M}_\odot}\right)^{-0.25}. \tag{6}$$

We note that this expression is only approximate, as it resulted from a suite of hydrodynamic simulations without outflows and without metal-line cooling, performed at redshifts smaller than our era of interest. We shall see that it only has a modest effect on the results.

[2] In particular, our model fails at the low-mass end when haloes of a given mass have a declining number density, due to their incorporation into more and more massive haloes.

[3] Obviously, we require that $f_\mathrm{shock} \leq 1$.

2.4 Merger-driven growth

A fraction of the gas available to high-z haloes arrives in the form of mergers. While this has not been measured directly in the relevant redshift regime, at moderate redshifts (and in more massive haloes), the fraction is ∼20 per cent (Goerdt et al. 2015). This induces considerable scatter in the effective accretion rate (and hence instantaneous star formation rate) within haloes. As a contrast to our fiducial assumption of completely smooth growth, we therefore consider an alternative model in which *all* accretion is through major mergers.

With this assumption, and using the Trac et al. (2015) accretion rates, we find that the number of major mergers per halo per Hubble time $t_{\rm H} = H^{-1}(z)$ is

$$N_{\rm merge} \approx 2.75 \left(\frac{1+z}{7}\right), \tag{7}$$

where we have neglected the weak mass dependence in the specific accretion rates measured by Trac et al. (2015). If we further assume that each merger is followed by star formation over a time-scale equal to the dynamical time of the host halo ($\approx t_{\rm H}/\sqrt{\Delta_{\rm vir}}$, where $\Delta_{\rm vir} = 18\pi^2$ is the virial overdensity), the fraction of haloes actively forming stars at any given time is

$$f_{\rm merge} \approx 0.2 \left(\frac{1+z}{7}\right). \tag{8}$$

Interestingly, these high-z haloes are growing extremely fast, so that even in this extreme picture each one is forming stars for a large fraction of the Universe's history. (Clearly this picture breaks down at sufficiently early times, when $f_{\rm merge} > 1$, but we will not concern ourselves with such early star formation here.)

To model this possibility, we assume that a fraction $f_{\rm merge}$ of haloes at each mass are actively forming stars, but during those episodes they accrete gas at a rate $1/f_{\rm merge}$ larger than the value provided by our abundance-matching prescription.

3 FEEDBACK-REGULATED STAR FORMATION

We make the simple ansatz that the star formation rate in a galaxy, $\dot{m}_\star$, is a balance between gas accretion and stellar feedback, which could arise from radiation pressure, supernovae, or some other process like grain heating. We therefore write

$$\dot{m}_{\rm b} = \dot{m}_\star + \dot{m}_{\rm w}, \tag{9}$$

where $\dot{m}_{\rm b}$ is the rate at which baryons accrete on to the halo (which we assume to be $\dot{m}_{\rm b} = [\Omega_{\rm b}/\Omega_{\rm m}]\dot{m}_{\rm h}$) and $\dot{m}_{\rm w}$ is the rate at which baryons are expelled. We then define the star formation efficiency as $f_\star = \dot{m}_\star/\dot{m}_{\rm b}$ and assume that the rate at which gas is expelled by feedback is proportional to the star formation rate, so that $\dot{m}_{\rm w} = \eta\dot{m}_\star$, where in general η can be a function of halo mass and redshift. Then

$$f_\star = \frac{f_{\rm shock}}{1+\eta(m_{\rm h}, z)}. \tag{10}$$

Here, we have inserted $f_{\rm shock}$ into the numerator in order to include the suppression of accretion by the virial shock; our $f_\star$ therefore represents the fraction of gas that is transformed into stars, relative to that which one naively assumes to accrete on to a halo.

This model is certainly a simplification. For example, it ignores gas residing in the galaxy's interstellar medium (ISM). In the 'bathtub' model for global galaxy growth, galaxies grow through a transient adjustment phase towards a quasi-equilibrium in which the ISM has a roughly constant mass (Dekel & Mandelker 2014). Once that equilibrium is reached, our model accurately describes the galaxy's growth, even though gas cycles through the ISM before it forms stars. If instead that component grows in proportion to $m_\star$, we simply change the constant term in the denominator of equation (10), while if it grows in proportional to $\dot{m}_{\rm b}$, we would simply change the overall normalization of $f_\star$.

Although we have motivated our model for $f_\star$ with explicit reference to stellar feedback, we note that when fit to observations (as in Sun & Furlanetto 2016), such a parameter *cannot* be directly translated into the efficiency with which galaxies cycle their gas into stars, as it will be degenerate with any other limitation on star formation – or gas accretion. For example, heating during reionization can suppress accretion on to small haloes; in our model, this would manifest as a decline in $f_\star$ at small halo masses. A reduction in the gas fraction in large galaxies due to mergers would also appear as a decline in $f_\star$ in these systems. Below we will therefore allow a broader parametrization of η to allow for deviations from our naive expectations.

The key assumption so far is that there is no limitation on the star formation rate other than feedback. Because our expression for $f_\star$ is an increasing function of halo mass, this is certainly a poor assumption at high enough halo masses, where $f_\star$ eventually approaches unity. In the calculations that follow we therefore impose a maximum $f_{\star,\,\rm max}$, which for now we leave as a free parameter. Our results below will depend on the derivative of $f_\star$ with respect to mass, so we impose this maximum through

$$f_\star = \frac{f_{\rm shock}}{f_{\star,\rm max}^{-1} + \eta(m_{\rm h}, z)} \tag{11}$$

in order to maintain continuity.

Note that our model specifies the *instantaneous star formation efficiency*, or the fraction of accreting gas converted into stars. The total star formation rate will be $f_\star\dot{m}_{\rm b}$: even though our prescription applies equally well to smooth accretion and merger events, the latter will have higher overall star formation rates at fixed $f_\star$ because of their increased (temporary) accretion rates.

3.1 Models of feedback

In our fiducial model, we obtain η by assuming that the star formation rate is set by the balance between feedback and the forces binding the gas to a galaxy. However, the coupling between the feedback source and the galaxy's ISM is not yet clear, so we parametrize the feedback mechanism in a flexible manner.

One scenario appeals to supernovae to control the star formation rate, assuming that supernova blastwaves retain their energy long enough to disturb the accreting gas, so that we can balance the rate of energy input from supernovae with the rate at which the accreting gas acquires binding energy. Then

$$\frac{1}{2}\dot{m}_{\rm w} v_{\rm esc}^2 = \dot{m}_\star \epsilon_K \omega_{\rm SN}, \tag{12}$$

where $v_{\rm esc}$ is the halo escape velocity, $\omega_{\rm SN} = 10^{49}\omega_{49}$ erg $\rm M_\odot^{-1}$ is the energy released in supernovae per unit mass of star formation (determined by the stellar IMF and metallicity) and ϵ_K is the fraction of that energy released in the wind. Here, ω_{49} is of order unity for a typical initial mass function (IMF). This sets our fiducial feedback parameter to be

$$\eta_{\rm E} = 10\epsilon_K\omega_{49}\left(\frac{10^{11.5}\,\rm M_\odot}{m_{\rm h}}\right)^{2/3}\left(\frac{9}{1+z}\right). \tag{13}$$

The prescription in equation (12) assumes that a fixed fraction ϵ_K of the supernova kinetic energy is available to lift gas out of the dark matter halo. In fact, the high densities and temperatures of supernova blastwaves at these early times, as well as collisions between nearby ejecta, imply that some fraction of the energy will be lost to radiative cooling or other processes. If this fraction is large, feedback is much less efficient. A more conservative limit on the star formation rate is therefore provided by momentum conservation. We compare the momentum released in supernovae (or other feedback mechanisms, like radiation pressure) to the momentum required to lift the gas out of the halo at the escape velocity. We write the momentum injection rate as

$$\dot{P} = \pi_{\rm fid} \dot{P}_0 \left(\frac{\dot{m}_\star}{\rm M_\odot\ yr^{-1}} \right), \tag{14}$$

where $\pi_{\rm fid}$ is of order unity for a typical IMF and $\dot{P}_0 = 2 \times 10^{33}$ g cm s^{-2} (which equals the momentum input from a Salpeter IMF with solar metallicity). If a fraction $\epsilon_{\rm p}$ of this momentum is used to drive a wind, we have $\epsilon_{\rm p} \pi_{\rm fid} \dot{P}_0 \dot{m}_\star = \dot{m}_{\rm b} v_{\rm esc}$, or

$$\eta_{\rm p} = \epsilon_{\rm p} \pi_{\rm fid} \left(\frac{10^{11.5}\ {\rm M_\odot}}{m_{\rm h}} \right)^{1/3} \left(\frac{9}{1+z} \right)^{1/2}. \tag{15}$$

Because feedback transitions between these two regimes, and because other physics is most certainly relevant as well, we will also allow for a more general form and parametrize this factor as a power law in mass and redshift,

$$\eta = C \left(\frac{10^{11.5}\ {\rm M_\odot}}{m_{\rm h}} \right)^{\xi} \left(\frac{9}{1+z} \right)^{\sigma}, \tag{16}$$

where C is a normalization constant that can be fixed by comparison to observations for a given choice of power-law indices. By performing a fit to the galaxy luminosity functions at $z \sim 6$–8, Sun & Furlanetto (2016) found that $\xi \sim 1/3$–$2/3$ (bracketed by our energy and momentum conservation models) provides an adequate fit to the observed luminosity functions at $z \gtrsim 6$. Mirocha et al. (2017) showed that an even broader range of mass dependence is allowed and that the redshift evolution at $z \gtrsim 6$ cannot yet be constrained empirically. We therefore introduce equation (16) to allow for additional physics in the feedback cycle (or in any other process that affects the star formation efficiency but is not captured explicitly in our model).

Fig. 2 shows the star formation efficiency in several models that provide adequate fits to the existing observational data (see below). Note that, at small masses, $f_\star$ increases with redshift because the binding energy of the host halo does as well, allowing more star formation before the wind breaks out. The momentum-regulated wind model also allows more star formation in small haloes. On the other hand, a larger value of ϵ_K decreases the overall star formation rate because more of the supernova energy is used to drive winds. Finally, note that virial heating only affects the largest masses. (It is this quenching that imposes the kink in $f_\star$ above $10^{11}\ {\rm M_\odot}$.)

3.2 The integrated star formation efficiency

Our star formation efficiency parameter $f_\star$ describes the fraction of accreted gas turned to stars over a short time-scale (formally, instantaneously). A halo's efficiency is often defined as the fraction

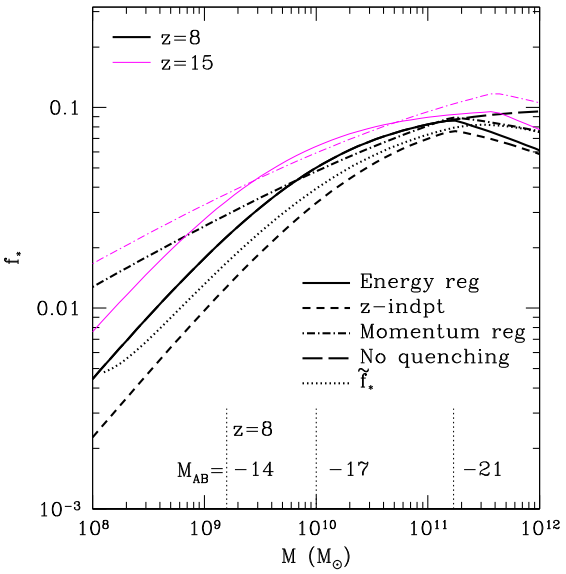

Figure 2. Star formation efficiency $f_\star$ in several models providing reasonable fits to the observed luminosity functions. The thick (thin) curves show $f_\star$ at $z = 8$ (15). The solid curves use our energy-driven wind model with $\epsilon_K = 0.1$ and $f_{\star,\rm max} = 0.1$. The dot–dashed curves use our momentum-driven wind model, with $\epsilon_{\rm p} = 0.2$ and $f_{\star,\rm max} = 0.2$. The long-dashed line ignores virial shock quenching in the energy-driven model. The short-dashed line takes equation (16) with $\xi = 2/3$ and $\sigma = 0$ (i.e. assuming it is independent of redshift). The dotted curve shows $\tilde{f}_\star$, or the average star formation efficiency over a halo's history, in our energy-driven wind model at $z = 8$. Along the bottom axis we also show how these masses translate into absolute magnitudes in our energy-driven wind model at $z = 8$.

of the baryons associated with a halo that has been transformed into stars:[4]

$$\frac{m_\star}{m_{\rm h}} = \tilde{f}_\star \frac{\Omega_{\rm b}}{\Omega_{\rm m}}. \tag{17}$$

We can compare $f_\star$ and $\tilde{f}_\star$ through direct integration of the accretion rate and star formation rate for each halo. We can obtain an approximate form by using equation (3) with $\mu = 1$ (i.e. a mass-independent specific accretion rate), assuming that the accreted mass is much larger than the initial mass, taking equation (16) with $\eta \gg 1$ for the feedback efficiency, and ignoring $f_{\star,\,\rm max}$. Then

$$\tilde{f}_\star(m_{\rm h}, z) \approx \frac{f_\star(m_{\rm h}, z)}{1+\xi}. \tag{18}$$

For our energy-driven and momentum-driven models, this yields $\tilde{f}_\star \approx (0.6$–$0.75) f_\star$. The net star formation efficiency is smaller than the instantaneous version because $f_\star$ is an increasing function of the exponentially growing mass.

Fig. 2 shows this integrated efficiency at $z = 8$ in our fiducial energy-driven wind model with the dotted curve. Our approximation is reasonably accurate except for the smallest masses (where we have assumed a pre-existing mass of stars equal to the instantaneous $f_\star$) and near the peak efficiency (where quenching and $f_{\star,\rm max}$ introduce complexities).

We emphasize that both versions have (approximately) the same mass and redshift dependence; because the normalization factor is

[4] Note that this is also not quite the fraction of a halo's nominal baryonic mass that is in stars at any given time, because some fraction of the stars would have already completed their core fusion life cycles.

uncertain by at least a factor of 2, the distinction between $f_\star$ and $\tilde{f}_\star$ is likely not terribly important in comparison to observations.

3.3 From star formation to luminosity

Finally, in order to compare our model to observations we must translate our star formation rates into luminosities. Because the UV luminosity results from massive, short-lived stars, it is a good tracer of the star formation rate. We there take the standard conversion

$$\dot{m}_\star = \mathcal{K}_{\rm UV} \times L_{\rm UV}, \tag{19}$$

where $L_{\rm UV}$ is the intrinsic (i.e. without dust) luminosity in the rest-frame continuum (1500–2800 Å) and $\mathcal{K}_{\rm UV}$ is a proportionality constant that depends on the IMF, star formation history, metallicity, binarity, etc. Following Madau & Dickinson (2014), we take $\mathcal{K}_{\rm UV} = 1.15 \times 10^{-28}\,{\rm M}_\odot\,{\rm yr}^{-1}/({\rm ergs\,s}^{-1}\,{\rm Hz}^{-1})$. Those authors chose this as a reasonable compromise for models with a Salpeter IMF (from 0.1 to 100 $\rm M_\odot$) and long periods of continuous star formation ($\gtrsim$300 Myr). It is not particularly sensitive to metallicity for such IMFs, but (according to the Conroy, Gunn & White 2009 models) it nominally corresponds to a metallicity slightly above $0.3\,{\rm Z}_\odot$. The assumption of long periods of star formation is not accurate in this high-redshift regime, but we have retained this simple conversion for easy comparison with existing results. We also note that this corresponds to a population with $Z = 0.05\,{\rm Z}_\odot$, a Salpeter IMF, and no binaries according to the BPASS v1.0 models (Eldridge & Stanway 2009; Mirocha et al. 2017), which gives a sense for the uncertainty in stellar models. We note that errors in this conversion affect the overall amplitude of the star formation efficiency, which is therefore uncertain by a factor of a few.

We ignore dust, both in our abundance matching procedure and in our luminosity estimates from the star formation rate. Based on recent measurements, the galaxy dust correction appears to be declining rapidly in the $z \sim 6$ regime (Dunlop et al. 2013; Bouwens et al. 2014). Bright galaxies have some evidence for dust at this time, but fainter ones do not appear to require it. At even higher redshifts, there is no evidence for dust even in the bright populations. As a result, other studies that have included dust find it provides only a modest increase to the estimated star formation rates at $z > 6$ (see e.g. fig. 7 of Sun & Furlanetto 2016 and fig. 10 of Mason et al. 2015b) except in the brightest galaxies. Nevertheless, we have neglected it here because we have found that, in this limit where the dust is evolving rapidly, the correction depends sensitively on its parametrization and requires careful consideration (see Smit et al. 2012 for the standard method). In the spirit of our very simple treatment, we have ignored it here, but we note that it could affect our interpretation of the bright end of the galaxy luminosity function. These uncertainties about parametrizing the dust content have also provided one of our motivations for confining our comparisons to $z \geq 6$.

Along the bottom axis of Fig. 2, we show the mass–luminosity relation corresponding to the solid curve at $z = 8$. Large volume surveys currently reach $M_{\rm AB} \approx -18$ at this redshift, so current observations only probe halo masses near the peaks of these curves, emphasizing the importance of a physical model to guide extrapolation to fainter luminosities.

4 RESULTS

The simple tools in the preceding section allow us to compute model star formation histories of the universe. In this section, we compare our results to the galaxy populations at $z \gtrsim 6$.

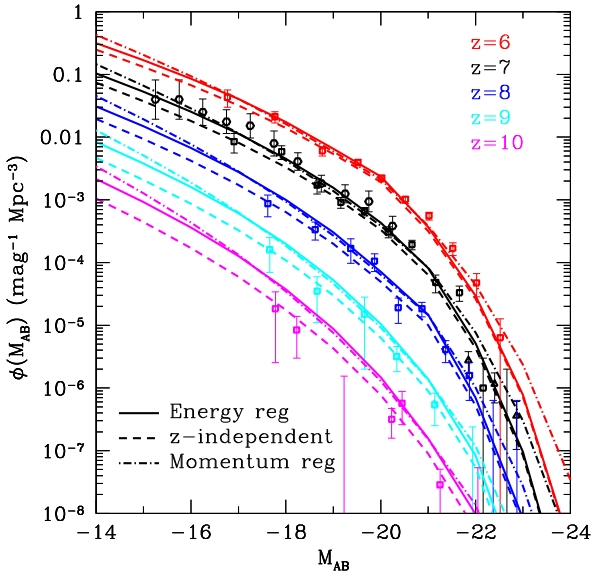

Figure 3. Comparisons of our model luminosity functions with the data at $z = 6$–10, including our fiducial energy-regulated model, momentum-regulated model and redshift-independent model (solid, dot–dashed and dashed curves, respectively). See the text for detailed parameter choices. In each panel, the $z = 6, 7, 8, 9$ and 10 luminosity functions are displaced by $+0.5, 0, -0.5, -1$ and -1.5 in logarithmic units along the ordinate for ease of display. We compare to luminosity functions from Bouwens et al. (2015a), supplementing with data from Oesch et al. (2013) and Bouwens et al. (2016) at $z = 9$–10 and from Bowler et al. (2017) and Atek et al. (2015a) at $z = 7$.

4.1 Comparison to measured luminosity functions

We first consider whether our simple model provides a reasonable fit to observed luminosity functions at $z \sim 6$–10 and use those observations to calibrate the free parameters in each model (the efficiency of feedback, ϵ_K or $\epsilon_{\rm p}$, and the maximal star formation efficiency $f_{\star,\,\rm max}$). Given the simplicity of our models and the many uncertainties in their implementation, we will not attempt a rigorous statistical test (better suited to more flexible empirical models; see Sun & Furlanetto 2016; Mirocha et al. 2017).

For concreteness, we will primarily compare our results to luminosity functions measured in the range $z = 6$–10 by Bouwens et al. (2015a), supplemented by Oesch et al. (2013) and Bouwens et al. (2016) at the highest redshifts, Bowler et al. (2017) at the bright end for $z \sim 7$ and Atek et al. (2015a) at the faint end for $z \sim 7$. We note that several other groups have produced luminosity functions in this range (e.g. McLure et al. 2013; Schenker et al. 2013; Finkelstein et al. 2015) but we have chosen one group with broad redshift coverage for consistency. The other measurements yield similar results.[5]

Fig. 3 compares our model to this data. To roughly match the data, we have set $\epsilon_K = 0.1$ and $f_{\star,\rm max} = 0.1$ for our energy-regulated model in the solid curves. These parameters do an excellent job reproducing the data at $z = 6$–8, but they lay at the upper end of the

[5] The Finkelstein et al. (2015) measurements are somewhat lower in overall number density than the Bouwens et al. (2015a), but as this amplitude is degenerate with the normalization of the star formation efficiency the difference does not affect our overall conclusions. The shapes of the luminosity functions are quite similar, as is their redshift evolution. See Mirocha et al. (2017) and Mason et al. (2015b) for more explicit comparisons between the two data sets.

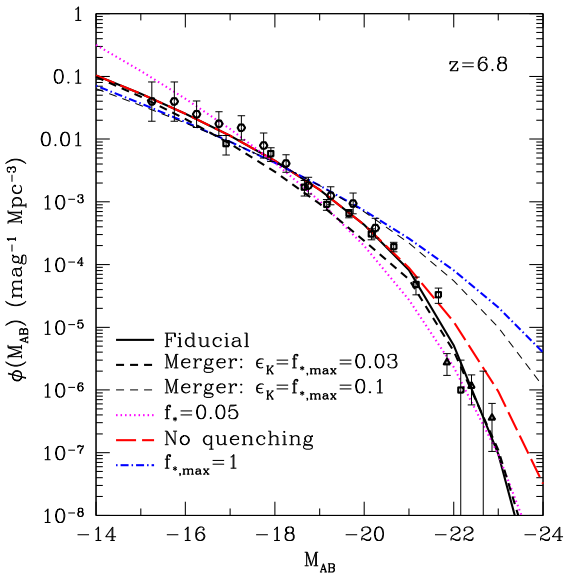

Figure 4. Illustration of the parameter dependence of our model luminosity functions. We show results at $z = 6.8$ and compare to data from Bouwens et al. (2015a), Bowler et al. (2017) and Atek et al. (2015a). The solid curve is our fiducial energy-regulated model ($\epsilon_K = 0.1$ and $f_{\star,\rm max} = 0.1$). The long-dashed curve ignores quenching of accretion, while the dot–dashed curve also takes $f_{\star,\rm max} = 1$ and $\epsilon_K = 0.2$. The dotted curve assumes a constant $f_\star = 0.05$. Finally, the short-dashed curves use our merger model. The thin, short-dashed curve follows our fiducial parameters, without quenching, while the thick short-dashed curve provides a better fit to observations by setting $\epsilon_K = 0.03$ and $f_{\star,\rm max} = 0.03$.

allowed range at the higher redshifts. This model has fairly steep redshift dependence, because the binding energy of haloes (at fixed halo mass) increases as $(1 + z)$.

To try to better match the different redshifts, the dashed curves set $\xi = 2/3$ and $\sigma = 0$ in equation 16). We set the normalization constant to be equivalent to $\epsilon_K = 0.2$ at $z = 8$ but keep $f_{\star,\,\rm max} = 0.1$. This underpredicts the abundance of moderately bright galaxies at $z = 6$ but provides a much better match to the data at $z > 8$.

The dot–dashed curves use our momentum-regulated model of equation (15), with $\epsilon_{\rm p} = 5$ and $f_{\star,\,\rm max} = 0.2$. This choice also provides a reasonably good fit to the data, though it too somewhat overpredicts the amplitude at $z > 8$. Thus, even in very simple models, at present it is difficult to conclude anything about the processes regulating star formation at high redshifts. However, the redshift dependence is already suggestive.

Fig. 4 explores some details of our model. We focus here on $z = 6.8$. The dotted line assumes a constant $f_\star = 0.05$, so that the luminosity function is close to a rescaling of the halo mass function (because $\dot{m}_h \propto m_h$, at least roughly; see equation 3). This fails badly in reproducing the shape of the luminosity function, demonstrating why models require $f_\star$ to decline towards small halo mass.

Two other curves show how our parameters affect the bright end. The long-dashed curve ignores the quenching of accretion: while not an enormous effect at this redshift, some kind of cutoff is nevertheless important to match the bright end. The dot–dashed curve sets $f_{\star,\rm max} = 1$ in addition to ignoring quenching: this dramatically overpredicts the number of bright galaxies. (This model also takes $\epsilon_K = 0.2$ to ensure a reasonable fit to the faint end.) We also note that these bright galaxies very likely have dust, which we have not modelled. As it will preferentially extinguish light from the brightest sources, its presence will certainly mitigate (if not eliminate) the need for quenching mechanisms. For example, Sun & Furlanetto (2016) found, using a standard dust correction, that dust can reduce the UV luminosity of the brightest galaxies at $z \lesssim 8$ by $\sim$2 magnitudes. Another possible explanation is a decrease in the gas accretion rate of massive galaxies, if they are built primarily from mergers of smaller systems that have consumed some of their gas.

A particularly interesting result at the bright end is the comparison to the Bowler et al. (2017), who favoured a double power law over a Schechter function at the bright end. Our models (which are ultimately sourced by the halo mass function) have no trouble fitting the bright end, and in fact – given the abundance of faint galaxies – demonstrate that, just as at low redshifts, we must impose restrictions on star formation in massive haloes in order to reproduce the data. Thus it is not surprising – at least from a theoretical perspective – that a double power law is preferred.

The other two curves in Fig. 4 explore our merger prescription. First, the thin short-dashed curve uses our fiducial parameters from the energy-regulated model but assumes that all star formation occurs in merger events. This dramatically overpredicts the abundance of bright galaxies. Recall that our extreme merger picture has the same *average* accretion rate as the smooth models, but it assumes that the gas inflow is large during brief bursts and off otherwise. Thus, when a galaxy is active, it has a much higher total star formation rate. Using our fiducial parameters therefore overpopulates the bright end by shifting smaller haloes towards larger luminosities. A better fit can be obtained by reducing the maximal efficiency of star formation while also decreasing the importance of feedback ($\epsilon_K = 0.03$ and $f_{\star,\rm max} = 0.03$).

Recently, Oesch et al. (2016) discovered a very bright galaxy ($M_{\rm AB} \sim -22.1$) with an inferred redshift of $z \sim 11.1$. Unless this object proves to be a fluke, it presents a serious challenge to the treatment of very bright galaxies in our models. Given their survey volume, the apparent number density of sources similar to this is $\sim 8 \times 10^{-7}$ Mpc^{-3}, albeit with very large errors from the single object detected. Our energy-regulated model, the most optimistic of our fiducial choices, falls about a factor of 10 short of this inferred number density. Although this galaxy is very bright, it is not massive enough to be subject to our quenching prescription (equation 6), and the only way to significantly increase the abundance at this luminosity is to remove the saturation level of star formation. We also note a tension with data from slightly lower redshifts: recall that our fiducial models already tend to overpredict the galaxy number density at $z \sim 9$–10. Adjusting our parameters to fit this $z \sim 11.1$ galaxy would dramatically increase the discrepancy at slightly lower redshifts: indeed, our redshift-independent model, which fits the $z \sim 9$–10 data nicely, is strongly discrepant with the $z = 11$ object. For now, this object remains very puzzling, though we note that some numerical simulations have found very bright objects to be more common than our models predict (see below; Mutch et al. 2016a; Waters et al. 2016).

4.2 The faint end of the luminosity function

With the advent of the Hubble Frontier Fields, observations have begun to probe intrinsically faint galaxies during the reionization era (Atek et al. 2015a; Kawamata et al. 2016; Livermore, Finkelstein & Lotz 2017). This is a particularly interesting question both because most of the light appears to be in this faint population (Robertson et al. 2013) and because reionization itself may affect these galaxies (e.g. Noh & McQuinn 2014). The open hexagons in Fig. 3 show the measurements of Atek et al. (2015a) at $z \sim 7$, which extend

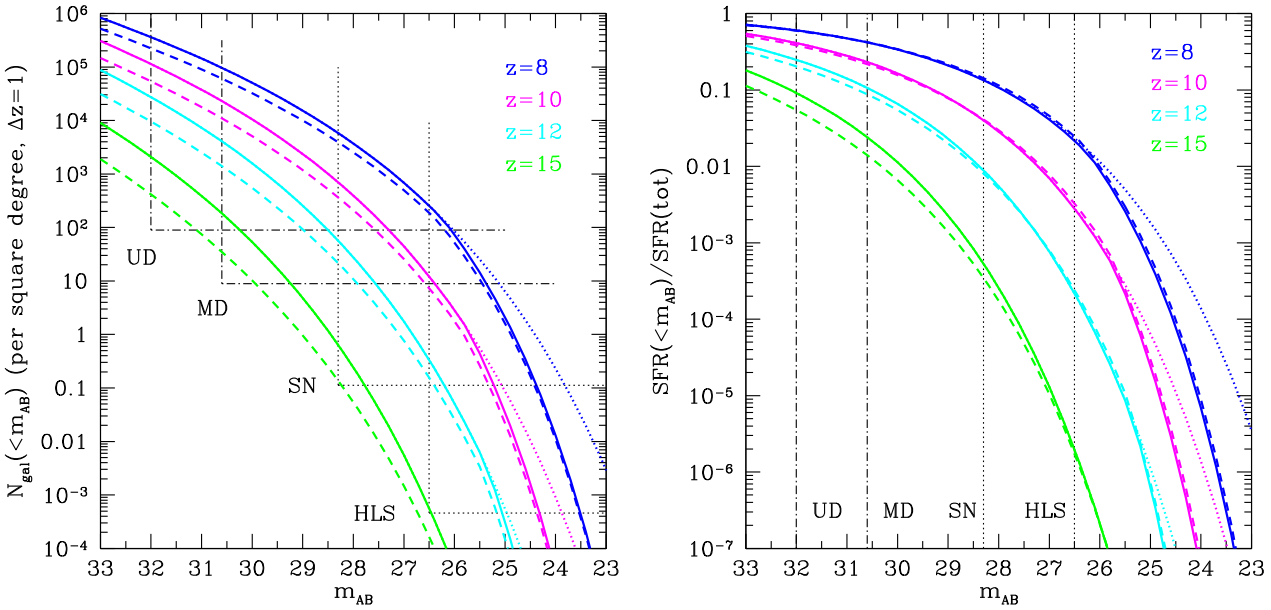

Figure 5. *Left:* Surface density of high-z galaxies in our models, with comparisons to several potential surveys. The curves show the number of galaxies more luminous than the specified apparent magnitude (in AB units) per square degree in surveys spanning $(z - 0.5, z + 0.5)$, where $z = 8, 10, 12$ and 15, from top to bottom. Within each set, the solid and dashed curves show our fiducial energy-regulated and redshift-independent models. The dotted curves use our energy-regulated model but ignore quenching. We compare these surface densities to four potential space-based surveys. The UD and MD surveys each assume 800 h total survey time with *JWST*, while SN and HLS refer to surveys similar to those in WFIRST's cosmology program. In each case, we show the approximate limiting magnitude (vertical lines) and the inverse of the survey's area. *Right:* Fraction of total star formation inside galaxies brighter than the specified limiting magnitude. The curves are identical to those in the left-hand panel. We assume that star formation extends to the atomic cooling threshold at all redshifts. Limiting depths of surveys are indicated by vertical lines.

about two magnitudes below the unlensed data. Nevertheless, they are still consistent with our models, indicating that they require no additional physics to explain them. In our models, $M \sim -15$ typically corresponds to a halo mass $m \sim 4 \times 10^9\,M_\odot$, just beginning to touch the range in which reionization might affect galaxy formation. Livermore et al. (2017) extend searches even further, to $M \sim -12$ at $z \sim 6$, albeit with large errors for the faintest galaxies. They find that their results are consistent with a faint-end slope ~ -2, very similar to inferences from unlensed samples. All of the models in Fig. 3 are also consistent with their measurements, again indicating no need to invoke reionization feedback at this time. However, our models do suggest that observations are not far from this interesting limit.

4.3 Future surveys

One of the key advantages of a physical model for galaxy formation is its utility in forecasting the results of future surveys. Our models are calibrated to the known galaxy population and essentially provide a conservative extrapolation of the existing data. The most interesting result from future surveys will be to discover deviations from these expectations.

To develop a baseline prediction, we next use our models to estimate the number of galaxies that will be observed at a variety of redshifts with two important future instruments, *JWST* and *WFIRST*. We consider four fiducial surveys, two with each observatory. For *WFIRST*, we include the High Latitude Survey (HLS), which will span 2227 deg^2 to an approximate infrared depth of $m \approx 26.5$, as well as a nine square degree deep field reaching a depth of $m \approx 28.3$ (SN). Both of these surveys approximate core parts of the mission that are structured for cosmological observations (lensing, baryon acoustic oscillations and supernovae) but will prove useful for many other science questions. For *JWST*, we follow Mason et al. (2015b) and consider an 'ultradeep (UD) survey consisting of four 200 h pointings (each ≈10 arcmin), which reaches $m \approx 32.0$, and another 'medium deep (MD) survey consisting of forty 20 h pointings, which reaches $m \approx 30.6$.[6] Note that we do not attempt to model the detailed selection functions or individual filter depths of any of these potential surveys, as the *JWST* programs are purely hypothetical, while *WFIRST* has not yet settled on a detailed mission design. We also neglect gravitational lensing, which can substantially affect the bright end (Mason et al. 2015a). Also, note that *WFIRST* will only have coverage to ≈2 μm, so its maximum redshift will be limited.

The left-hand panel of Fig. 5 shows our model predictions for the surface density of galaxies from $z = 8$ to 15, assuming surveys spanning $(z - 0.5, z + 0.5)$.[7] We compare these predictions to the four surveys outlined above. We illustrate the parameter space each can probe by marking their AB limiting magnitudes and the inverse of their areal coverage (which corresponds to the surface density above which they are expected to detect *at least one* object).

In our models, the characteristic luminosity declines rapidly with redshift. By $z \sim 15$, objects bright enough to be seen by the shallowest surveys are very, very rare. Even our UD survey would detect, at best, a few dozen objects at $z \sim 15$. The solid and dashed curves in this figure roughly correspond to optimistic and pessimistic extrapolations from models calibrated to fit the observed data. The largest

[6] Mason et al. (2015b) calculated these depths using the *JWST* Exposure Time Calculator in its pre-Cycle 1 form. They also considered a wide survey with similar depth to the *WFIRST* SN survey, but with only about 1/5 the area.

[7] We assume a flat spectrum in f_ν to estimate the K-correction for the AB magnitude flux density.

differences, at the faintest luminosities and highest redshifts, are up to an order of magnitude, which shows the precision required to make useful constraints on the physics of galaxy formation during this era (and/or detect deviations from our expectations).

However, all four of these surveys would detect many thousands of galaxies at $z \sim 8$–12. None will span the entire range of the luminosity function, but in combination they can map out a dynamic range of up to eight magnitudes. As an additional, albeit crude, figure of merit, in the right-hand panel of Fig. 5, we show the fraction of the total star formation at each redshift occurring in haloes brighter than the specified limiting magnitude. In all cases, we assume that star formation extends to the minimum mass for atomic cooling. Because of the steepness of the luminosity function, the deeper surveys are better at probing the majority of the emission: our UD survey, for example, is sensitive to ∼60 per cent (10 per cent) of the total star formation at $z = 8$ (15).

Nevertheless, we note that even *JWST* will require either large lensing surveys or substantial extrapolation to fainter luminosities if we are to compute quantities depending on the *total* star formation rate density in the Universe (such as reionization). Even without lensing, precision constraints on the shape of the luminosity function will improve substantially, which will be essential for extrapolating to the faint galaxies responsible for most of the emission. In this respect, large-area surveys will be extremely useful in avoiding biases due to cosmic variance (Robertson 2010). They are also essential for probing the full range of galaxy luminosities in order to study how galaxies evolve in this early epoch (e.g. measuring how chemical enrichment occurs as galaxies grow) and for cross-correlation with, e.g. highly redshifted 21 cm measurements and intensity mapping, both of which require very large spatial volumes (Lidz et al. 2009, 2011).

4.4 The star formation history

We next consider the evolution of the integrated star formation rate (and hence UV emissivity) over a broad redshift range, as shown in Fig. 6. Here the upper set of curves show the total star formation rate density, assuming that all haloes above the atomic cooling threshold (at a virial temperature $T_{\rm vir} \approx 10^4$ K) are able to form stars. After calibrating to the observations, many of our models (energy- and momentum-regulated) show very similar behaviour, within a factor ∼2 of each other over the entire redshift range. In detail, the momentum-regulated case has more star formation at most redshifts (because it imposes a weaker constraint on small haloes), while the merger-driven model falls short of the others at $z \lesssim 10$ (because it underpredicts the abundance of moderate-luminosity galaxies; see Fig. 4). The dashed curve is the same redshift-independent model shown in Fig. 3, which fits the $z \sim 10$ data better. It decreases more steeply towards early times, because (unlike the other models) the star formation efficiency of small haloes does not increase at higher redshifts.

The lower set of thin curves of Fig. 6 shows the star formation rate density in bright galaxies, here including all those with $M_{\rm AB} < -17.7$, as often calculated from observations. Unsurprisingly, this evolves much more rapidly than the total star formation rate density, as haloes massive enough to host bright galaxies become exponentially more rare towards early times. This is also clear from the right-hand panel of Fig. 5: at $z = 8$ (15), $M_{\rm AB} = -17.7$ corresponds to $m_{\rm AB} \approx 29.5$ (30.4). All of our models have similar behaviour in this respect.

We also include a comparison to measurements from several high-redshift surveys. All of these are consistent with our fiducial models, showing that our general framework describes bright galaxies quite well. However, we note that we have not applied dust corrections to these measurements, which can increase the overall star formation rate by a factor of a couple at $z \sim 6$ (Bouwens et al. 2015a).

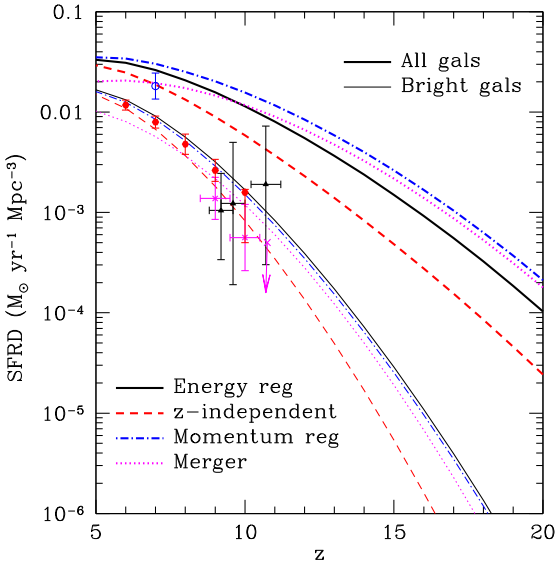

Figure 6. Integrated star formation histories of our models. The upper set of curves include galaxies down to the minimum virial temperature for atomic cooling, 10^4 K. The lower set of thin curves include only galaxies with $M_{\rm AB} < -17.7$. Within each set, the solid, dot–dashed, dotted and dot–dashed curves use our energy-regulated, merger-driven, momentum-regulated and redshift-independent models, respectively. We also compare our results to a number of surveys: CLASH cluster searches (Zheng et al. 2012; Coe et al. 2013; triangles), CANDELS/GOODS/HUDF (Oesch et al. 2014; crosses) and from composite measurements (McLeod et al. 2015; filled hexagons). These searches for luminous galaxies should be compared to the thin curves. The open circle shows the Atek et al. (2015b) measurement, which integrates down to $M_{\rm AB} = -15$. Error bars in each case are 1σ limits.

For completeness, we also note that, although we have allowed star formation to persist to very small halo masses (∼10^8 M$_\odot$), at $z \sim 10$ the smallest haloes produce a smaller fraction of the star formation than in many previous treatments: at $z = 8$, haloes with $M \gtrsim 10^9$ (10^{10}) M$_\odot$ produce $\gtrsim$75 per cent (30 per cent) of the total star formation (see the right-hand panel of Fig. 5). This is because internal feedback strongly suppresses star formation in shallow potential wells.

4.5 Comparison to existing models

We next consider our approach in the context of the many other models of the high-z galaxy luminosity function. Most obviously, our approach relies on a theoretically motivated framework to model the galaxy populations rather than providing empirical fits to the measurements. Several groups have used empirical fits to the high-z luminosity functions, with a range of parametrizations, to extrapolate the results to higher redshifts. Robertson et al. (2013, 2015) and Oesch et al. (2013) provided fits to the global star formation rate. Their results demonstrate the difficulty of empirical extrapolations, as the extrapolations are dramatically different because of different assumptions about the continuity of the imposed functional form. Of course, more detailed observations provide more basis for such empirical extrapolations, and the improving measurements of the galaxy luminosity functions in this high-z regime have led to

several recent attempts at more sophisticated empirical fits (e.g. Mason et al. 2015b; Visbal et al. 2015; Sun & Furlanetto 2016; Mirocha et al. 2017). All of these models agree that $f_\star(m)$ peaks at $m \sim 10^{11}$ M$_\odot$ and declines towards smaller masses, but they are also subject to systematic uncertainties in the way in which that function is parametrized (e.g. whether or not redshift-dependence is allowed). Sun & Furlanetto (2016) and Mirocha et al. (2017) included error estimates on their fit parameters, and our example models are all within the bounds allowed by those estimates.

The key difference between our approach and these other works is our reliance on an underlying theoretical model. There are clear advantages to both methods. To the extent that our model's fundamental ingredients are trustworthy, it provides more reliable extrapolations to higher redshifts, because it 'builds in' the proper redshift evolution. We have found that the remaining uncertainties in the data allow a fairly wide range of star formation prescriptions (in our case, expressed through the feedback model), but that normalizing to the $z \lesssim 10$ data nevertheless significantly limits the range of extrapolated star formation densities. Moreover, we have found that the bright galaxy population becomes an increasingly poor proxy for the total star formation rate. Empirical predictions based on the observed bright galaxies are therefore subject to large systematic uncertainties as they are extrapolated to higher redshifts.

Trenti et al. (2010) and Tacchella et al. (2013) developed a physical framework for galaxy formation using similar assumptions to ours, by tying star formation to the growth of dark matter haloes. They used the extended Press-Schechter approach (Lacey & Cole 1993) to estimate accretion rates and calibrated the overall star formation efficiency to the $z = 4$ galaxy luminosity function. Unlike our model, they included two separate star formation modes for bursts and secular growth. They found reasonable fits to the luminosity function from $z = 0$ to 6. Their overall star formation efficiency had similar mass dependence to our feedback-regulated models.

Dayal et al. (2014) also modelled the high-z luminosity and stellar mass functions. They used a merger tree model and assumed that star formation is regulated by feedback in a manner similar to our energy-regulated model. They also found a good match to the observed luminosity function over the redshifts we consider. Qualitatively, they found that mergers are responsible for most of the star formation, especially in massive haloes. Our model essentially averages over merger-induced bursts, but as we have argued mergers are so common at these early times that the distinction may not be clearcut.

Behroozi & Silk (2015) used an abundance matching model to predict high-redshift galaxy populations. Their model, originally developed to describe galaxies at low and moderate redshifts, uses a more complex abundance matching procedure based on stellar mass measurements and accounting for subhaloes, scatter in the mass–luminosity relation and other effects. Though they model the dark matter physics explicitly, they do not attempt to parametrize the star formation laws as we do. Because we calibrate our model exclusively at $z \geq 6$, we have focused on the well-measured UV luminosity functions, and we have shown that the data do not yet demand more complex treatments of the halo populations. Nevertheless, we find qualitatively similar results to Behroozi & Silk (2015).

Mason et al. (2015b) and Mashian et al. (2016) use abundance matching to make empirical fits to the star formation efficiency as a function of stellar mass and extrapolate to higher redshifts (see also Trenti et al. 2010). They include slightly different sets of effects in their modelling than us (e.g. Mashian et al. 2016 allows scatter in the halo–luminosity relation). Neither finds evidence for redshift evolution in this relation (though they also do not rule it out at the level implied by our model or by Behroozi & Silk 2015); however, Mason et al. (2015b) find an increase in *luminosity* with redshift at a fixed halo mass because of the increased growth rate of haloes at earlier times, similar to our model's increased accretion rate. Mason et al. (2015b) has the advantage of modelling a galaxy's luminosity more explicitly in terms of evolving stellar populations, but they find that only the most recent star formation episode contributes substantially to the overall luminosity, validating our simple approach (see their fig. 4). Like our model, these studies also predict rapid declines in the galaxy populations at $z \gtrsim 12$, making it difficult for *JWST* to observe significant numbers of galaxies at very high redshifts.

There are also a variety of more computationally intensive approaches to modelling the high-z galaxy population. For example, the DRAGONS program has constructed a semi-analytic model, embedded in numerical simulations of structure formation and reionization, to describe these objects (Liu et al. 2016; Mutch et al. 2016b). It includes much more of the detailed physics of galaxy formation, including mergers, infall, a multiphase ISM and recycling. The various free parameters of the model are calibrated by matching to the observed stellar mass function at $z = 5$–7. The stellar mass–halo mass relation from this model agrees reasonably well with empirical fits (see fig. 11 of Liu et al. 2016), except at the brightest end and so also agrees with our model reasonably well. That team has not fully explored the galaxy population at $z > 10$, but they have found significantly more bright galaxies at $z \sim 11$ than our model predicts (Mutch et al. 2016a), in better agreement with the object from Oesch et al. (2016).

Additionally, several numerical simulations have now studied high-z galaxy populations (e.g. Jaacks et al. 2012; Trac et al. 2015; Feng et al. 2016; Gnedin 2016). In general, these have subgrid star formation prescriptions calibrated to reproduce some set of observations, often at lower redshift, so their physical inputs are difficult to compare in detail with our model. We will focus on a couple of the more informative comparisons here. We have drawn on the dark matter measurements of Trac et al. (2015), who also use abundance matching to make predictions within their simulation framework, with results generally consistent with ours. Feng et al. (2016) studied distant galaxies in the BlueTides simulation (notably, they also included AGN), finding that their stellar feedback prescription was the most important tunable parameter in regulating the star formation rate of faint galaxies. Their predictions at very high redshifts follow the same qualitative trends as ours, though they do not decline at high redshifts as fast as some of our models (Waters et al. 2016). They also find that the bright end declines significantly less rapidly than a Schechter function, reducing the need for quenching in our model. This implies that surveys of bright objects, as will be possible with *WFIRST*, will be instrumental in understanding the complex astrophysics of massive, high-z galaxies.

Finally, detailed simulations of individual galaxies with state-of-the-art feedback prescriptions produce time-averaged star formation efficiencies qualitatively similar to our feedback models. Notably, Kimm & Cen (2014) simulated low-mass galaxies at $z \gtrsim 6$ using a detailed prescription for internal feedback. They found that (in our language) $\tilde{f}_\star \propto m^{1/2}$, within the range of feedback models we have prescribed. However, they also found that star formation was very bursty in low-mass systems, as the star formation 'overshot and the subsequent supernova suppressed star formation for a substantial period. Within our implementation, this is similar to our merger model, which we have shown also fits the data reasonably well,

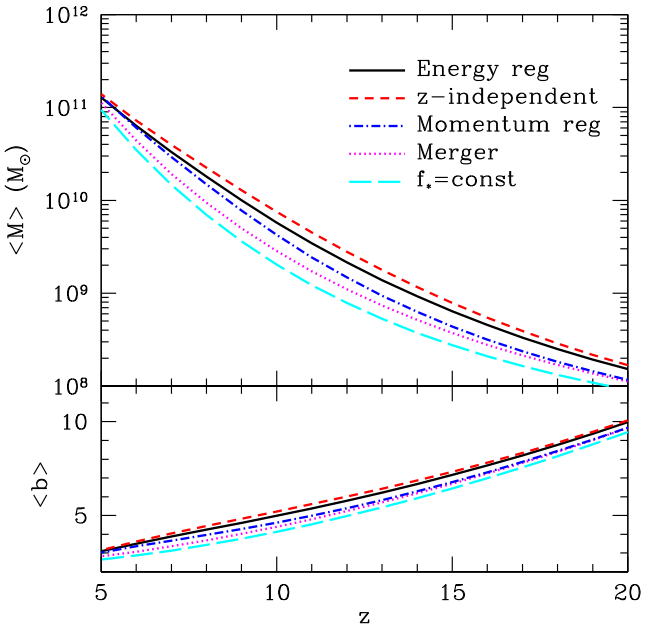

Figure 7. Luminosity-weighted average mass (upper panel) and bias (lower panel) of star-forming galaxies as a function of redshift. The solid, dotted, dashed and dash–dotted curves show our model predictions at $z = 15$ for our energy-driven feedback, momentum-driven feedback, merger-driven and redshift-independent models.

though of course it affects our parameter interpretation. More study is therefore needed in order to understand this star formation phase and the transition from bursty to smooth star formation.

5 GALAXIES AT $z > 6$

5.1 Average properties of star-forming haloes

Within our suite of models, it is useful to consider the aggregate properties of the galaxy population. The luminosity-weighted average halo masses in our fiducial models are shown in the upper panel of Fig. 7. The increasing star formation efficiency with mass (at least up to $m \approx 10^{11}\,\mathrm{M}_\odot$ increases the characteristic mass by a factor of several over a case in which $f_\star$ is independent of mass, a fairly modest increase considering the wide range of halo masses available for star-forming galaxies. This simply reflects the rapid decrease in halo number density with mass over most of the relevant parameter space.

The lower panel shows the luminosity-weighted average linear bias. In all of the models, this decreases from $\sim$10 to $\sim$3 over this redshift interval. In this picture of high-z galaxy formation, the sources contributing most of the luminosity density are always highly clustered. Again, the details of the feedback prescription make only a small difference to this bias.

Our different models yield similar predictions for the luminosity-weighted average halo mass (and hence bias) of the galaxy population: as we found with the star formation history, calibrating to the $z \lesssim 8$ luminosity functions leaves relatively little freedom in the extrapolation to higher redshifts. In this case, the average is fixed by the competition between the steeply falling halo mass function and the rising star formation efficiency (at small halo masses), so even the redshift-independent star formation efficiency model is only modestly different. Note, however, how rapidly $\langle M \rangle$ falls with redshift: it is $\lesssim 10^9\,\mathrm{M}_\odot$ by $z \sim 15$, which corresponds to *stellar* masses $\sim 3 \times 10^6\,\mathrm{M}_\odot$, illustrating the difficulty of detecting the majority of the light at extremely high redshifts.

5.2 The star formation history of individual $z > 6$ galaxies

Because our model tracks the growth of individual haloes, it allows us to follow the star formation histories of these haloes over time. Fig. 8 shows several examples. We follow haloes that begin forming stars at the atomic cooling threshold at a specified redshift and grow according to our abundance matching model to $z = 6$. The right-hand panel, which shows $\tilde{f}_\star$, includes a subset of these models. As haloes grow and their potential wells deepen, they become more resistant to feedback, and their star formation efficiencies increase by a factor of several. In our fiducial energy-regulated and momentum-regulated models, the increase is moderated somewhat by the larger binding energy (at fixed halo mass) at earlier times. (The flatness of the curves in the right-hand panel near their origin points is because we assume that haloes immediately reach their maximal star formation rate as soon as they pass the atomic cooling threshold, so this transient behaviour is an artefact of our simple model.) Note the wide disparity in initial star formation efficiencies within the model: the weaker feedback in the momentum-regulated model allows about twice as many stars to be formed in the lowest mass haloes, while the redshift-independent model falls far below the others at these very small masses.

Turning back to the left-hand panel in Fig. 8, we see that rapid halo growth (see Fig. 1) and the increasing star formation rates with halo mass combine to make the stellar populations grow extremely rapidly. All of our models are at least roughly exponentially growing, with the stellar mass-doubling time corresponds to just $\Delta z \sim 1$–2. This is far faster than the gentle increase in the star formation efficiency and is driven by the growth of haloes: early generations of galaxies have exponentially *growing* star formation histories, if accretion is smooth and steady. Even in the case of merger-driven models (or bursty feedback-driven star formation), the rapid merger rate suggests that exponentially increasing histories are a reasonable approximation, unless the individual episodes can be resolved through high-quality observations.

This extremely rapid evolution does introduce one caveat to our model: we have assumed that feedback-regulation imposes a quasi-steady state on star formation, but in reality that will be hard to achieve as the haloes evolve. Instead, the process of accretion on to the ISM and star formation inside clouds likely induces a delay into the response which will manifest itself as an offset from the expected $f_\star$. However, given the other uncertainties in our feedback prescriptions, we have chosen not to attempt to model this here.

6 IMPLICATIONS OF FEEDBACK-REGULATED STAR FORMATION

In this section, we place our picture in the context of some of the markers of and global events induced by the first generations of galaxies in the Universe.

6.1 The reionization history

The hallmark event of this era is the reionization of the IGM. In addition to our source model, a reionization calculation requires a prescription for the escape of ionizing photons into the IGM and a model for recombinations within the IGM. We also require assumptions about the stellar populations (including their metallicity,

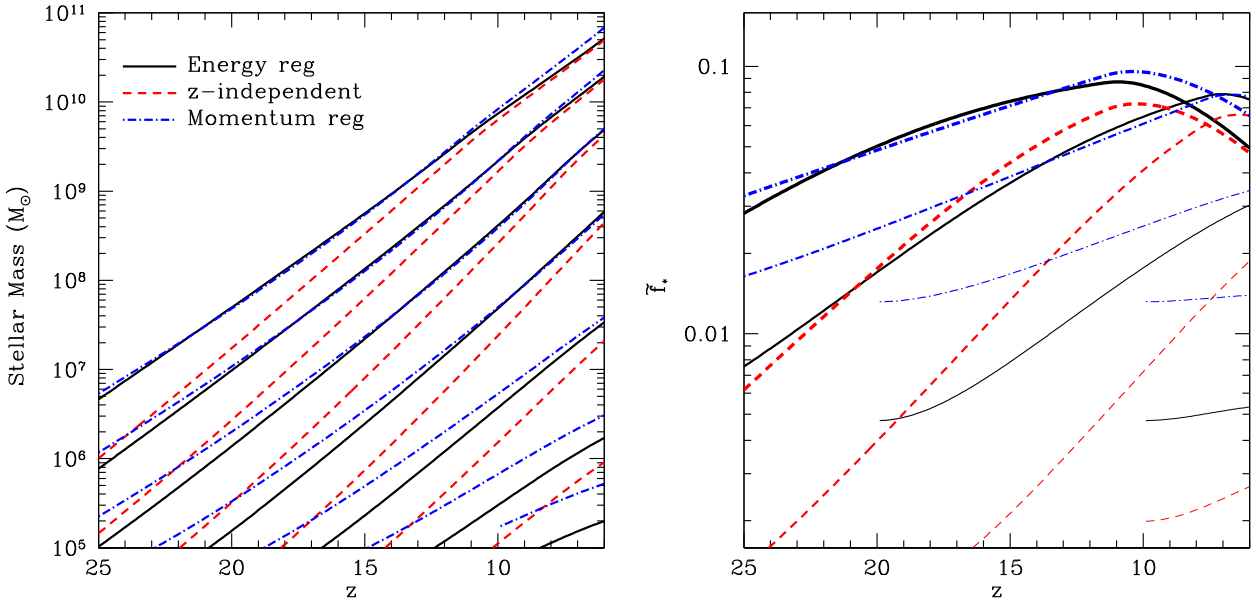

Figure 8. Star formation histories of individual haloes. *Left:* Stellar mass growth. The curves are initialized with masses at the atomic cooling threshold and allowed to grow according to our star formation prescriptions. We show curves beginning at $z = 40, 35, 30, 25, 20, 15$ and 10, which have final halo masses of $M = (664, 196, 41, 6.0, 0.71, 0.10, 0.024) \times 10^{10}\,\mathrm{M}_\odot$ at $z = 6$, respectively. The solid, dot–dashed and dashed curves show results for our energy-driven feedback, momentum-driven feedback and redshift-independent models, respectively. *Right:* Ratio of stellar mass to expected baryonic halo mass, $\tilde{f}_\star$, for a selection of haloes shown at left. For clarity, we only show haloes initialized at $z = 40, 30, 20$ and 10.

binarity and IMF; e.g. Robertson et al. 2013; Mirocha et al. 2017). We will write

$$\dot{n}_{\rm ion} = N_{\rm i}\,\langle f_\star\rangle\, f_{\rm coll}\bar{n}_{\rm H,0}, \tag{20}$$

where $\dot{n}_{\rm ion}$ is the (comoving) number density of ionizing photons produced per second,[8] $N_{\rm i}$ is the number of ionizing photons produced per baryon in stars, $\bar{n}_{\rm H,0}$ is the comoving number density of hydrogen and the average $\langle f_\star\rangle$ is over the halo mass function. We take $N_{\rm i} = 6000$ (see Section 3.3 and fig. A1 of Mirocha et al. 2017). Uncertainty in the IMF and metallicity lead to at least a factor of 2 uncertainty in this number, though with our procedure the effect on our models is much smaller because the star formation efficiency is also calibrated to the UV (Mirocha et al. 2017).

The second ingredient is usually parametrized by the escape fraction of UV photons, $f_{\rm esc}$. This appears to fluctuate rapidly within and between galaxies, and there is no direct observational constraint on the high-z galaxy population. In the spirit of our simple approach, we will assume that it is a constant (but see below).

For the IGM, we will follow the usual clumping factor prescription, assuming that the overall recombination rate is enhanced relative to that in a uniform-density IGM by a factor $C = \langle n_{\rm e}^2\rangle / \langle n_{\rm e}\rangle^2$, where $n_{\rm e}$ is the electron density and the average is taken over all ionized regions. We will assume $C = 3$ for simplicity, which is a reasonable match to simulations in the relevant redshift range (Pawlik, Schaye & van Scherpenzeel 2009). We also take a fixed IGM temperature of $T = 2 \times 10^4$ K, and we use the case-B recombination rate within the ionized gas.

In that case, we can compute the ionization history via

$$\frac{{\rm d}Q_{\rm i}}{{\rm d}t} = \frac{f_{\rm esc}\dot{n}_{\rm ion}}{\bar{n}_{\rm H,0}} - \frac{Q_{\rm i}}{t_{\rm rec}}, \tag{21}$$

[8] We also include a correction factor for helium, $A_{\rm He} = 4/(4 - 3Y_{\rm p}) = 1.22$, which we assume to be singly ionized along with hydrogen.

where Q_i is the mass fraction of ionized gas and $t_{\rm rec}$ is the recombination time.

An important constraint on the reionization history comes from the cumulative optical depth to CMB scattering:

$$\tau_{\rm e} = c\sigma_{\rm T}\bar{n}_{\rm H,0}\int_0^z f_{\rm e}(z')Q_{\rm i}(z')\frac{(1+z')^2}{H(z')}{\rm d}z', \tag{22}$$

where $\sigma_{\rm T}$ is the Thomson cross-section and $f_{\rm e}(z)$ is the number of free electrons per hydrogen atom. We assume that helium is doubly ionized at $z < 4$ and that the fraction of singly-ionized helium traces Q_i at higher redshifts. Most recently, the Planck Collaboration XLVII (2016) measured $\tau_{\rm e} = 0.055 \pm 0.009$.

We compute the reionization histories corresponding to our star formation models and show some results in Fig. 9. For a fair comparison, in each case we choose $f_{\rm esc}$ such that $\tau_{\rm e} = 0.055$, the best-fitting value from Planck's measurement. While the resulting $f_{\rm esc}$ values are illustrative of the *relative* production rate of ionizing photons in our models, their numerical values should not be taken too seriously, as they are degenerate with uncertainties in, for example, the stellar populations (through $N_{\rm i}$ in equation 20). For comparison, we also show the empirically motivated model history from Robertson et al. (2015).

An additional constraint on the ionization history comes from the Ly$\,\alpha$ forest, from which we can measure the overall ionizing emissivity at times just beyond reionization (Kuhlen & Faucher-Giguère 2012; Becker & Bolton 2013). Because the inferred ionizing emissivity is relatively small at the end of reionization, it can be hard to reconcile these measurements with reionization models based on the rapidly increasing galaxy luminosity function (e.g. Kuhlen & Faucher-Giguère 2012; Alvarez, Finlator & Trenti 2012; Sun & Furlanetto 2016) without introducing additional physics. We therefore also show in Fig. 9 a variant of our energy-regulated model in which $f_{\rm esc} = 0.067[(1+z)/7]$, which again matches the Planck optical depth. However, we also note that all of the models shown here are within the 1σ bounds on the ionizing emissivity

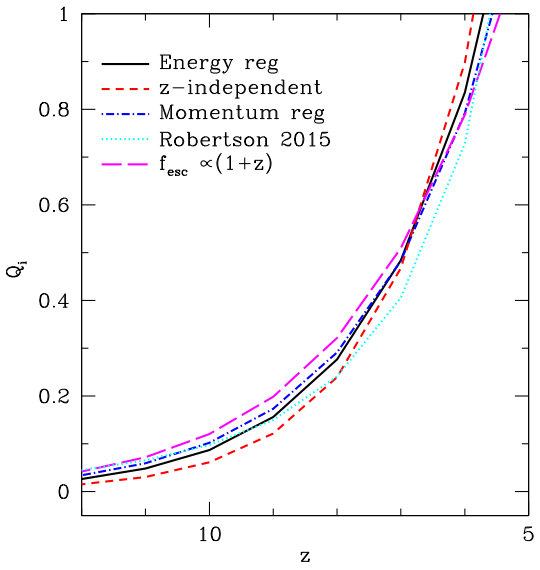

Figure 9. Reionization histories in our models of star formation, all normalized to have $\tau_e = 0.055$. The solid, dashed and dot–dashed curves are our energy-regulated, redshift-independent and momentum-regulated models, while the dotted curve shows the empirical fit from Robertson et al. (2015). We have varied the escape fraction in order to normalize τ_e, choosing $f_{esc} \approx 0.096, 0.084, 0.13$ and 0.09, respectively, for these curves. In addition, we show with the long-dashed line an energy-regulated model in which we assume that $f_{esc} \propto (1 + z)$.

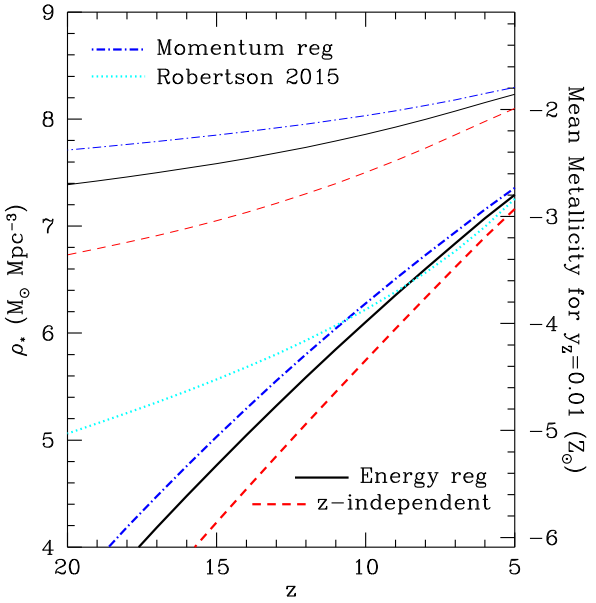

Figure 10. Stellar mass density in our models (*left axis*) and corresponding mean metallicities (*right axis*). The latter assumes yields typical of Population II supernovae, $y = 0.01$, with a standard Salpeter IMF; Population I stars would increase the overall metallicity by a factor $\lesssim$3, while top-heavy IMFs would increase it. The thick lines show the stellar mass density as well as the mean metallicity *of the Universe*. The thin curves show the mean metallicity of collapsed objects, *if they retain all their metals*. The solid, dashed and dot–dashed curves are our energy-regulated, redshift-independent and momentum-regulated models, while the dotted curve shows the empirical fit to the star formation history from Robertson et al. (2015).

at $z = 5$ from Becker & Bolton (2013). Additionally, the $z > 5$ ionizing background may be more complex than assumed in these measurements (D'Aloisio et al. 2016).

Given the overall similarities in the star formation histories of these models, it is not surprising to see that they imply very similar reionization histories as well. The redshift-independent model requires a higher escape fraction and has a steeper evolution: note that at $z = 12$, it has roughly half the ionized fraction of the Robertson model. We conclude that – as others have shown – the star formation history inferred from galaxy observations is consistent with current optical depth constraints so long as a modestly higher escape fraction is assumed than is measured in lower z samples (Robertson et al. 2013, 2015; Bouwens et al. 2015b; Sun & Furlanetto 2016). The differences amongst our models become large only at very high redshifts, when Q_i is so small that it will be very difficult to measure directly.

Finally, we note that the curve with a variable f_{esc} probably ends reionization too late ($z \sim 5.5$) in comparison to observations, because its higher ionized fraction early on contributes a non-negligible optical depth. In practice, such a prescription would only complete reionization by $z \sim 6$ if τ_e were larger than the best-fitting value from Planck.

6.2 Metal production and enrichment

A final, straightforward implication of the star formation history is a baseline chemical enrichment history. The time-dependent metal yield $y_Z(t)$, or the fraction of stellar mass returned to the ISM as metals, depends upon the IMF, stellar evolution parameters (metallicity, binarity, etc.), stellar winds and supernova properties. It is a function of the time delay since a stellar population formed, but after several million years $y_Z \sim 0.01$ in most cases. With it, one can compute the rate at which metals are produced inside our galaxies. Again, for a simple estimate we will ignore the time-dependent factor by including only metals from supernovae (the instantaneous recycling approximation). In this case, the mean metallicity of the Universe can be written

$$\langle Z(z) \rangle = \frac{y_Z \rho_\star(z)}{\bar{\rho}_b}, \tag{23}$$

where $\rho_\star$ is simply the integral of the star formation rate density shown in Fig. 6 and $\bar{\rho}_b$ is the mean baryon density.

Fig. 10 illustrates the evolution of the mean metallicity in our models. The thick curves show the total stellar mass density (left vertical axis). We compare these to the empirical fit to the star formation history from Robertson et al. (2015). Given our normalization to the luminosity function, it is not surprising that our results are also in agreement with stellar mass measurements in this era, though those still have large errors (e.g. Stark et al. 2013).

On the right axis, we then convert $\rho_\star(z)$ into the mean metallicity. Here we assume $y_Z = 0.01$. This is typical of the metal yield for Population II stars with a 'normal' IMF; higher metallicity stars will typically produce more metals overall ($y \approx 0.03$ for Population I stars) (Benson 2010), so these values are somewhat conservative. (Moreover, a top-heavy IMF would of course produce more supernova explosions and therefore more metals.) The thick curves show the overall metallicity of the Universe, averaged over all the baryons. Our models have $\langle Z \rangle \lesssim 10^{-3}\ Z_\odot$ at $z \sim 6$, which is typical of any model that reionizes the Universe at that time, assuming an escape fraction ~0.1. This is because the massive stars responsible for reionizing the IGM are (mostly) the same ones that explode as supernova and hence enrich the Universe (Oh, Cooray & Kamionkowski 2003), so (given present observational constraints

on reionization) the overall metallicity cannot differ dramatically from this value at $z \sim 6$.

This mean value is only part of the story, however, as the real distribution is highly inhomogeneous. The thin curves show the average metallicity of star-forming haloes, assuming that those haloes *retain all of their gas and metals*: this is therefore a *maximal* estimate of the metallicity in these regions. To obtain the thin curves, we have simply divided the thick curves by the fraction of baryons nominally inside of star-forming haloes, $f_{\rm coll}$.

Overall, we find that the mean metallicity of collapsed objects is $Z \sim 0.001$–$0.01\,{\rm Z}_\odot$, even if they retain all their metals. This is because most of the collapsed matter is inside very small haloes, where feedback strongly suppresses star formation. Of course the actual metallicity of any given halo depends on its star formation history (and hence mass in our model): applying the analogue of equation (23) to an individual galaxy, we would have $Z_{\rm gal} = 2(y_Z/0.01)\tilde{f}_\star\,{\rm Z}_\odot$. Thus, from a glance at Fig. 8, it is clear that *massive* haloes will have $Z \gtrsim 0.1\,{\rm Z}_\odot$ by $z \sim 10$.

A related, but much more difficult, question is how these metals are dispersed through the IGM. In the standard picture, galactic winds – driven by the same feedback mechanisms we have used to regulate star formation – eject some fraction of the metals from galaxies and advect them through the IGM. However, the efficacy of this process depends on the wind energetics and mechanism. Strictly following the assumptions of our energy-regulated model, the wind energy exactly balances the binding energy of the halo, and the wind material will only barely escape. Of course, these winds are likely complex phenomena with wide velocity and density distributions within the wind material, so some of it can escape. However, even a simple estimate of the extent of these winds requires additional assumptions.

We will therefore attempt to bracket the importance of metal enrichment with some very simple estimates, which we illustrate in Fig. 11. Each thick curve shows the fraction of the Universe's baryons that have been exposed to enriched material. To compute this fraction, we integrate the enriched volume $Q_{\rm e}$ over all star-forming galaxies, according to the following wind prescriptions. We then allow for overlap between the sources by plotting $(1 - \exp^{-Q_{\rm e}})$; in reality, overlap is more complicated because of source clustering (Furlanetto & Loeb 2005). For comparison, we also show the mass fraction of material that has been incorporated into star-forming haloes with the thin dotted curve.[9]

First, in our simplest model we suppose that the material is ejected from its source halo at that halo's escape velocity. If it suffers no further deceleration, the comoving distance that such material would reach is

$$r \approx 0.14\ {\rm Mpc}\left(\frac{7}{1+z}\right)^{1/2}\left(\frac{v_{\rm esc}}{40\ {\rm km\ s^{-1}}}\right). \quad (24)$$

We show the resulting filling fraction with the short dashed line in Fig. 11 (note that this prescription does not use any of the details of our star formation prescriptions). These kinds of slow winds near the escape velocity result in only very modest enrichment.

However, the energy available from feedback can lead to much faster expansion, if it has only a few times more energy than required in our model. The dot–dashed curve in Fig. 11 is the most extreme case we consider. Here, we assume that all the energy generated by feedback is used to drive a blastwave into the surrounding IGM and neglect the gravitational potential of the host halo, so that the wind follows the energy-conserving Sedov–Taylor–von Neumann solution. In this case, $r \propto (Et^2/\bar{\rho})$. For simplicity, we assume the wind propagates into a medium at the mean density of the Universe, $\bar{\rho}(z)$, in which the wind has been active for half the age of the Universe. In this case, the comoving wind radius in our energy-regulated model is (Voit 1996)

$$r \approx 0.25\ {\rm Mpc}\left(\frac{m_{\rm h}}{10^8\ {\rm M}_\odot}\right)^{1/3}, \quad (25)$$

Interestingly, in this case the enriched mass is independent of the source's redshift and proportional to the source's mass, with each halo enriching ∼25 times its own mass.[10]

This is certainly an overestimate, however, because the blastwave will undergo radiative (and other) losses as it propagates. Detailed wind models are beyond the scope of this paper (see e.g. Ostriker & McKee 1988; Tegmark, Silk & Evrard 1993; Furlanetto & Loeb 2003), but we provide two simple corrections. First, the blastwaves will inevitably lose energy to Compton cooling off the CMB, for which the cooling time is $t_{\rm Comp} \approx 1.2 \times 10^8[10/(1+z)]^4$ yr. We approximate this effect by limiting winds to expand for no more than $t_{\rm Comp}$ and show the result with the solid curve in Fig. 11. This sharply reduces the enriched volume at high redshifts, when Compton cooling is efficient. Finally, detailed wind models show that our simple treatment typically overestimates the maximum radius by a factor of ∼2 due to other

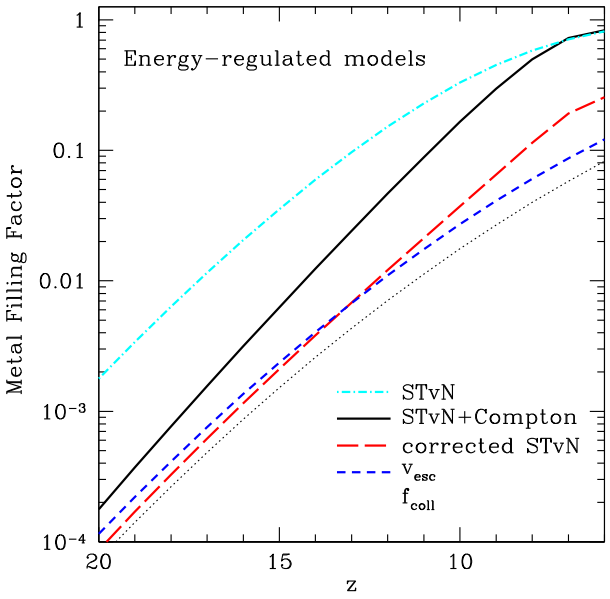

Figure 11. Simple estimates of metal enrichment in our star formation models. The thin dotted line shows the fraction of matter inside collapsed haloes according to the Trac et al. (2015) mass function. The short-dashed curve assumes that metals propagate at their source halo's $v_{\rm esc}$ for half the age of the Universe. The other curves use our energy-regulated feedback model and the Sedov–Taylor–von Neumann blastwave solution. The dot–dashed line shows the maximum expansion allowed over half the age of the Universe, assuming no radiative or other losses. The solid curve includes energy loss due to Compton cooling, and the long-dashed curve includes a further approximate correction to match the results of more detailed calculations at late times.

[9] Of course, the *volume* fraction of collapsed material is much smaller than this, because it is at an overdensity of ∼200.

[10] Note that a very similar estimate results from assuming that the blastwaves expand until they have accelerated all their material to the local Hubble flow velocity (Furlanetto & Loeb 2001).

losses (Furlanetto & Loeb 2003). The long-dashed curve in Fig. 11 shows this corrected filling factor. Note that the fraction does not always fall by a factor of 8 because we always assume that material inside star-forming haloes is enriched.

We see that the enrichment process is likely very inhomogeneous. Only at $z \lesssim 10$ can a substantial fraction of the volume be enriched, and in our 'best guess' models the fraction is still $\lesssim$20 per cent.

7 DISCUSSION

We have presented a simple framework for modelling the high-redshift galaxy population. The three ingredients are the dark matter halo abundance, their average accretion rates and a prescription for setting the star formation efficiency $f_\star$ through stellar feedback. We roughly calibrate the free parameters in the feedback scheme by comparison to the measured galaxy luminosity functions at $z = 6$–10, finding that a variety of feedback prescriptions fit the currently available data with reasonable parameter choices. In all cases, we find that $f_\star$ peaks at $m_\mathrm{h} \sim 10^{11}$–$10^{12}\,\mathrm{M}_\odot$, possibly declining towards larger masses and certainly declining rapidly at lower masses. This is consistent with empirical fits using similar schemes (e.g. Mason et al. 2015b; Mashian et al. 2016; Sun & Furlanetto 2016; Mirocha et al. 2017). Overall we find that haloes turn ∼1–10 per cent of their total baryonic mass into stars, with a strong dependence on halo mass.

With our model, we then extrapolate to higher redshifts and fainter galaxies with a clear understanding of the physical meaning of the extrapolation. Necessarily, our extrapolation implicitly assumes that the underlying mechanisms of galaxy and star formation do not evolve with redshift, which is certainly too simplistic. But it provides a baseline against which evidence for new physics can be evaluated. Interestingly, the data are already good enough to estimate the amount of star formation beyond $z \sim 10$, even allowing for variations in our model's parameters (provided again that the underlying physics does not change). Within our parametrization of feedback, the most significant question is whether the processes controlling star formation depend only on halo mass or if we allow explicit dependence on redshift. The $z < 10$ data are consistent with a redshift-independent solution, and in fact the 'natural' redshift dependence of feedback models (in which the star formation efficiency of haloes at a fixed mass increases with redshift, as the haloes are more tightly bound at early times) appears to overpredict modestly the number of $z \sim 9$–10 galaxies.

However, Oesch et al. (2016) recently discovered an extremely bright source at $z = 11.1$. The implied number density of similar objects is very large (albeit with significant errors for a single detection), and even the most optimistic of our models – which assumes strong redshift dependence within the feedback – struggles to reproduce this object, which requires very strong star formation in massive haloes. However, it is worth noting that this object tells us little about the dominant star formation mode, as it is such a massive galaxy that it is likely unaffected by stellar feedback (which we assume controls the star formation rate in the much more abundant smaller haloes).

We have shown (see Fig. 5) that future space missions are ideally suited to measure precisely the high-z luminosity function of galaxies over a factor of nearly 10^4 in luminosity, at least at $z \lesssim 15$. While deep surveys with *JWST* will uncover the bulk of the galaxy emission, *WFIRST* will be essential for exploring the growth of the most massive galaxies. However, in our model the galaxy luminosity function decreases rapidly enough that by $z \sim 15$, very few sources will be detected even in very deep surveys. Detections in this era and beyond will therefore indicate the presence of new stellar populations or formation mechanisms.

We have further explored some of the basic implications of our model. Within feedback-regulated models the star formation efficiency increases with halo mass, so most of the star formation occurs in haloes with virial temperatures well above the atomic cooling threshold, especially at $z \lesssim 10$. This implies that the smallest haloes are less important than previously assumed and that 'photoheating feedback' during reionization is less important as well. Even though the star formation processes evolve only slowly (if at all) with redshift, the characteristic halo mass still evolves quickly, because massive haloes are assembling so rapidly during this era. We have also found that our models produce reasonable reionization histories, consistent with available data, if we set $f_\mathrm{esc} \sim 0.1$, and that they likely enrich only a small fraction of space through galactic winds.

Our model is clearly far too simplistic to provide an accurate understanding of the earliest generations of galaxies. We ignore, for example, evolution in the IMF, chemistry and spatial distribution of star formation. However, we have used its simple physical principles to model the available data and shown how it can easily be extrapolated to higher redshifts while understanding the detailed physical implications of this extrapolation. We hope that this model will be useful in rapid explorations of the parameter space allowed to high-z galaxies, in forecasting future surveys and in qualitatively understanding the nature of these sources. In the future, we will explore further implications of our model for distant galaxies as well as improve its physical inputs.

ACKNOWLEDGEMENTS

SRF thanks A. Benson, A. Dekel and B. Robertson for helpful conversations. This work was supported by the National Science Foundation through award AST-1440343 and by National Aeronautics and Space Administration (NASA) through award NNX15AK80G. We also acknowledge a NASA contract supporting the 'WFIRST Extragalactic Potential Observations (EXPO) Science Investigation Team' (15-WFIRST15-0004), administered by GSFC. SRF was partially supported by a Simons Fellowship in Theoretical Physics and thanks the Observatories of the Carnegie Institute of Washington for hospitality while much of this work was completed. This research was also completed as part of the University of California Cosmic Dawn Initiative. We acknowledge support from the University of California Office of the President Multicampus Research Programs and Initiatives through award MR-15-328388.

REFERENCES

Alvarez M. A., Finlator K., Trenti M., 2012, ApJ, 759, L38
Atek H. et al., 2015a, ApJ, 800, 18
Atek H. et al., 2015b, ApJ, 814, 69
Becker G. D., Bolton J. S., 2013, MNRAS, 436, 1023
Behroozi P. S., Silk J., 2015, ApJ, 799, 32
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Benson A. J., 2010, Phys. Rep., 495, 33
Bouwens R. J. et al., 2014, ApJ, 793, 115
Bouwens R. J. et al., 2015a, ApJ, 803, 34
Bouwens R. J., Illingworth G. D., Oesch P. A., Caruana J., Holwerda B., Smit R., Wilkins S., 2015b, ApJ, 811, 140
Bouwens R. J. et al., 2016, ApJ, 830, 67
Bowler R. A. A., Dunlop J. S., McLure R. J., McLeod D. J., 2017, MNRAS, 466, 3612

Cai Z.-Y., Lapi A., Bressan A., De Zotti G., Negrello M., Danese L., 2014, ApJ, 785, 65
Capak P. L. et al., 2015, Nature, 522, 455
Coe D. et al., 2013, ApJ, 762, 32
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
D'Aloisio A., McQuinn M., Davies F. B., Furlanetto S. R., 2016, MNRAS, preprint (arXiv:1611.02711)
Dayal P., Ferrara A., Dunlop J. S., Pacucci F., 2014, MNRAS, 445, 2545
Dekel A., Mandelker N., 2014, MNRAS, 444, 2071
Dekel A., Zolotov A., Tweed D., Cacciato M., Ceverino D., Primack J. R., 2013, MNRAS, 435, 999
Dunlop J. S., McLure R. J., Robertson B. E., Ellis R. S., Stark D. P., Cirasuolo M., de Ravel L., 2012, MNRAS, 420, 901
Dunlop J. S. et al., 2013, MNRAS, 432, 3520
Eldridge J. J., Stanway E. R., 2009, MNRAS, 400, 1019
Faucher-Giguère C.-A., Kereš D., Ma C.-P., 2011, MNRAS, 417, 2982
Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778
Finkelstein S. L. et al., 2015, ApJ, 810, 71
Furlanetto S. R., Loeb A., 2001, ApJ, 556, 619
Furlanetto S. R., Loeb A., 2003, ApJ, 588, 18
Furlanetto S. R., Loeb A., 2005, ApJ, 634, 1
Genel S. et al., 2014, MNRAS, 445, 175
Gnedin N. Y., 2016, ApJ, 825, L17
Goerdt T., Ceverino D., Dekel A., Teyssier R., 2015, MNRAS, 454, 637
Ishigaki M., Kawamata R., Ouchi M., Oguri M., Shimasaku K., Ono Y., 2015, ApJ, 799, 12
Jaacks J., Choi J.-H., Nagamine K., Thompson R., Varghese S., 2012, MNRAS, 420, 1606
Kawamata R., Oguri M., Ishigaki M., Shimasaku K., Ouchi M., 2016, ApJ, 819, 114
Kimm T., Cen R., 2014, ApJ, 788, 121
Kuhlen M., Faucher-Giguère C.-A., 2012, MNRAS, 423, 862
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lidz A., Zahn O., Furlanetto S. R., McQuinn M., Hernquist L., Zaldarriaga M., 2009, ApJ, 690, 252
Lidz A., Furlanetto S. R., Oh S. P., Aguirre J., Chang T.-C., Doré O., Pritchard J. R., 2011, ApJ, 741, 70
Liu C., Mutch S. J., Angel P. W., Duffy A. R., Geil P. M., Poole G. B., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 235
Livermore R. C., Finkelstein S. L., Lotz J. M., 2017, ApJ, 835, 113
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Mashian N., Oesch P. A., Loeb A., 2016, MNRAS, 455, 2101
Mason C. A. et al., 2015a, ApJ, 805, 79
Mason C. A., Trenti M., Treu T., 2015b, ApJ, 813, 21
McBride J., Fakhouri O., Ma C.-P., 2009, MNRAS, 398, 1858
McLeod D. J., McLure R. J., Dunlop J. S., Robertson B. E., Ellis R. S., Targett T. A., 2015, MNRAS, 450, 3032
McLeod D. J., McLure R. J., Dunlop J. S., 2016, MNRAS, 459, 3812
McLure R. J. et al., 2013, MNRAS, 432, 2696
Mirocha J., Furlanetto S. R., Sun G., 2017, MNRAS, 464, 1365
Muñoz J. A., Furlanetto S. R., 2014, MNRAS, 438, 2483
Mutch S. J. et al., 2016a, MNRAS, 463, 3556
Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016b, MNRAS, 462, 250
Neistein E., Dekel A., 2008, MNRAS, 388, 1792
Neistein E., van den Bosch F. C., Dekel A., 2006, MNRAS, 372, 933
Noh Y., McQuinn M., 2014, MNRAS, 444, 503
Ocvirk P. et al., 2016, MNRAS, 463, 1462
Oesch P. A. et al., 2013, ApJ, 773, 75
Oesch P. A. et al., 2014, ApJ, 786, 108
Oesch P. A., Bouwens R. J., Illingworth G. D., Franx M., Ammons S. M., van Dokkum P. G., Trenti M., Labbé I., 2015, ApJ, 808, 104
Oesch P. A. et al., 2016, ApJ, 819, 129
Oh S. P., Cooray A., Kamionkowski M., 2003, MNRAS, 342, L20
Ostriker J. P., McKee C. F., 1988, Rev. Mod. Phys., 60, 1
Pawlik A. H., Schaye J., van Scherpenzeel E., 2009, MNRAS, 394, 1812
Planck Collaboration XIII, 2016, A&A, 594, A13
Planck Collaboration XLVII, 2016, A&A, 596, A108
Robertson B. E., 2010, ApJ, 713, 1266
Robertson B. E. et al., 2013, ApJ, 768, 71
Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, ApJ, 802, L19
Salvaterra R., Maio U., Ciardi B., Campisi M. A., 2013, MNRAS, 429, 2718
Schenker M. A. et al., 2013, ApJ, 768, 196
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
Smit R., Bouwens R. J., Franx M., Illingworth G. D., Labbé I., Oesch P. A., van Dokkum P. G., 2012, ApJ, 756, 14
Stark D. P., Schenker M. A., Ellis R., Robertson B., McLure R., Dunlop J., 2013, ApJ, 763, 129
Sun G., Furlanetto S. R., 2016, MNRAS, 460, 417
Tacchella S., Trenti M., Carollo C. M., 2013, ApJ, 768, L37
Tegmark M., Silk J., Evrard A., 1993, ApJ, 417, 54
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
Trac H., Cen R., Mansfield P., 2015, ApJ, 813, 54
Trenti M., Stiavelli M., Bouwens R. J., Oesch P., Shull J. M., Illingworth G. D., Bradley L. D., Carollo C. M., 2010, ApJ, 714, L202
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
van Dokkum P. G. et al., 2010, ApJ, 709, 1018
Visbal E., Haiman Z., Bryan G. L., 2015, MNRAS, 453, 4456
Voit G. M., 1996, ApJ, 465, 548
Waters D., Di Matteo T., Feng Y., Wilkins S. M., Croft R. A. C., 2016, MNRAS, 461, L51
Wellons S., Torrey P., 2017, MNRAS, 467, 3887
Yue B., Ferrara A., Xu Y., 2016, MNRAS, 463, 1968
Zheng W. et al., 2012, Nature, 489, 406

APPENDIX A: THE FORMATION OF NEW GALAXIES

Ongoing accretion on to existing galaxies is not the only channel through which stars can form at very high redshifts: in particular, if one imposes a mass threshold $m_{\rm min}$, the steepness of the mass function implies that haloes passing that threshold can carry a non-negligible fraction of the overall increase in galaxy mass density. To see this, note that in our picture the total rate at which the collapse fraction $f_{\rm coll}$ (or the fraction of mass inside collapsed objects) changes is the sum of two terms:

$$\bar{\rho}\frac{{\rm d}f_{\rm coll}}{{\rm d}t} = \int_{m_{\rm min}}^{\infty} {\rm d}m \left(\frac{\dot{m}}{\bar{\rho}}\right) n_{\rm h}(m,z) + [\dot{m}(mn)]_{m_{\rm min}}\,, \tag{A1}$$

where the first term on the right-hand side describes accretion on to haloes already above the threshold and the second term describes haloes that just cross the threshold.[11] This expression follows from taking the total time derivative of the integral and noting that the total number of haloes is conserved in our 'abundance-matching' picture, with only the mass of each individual halo evolving over time.

Fig. A1 shows the importance of these two terms: the solid line shows ${\rm d}f_{\rm coll}/{\rm d}z$, the dashed line – which is almost indistinguishable from the solid line – shows the sum of the two terms on the right-hand side of equation (A1),[12] and the dotted line shows the boundary term alone. This is a small fraction of the overall evolution at $z \sim 6$,

[11] For clarity of presentation, in the remainder of this section we will assume that $m_{\rm min}$ is constant with redshift, although in practice it varies slowly with time.

[12] Note that our abundance matching prescription is not guaranteed to reproduce the overall evolution, because it relates only integrated quantities. However, this comparison shows that it is nevertheless very close to self-consistent.

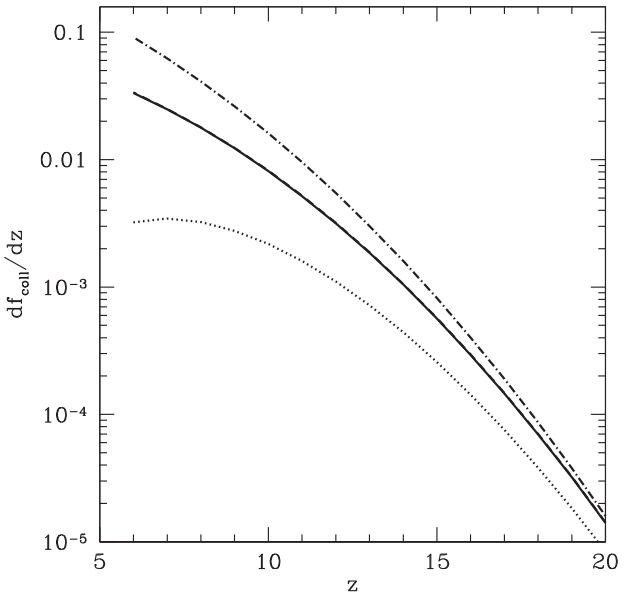

Figure A1. The importance of haloes crossing the minimum mass threshold. The solid line shows $df_{\rm coll}/dz$, the dashed line (nearly indistinguishable from the solid line) shows the sum of the two terms on the right-hand side of equation (A1) and the dotted line shows the boundary term alone. For reference, the dot–dashed line shows $f_{\rm coll}(z)$ in our model. All curves assume that $m_{\rm min} = 10^8\,{\rm M}_\odot$.

but by $z \sim 10$ it is non-negligible, and at higher redshifts it provides about half the new mass. For reference, the dot–dashed line shows $f_{\rm coll}(z)$ in our model, assuming $m_{\rm min} = 10^8\,{\rm M}_\odot$.

In the main text, we ignore the boundary term when computing star formation rates, because it is not clear that our approximations or physical picture make sense in these newly formed haloes. For example, many will have already undergone bursts of star formation seeded by Population III stars, so their initial stellar populations and gas contents will have already been disturbed. These objects are very unlikely to substantially affect our luminosity function predictions, because the haloes are so fragile to feedback. But they will affect the global star formation rate in some way, especially at $z \gtrsim 15$ where they carry roughly half of the overall increase in collapsed mass.

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