



Balance among gravitational instability, star formation and accretion determines the structure and evolution of disc galaxies

John C. Forbes,^{1★} Mark R. Krumholz,¹ Andreas Burkert^{2,3†} and Avishai Dekel⁴

¹*Department of Astronomy & Astrophysics, University of California, Santa Cruz, CA 95064, USA*

²*University Observatory Munich (USM), Scheinerstrasse 1, D-81679 Munich, Germany*

³*Max-Planck-Institut fuer extraterrestrische Physik, Giessenbachstrasse 1, D-85758 Garching, Germany*

⁴*Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel*

Accepted 2013 November 26. Received 2013 November 25; in original form 2013 May 13

ABSTRACT

Over the past 10 Gyr, star-forming galaxies have changed dramatically, from clumpy and gas rich, to rather quiescent stellar-dominated discs with specific star formation rates lower by factors of a few tens. We present a general theoretical model for how this transition occurs, and what physical processes drive it, making use of 1D axisymmetric thin disc simulations with an improved version of the Gravitational Instability-Dominated Galaxy Evolution Tool (GIDGET) code. We show that at every radius galaxies tend to be in a slowly evolving equilibrium state wherein new accretion is balanced by star formation, galactic winds and radial transport of gas through the disc by gravitational instability-driven torques. The gas surface density profile is determined by which of these terms are in balance at a given radius – direct accretion is balanced by star formation and galactic winds near galactic centres, and by transport at larger radii. We predict that galaxies undergo a smooth transition from a violent disc instability phase to secular evolution. This model provides a natural explanation for the high velocity dispersions and large clumps in $z \sim 2$ galaxies, the growth and subsequent quenching of bulges, and features of the neutral gas profiles of local spiral galaxies.

Key words: galaxies: evolution – galaxies: ISM – galaxies: kinematics and dynamics – galaxies: structure.

1 INTRODUCTION

Historically astronomers have studied the evolution of galaxies through changes in their stellar populations. The real action, though, takes place in the gas phase. However, it is only recently that observations in the radio have had sufficient sensitivity to detect molecular gas in emission at high redshift, and sufficient resolution to map both molecular and atomic gas in great detail for nearby galaxies. Integral field and grism spectroscopy of H α have also opened a new view on the spatial distribution of star formation (SF) and gas kinematics at $z \sim 1$ –2.

Numerous surveys have shown that the specific star formation rates (sSFRs, the star formation rate divided by the stellar mass) of Milky Way (MW) mass galaxies have decreased by roughly a factor of 20 since $z = 2$. With the wide acceptance of Λ cold dark matter (Λ CDM) cosmology, which entails the hierarchical growth of dark matter haloes, it became common lore that mergers were a major driver of this dramatic change in the nature of galaxies.

More recently though, the small scatter in the correlation between the stellar mass M_* and the star formation rate (SFR, the star-forming main sequence) for galaxies out to $z = 2$ has suggested that most stellar mass growth occurs in galaxies that are not undergoing dramatic merger events, but rather in typical-looking discs (e.g. Noeske et al. 2007; Rodighiero et al. 2011; Kaviraj et al. 2013). Maps of H α emission in main-sequence galaxies confirm that SF occurs in radially extended discs at $z \sim 1$ (Nelson et al. 2013).

Even though the higher SFRs at $z \sim 2$ are unlikely to be caused by mergers, galaxies where the sSFRs are so much higher than in local galaxies must be dramatically different. This has been verified directly by gas-phase observations, which show that these galaxies are gas rich (Tacconi et al. 2010, 2013), highly turbulent (Cresci et al. 2009; Förster Schreiber et al. 2009) and gravitationally unstable (Burkert et al. 2010; Genzel et al. 2011). These differences are also reflected in the optical morphologies, which are distinctly clumpy (Elmegreen, Elmegreen & Hirst 2004; Elmegreen et al. 2005).

High-resolution hydrodynamical simulations (Bournaud, Elmegreen & Martig 2009; Ceverino, Dekel & Bournaud 2010) have strongly suggested that the reason why these galaxies are so different from low-redshift discs is rapid gas accretion from the cosmic web through cold dense filaments (Dekel et al. 2009a),

★ E-mail: jforbes@ucolick.org

† Max Planck Fellow.

which in turn leads to galaxies with low values of the Toomre Q parameter (Toomre 1964),

$$Q_{\text{Toomre}} = \frac{\kappa \sigma_d}{\pi G \Sigma_d}. \quad (1)$$

Here $\kappa(r) = \sqrt{2(\beta(r) + 1)}\Omega(r)$ is the epicyclic frequency, which is roughly comparable to the angular frequency Ω , depending on the local power-law slope of the rotation curve, $\beta = d \ln v_\phi / d \ln r$. The velocity dispersion and surface density of the disc material are σ_d and Σ_d , respectively. This instability has dramatic effects on the dynamics of the disc (Dekel, Sari & Ceverino 2009b). In regions where $Q_{\text{Toomre}} \lesssim 1$, the disc is unstable to axisymmetric perturbations on a scale $\lambda \sim \sigma_d^2 / G \Sigma_d$, leading to clumps of this characteristic size. The clumpiness of the disc will in turn drive turbulence through the random torques exerted by the inhomogeneous gravitational field on material in the disc. The ultimate source of this kinetic energy is the gravitational potential of the galaxy, so mass must flow inwards (e.g. Gammie 2001; Dekel et al. 2009b) (though some will flow outwards to conserve angular momentum). As a result of this, the turbulent velocity dispersion σ_d , and hence Q_{Toomre} , is increased, so given a sufficient gas supply, the value of Q_{Toomre} will be self-regulated to a marginally stable value of order unity.

Alternative scenarios for driving the turbulence and producing clumps have been explored by other authors. Genel, Dekel & Cacciato (2012) constructed a simple model for the scenario in which the turbulence is driven by the kinetic energy of material as it accretes on to the disc (see also Elmegreen & Burkert 2010). The details of the origin of the clumpy morphologies have also come under recent theoretical and observational investigation, and the importance of ex situ clumps from minor mergers is not negligible (Mandelker et al. 2013). Supernovae (SNe; Joung, Mac Low & Bryan 2009), radiation pressure (Krumholz & Thompson 2012, 2013) and the two working in tandem (Agertz et al. 2013), being the primary sources of energy outside of gravitational potential energy, have also been studied as drivers of turbulence and outflows.

Undoubtedly all of these processes occur. All of the sources of stellar feedback suffer from a great deal of uncertainty in the degree to which they couple with the interstellar medium (ISM), and typically require extremely high-resolution hydrodynamical simulations to model properly. The highest resolution simulations to date, those of Krumholz & Thompson (2012, 2013) for radiation pressure and those of Joung et al. (2009) for SNe, suggest that these sources of turbulence are unable to produce the high velocity dispersions observed in $z \sim 2$ discs. The gravitational instability (GI) scenario has the advantage that it is difficult to avoid; if $Q_{\text{Toomre}} \lesssim 1$, gas will collapse and drive turbulence. In fact, high-resolution hydrodynamic simulations identify gravitational instability as the primary regulator of the disc and ISM (starting with Bournaud et al. 2010). Simple analytic arguments also suggest that the GI scenario leads to the correct behaviour of σ/v_{circ} over time, whereas the direct kinetic energy injection scenario does not (Genel et al. 2012). Moreover, even $z = 0$ disc galaxies have values of Q_{Toomre} (when corrected for multiple components and finite disc thickness) of order unity (Romeo & Wiegert 2011).

In this work, we build on the physical picture presented in simple toy models (Dekel et al. 2009b; Cacciato, Dekel & Genel 2012) of the GI and how it evolves over time. Krumholz & Burkert (2010) developed a formalism to show how gravitationally unstable discs behave as a function of radius in a steady state and how quickly the discs approach the steady state. In Forbes, Krumholz & Burkert (2012, hereafter F12), we extended the time-dependent numerical

model of Krumholz & Burkert (2010) to include SF, stellar migration and metallicity evolution to give a realistic picture for how galaxies evolve over cosmological times with all these processes. In this work, rather than focusing on the stellar populations, we explore what sets the gas distribution. Our model includes a number of improvements over the models presented in F12 which we discuss in detail in Appendix A, and a new stellar migration formalism (Appendix B).

One of our goals here is to understand the connection between the high-redshift star-forming galaxies and their $z = 0$ descendants. The two galaxy populations are vastly different in terms of their gas fractions and sSFRs, yet remarkably similar in morphology. Recent $z = 0$ measurements of the structure of gas in nearby spirals, The H I Nearby Galaxy Survey (Walter et al. 2008) and the HERA CO-Line Extragalactic Survey (Leroy et al. 2009), have provided unprecedented high spatial resolution data. These data have been fundamental in our understanding of SF, and Bigiel & Blitz (2012) recently showed that these galaxies exhibit a universal gas surface density profile with remarkably small scatter.

The general problem of how to connect high-redshift galaxy populations to their low-redshift counterparts has been approached for the past few decades with semi-analytical models (SAMs). These models are generally built on top of dark matter merger trees constructed from N -body cosmological simulations. Each galaxy is typically treated as a simple system described by a few quantities, e.g. cold and hot gas mass, stellar mass, black hole mass and the entire population evolves according to parametrized recipes for gas cooling, SF, stellar feedback, black hole growth, mergers, etc. With a few exceptions [van den Bosch & Swaters (2001) with subsequent work by Dutton et al. (2007), Dutton & van den Bosch (2009) and the simpler Fu et al. (2010)], SAMs have not tracked quantities as a function of radius (or more accurately specific angular momentum). The only model where matter can change its specific angular momentum (Fu et al. 2013) does so in an ad hoc way with no physical justification. This work attempts to fill this void without resorting to extremely expensive 3D hydrodynamical simulations, which must necessarily be either of very low resolution to see a large number of galaxies (e.g. Davé, Oppenheimer & Finlator 2011) or one galaxy at a time (e.g. Guedes et al. 2011).

In Section 2, we review the equations solved by our 1D code. The results of a wide range of simulations done with the code are presented in Section 3. We discuss the implications in Section 4 and summarize in Section 5.

2 THE GIDGET CODE

Our one-dimensional disc galaxy evolution code, Gravitational Instability-Dominated Galaxy Evolution Tool (GIDGET),¹ is described in more detail in F12. The code tracks the surface density, velocity dispersion and metallicity of one gas component and one or more stellar components, as a function of radius and time. The following subsections will describe the evolution equations for these quantities in some detail; we include a comprehensive list of all parameters used in this study, defined below, in Table 1. The most important physical ingredients are SF, external accretion on to the disc and radial transport of gas through the disc.

¹ The source code repository is freely available from <http://www.johnforbes.com/gidget.html>

Table 1. An exhaustive list of all parameters used in this study.

Parameter	Fiducial value	Plausible range	Description
<i>Gas migration</i> (Section 2.1)			
η	1.5	0.5–4.5	(3/2) kinetic energy dissipation rate per scaleheight crossing time
Q_{GI}	2	1–3	Marginally stable value of Q
T_{gas}	7000 K	3000–10 ⁴	Gas temperature; sets the minimum gas velocity dispersion
α_{MRI}	0.01	0–0.1	Value of $T_{\text{r}\phi}/\rho\sigma_{\text{sf}}^2$ without GI
<i>Rotation curve</i> (Section 2.2)			
v_{circ}	220 km s ^{−1}	180–250	Circular velocity in flat part of the rotation curve
r_{b}	3 kpc	0–10 kpc	Radius where the rotation curve transitions from power law to flat
β_0	0.5	0–1	Power-law slope of $v_{\phi}(r)$ at small radii
n	2	1–5	Sharpness of the transition in the rotation curve
<i>Star formation</i> (Section 2.3)			
ϵ_{ff}	0.01	0.003–0.03	SF efficiency per freefall time in the Toomre regime
$f_{\text{H}_2, \text{min}}$	0.03	0.01–0.1	Minimum f_{H_2}
t_{SC}	2 Gyr	1–3 Gyr	Depletion time of H ₂ in the single-cloud regime
f_{R}	0.54	0.4 +	Mass fraction of a zero-age stellar population not recycled to the ISM
μ	0.5	0–2	Galactic winds’ mass loading factor
<i>Metallicity</i> (Section 2.4)			
y	0.054	0.05–0.07	Mass of metals yielded per mass locked in stellar remnants
ξ	0	0–1	Metallicity enhancement of galactic winds
Z_{IGM}	$0.1 Z_{\odot} = 0.002$	$(0.01–1) Z_{\odot}$	Metallicity of initial and infalling baryons
kZ	1	0.3–3	Amplitude of metallicity diffusion relative to Yang & Krumholz (2012)
<i>Stellar migration</i> (Appendix B)			
Q_{lim}	2.5	2–3	Value of Q_* below which spiral instabilities will heat the stars
T_{mig}	4	2–5	Number of local orbital times over which stars are heated by spiral instabilities
<i>Accretion</i> (Section 2.5)			
$M_{\text{h},0}$	$10^{12} M_{\odot}$	–	Halo mass at $z = 0$
$\Delta\omega$	0.5	0.1–1	Interval of $\omega \sim z$ over which the accretion rate is constant
$r_{\text{acc}}(z = 0)$	6.9 kpc	3–20 kpc	Scalelength of new infalling gas
β_z	0.38	0–1	Scaling of efficiency with $(1 + z)$
$\beta_{M_{\text{h}}}$	−0.25	−1 to 0	Scaling of efficiency with halo mass
ϵ_0	0.31	~0 to 0.5	Efficiency at $M_{\text{h}} = 10^{12} M_{\odot}$, $z = 0$
ϵ_{max}	1	0.5–1	Maximum value of efficiency
<i>Initial conditions</i> (Section 2.6)			
α_{r}	1/3	0–1	Scaling of accretion scalelength with halo mass
$f_{\text{g},0}$	0.5	0.2–0.7	Initial gas fraction
f_{cool}	1	0.4–1	Fraction of $f_{\text{b}} M_{\text{h}}(z = z_{\text{relax}})$ contained in the initial disc
z_{relax}	2.5	2–3	z at which the simulation is initialized
ϕ_0	1	1–5	Initial ratio of stellar to gaseous velocity dispersion
<i>Computational domain</i> (see F12)			
x_0	0.004	$0 < x_0 \ll 1$	Inner edge of domain as a function of R
R	40 kpc	10–100 kpc	Outer edge of domain
n_x	200	$\gtrsim 100$	Number of radial cells
tol	10^{-4}	$10^{-5} - x_0$	Fastest change allowed in state variables, per orbital time at $r = R$
<i>Cosmology</i> ^a (Section 2.5)			
Ω_{m}	0.258	–	Average present-day matter energy density as a function of the critical density
$1 - \Omega_{\text{m}} - \Omega_{\Lambda}$	0	–	Deviation from a flat universe
f_{b}	0.17	–	Universal baryon fraction
H_0	$72 \text{ km s}^{-1} \text{ Mpc}^{-1}$	–	Hubble’s constant
σ_8	0.796	–	Normalization of the dark matter power spectrum

^aWe are restricted to using the *WMAP5* cosmology because the stochastic accretion histories use fits to N -body simulations with those cosmological parameters.

2.1 Gas transport and cooling

GIDGET solves the full equations of hydrodynamics in the limit of a thin, axisymmetric, rotationally supported disc, supported vertically by supersonic turbulent pressure. In this limit, the state of the gas at a particular time is described by a surface density $\Sigma(r)$ and a velocity dispersion $\sigma(r) = \sqrt{\sigma_{\text{turb}}^2(r) + \sigma_{\text{sf}}^2}$ with a turbulent and a component supported purely by stellar feedback.

The change in gas surface density at a given radius is described by a simple continuity equation accounting for mass flow through the disc, with source terms for SF, recycling of gas by stellar mass-loss, galactic winds and cosmological accretion,

$$\frac{\partial \Sigma}{\partial t} = \frac{1}{2\pi r} \frac{\partial}{\partial r} \dot{M} - (f_{\text{R}} + \mu) \dot{\Sigma}_{\text{SF}} + \dot{\Sigma}_{\text{cos}}. \quad (2)$$

The first term represents the flow of mass within the disc, where \dot{M} is defined as the net gas mass per unit time moving towards the centre of the disc across cylindrical radius r . Typically $\dot{M} > 0$, representing inward mass flux, but negative values at large radii in the disc are generally necessary to conserve angular momentum. The second term of the continuity equation represents gas-forming stars. Only a fraction f_R of that gas will remain in stellar remnants, while the remainder will be recycled to the ISM; we approximate this process as instantaneous as suggested by Tinsley (1980). Mass is also ejected at each radius in galactic-scale winds in proportion to the SFR, with mass loading factor μ . Finally, $\dot{\Sigma}_{\text{cos}}$ represents the rate of cosmological accretion on to the disc. The winds are assumed to escape the galaxy, though in principle they could be re-accreted later through this final term.

To evolve the velocity dispersion of the gas, we employ the energy equation added to the dot product of \mathbf{v} with the momentum equation, yielding a total (kinetic + internal) energy equation,

$$\frac{\partial \sigma}{\partial t} = \frac{\mathcal{G} - \mathcal{L}}{3\sigma\Sigma} + \frac{\sigma}{6\pi r\Sigma} \frac{\partial}{\partial r} \dot{M} + \frac{5(\partial\sigma/\partial r)}{6\pi r\Sigma} \dot{M} + \frac{(\beta - 1)v_\phi}{6\pi r^3\Sigma\sigma} \mathcal{T}. \quad (3)$$

Radiative gains and losses per unit area, respectively, \mathcal{G} and \mathcal{L} , are encompassed in the first term. The second and third terms account for the advection of kinetic energy as the gas moves through the disc. The torques which move gas radially in the disc, included in the final term, transfer energy between the galactic potential and the turbulent velocity dispersion. Here $\mathcal{T} = \int 2\pi r^2 T_{r\phi} dz$ is the vertically integrated effective viscous torque. Note that physically $T_{r\phi} \leq 0$, and for rotation curves flatter than solid body $\beta < 1$, so this final term adds kinetic energy to the gas. We also note that we do not explicitly include any terms related to energy input by cosmological accretion, as this is expected to be subdominant, certainly below $z = 2$ (Genel, Dekel & Cacciato 2012; Gabor & Bournaud 2013; Hopkins, Keres & Murray 2013).

The viscous torque is related to the mass flux via the conservation of angular momentum, as derived from the ϕ -component of the Navier–Stokes equation:

$$\dot{M} \equiv -2\pi r\Sigma v_r = -\frac{1}{v_\phi(1 + \beta)} \frac{\partial \mathcal{T}}{\partial r}. \quad (4)$$

The mass flux, the gas velocity in the radial direction and the torque are not known ab initio. To calculate them modellers have historically, since Shakura & Sunyaev (1973), appealed to an order-of-magnitude argument, namely that $T_{r\phi} = \alpha\rho\sigma^2$, or equivalently $\nu = \alpha\sigma H$, where α is a parameter that might be measured from hydrodynamical simulations, ν is the resultant effective turbulent viscosity and H is the scaleheight. Physical causes for the turbulence include the magnetorotational instability (MRI), the GI and misalignment of the angular momentum of accreting material. The value of α measured from simulations of the MRI varies by orders of magnitude, but is generally less than 0.1, particularly if the magnetic field is not forced to be vertical (Balbus & Hawley 1998). To distinguish between GI, which we model in a more consistent way, and the MRI or any other source of turbulence, which we include for comparison, we split our variables related to the torque into two components, $\mathcal{T} = \mathcal{T}_{\text{GI}} + \mathcal{T}_{\text{MRI}}$, and similarly for v_r , $T_{r\phi}$, \dot{M} and α (the effects can just be added together since all of our equations are linear in these quantities). We neglect any mismatch in angular momentum between the disc and the infalling material, both for simplicity, and since for likely sources of accretion (cold streams, cooling from the hot halo and re-accreted galactic winds), the mismatch is unlikely to be large.

Rather than picking a constant value of α_{GI} , we calculate at every time step the value of $\mathcal{T}_{\text{GI}}(r)$ such that in regions where $Q \leq Q_{\text{GI}}$, the torques will act to move and heat the gas so that $dQ/dt = 0$. In regions of the disc where $Q > Q_{\text{GI}}$, $\mathcal{T}_{\text{GI}} = \dot{M}_{\text{GI}} = v_{r,\text{GI}} = 0$. This also serves as both the inner and outer boundary conditions, i.e. we assume that the disc is gravitationally stable outside the computational domain, which leaves gas free to flow off either boundary if the innermost or outermost cell has non-zero torque.

To see how this works, consider the rate of change of Q with time,

$$\begin{aligned} \frac{dQ}{dt} &= \frac{\partial \Sigma}{\partial t} \frac{\partial Q}{\partial \Sigma} + \frac{\partial \sigma}{\partial t} \frac{\partial Q}{\partial \sigma} \\ &\quad + \frac{\partial \Sigma_*}{\partial t} \frac{\partial Q}{\partial \Sigma_*} + \frac{\partial \sigma_{rr}}{\partial t} \frac{\partial Q}{\partial \sigma_{rr}} + \frac{\partial \sigma_{zz}}{\partial t} \frac{\partial Q}{\partial \sigma_{zz}} \\ &= f_{\text{transport}} \left(\Sigma, \sigma, \Sigma_*, \sigma_{rr}, \sigma_{zz}, \mathcal{T}_{\text{GI}}, \frac{\partial \mathcal{T}_{\text{GI}}}{\partial r}, \frac{\partial^2 \mathcal{T}_{\text{GI}}}{\partial r^2} \right) \\ &\quad + f_{\text{source}}(\Sigma, \sigma, \Sigma_*, \sigma_{rr}, \sigma_{zz}, Z). \end{aligned} \quad (5)$$

The first equation is simply an application of the chain rule, while the second is just a definition, wherein we split all the terms into those which depend on \mathcal{T}_{GI} and those which do not. Note that the source term includes the terms related to the α_{MRI} viscosity, SF and radiative cooling. The function $f_{\text{transport}}$ has the nice property that it is linear in \mathcal{T}_{GI} and its spatial derivatives, so when $\mathcal{T}_{\text{GI}} = 0$, $f_{\text{transport}} = 0$ and we are left with $dQ/dt = f_{\text{source}}$. Meanwhile in regions where $Q < Q_{\text{GI}}$ (by some small amount), we solve the equation $f_{\text{transport}} = -f_{\text{source}}$, i.e. we force $dQ/dt = 0$. Because $f_{\text{transport}}$ is linear, this equation may be solved efficiently for \mathcal{T}_{GI} by the inversion of a tridiagonal matrix.

This treatment raises a key question. If $dQ/dt = 0$, how can the disc ever stabilize? In the course of solving $f_{\text{transport}} = -f_{\text{source}}$, sometimes a non-physical value of \mathcal{T} will be obtained. In particular, since viscous heating $\propto -\mathcal{T}$ for reasonable rotation curves $\beta < 1$, it must be the case that $\mathcal{T} \leq 0$ to satisfy the second law of thermodynamics (turbulence should not decay into large-scale coherent motions). If this condition is not satisfied by the solution of $f_{\text{transport}} = -f_{\text{source}}$, then we set $\mathcal{T} = 0$ in that cell. Under this circumstance the cell behaves exactly as if it has stabilized, and Q in that cell will obey $dQ/dt = f_{\text{source}}$. Typically, the reason why a cell falls into this situation is that $f_{\text{source}} > 0$ and no physical value of $f_{\text{transport}}$ can cancel this effect, so Q is allowed to rise in that cell.

The gravitational stability of discs to linear axisymmetric perturbations is roughly determined by the value of Q_{Toomre} . Modern versions of this parameter take into account both gas and stars (e.g. Rafikov 2001; Romeo & Falstad 2013), the finite thickness of the disc (Shu 1968; Romeo 1992, 1994; Elmegreen 2011), gas turbulence (Hoffmann & Romeo 2012) and the fact that gas which can cool to arbitrarily small scales is never formally stable (Elmegreen 2011). Romeo & Wiegert (2011) have developed an approximate, but analytic, formula for Q taking into account two components of finite thickness, i.e. both gas and stars. In this way the gravitational effects of the stars are included in the instability, subject to what we assume about how the stars self-regulate Q_* (see Appendix B). To account for the final complication, we demarcate the stable from the unstable values of Q at $Q_{\text{GI}} = 2$, rather than the canonical value of unity, as suggested by Elmegreen (2011). This approximation to Q and its partial derivatives with respect to Σ , σ , Σ_* and σ_* is extremely cheap to compute, which is advantageous since all of these values must be computed at each (unstable) radius and time to solve $f_{\text{transport}} = -f_{\text{source}}$.

Numerical experiments (Mac Low et al. 1998; Stone, Ostriker & Gammie 1998) of turbulent gas in periodic boxes have shown that the turbulence decays in roughly a crossing time of the turbulent driving scale. For the purposes of our simulations, we assume that the driving scale is the scaleheight of the disc, in which case the kinetic energy surface density $(3/2)\Sigma\sigma_{\text{turb}}^2$ will decay at a rate,

$$\mathcal{L} = \eta \Sigma \sigma^2 \kappa Q_g^{-1} \left(1 + \frac{\sigma \Sigma_*}{\sigma_{zz} \Sigma}\right) \left(1 - \frac{\sigma_{\text{sf}}^2}{\sigma^2}\right)^{3/2}, \quad (6)$$

where η is a free parameter which would be 3/2 if the decay time were exactly one scaleheight crossing time. As a result of the final factor, $\mathcal{L} \rightarrow 0$ as $\sigma \rightarrow \sigma_{\text{sf}}$, i.e. when the gas reaches the velocity dispersion induced by various forms of SF feedback, it will no longer lose any net energy. The value of σ_{sf} is set to agree with the gas kinetic temperature in the warm neutral medium of the MW (7000 K), which is in the same range as the maximum velocity dispersion achievable by SN feedback in simulations (Joung et al. 2009) and the velocity dispersion caused by far-ultraviolet (FUV) heating in MW-like galaxies (Ostriker, McKee & Leroy 2010), and is consistent with both neutral and molecular gas velocity dispersions in local disc galaxies (Caldú-Primo et al. 2013).

2.2 Rotation curve

In order to derive the evolution equations shown in the previous section, we assumed that the potential and rotation curve of the disc are constant in time. The primary reason for this is that to self-consistently calculate v_ϕ would require knowledge of the dark matter. While N -body simulations assuming Λ CDM cosmology consistently produce dark matter haloes with well-characterized density profiles, the effects of baryons are highly controversial. Moreover, if one were to calculate the rotation curve simply from the dark matter (e.g. Cacciato et al. 2012), the circular velocity would decrease with time since $z = 1$ (at $z = 2$, $v_{\text{circ}} \approx 185 \text{ km s}^{-1}$, increasing to $\approx 200 \text{ km s}^{-1}$ at $z = 1$, and falling back to $\approx 190 \text{ km s}^{-1}$ at $z = 0$), whereas observations (Kassin et al. 2012) show that (at fixed stellar mass) the circular velocity actually increases from $z = 1$ to the present. Therefore, rather than constructing a model for the rotation curve which depends on the poorly constrained interactions between baryons and dark matter, we adopt a simple functional form

$$v_\phi(r) = v_{\text{circ}} \left(1 + (r_b/r)^{|\beta_0 n|}\right)^{-\text{sign}(\beta_0)/n}. \quad (7)$$

This is designed to represent a smooth transition from power law to flat, where r_b is the characteristic radius where the rotation curve turns over. Within this radius, the velocity approaches a power law with index β_0 , and the sharpness of the transition between power law and flat increases with increasing n . The disadvantages of this approach are that we are restricted to evolving our galaxies over periods during which the circular velocity does not change very much ($z \sim 2-0$), and changes to the potential owing to the movement of baryons are not reflected in the rotation curve.

2.3 Star formation

Stars form with a constant efficiency per freefall time ϵ_{ff} from molecular gas, so that $\dot{\Sigma}_*^{\text{SF}} \sim \epsilon_{\text{ff}} f_{\text{H}_2} \Sigma / t_{\text{ff}}$, where t_{ff} is the freefall time and f_{H_2} is the molecular fraction (Krumholz & Tan 2007; Krumholz, Dekel & McKee 2012). Following Krumholz et al. (2012), we posit that there are two regimes: one in which the appropriate time-scale is the freefall time of gas distributed over

the full scaleheight H of the disc, namely $t_{\text{ff}} = \sqrt{3\pi/32 G \rho} \approx \sqrt{3\pi H/32 G \Sigma}$, which we call the ‘Toomre regime’ and one in which the time-scale is determined by the freefall time of individual molecular clouds, which observations suggest is $t_{\text{ff}}/\epsilon_{\text{ff}} = \Sigma_{\text{H}_2}/\dot{\Sigma}_*^{\text{SF}} \equiv t_{\text{SC}} \approx 2 \text{ Gyr}$ (Bigiel et al. 2011), the ‘single-cloud regime’. Then the SFR is simply set by which of these two time-scales is shorter,²

$$\dot{\Sigma}_*^{\text{SF}} = \max \left(\epsilon_{\text{ff}} f_{\text{H}_2} \Sigma \kappa \frac{\sqrt{32/3}}{Q_g \pi} \left(1 + \frac{\sigma \Sigma_*}{\sigma_{zz} \Sigma}\right)^{1/2}, f_{\text{H}_2} \frac{\Sigma}{t_{\text{SC}}} \right). \quad (8)$$

Typically, the first regime is relevant at small radii since $\kappa \propto 1/r$, and the transition tends to be fairly constant in time, since the rotation curve is fixed in our model, and both terms are proportional to f_{H_2} and Σ , though Q_g can change by an order of magnitude or more if the disc has stabilized.

The molecular fraction f_{H_2} is calculated according to the analytic formula of Krumholz, McKee & Tumlinson (2009). Their formula predicts f_{H_2} as a function of Σ and Z . Roughly speaking, $f_{\text{H}_2} \rightarrow 1$ at high surface densities, and below some transition surface density, there is a sharp cutoff where f_{H_2} rapidly approaches zero. This transition is metallicity dependent, roughly $5 M_\odot \text{ pc}^{-2} (Z/Z_\odot)^{-1}$. We include the slight modifications to this formula we used in F12, namely a floor of $f_{\text{H}_2} \geq 0.03$ to account for the fact that SF is observed even at very low surface densities (Bigiel et al. 2010; Schrubba et al. 2011), likely as a result of the requirement that the FUV flux not fall below a certain floor in order for two-phase equilibrium in the atomic ISM to be possible (Ostriker et al. 2010).

At each time step, a new population of stars is formed with surface density $\dot{\Sigma}_*^{\text{SF}} dt$, where dt is the duration of the time step. The velocity dispersion of this population is the maximum of σ_{turb} and $\sigma_{*,\text{min}}$. Physically this floor might correspond to some combination of cloud-to-cloud velocity dispersion or the internal velocity dispersion of a cloud, roughly 2 km s^{-1} . The newly formed stars are then merged with the extant population while conserving mass and kinetic energy, meaning $\dot{\Sigma}_*^{\text{SF}} dt$ is added to Σ_* , and the velocity dispersion of the extant population is updated so that $(\Sigma_* \sigma_*^2)^{\text{new}} = (\Sigma_* \sigma_*^2)^{\text{old}} + (\dot{\Sigma}_*^{\text{SF}} dt \max(\sigma_{\text{turb}}^2, \sigma_{*,\text{min}}^2))$.

Once stars form, they also migrate. In our model, this is treated quite similarly to the gas migration discussed in Section 2.1, namely the stars experience torques if they are gravitationally unstable to spiral instabilities. Our prescription has improved significantly since F12, so we discuss the new governing equations in Appendix B. Overall this typically has a minor effect on the dynamics of the disc, although it can strongly influence the stellar velocity dispersions particularly at small radii.

2.4 Metallicity

In addition to its dynamical effects, SF is responsible for the production of metals. We approximate this process as instantaneous, in which case the production of metals is proportional to the SFR. In each cell, the mass in metals is evolved according to

$$\begin{aligned} \frac{\partial M_Z}{\partial t} = & \Delta r \frac{\partial \dot{M}_Z}{\partial r} + (y f_R - f_R Z - \mu Z_w) \dot{M}_*^{\text{SF}} \\ & + \dot{M}_{\text{acc}} Z_{\text{IGM}} + \frac{\partial}{\partial r} \kappa_Z \frac{\partial}{\partial r} M_Z. \end{aligned} \quad (9)$$

² Note that F12 omitted the factor of $1/Q_g \pi$, though the code and Appendix A with the dimensionless version were correct.

The first term accounts for metals advected from other parts of the disc; Δr is defined as the width of the cell under consideration. The next term includes three effects which occur in proportion to the SFR in that cell, $\dot{M}_*^{\text{SF}} \equiv \pi(r_{i+1/2}^2 - r_{i-1/2}^2)\dot{\Sigma}_*^{\text{SF}}$ – here $r_{i+1/2} \equiv \sqrt{r_i r_{i+1}}$, the location of the boundary between cells i and $i+1$ on our logarithmic grid. The first is the production of new metals through the course of stellar evolution, which occurs in proportion to y , defined as the mass of metals produced per unit mass (of all gas) locked in stars. Next is the mass of metals locked in stellar remnants. The final term proportional to the SFR is the mass of metals ejected in galactic winds with mass loading factor μ . Defining $\dot{M}_{\text{acc}} \equiv \pi(r_{i+1/2}^2 - r_{i-1/2}^2)\dot{\Sigma}_{\text{cos}}$, the next term is simply the mass of metals accreting from the intergalactic medium (IGM). The final term, metal diffusion, will be discussed momentarily.

The metallicity of the wind is given by Z_w . Many authors assume that $Z_w = Z$, the metallicity of the gas in the disc. It is worth pointing out that this is probably a lower bound, but there is also an upper bound. In the limit of small mass loading factor μ , the maximum metallicity is the mass in metals expelled by stellar winds and SNe: $(y f_R + (1 - f_R)Z)\Delta M_*$ divided by the total mass ejected, $(1 - f_R)\Delta M_*$. When the mass loading factor is larger than $1 - f_R$, some additional mass from the ISM must also be swept up, thereby decreasing the maximum metallicity. The metallicity of the wind must therefore be

$$Z < Z_w < \begin{cases} Z + y f_R / (1 - f_R) & \text{if } \mu \leq 1 - f_R \\ Z + y f_R / \mu & \text{if } \mu > 1 - f_R \end{cases}. \quad (10)$$

We therefore define a new parameter ξ , similar in spirit to e.g. the metal loss factor in Krumholz & Dekel (2012), so that

$$Z_w = Z + \xi \frac{y f_R}{\max(\mu, 1 - f_R)}. \quad (11)$$

Here ξ may vary between 0 and 1, with 0 representing the usual assumption of perfect mixing of stellar ejecta and galactic outflows, and 1 representing the minimal possible mixing.

The diffusion of metals has received relatively little attention until recently. In F12, we included this diffusion term to prevent the metallicity gradient from steepening excessively, tuning the value of κ_Z to yield a reasonable gradient. Since then, Yang & Krumholz (2012) have measured the value of κ_Z in a 2D shearing box simulation with turbulence driven by thermal instability. They show that to a reasonable approximation $\kappa_Z \propto r_{\text{inj}}^2 / t_{\text{orb}}$, where r_{inj} is the initial wavelength of the metallicity perturbation and t_{orb} is the orbital time. Here we make the approximation that $r_{\text{inj}} \approx \lambda_J = \sigma^2 / G\Sigma$, the 2D Jeans length, since this should be similar to the spacing of the largest giant molecular clouds. We can therefore scale κ_Z in our simulation to their measured value as

$$\kappa_Z(r, t) = k_Z 1.2 \frac{\text{kpc}^2}{\text{Gyr}} \left(\frac{\sigma^2 / G\Sigma}{3.1 \text{ kpc}} \right)^2 \frac{\kappa}{\sqrt{2} (26 \text{ km s}^{-1} \text{ kpc}^{-1})}. \quad (12)$$

The numerical values are the measured κ_Z and the input parameters r_{inj} and Ω quoted for one of their simulations. We also include a free parameter $k_Z \approx 1$, recognizing that there is some uncertainty in this result. The numerical implementation of the diffusion term is operator split from the rest of the terms, implicit and computed in terms of fluxes so that metal mass is explicitly conserved. We also enforce $\kappa_Z < v_{\text{circ}} R$, the largest velocity and radius in the problem, which is not guaranteed by equation (12) when Σ is very small. This essentially makes sure that the metal injection scale $r_{\text{inj}} \lesssim R$, the size of the system.

2.5 Accretion

In our model, gas accretes on to the disc at an externally prescribed rate \dot{M}_{ext} and a profile $\dot{\Sigma}_{\text{cos}}$ such that

$$\dot{M}_{\text{ext}}(t) = \int_0^\infty \dot{\Sigma}_{\text{cos}}(r, t) 2\pi r \, dr. \quad (13)$$

In our fiducial model, we take $\dot{\Sigma}_{\text{cos}} \propto \exp(-r/r_{\text{acc}}(z))$. The angular momentum of accreting gas is thereby entirely set by $r_{\text{acc}}(z)$, which is assumed to scale with halo mass so that

$$r_{\text{acc}}(z) = r_{\text{acc}}(z=0) \left(\frac{M_h(z)}{M_h(z=0)} \right)^{\alpha_r}, \quad (14)$$

with $r_{\text{acc}}(z=0)$ and α_r left as free parameters. A reasonable guess for α_r is $1/3$, which roughly corresponds to the assumption that $r_{\text{acc}} \propto R_{\text{vir}}$ (e.g. Mo, Mao & White 1998), while a reasonable guess for r_{acc} might be the size scale of local disc galaxies, which varies significantly at fixed mass but is of the order of 10 kpc.

To determine \dot{M}_{ext} at each time step in our simulation, we calculate $\dot{M}_h(t)$, the history of the dark matter halo mass, differentiate with respect to time and multiply by $f_b \epsilon_{\text{in}}(M_h, z)$, where $f_b \approx 0.17$ is the universal baryon fraction and ϵ_{in} is some efficiency. We take two separate approaches to calculating $\dot{M}_h(t)$. The first is to use an average dark matter accretion history (Neistein & Dekel 2008; Bouché et al. 2010), which estimates the average growth rate to be

$$\dot{M}_h = 39(M_h/10^{12} M_\odot)^{1.1} (1+z)^{2.2} M_\odot \text{ yr}^{-1}, \quad (15)$$

which agrees well with hydrodynamic simulations (Dekel et al. 2013). This approach allows us to quickly and clearly see the effects of changes in the physical parameters of the simulations without averaging over many galaxies with different accretion histories. The disadvantage is that in reality galaxies are likely to have stochastic accretion histories, and this will have a significant effect on the resultant galaxies. For instance, if a galaxy is fed at a steady rate, if a given region of the disc becomes stable to gravitational turbulence, it is unlikely to ever destabilize again, but an accretion history with variation about the median could be unstable at low redshifts or stable at high redshifts.

To capture the effects of variable accretion histories, we also generate accretion histories using the analytical formalism developed by Neistein & Dekel (2008) and Neistein, Maccio & Dekel (2010). The procedure is as follows. The desired final halo mass $M_{h,0}$ and redshift ($z=0$) are converted into their corresponding dimensionless values S and ω . We use the approximate relation from van den Bosch (2002)

$$S(M_h) = u^2 \left(\frac{c_0 \Gamma}{\Omega_m^{1/3}} \left(\frac{M_h}{1 M_\odot} \right)^{1/3} \right) \frac{\sigma_8^2}{u^2 (32 \Gamma)}. \quad (16)$$

The parameters c_0 and Γ are, respectively, 3.804×10^{-4} and 0.169. The function $u(x)$ is given by

$$u(x) = 64.087 (1 + 1.074 x^{0.3} - 1.581 x^{0.4} + 0.954 x^{0.5} - 0.185 x^{0.6})^{-10}. \quad (17)$$

Meanwhile, $\omega(z)$ may be computed approximately (Neistein & Dekel 2008) by

$$\omega(z) = 1.260 (1 + z + 0.09(1+z)^{-1} + 0.24e^{-1.16z}). \quad (18)$$

With these relations, we now have $S(M_h(z=0))$ and $\omega(z=0)$.

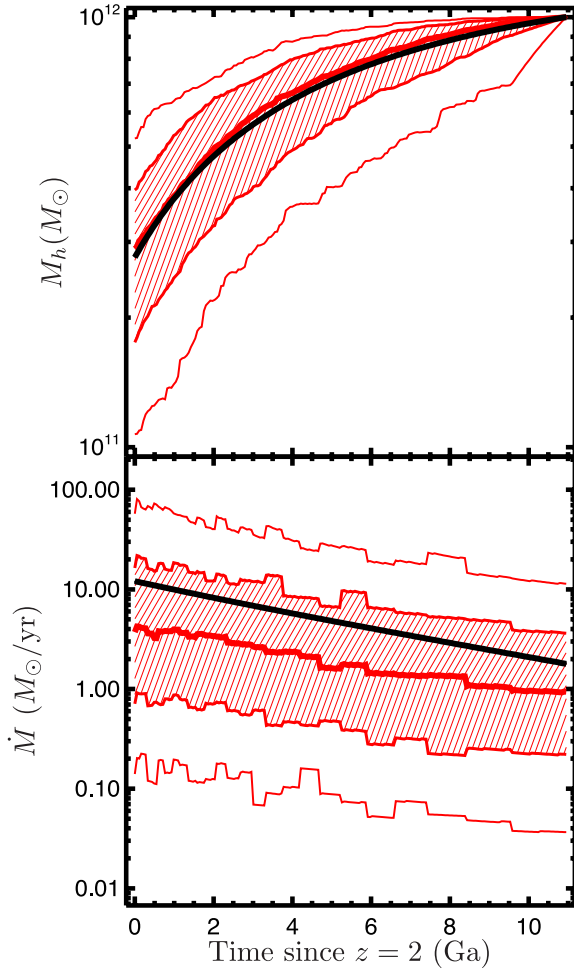


Figure 1. The growth of haloes. The top panel shows the evolution of the halo mass for the smooth accretion history (black) and the median, central 68 per cent (shaded), and central 95 per cent of 400 stochastic accretion histories (red). The corresponding distribution of the inferred baryon accretion rates, including the efficiency factor (equation 22), is shown in the bottom panel. The steps in \dot{M} correspond to our fixed interval $\Delta\omega = 0.086$.

The independent variable ω is steadily incremented by a fixed value $\Delta\omega$ until the entire desired redshift range is encompassed. At each step in ω , a new value of S is computed by adding

$$\Delta S = \exp(x\sigma_k + \mu_k), \quad (19)$$

where x is a value drawn from a normal distribution with zero mean and unity variance. We use a fixed $\Delta\omega = 0.1$, since this is the time step used in generating the fitting formulae for σ_k and μ_k in Neistein & Dekel (2008). The fact that we use a fixed $\Delta\omega$ rather than a distribution leads to the distinct steps in Fig. 1, where all of the accretion histories change at once.

The mean of the normal distribution to be exponentiated, μ_k , and its standard deviation, σ_k , depend on halo mass, and are fitted to the results of the Millennium Run (Springel et al. 2005),

$$\sigma_k = 1.367 + 0.012 \log_{10} S + 0.234 (\log_{10} S)^2 \quad (20)$$

$$\mu_k = -3.682 + 0.76 \log_{10} S - 0.36 (\log_{10} S)^2. \quad (21)$$

Converting each value of S back to M_h , one obtains a dark matter accretion history $M_h(\omega_j)$, where the ω_j are the sequence of ω 's obtained by incrementing ω by the fixed $\Delta\omega$, namely $\omega_j =$

$\omega_0 + j\Delta\omega$ for $j = 0, 1, 2, \dots$ and $\omega_0 = \omega(z = 0)$. We require that the change in M_h over a single step, $M_h(\omega_i) - M_h(\omega_{i+1})$, not exceed $M_h(\omega_{i+1})$ to avoid galaxies ‘accreting’ a larger mass than their own, i.e. becoming a satellite. Since equations (20) and (21) were obtained by a fit to the Millennium Run using a cosmology where $(\Omega_m, \sigma_8) = (0.25, 0.9)$, when converting between M_h and S with equation (16), we use the parameters from the Millennium Run. Once we have obtained $M_h(\omega_j)$ using this cosmology, we can transform it so that it agrees with the *Wilkinson Microwave Anisotropy Probe 5* (WMAP5; Komatsu et al. 2009) cosmology $(\Omega_m, \sigma_8) = (0.258, 0.796)$, which is much closer to the current best-fitting values. We use the scaling obtained in Neistein et al. (2010) via a comparison of merger trees from Millennium and an N -body simulation run with WMAP5 cosmology, namely we replace ω_j with $\tilde{\omega}_j = \omega_0 + 0.86j\Delta\omega$. The full dark matter mass history of the halo $M_h(t)$ is then obtained by converting $\tilde{\omega}_j$ to z (with equation 18) and subsequently to t , and linearly interpolating the sequence of halo masses $M_h(\tilde{\omega}_j)$ in time. The dark matter accretion history is then just the instantaneous derivative of $M_h(t)$.

The input $\dot{M}_{\text{ext}}(t)$ to our simulation is taken to be the average dark matter accretion rate at time t , generated either from the smooth accretion formula (15) or the lognormal one (19) times $f_b \epsilon_{\text{in}}$. For the efficiency, we use a reasonably general parametrization,

$$\epsilon_{\text{in}}(M_h, z) = \min \left(\epsilon_0 \left(\frac{M_h}{10^{12}} \right)^{\beta_{M_h}} (1+z)^{\beta_z}, \epsilon_{\text{max}} \right). \quad (22)$$

Faucher-Giguere, Keres & Ma (2011) fit the results of a cosmological smoothed particle hydrodynamic simulation with no feedback to find $(\epsilon_0, \beta_{M_h}, \beta_z, \epsilon_{\text{max}}) = (0.31, -0.25, 0.38, 1)$, though they explicitly only use this fit above $z = 2$.

Despite its success at high redshift, the paradigm of cold accretion is fairly uncertain for galaxies which have some hot coronal gas, like the MW, at low redshift. Moreover, Diemer, More & Kravtsov (2013) have pointed out that below $z \sim 1$, nearly all the growth in M_h for haloes with $M_h(z = 0) \sim 10^{12} M_\odot$ corresponds to the fact that the background density of the Universe is decreasing (roughly as $\rho_m \propto (1+z)^3$) while dark matter haloes are changing very little. Because the haloes are defined in simulations as having a spherical overdensity relative to the background of ~ 200 , relatively static haloes increase their mass merely because of this drop in the background density. Dekel et al. (2013) have verified that this is not a significant effect at $z > 1$.

A number of ideas have been proposed to explain how MW-like galaxies can maintain SFRs of the order of $2 M_\odot \text{ yr}^{-1}$ despite little evidence of cold accretion at anything near these rates. The gas may be accreting in an ionized phase, slightly hotter than the observed high-velocity clouds in H I (Joung et al. 2012). The process may be helped along by SN-induced accretion, where hot halo gas is supposed to condense in the wakes of cold clouds ejected by SN feedback from the disc of the galaxy (Marinacci et al. 2010). Alternatively, galaxies can be powered by gas recycled back to the ISM from stars (Leitner & Kravtsov 2011); while much of this process can be approximated as occurring instantaneously (the winds from and SNe of massive stars), a significant amount of mass is returned even from very old stellar populations (see also Martig & Bournaud 2010). Gas ejected by galactic winds often finds its way back to the star-forming disc (Oppenheimer et al. 2010), which may provide yet another way to provide star-forming gas to galaxies even if dark matter is not accreting.

Given the uncertainties in how gas is accreted at low redshift, our naive approach of setting $\dot{M}_{\text{ext}} = \dot{M}_h f_b \epsilon_{\text{in}}$ is not unreasonable. In our

fiducial model, an MW-mass galaxy accretes roughly $2 M_\odot \text{ yr}^{-1}$ at redshift zero, and so yields an SFR similar to observations, even if the physical mechanism for this accretion is unclear. We do retain, in varying the parameters of the accretion efficiency and the accretion profile, a considerable amount of flexibility in the model, which is appropriate given the uncertainties. In Fig. 1, we show $M_h(t)$ and the resulting \dot{M}_{ext} for the fiducial smooth model and the stochastic accretion model.

2.6 Initial conditions at $z \sim 2$

Having constructed the accretion history, we can now generate an initial condition. To do so, we first require that the total surface density in gas and stars equals some fraction f_{cool} of the total baryonic mass available, $f_b M_h(z = z_{\text{start}})$. For haloes which will host a single galaxy at redshift zero, it is reasonable to assume that at high redshift, $M_h(z = z_{\text{start}})$ will be small enough that the cooling time of halo gas is short, and that even if a galaxy has a stable virial shock, it may still be fed by cold streams, and so f_{cool} should be of order unity (Birnboim & Dekel 2003; Kereš et al. 2005; Dekel & Birnboim 2006; Ocvirk, Pichon & Teyssier 2008; Dekel et al. 2009a, 2013; Danovich et al. 2012).

We next make the fairly arbitrary decision to have a fixed initial gas fraction $f_{g,0}$, defined at each radius to be $f_g(r) = \Sigma / (\Sigma + \Sigma_*)$. Thus, Σ and Σ_* will have the same shape. Observations of main-sequence galaxies at high redshift show their stellar profiles to be exponential (Wuyts et al. 2011; van Dokkum et al. 2013), so we choose an initial exponential profile with scalelength $r_{\text{IC}} = r_{\text{acc}}(z = z_{\text{start}})$. With these requirements we arrive at the initial profile,

$$\Sigma = f_{g,0} f_{\text{cool}} f_b \frac{M_h(z = z_{\text{start}})}{2\pi r_{\text{IC}}^2} \exp(-r/r_{\text{IC}}) \frac{1}{1 - f_{\text{out}}}. \quad (23)$$

The final factor is a correction for the finite size of the computational domain. In particular, we want the initial mass of the disc to be independent of R , so f_{out} is the fraction of the mass profile which lies beyond the computational domain,

$$f_{\text{out}} = \frac{1}{2\pi r_{\text{IC}}^2} \int_R^\infty 2\pi r e^{-r/r_{\text{IC}}} dr. \quad (24)$$

Since the initial conditions are highly uncertain, it is more important to get the correct amount of mass in the computational domain than to make sure the profile has a particular normalization. Still, we typically set $R \gg r_{\text{IC}}$ so that this is a minor correction.

The other initial variables we need to specify are σ_r , σ_{zz} , Z and Z_* . For the metallicities, we simply set $Z = Z_* = Z_{\text{IGM}}$. For the velocity dispersions, we use $\sigma = \sigma_{rr} \phi_0^{-1} = \sigma_{zz} \phi_0^{-1} = \sigma_{\text{sf}}$, i.e. the value of our minimum velocity dispersion. We allow the velocity dispersion of the stars to be different (generally higher) than that of the gas, with a free parameter ϕ_0 . The low constant values of the velocity dispersion will often lead some parts of the disc to have $Q < Q_{\text{GI}}$, so in those regions we raise σ , $\sigma_{rr} \phi_0^{-1}$ and $\sigma_{zz} \phi_0^{-1}$ simultaneously (keeping them equal) until $Q = Q_{\text{GI}}$. We emphasize, though, that the gas velocity dispersion σ and the two stellar velocity dispersions σ_{rr} and σ_{zz} evolve separately throughout the simulation – their ratio is fixed only initially. The idea is that, since supersonic turbulence in the disc is generated exclusively by GI in our model, any region not subject to this instability will have $\sigma \approx \sigma_{\text{sf}}$.

Typically, our initial conditions have $Q = Q_{\text{GI}}$ in some annulus. At larger radii Σ drops off quickly so $Q \propto \Sigma^{-1}$ increases, while Q also increases at smaller radii through the dependence $\kappa \propto v_\phi/r$.

We discuss the (lack of) sensitivity of our results to these choices of initial conditions in Appendix C.

3 SIMULATION RESULTS

In this section, we discuss some generic features of the galaxies produced by our model. We begin by exploring models with smooth accretion histories and a fiducial choice of parameters, which we summarize in Table 1. These are compared with artificial, illustrative models where one important physical ingredient is turned off by hand. We then allow the accretion histories to vary stochastically in a cosmologically realistic way, illustrating the differences between galaxies with identical physical laws but different accretion histories, as one might expect for real galaxies. Finally, we compare our models with recent observational results.

3.1 Equilibria in smoothly accreting models

There are three terms in the continuity equation (equation 2). At a particular radius, gas arrives via $\dot{\Sigma}_{\text{cos}}$, departs via $(f_R + \mu)\dot{\Sigma}^{\text{SF}}$ and moves to or from other radii via $\dot{\Sigma}_{\text{tr}} \equiv (2\pi r)^{-1}(\partial \dot{M}/\partial r)$. The generic behaviour of this equation at a given radius in our fiducial model is that gas will build up, either via direct accretion or as mass arrives from somewhere else in the disc, until an equilibrium is reached such that $\dot{\Sigma} \approx 0$. This equilibrium will then slowly evolve with time as the global gas accretion rate \dot{M}_{ext} falls off.

To aid in understanding how this equilibrium emerges, we have run three simple models with identical smooth accretion histories: (i) the fiducial model – our best guess for physical parameters which will lead to something resembling the MW (see Table 1), (ii) the same model with no SF and (iii) the same model with no gravitational instability, i.e. $\mathcal{T}_{\text{GI}} = 0$ everywhere. The features of models (i) and (ii) are similar at large radii, while the features of (i) and (iii) bear some resemblance at small radii. This immediately suggests that GI transport is important at large radii and SF is important at small radii. The gas surface density distributions of each model are shown in Fig. 2 as a function of time. The gas is supplied via an exponential distribution, $\dot{\Sigma}_{\text{cos}} \propto e^{-r/r_{\text{acc}}}$. Without GI (model iii), SF carves out the inner parts of the distribution, leaving a hole in the gas at galactic centres, while without SF (ii), gas is redistributed into a power-law distribution, following roughly $\Sigma \propto 1/r$.

A useful way to understand what sets the surface density is to examine the relative effects of each term in the continuity equation. In particular, at each time and radius, we can divide each term by $A \equiv |\dot{\Sigma}_{\text{tr}}| + \dot{\Sigma}_{\text{cos}} + (\mu + f_R)\dot{\Sigma}^{\text{SF}}$. In Fig. 3, we compute these contributions, including the sign of their effect on the overall value of $\partial \Sigma / \partial t$, so at each radius the fraction of the coloured region occupied by red, orange and blue represents the fraction of A from SF, cosmological accretion and transport. The different shades of blue show which way the mass is flowing in the disc, i.e. the sign of \dot{M} – dark blue indicates gas flowing towards the centre of the disc and light blue outward motion.

When the coloured band in Fig. 3 stretches from -0.5 to 0.5 , that region of the disc has reached an equilibrium configuration. In each case shown here, the equilibration proceeds from inside outwards. This is a combination of two effects – the especially efficient SF in the centre of the disc and the fairly centrally concentrated distribution of accreting gas. The equilibrium does not last forever – at $z = 0$, there can be significant deviations as the disc processes past accretion and the instantaneous accretion rate falls owing to the

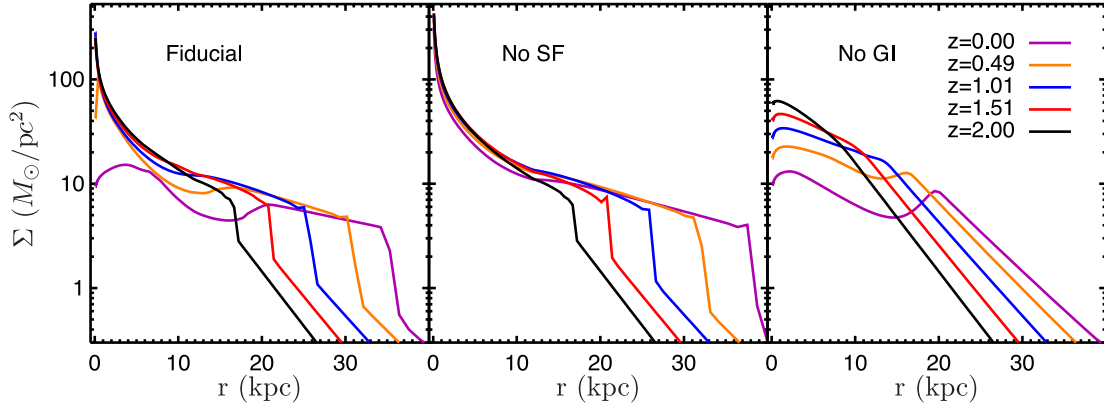


Figure 2. The three simplified models. For each of these simplified models, we show the evolution of the radial gas surface density distribution. We see that the evolution of the fiducial model is a non-trivial combination of the effects of SF and GI transport.

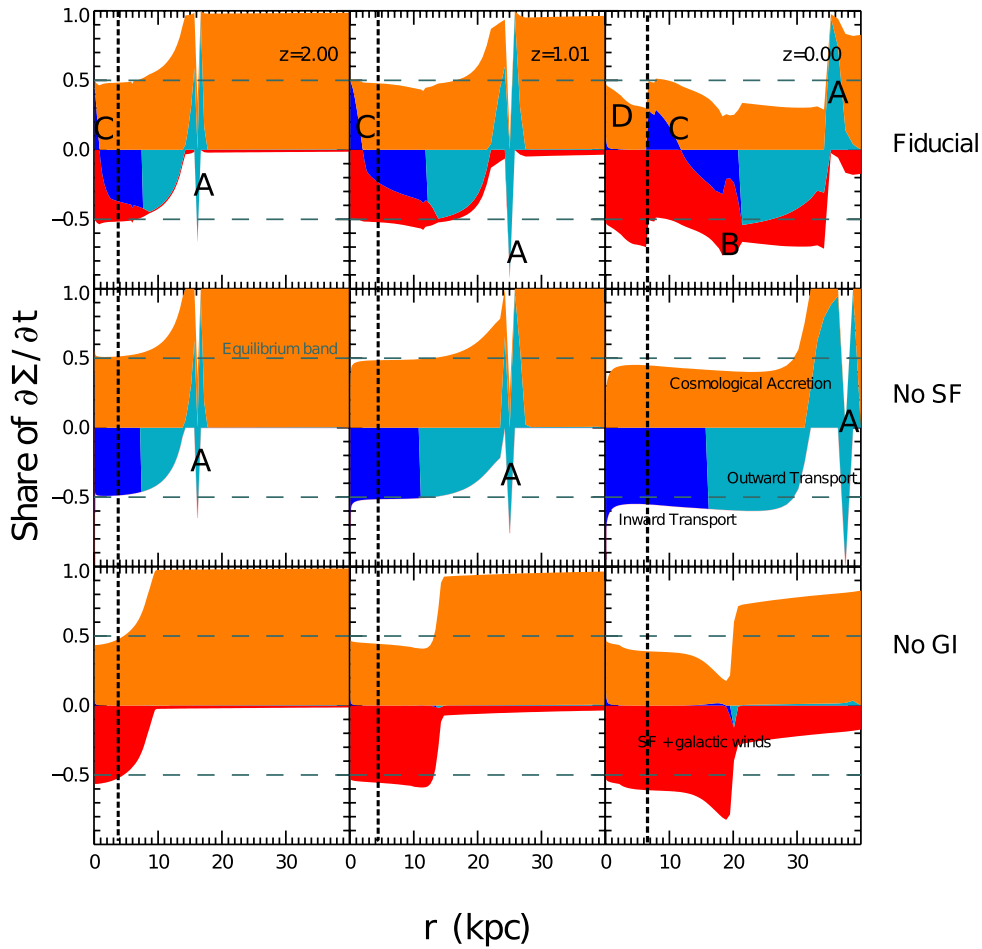


Figure 3. The balance of terms in the continuity equation. The terms contributing to $\partial\Sigma/\partial t$ are split into those which increase Σ at a particular radius and time and those which decrease Σ . The former appear above zero and the latter below. Each term is represented by a different colour – orange for $\dot{\Sigma}_{\text{cos}}$ (with the exponential scalelength marked as a vertical dashed line), red for $(\mu + f_R)\dot{\Sigma}_*^{\text{SF}}$ and blue for $\dot{\Sigma}_{\text{trans}}$. Light (dark) blue indicates gas being transported outwards (inwards). At each radius, the height of the coloured band is normalized to unity, and its position shows how close the disc is to equilibrium (equal positive and negative contributions) at that radius – a radius is in equilibrium if the coloured band falls exactly between the dashed lines labelled ‘equilibrium band’. The columns show different redshifts ($z = 2, 1, 0$) and the rows show different models: fiducial, no SF and no GI transport. The features labelled ‘A’ at e.g. $r = 15$ kpc and $z = 2$ come from gas from the unstable region heating when it piles up in a single stable cell. Real galaxies will not have such a sharp transition since there will be some breaking of axisymmetry and some overshoot from the unstable region, neither of which we model here. The feature labelled ‘B’ at $z = 0$ around 15–20 kpc in the fiducial model is caused by an enhancement in the surface density distribution – this section of the galaxy has had the direction of GI transport reverse from outwards to inwards. Features labelled by ‘C’ show where GI changes from removing gas from a given radius to adding it to that radius (see the text for more details), and ‘D’ shows GI quenching, wherein SF has exhausted the supply of inflowing gas.

expansion of the Universe on time-scales potentially shorter than the gas depletion time at these large radii.

3.1.1 Equilibrium between SF and accretion: the no GI model

We first focus on the model with no GI. In this model, at a given radius, gas builds up until the local $\text{SFR} \propto \Sigma$ can balance the incoming accretion. This happens first in the centre of the disc. Not only is the cosmological accretion rate per unit area larger there, but the SF time-scale is shortest (Fig. 4). In this model there is in fact

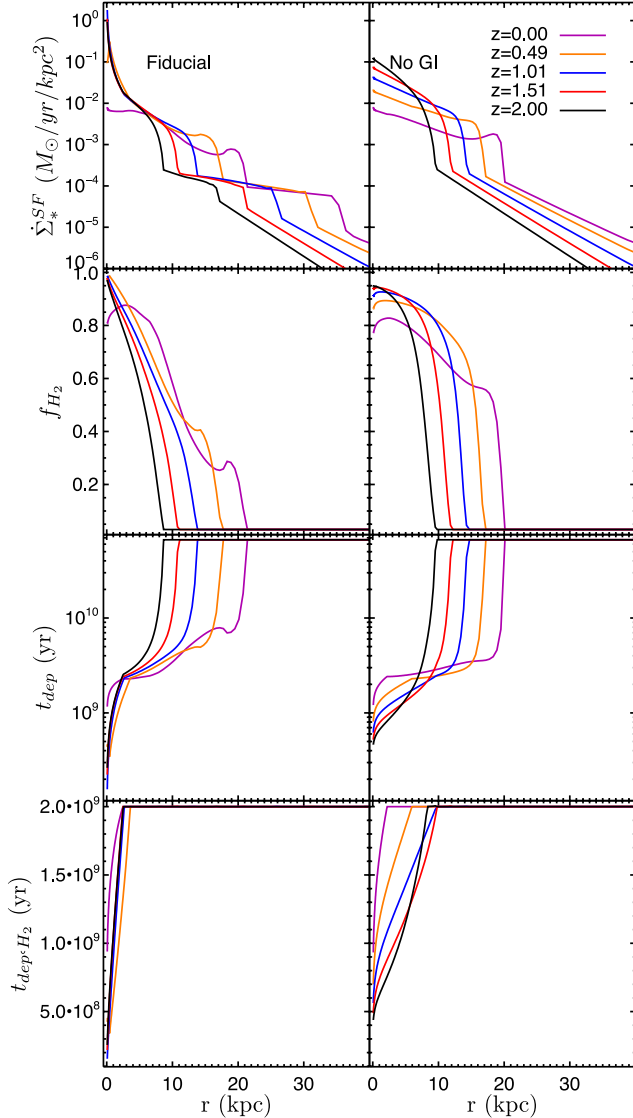


Figure 4. SF in the fiducial and no GI models. The SFR (top) is proportional to the surface density of gas modulated by other factors reflected in the depletion time, $t_{\text{dep}} = \Sigma / \dot{\Sigma}_{\text{SF}}^{\text{SF}}$ (second row). The first is the molecular fraction (third row), itself determined by Σ and Z , and the second is the regime of SF (bottom row), either single-cloud ($t_{\text{dep}, \text{H}_2} = 2$ Gyr) or Toomre ($t_{\text{dep}, \text{H}_2} \propto 1/\sqrt{G\rho} < 2$ Gyr). For each quantity, the left-hand panel shows the fiducial model, while the right shows the model with GI transport turned off. SF in the fiducial model is much more concentrated and reaches much higher surface densities $\dot{\Sigma}_{\text{SF}}^{\text{SF}}$ through the action of GI transport. The absence of GI causes so much gas to build up at larger radii that at high redshift the Toomre regime of SF extends to nearly 10 kpc, instead of just the inner few kpc.

a huge range of depletion times, from roughly 100 Myr at $z = 2$ at small radii to 60 Gyr in the outer disc. There are two effects driving this diversity. For depletion times between 100 Myr and 2 Gyr, the disc is in the Toomre regime of SF (see equation 8), for which the depletion time scales as κ^{-1} . This region is typically small, $\lesssim 3$ kpc, outside of which the time-scale would become longer than 2 Gyr if it continued to follow the κ^{-1} scaling. At this point, the disc transitions to the single-cloud regime of SF. At the transition, the disc still tends to be dominated by molecular gas. In the mostly molecular but still single-cloud regime, the depletion time is roughly 2 Gyr, the single-cloud molecular depletion time – this can be seen as a flattening in the t_{dep} distribution with radius. There is then a transition from molecular to atomic gas, which accounts for the difference between parts of the disc with a 2 Gyr depletion time and a 60 Gyr depletion time – this maximum depletion time is set by $f_{\text{H}_2, \text{min}}$, which is quite uncertain.

A generic feature of the no GI model is that at the edge of the star-forming region, SF occurs at a slightly faster rate than new gas is accreted at that radius (Fig. 3, bottom-right panel). All of the models, particularly at lower redshift, exhibit a slight tendency to fall just below the ‘equilibrium band’ after they have initially equilibrated at a given radius, since the accretion rate is externally imposed and falling monotonically. The feature at $z = 0$ in the no GI model goes beyond this, however, and may be explained by a small feedback loop in the SF law introduced by the dependence on metallicity. The demarcation between the star-forming part of the disc and the outskirts is set by the molecular to atomic transition. Typically, the SFR at a given radius is able to consume the incoming material only if the molecular fraction there is above the minimum allowed value – otherwise SF would be too slow. When enough gas has accumulated to satisfy $f_{\text{H}_2} > f_{\text{H}_2, \text{min}} = 0.03$, the SFR rises steeply with column density and new metals are produced, which in turn catalyse SF by reducing the amount of gas needed to maintain a molecular, star-forming phase. Thus, the extra gas, which is now no longer necessary for the SFR to balance the accretion rate, can be consumed, though this generally takes a significant amount of time, $t_{\text{dep}} \gtrsim 2$ Gyr.

3.1.2 Equilibrium between GI transport and accretion: the no SF model

We now turn to the no SF model to help us understand the importance of GI. In our model, when the disc has enough gas to be gravitationally unstable, it self-regulates to a marginally stable level, namely $Q = Q_{\text{GI}} = \text{const.}$, where Q_{GI} demarcates gravitational stability from instability. The value of Q depends on the surface densities and velocity dispersions of the gas and stars. In our numerical simulations, we account for these dependences using the formula from Romeo & Wiegert (2011), but this formula reduces to something quite similar to the much simpler Wang & Silk (1994) approximation when $\sigma \approx \sigma_{\text{rr}} \approx \sigma_{\text{zz}}$, namely $Q^{-1} \sim (2/3)(Q_{\text{g}}^{-1} + Q_{\text{*}}^{-1})$. In our model, the situation can be simplified even further by the fact that $Q_{\text{*}}$ is separately self-regulated by stellar migration via transient spiral heating, so that $Q \sim (3/2)Q_{\text{g}}$. In this case, the $Q = Q_{\text{GI}}$ condition may be re-written as

$$\Sigma \approx \Sigma_{\text{GI}} \equiv \frac{3}{2} \frac{\sqrt{2(\beta+1)}v_{\phi}\sigma}{\pi G r Q_{\text{GI}}}. \quad (25)$$

At a given radius, β , v_{ϕ} , r and Q_{GI} are all fixed, so equation (25) may be considered a direct mapping between Σ and σ . If σ does not vary by much and the velocity dispersions of the gas and stars

are similar, then Σ will simply follow a $1/r$ power law over a wide range of radii.

The velocity dispersion and hence $r\Sigma_{\text{GI}}$ is restricted to a relatively narrow range because there is both a minimum and maximum velocity dispersion. The minimum is set by the feedback velocity dispersion, σ_{sf} – the gas cannot get colder than when its turbulent velocity dispersion is zero. We can therefore say that in a gravitationally unstable region,

$$\Sigma \gtrsim \Sigma_{\text{crit}} \equiv \frac{3}{2} \frac{\sqrt{2(\beta+1)}v_\phi\sigma_{\text{sf}}}{\pi G r Q_{\text{GI}}}. \quad (26)$$

The maximum is determined by the gas supply – for a given \dot{M}_{ext} to be transported to the centre of the disc in a quasi-steady state, it must dissipate the gravitational potential energy between where it arrives and the centre of the galaxy, and it must experience enough torque to lose its angular momentum. In a steady state, local heating by torques balances local cooling by turbulent dissipation (see Section 3.1.4). Note that ‘heating’ and ‘cooling’ refer to changing the turbulent velocity dispersion of the gas, not its kinetic temperature. The rate at which the gas cools (and hence experiences torques) \mathcal{L} depends on the velocity dispersion. The maximum velocity dispersion is therefore set by assuming that 100 per cent of the gas arriving from an external source flows towards the centre in a steady state. Since some gas never reaches the centre because of SF, and other gas moves outwards rather than inwards, this is an upper limit. As shown in Section 4.1, at $z \sim 2$ for galaxies accreting at $\sim 10 M_\odot \text{ yr}^{-1}$ the velocity dispersion is restricted to $8 < \sigma \lesssim 20 \text{ km s}^{-1}$. This value is low compared to the measured velocity dispersions in the SINS galaxies (Förster-Schreiber et al. 2009). As we will see in Section 4.1, some small fraction of MW progenitors do have much higher accretion rates in our stochastic accretion model. Moreover, the SINS galaxies are likely somewhat more massive than the MW progenitors we consider here.

As more gas arrives at a region of the disc in a marginally unstable state, the surface density is fixed in the profile given by Σ_{GI} . Since there is a maximum velocity dispersion for a fixed accretion rate, gas is not allowed to accumulate, lest $\sigma \propto \Sigma_{\text{GI}}$ exceed this maximum, so the only thing the gas can do is move elsewhere. The gas will then be transported away from where it arrives until it reaches part of the disc which is stable, where it will pile up until that region too becomes unstable. This ‘wave’ of GI can be seen propagating outwards in

Fig. 3 in both the fiducial model and the model without SF, until essentially the entire disc is unstable. The equilibrium between GI transport and accretion appears originally at $r \leq 7 \text{ kpc}$ both with and without SF. This location is picked out by the maximum in $\dot{\Sigma}_{\text{acc}}/\Sigma_{\text{GI}} \propto r \exp(-r/r_{\text{acc}})$, i.e. where gas piles up fastest relative to the amount necessary to be gravitationally unstable, which occurs at r_{acc} for a flat rotation curve.

3.1.3 The fiducial model

Having examined the simplified models where we disabled GI transport or SF, we now turn to our fiducial model which includes both. Recalling the surface density distributions shown in Fig. 2, it seems that the fiducial model behaves largely like a superposition of the model without SF and the model without GI.

In the previous section, we point out that an equilibrium between GI transport and infalling accretion arises when $\Sigma \approx \Sigma_{\text{GI}} > \Sigma_{\text{crit}}$ (see Fig. 5) and more gas is added. The new gas will be whisked away until it piles up somewhere in the disc that is not yet unstable. If we also include SF, then rather than being pushed out into a stable region, the gas can be consumed by SF. Comparing the model without SF to the fiducial model at $z = 2$ and 1 in Fig. 3, we can see this effect in action. Gas arrives around r_{acc} , and on its way inwards it is consumed by SF. The balance is then between cosmological accretion and both SF and GI transport, rather than just GI transport alone. In other words, if the disc can get rid of some gas via SF, it no longer has to transport it away as fast to maintain $\Sigma \approx \Sigma_{\text{GI}}$. Eventually, all of the infalling gas at a given radius can be consumed by SF, and GI transport briefly has no net effect. Just interior to this point though, the cosmological accretion rate is low enough and the SFR is fast enough that accretion alone can no longer supply the SF at that radius, and the stars start forming not from material falling directly on to that radius, but from gas arriving from other parts of the disc via GI transport. These are the points in Fig. 3 (labelled by ‘C’) where $\dot{\Sigma}_{\text{trans}}$ goes from negative to positive. Visually, it is clear that the SF (red) is being supplied by inflowing material (dark blue).

In this situation, where the SF is depleting the inflowing gas, the surface density is affected but not necessarily drastically. In a steady state, the surface density and velocity dispersion (related via $\Sigma \approx \Sigma_{\text{GI}}$) are primarily set by the amount of energy that needs to be dissipated by turbulence, which is set by the amount of torque

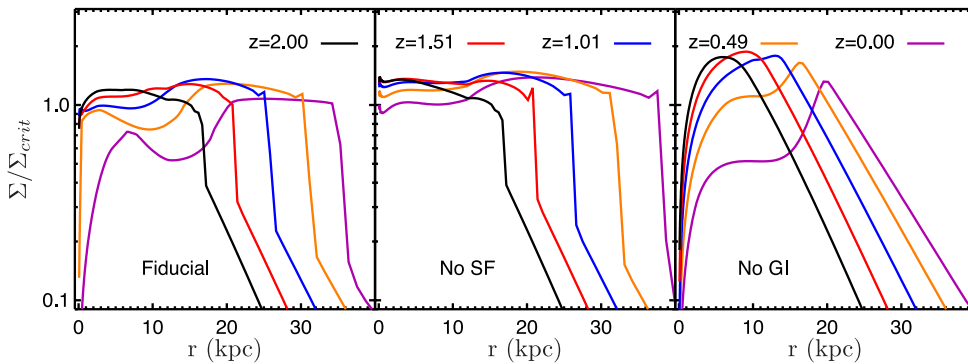


Figure 5. The ratio of Σ to the minimum surface density necessary for GI, Σ_{crit} . Gas spreads out to keep the surface density above this critical value but below the maximum value, Σ_{GI} evaluated at σ_{max} . This ratio falls below ~ 1 where the disc has not yet destabilized (two left-hand panels, large radii) or has stabilized due to GI quenching (left-hand panel, small radii, low redshift). Note however that in the fiducial model, especially at low redshift, Σ can fall slightly below Σ_{crit} even in gravitationally unstable regions because in deriving Σ_{GI} and Σ_{crit} we assumed that $\sigma \approx \sigma_{rr} \approx \sigma_{zz}$, which is no longer true at low redshift. This is a factor of 2 level effect – the gravitationally stable regions always have Σ well below Σ_{crit} . Interestingly, even the ‘no GI’ run does not reach values far larger than $\Sigma/\Sigma_{\text{crit}} = 1$ (although it reaches significantly higher values than the other two models), since the SFR increases as Q_g decreases – essentially the gas is compressed under its own weight and forms stars faster.

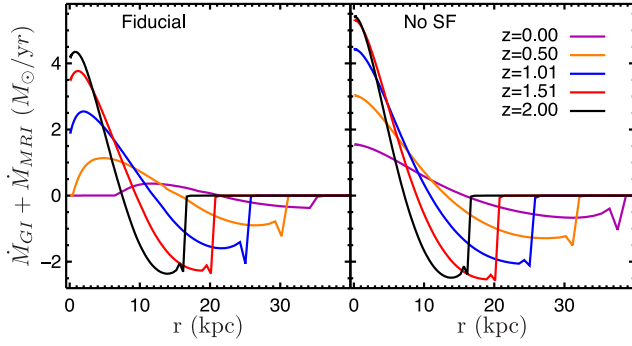


Figure 6. The inward mass flux through a given radius. Negative values are outward flux. We compare the fiducial model (left) to the same model with SF turned off (right). As usual, black, red, blue, orange and purple are $z = 2.0, 1.5, 1.0, 0.5, 0.0$. The outward mass flow is modestly affected by SF at late times, whereas the inward flow is completely consumed by low redshift. The disc stabilizes in the centre, i.e. $\dot{M} \approx 0$, simply because all of the available mass has been consumed by stars.

which must be exerted on the gas to maintain the steady state of matter flowing through the disc at rate \dot{M} (see Fig. 6), which is set by the profile and rate of external accretion. If SF is removing some of this gas supply, less energy needs to be dissipated and both Σ and σ will decrease. Eventually, if the SFR is fast enough, the inflowing gas (plus the much smaller supply of directly accreting gas) will be entirely depleted and GI will be shut down within that radius. The MRI or some other torque may operate within that radius, and there is certainly still gas within that radius. For $\alpha_{\text{MRI}} \lesssim 0.1$, the supply of gas from transport is essentially negligible compared to the supply from continued cosmological accretion. Once the gas supply is shut off in this manner, the gas will burn through the previously $\sim 1/r$ surface density until it reaches equilibrium with the infalling material. At this point newly accreted material is immediately consumed by SF, and it would take a large burst of accretion to re-activate the GI. In the fiducial model, this shutoff occurs between $z = 1$ and 0 . For quantitative estimates of when this is important, see Section 4.2.

The fiducial model also shows a peculiar peak in the SFR around $r = 17$ kpc at $z = 0$ (visible in Figs 3 and 4). This corresponds to a peak in the surface density where gas has built up in a ring, which in turn is caused by the fact that the stagnation point in the GI transport flow (i.e. where $\dot{M} = 0$) passes through this region. At

first gas arrives at this radius from a smaller radius, but at late times it arrives from a larger radius. The location of this stagnation point is set by the boundaries of the GI region, which move outwards with time (as a result of GI quenching and the steady viscous spread of the disc), and the particular choice of accretion profile. We therefore expect this feature to exist in many galaxies, but its location and prominence are quite parameter dependent in our model.

3.1.4 Energy equilibrium

Thus far we have been concerned mostly with the surface density distribution. It is clear that GI transport plays a significant role in setting this surface density. For regions of the disc which are gravitationally unstable, we have asserted that $\Sigma \approx \Sigma_{\text{GI}} \propto \sigma/r$. In Section 4.1, we will show that there is a maximum velocity dispersion set by the mass accretion rate; there is also a minimum velocity dispersion, σ_{sf} , set by the temperature of the gas. This is an adequate first-order understanding of what sets the surface density in the gravitationally unstable regions, but we have yet to explore what sets σ and hence Σ between the minimum and maximum values.

Just as with the surface densities, we can show which terms dominate the evolution of σ as a function of radius and time (Fig. 7) for the fiducial model. The equilibrium here is even more striking than for the surface densities. Nearly everywhere in the disc, the advection terms (blue and orange) are negligible, and the disc equilibrates between local heating via GI and MRI torques $\propto \mathcal{T}$ and cooling $\propto \mathcal{L}$ from turbulent dissipation. The exception is at the wave of gas moving outwards to maintain $\Sigma \approx \Sigma_{\text{GI}}$ in the inner disc. Here advection becomes important because gas is being transferred from an unstable cell to a stable one with much lower surface density. This stable cell does not pass any mass to the next cell since both have $\mathcal{T} \approx 0$, so $\partial \dot{M}/\partial r$ can be quite large. In reality, the radius separating the gravitationally unstable region from a stable region would be much less well defined, both because real galaxies are not axisymmetric and because there may be some ‘overshoot’. Our model overlooks these effects, so our transition is quite sharp – a single cell in our simulation. This is the cause of the spikiness, not only in Fig. 7, but also in Figs 3, 6 and 9.

Another exception to the otherwise good approximation that local heating balances local cooling is at $z = 2$ at large radii, where the disc is not gravitationally unstable and the only torque comes from the MRI. This region takes a long time to equilibrate because the

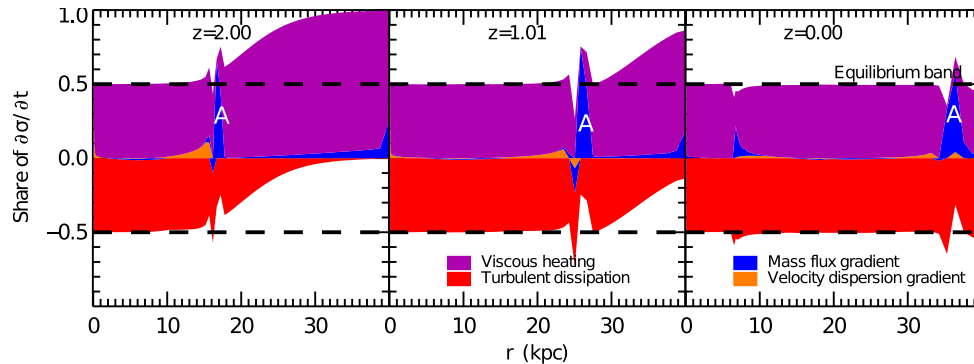


Figure 7. The balance of terms in the energy equation for the fiducial model. For each term in $\partial\sigma/\partial t$ (equation 3), the contribution it makes relative to all the other terms is shown as a function of radius and time. The red region shows cooling $\propto \mathcal{L}$, purple is heating $\propto \mathcal{T}$, orange and blue are terms associated with advection, $\propto \partial\sigma/\partial r$ and $\partial\dot{M}/\partial r$, respectively. Within the regions which are gravitationally unstable, heating and cooling balance almost perfectly. The advection terms are only relevant right where the gravitationally unstable region borders a stable region, where the velocity dispersion and especially the mass flux change dramatically, leading to the spikes at the locations labelled ‘A’ in the figure.

dynamical time is quite long, and the MRI is weak, so building up enough turbulent velocity dispersion to be countered by turbulent dissipation takes a few Gyr. Note that this is not the case in the central region at $z = 0$ where the disc is again gravitationally stable, but this time the dynamical time is short. Note also that our model implicitly assumes that gas near $\sigma \approx \sigma_{\text{sf}}$ is in equilibrium between radiative cooling and heating, so the terms we do not show here, e.g. cooling due to metal lines or heating due to the grain photoelectric effect, may dominate in the regions stable to GI.

Based on Fig. 7, it is safe to approximate the energy balance as entirely local, i.e. to neglect the advection terms, in regions of the disc where GI transport is important. Though our simulations keep all of the relevant terms, we will make this approximation in Section 4.1 to understand exactly what sets σ and Σ in gravitationally unstable regions.

3.2 Stochastic accretion

From the previous section, we have seen that a lot depends on the rate of new material being added to the galaxy. This is the term in the continuity equation which increases Σ , and the disc tends to adjust its available sinks – SF (plus galactic winds) and GI transport – to cancel this out. One may also be concerned that if galaxies do not accrete smoothly at the average rate, the intuition we have built up about a slowly evolving equilibrium in the previous section may not be applicable to real galaxies. In this section, we explore the effect of varying the accretion history stochastically.

Fig. 8 shows the distribution of surface densities for the same 400 galaxies whose accretion histories were shown in Fig. 1, plus the fiducial smooth model for reference. These galaxies all have the same radial scale, namely $r_{\text{acc}} = 6.9$ kpc. At high redshift, the galaxies have similar profiles – $1/r$ profiles at small radii and exponential profiles at large radii. The variation is mostly due to the different gas masses of each galaxy, largely the result of the variation in initial halo mass. Regions of galaxies that are gravitationally unstable have similar $\Sigma \approx \Sigma_{\text{GI}}$, since Σ_{GI} varies only weakly with accretion rate (see Section 4.1). As a consequence, the radii over which the galaxy is gravitationally unstable are just a matter of how far the gas needs to be pushed away from where it arrives to maintain $\Sigma \approx \Sigma_{\text{GI}}$.

By low redshift, the galaxies have become remarkably similar at large radii but with more than an order-of-magnitude variation near the centre. At large radii, the disc tends to be gravitationally unsta-

ble, but in contrast to the high-redshift case, these galaxies all have the same halo mass and so are quite similar in terms of the available gas budget. Meanwhile at small radii, some galaxies, namely those with a recent burst of accretion, are still gravitationally unstable and so exhibit the same $1/r$ profiles seen at high redshift, while others have stabilized and are in an equilibrium between infalling gas and SF. Thus, GI transport greatly magnifies the different accretion rates, causing a wide range of column densities near the centre of the galaxy, but at the same time GI enforces remarkable similarity at large radii.

Whether the galaxies are in equilibrium is shown explicitly in Fig. 9. As with the fiducial model, the ensemble of discs tends to equilibrate from the inside out. The most remarkable difference is the significant fraction of galaxies which are out of equilibrium, not because they are building up gas, but because they are burning through excess gas. These are galaxies which had a burst of accretion followed by a lull. Most galaxies in our stochastic sample are in this state because of the lognormal distribution of accretion rates, which vary on time-scales that are typically short compared to the depletion time. At any given time, a galaxy is therefore likely to be accreting gas slowly but still working through gas that was accreted in a recent burst.

3.3 Comparison with observations

Using high-resolution and high-sensitivity data to infer the H I and H₂ distributions in nearby spiral galaxies, Bigiel & Blitz (2012) found that these galaxies have neutral gas surface density profiles well approximated by a simple exponential,

$$\Sigma_{\text{UP}} = 2.1 \Sigma_{\text{tr}} e^{-1.65r/r_{25}}. \quad (27)$$

Here Σ_{tr} and r_{25} are empirical quantities derived from the data, respectively, the surface density at which a particular galaxy has $\Sigma_{\text{H I}} = \Sigma_{\text{H}_2}$ and the radius of the 25 mag arcsec⁻² *B*-band isophote. To compare to our simulations, we need to determine these quantities in our own simulated data. We can find Σ_{tr} in our simulations by searching for the location where $f_{\text{H}_2} = 0.5$. In our model this is determined by the Krumholz et al. (2009) formula, in which this transition surface density is set by the metallicity. The value we should use for r_{25} is somewhat more ambiguous. *B*-band luminosities are, roughly speaking, set by the SFR averaged over at least gigayear time-scales, and the exact luminosity derived for a particular SF history is somewhat model dependent. To avoid this issue,

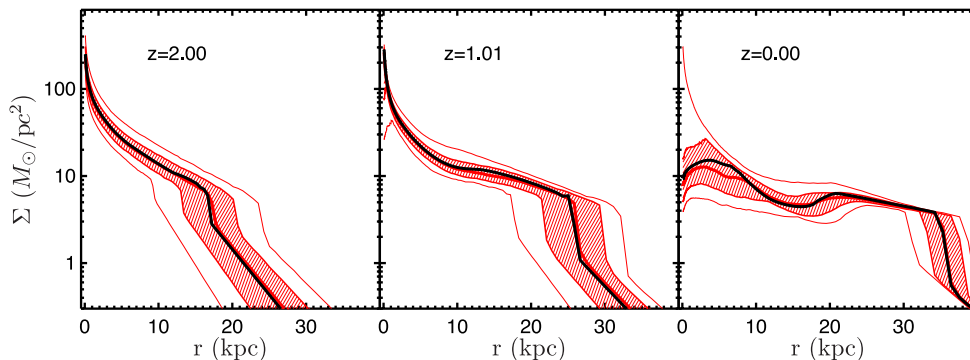


Figure 8. The radial surface density profile of gas for different redshifts. The black line shows the fiducial model, and the red lines show the median and central 68 per cent (shaded) and 95 per cent of the models with stochastic accretion histories. The variation between surface density profiles at a given radius and time depends mostly on whether the galaxy is gravitationally unstable there. The variation in external accretion rate is largely responsible for the differences between galaxies in regions of the disc which are gravitationally unstable. Additional dependences on parameters of the model and physical properties of the galaxy are shown in Appendix C.

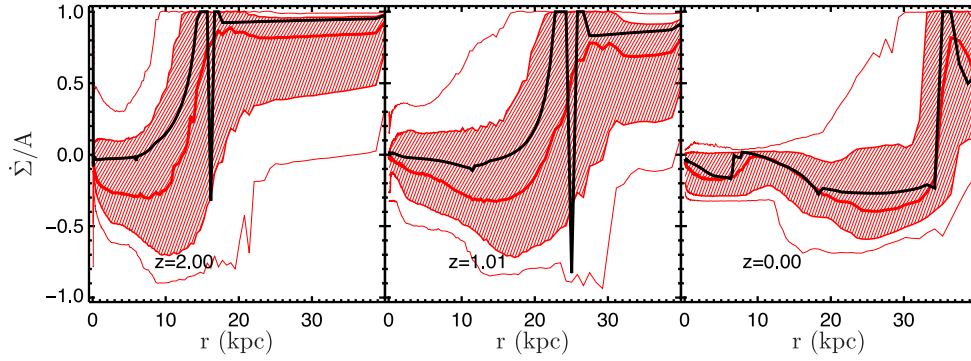


Figure 9. Inside-out equilibration. Here we show, for the smooth accretion model (black) and the median, central 68 per cent (shaded) and central 95 per cent of the stochastically accreting ensemble of galaxies (red), the radial distribution of $\dot{\Sigma}/A$, where A is the sum of the absolute value of each term contributing to $\dot{\Sigma}$. Values of $\dot{\Sigma}/A$ near 1 or -1 indicate that the surface density is changing entirely due to a single term in the equation, while values near zero mean terms of opposing sign are cancelling and the surface density is close to equilibrium. Equilibration occurs from inside out, though significant deviations from equilibrium are possible – in fact the typical galaxy is in a low-accretion-rate state and burning through the gas from a past accretion event. Galaxies are also out of equilibrium at large radii where the gas is mostly atomic and hence SF is slow.

we note that if the universal profile is correct, it can be written just as well

$$\Sigma_{\text{UP}} = 2.1 \Sigma_{\text{tr}} \exp(-0.74r/r_{\text{tr}}), \quad (28)$$

where r_{tr} is the radius at which $f_{\text{H}_2} = 0.5$. This is because $\Sigma_{\text{UP}} = \Sigma_{\text{tr}}$ at $r = r_{\text{tr}} = 0.45r_{25}$. In this way, we avoid the modelling uncertainty in converting between an SF history and a B -band luminosity, and the uncertainty in our SF prescription at low surface densities, or equivalently the uncertainty in the value of f_{H_2} .

For each of our galaxies, we can easily compute r_{tr} and Σ_{tr} (Fig. 10), each as a function of time, to construct the corresponding Σ_{UP} (Fig. 11). The agreement is reasonable, within a factor of 2 of the empirical relation at $z = 0$ for most of the simulated galaxies. At large radii, the effects of photoionization may be important – namely the observations are sensitive only to neutral gas, whereas for the low surface densities $\sim 1 \text{ M}_{\odot} \text{ pc}^{-2}$, UV radiation may ionize a significant portion of the gas. As in the observed galaxies, the largest scatter occurs within the central region. We argue that this is a consequence of variations in the accretion histories which allow some galaxies to continue to transport gas to their centres via GI torques, while others have stabilized.

The agreement between Σ and Σ_{UP} is not a trivial consequence of the exponential cosmological accretion profile we use. In particular, the universal profile predicts that the gas surface density profile should have a scalelength equal to $r_{25}/1.65$. Reading off from Fig. 10, we see that $r_{25}/1.65 \sim 12\text{--}18 \text{ kpc}$, whereas in our fiducial model, the scalelength of the exponential accretion only reaches 6.9 kpc at $z = 0$, and is smaller at higher redshift. In other words, the scalelength of the accretion is always substantially smaller than the universal profile scalelength in our simulations. Therefore, SF and GI must be responsible for altering the profile such that we find reasonable agreement with the observations.

4 DISCUSSION

One of the striking results of our models is the equilibrium that develops between different terms in the continuity equation. In retrospect this is not surprising, especially near the centre of the galaxy, where the SF time is short and the accretion rate is high. The former allows SF to quickly adjust to whatever supply of gas is available to it, while high accretion rates mean enough gas can build up to make the disc gravitationally unstable which allows the

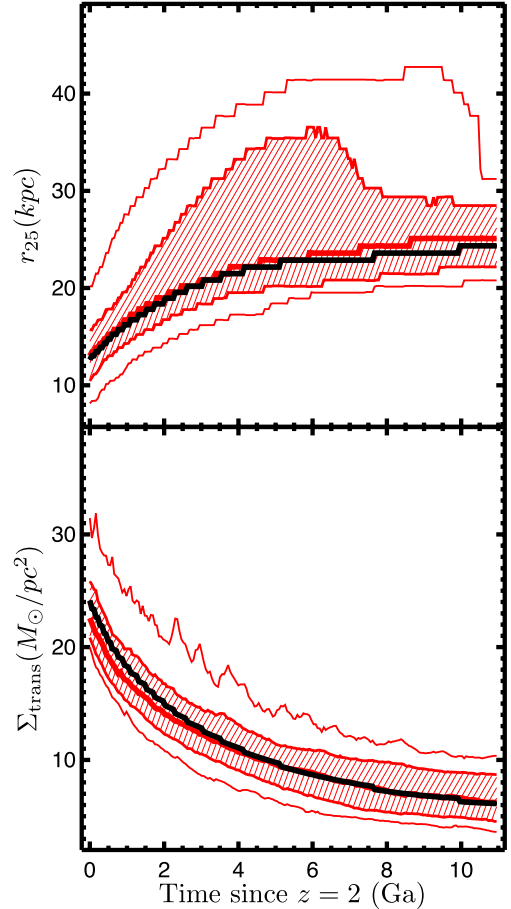


Figure 10. The parameters that define the universal profile. The median, central 68 per cent (shaded) and 95 per cent of the stochastic ensemble of galaxies are shown in red, with the smooth accretion model (black) for comparison. As the metallicity of the galaxies increases, the column density Σ_{tr} at which $f_{\text{H}_2} = 0.5$ falls. As metals build up in the outer disc from local SF, and advection and diffusion from SF nearer to the centre, the radius at which the molecular–atomic transition occurs, r_{tr} , and hence $r_{25} = r_{\text{tr}}/0.45$, steadily increases.

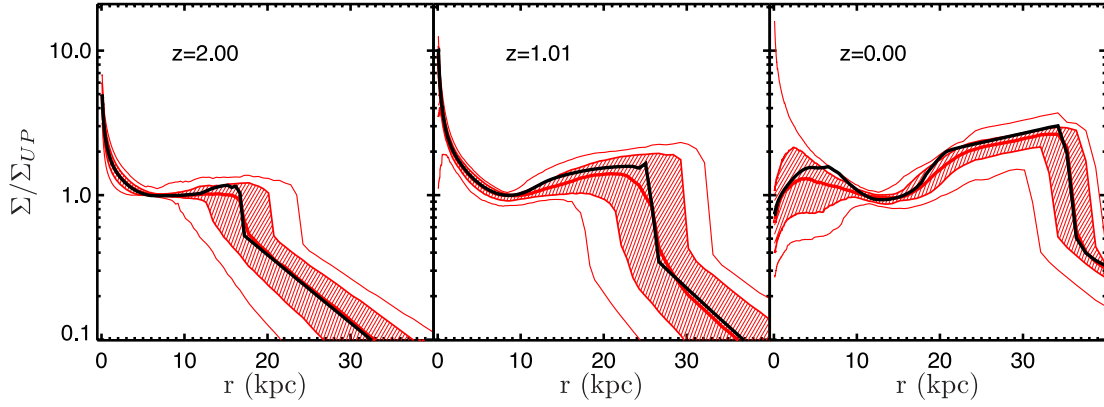


Figure 11. The ratio of the surface density to the universal profile inferred by measuring r_{tr} and Σ_{tr} for each simulation for each time. Black shows the smooth accretion model, while red shows the median, 68 per cent (shaded) and 95 per cent of the distribution for a sample of 400 stochastically accreting galaxies.

disc to redistribute the gas and prevent it from piling up wherever it happens to land.

We discuss, roughly in chronological order, or more to the point, in order of decreasing external accretion rate the implications of this slowly evolving equilibrium. At high redshift, the galaxy experiences the maximum surface density it can obtain via an equilibrium between cosmological accretion and GI transport (Section 4.1). GI transport is eventually shut off via SF (Section 4.2), after which each annulus near the centre of the disc reaches an equilibrium between local gas supply and local SF (Section 4.4).

4.1 Maximum velocity dispersion

Conservation of angular momentum requires that $\partial T / \partial r = -\dot{M} v_\phi (1 + \beta)$ (equation 4). At a particular time, we see that the torque at a given radius can be calculated by integrating

$$T(r) = T(r = r_0) - \int_{r_0}^r \dot{M}(r') v_\phi(r') (1 + \beta(r')) dr'. \quad (29)$$

In our numerical model, the rotation curve and hence v_ϕ and β are fixed in time, as is the inner boundary condition, $T(r = r_0) = 0$. Thus, the torque as a function of radius is exactly mapped to $\dot{M}(r)$. In a steady state, we also know that $\dot{M} < \dot{M}_{\text{ext}}$, since otherwise the surface density would be decreasing somewhere to increase it somewhere else. For the moment, we can specialize to a flat rotation curve for which

$$|T| < T_{\text{max}} \equiv \dot{M}_{\text{ext}} v_{\text{circ}} r. \quad (30)$$

This relation will still hold approximately for somewhat flat rotation curves, since, given the finite supply of new gas \dot{M}_{ext} , typically \dot{M} will be significantly less than \dot{M}_{ext} owing to the effects of SF and outward mass flow, necessary to conserve angular momentum.

We now employ the assumption of local energy balance, i.e. that the value of σ is set by local heating and local cooling with negligible contribution from advection. This assumption is well satisfied in gravitationally unstable regions of our simulations. Under this assumption,

$$\frac{1}{3} \eta \Sigma \sigma^2 \kappa (1 - \sigma_{\text{sf}}^2 / \sigma^2)^{3/2} = \frac{(\beta - 1) v_\phi}{6 \pi r^3} T. \quad (31)$$

Rearranging and approximating $\Sigma \approx \Sigma_{\text{GI}}$,

$$T = 6 r \eta (\beta + 1) v_\phi \sigma_{\text{sf}}^3 (\sigma^2 / \sigma_{\text{sf}}^2 - 1)^{3/2} / ((\beta - 1) G Q_{\text{GI}}). \quad (32)$$

Again specializing to a flat rotation curve and defining the dimensionless number $\mathcal{N} \equiv Q_{\text{GI}} G \dot{M}_{\text{ext}} / 6 \eta \sigma_{\text{sf}}^3 = 1.8 \dot{M}_{\text{ext}} / (1 \text{ M}_\odot \text{ yr}^{-1})$

and imposing the requirement that $-T \lesssim T_{\text{max}}$, we arrive at the condition

$$\sigma \lesssim \sigma_{\text{max}} \equiv \sigma_{\text{sf}} (\mathcal{N}^{2/3} + 1)^{1/2}. \quad (33)$$

Thus, we see that the velocity dispersions of galactic discs are a direct consequence of cosmological accretion and energy equilibrium. We compare this prediction with the maximum measured values of σ in our simulations in Fig. 12. From the decay of the spikes in the bottom panel, we see that the time-scale to reach the steady state assumed in our derivation can be of the order of a Gyr. We also see that the central value of $\max(\sigma) / \sigma_{\text{max}}$ is remarkably close to unity, meaning that σ_{max} is more of an estimate of $\max(\sigma)$ than an upper limit. We note that the measured $\max(\sigma)$ can exceed the predicted maximum slightly even for the smooth accretion model because the assumptions we made in deriving the limit are only approximately true – in particular $\Sigma \approx \Sigma_{\text{GI}}$ becomes a worse approximation as the stellar and gaseous velocity dispersions diverge from each other. Meanwhile the stochastic histories are likely to have $\max(\sigma) > \sigma_{\text{max}}$. This is because σ_{max} depends on the instantaneous accretion rate only, but since the accretion rate changes quickly, the galaxy is likely to still be adjusting to a past burst of accretion.

Even the most extreme galaxies in our population only have $\dot{M}_{\text{ext}} \sim 100 \text{ M}_\odot \text{ yr}^{-1}$, implying $\sigma / \sigma_{\text{sf}} \lesssim 5.7$, while a more typical $z = 2$ galaxy might only have $\dot{M}_{\text{ext}} \sim 10 \text{ M}_\odot \text{ yr}^{-1}$, implying $\sigma / \sigma_{\text{sf}} \lesssim 2.4$. Since of course $\sigma / \sigma_{\text{sf}} \geq 1$, the surface density in gravitationally unstable regions can typically only vary by a factor of a few at a fixed radius and v_{circ} . We note that the velocity dispersions we show here are somewhat smaller than those observed in the SINS galaxies; however, our MW-progenitor models likely have lower masses than the observed galaxies, and we have included no drivers of turbulence besides GI.

The exact way that σ varies between σ_{sf} and σ_{max} (Fig. 13) depends on the particular accretion profile feeding the galaxy (which roughly determines the shape of the $\sigma(r)$ profile), the total amount of gas accreted previously (which sets the outer boundary of the GI region) and SF (which sets the inner boundary). Qualitatively, the velocity dispersion is highest near the centre of the galaxy, since most of the accreted mass arrives near the centre of the galaxy and flows inwards. At low redshift, this is no longer true because the centre of the galaxy becomes gravitationally stable, so the velocity dispersion is forced towards its value from star formation feedback σ_{sf} . The outer edge of the unstable region moves outwards as well, since GI transport will always move some gas outwards to conserve

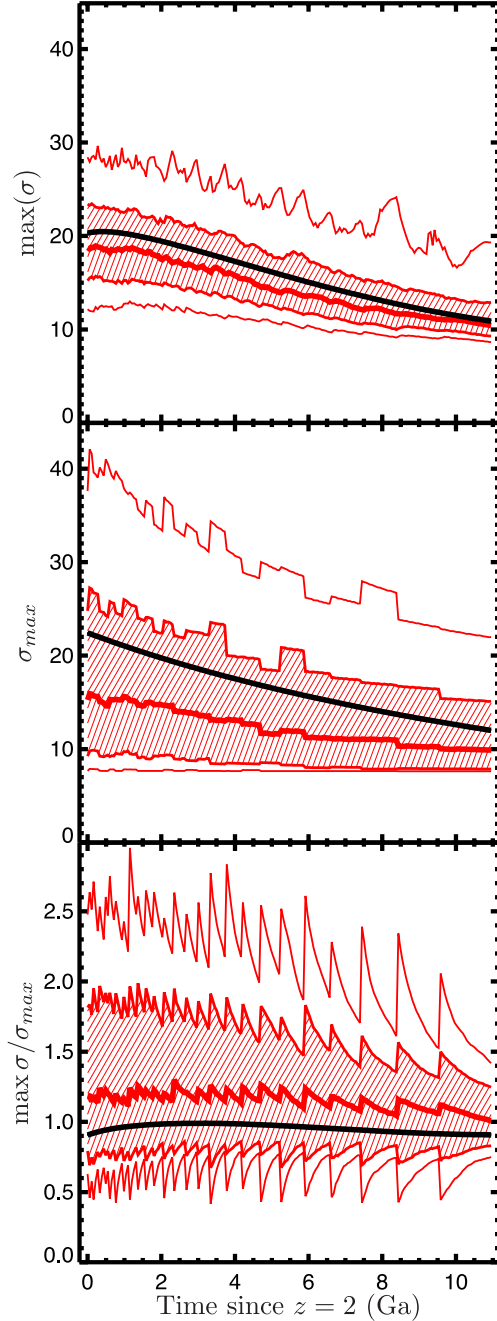


Figure 12. Simulated versus predicted velocity dispersion. Here we show the maximum value of σ measured in our simulations (top), the value of σ_{\max} predicted by equation (33) (middle panel) and their ratio, $\max(\sigma)/\sigma_{\max}$ (bottom). At every change in the accretion rates, the predicted σ_{\max} jumps and it takes some time for each galaxy to adjust to its new accretion rate. As usual the black line shows the fiducial model and the red lines show the median, central 68 per cent and 95 per cent of the distribution for the stochastically accreting models.

angular momentum. This gas is barely touched by SF given the low molecular fraction at large radii, so over cosmological time that gas will continue to build up and the edge of the gravitationally unstable region will march outwards. Stabilization at small radii and destabilization at large radii lead the whole unstable region to move outwards in time. The lower velocity dispersions in the unstable region, the result of the decreasing cosmological accretion

rate, leads to lower characteristic clump masses as estimated by the 2D Jeans mass, $M_J = \sigma^4/G^2\Sigma$, shown in Fig. 14.

The maximum value of σ immediately implies a maximum surface density for a flat rotation curve,

$$\Sigma \lesssim \frac{(3/2)v_{\text{circ}}\sigma_{\text{sf}}}{\pi G r Q_{\text{GI}}} (\mathcal{N}^{2/3} + 1)^{1/2} = \Sigma_{\text{crit}} \frac{\sigma_{\text{max}}}{\sigma_{\text{sf}}}. \quad (34)$$

Since Σ_{crit} for a given model is a fixed function of radius, we immediately see that at a given radius Σ in a gravitationally unstable region will also only vary by a factor of a few. However Σ , unlike σ , may fall below the value corresponding to $\sigma = \sigma_{\text{sf}}$. This typically happens because some process has shut off GI transport (Section 4.2), at which point the disc will equilibrate to a new, lower value of Σ (Section 4.4). We also note that, at least for galactic discs, this maximum column density is likely to be much more restrictive than the one proposed by Scannapieco (2013), which is based upon the requirement that the rate of turbulent energy dissipation must be removable by radiative cooling.

4.2 GI quenching

GI transport shuts off when SF can consume all of the transported gas. To get an idea of where this happens, we can compare the rate at which a region of the disc, between inner radius r_A and outer radius r_B , is resupplied to the rate at which stars are formed within this region,

$$\frac{\dot{M}_{\text{supply}}}{\dot{M}_{\text{SF}}} \approx \frac{\dot{M}(r_B)}{\int_{r_A}^{r_B} 2\pi r (f_R + \mu) \dot{\Sigma}_{\text{SF}}^{\text{SF}} dr}. \quad (35)$$

When this ratio is $\ll 1$, the region in question would easily deplete the gas supply and shut down GI transport, while when it is $\gg 1$, SF makes no difference and gas flows through the region unharmed. To evaluate this ratio, we use the SFR for the Toomre regime, on the grounds that once SF is slow enough to be in the single-cloud regime, it is unlikely to be hugely important anyway and this ratio will just be $\gg 1$. On similar grounds, we can also assume $f_{\text{H}_2} \approx 1$, $\Sigma \approx \Sigma_{\text{GI}}$ and $Q_g \approx (2/3)Q_{\text{GI}}$. In that case, our ratio becomes

$$\frac{\dot{M}_{\text{supply}}}{\dot{M}_{\text{SF}}} \approx \frac{\dot{M}(r_B) G Q_{\text{GI}}^2 \pi (1 + 2Q_{\text{GI}}/3Q_{\text{lim}})^{-1/2}}{\int_{r_A}^{r_B} 36\sqrt{2/3} (f_R + \mu) \epsilon_{\text{ff}} (\beta + 1) v_{\phi}^2 \sigma r^{-1} dr}, \quad (36)$$

and we have restricted ourselves to regions where SF is efficient. In practice, this means that r_B can be at most a few kpc. As usual, for simplicity's sake we will specialize to a flat rotation curve, for which we can easily evaluate the integral in the denominator assuming $\sigma \sim \sigma_{\max}$, leaving $\int_{r_A}^{r_B} r^{-1} dr = \ln(r_B/r_A)$. Recall that σ_{\max} depends on the external accretion to roughly the 1/3 power, so unsurprisingly our ratio will decrease with decreasing \dot{M}_{ext} , meaning that all else equal, for a low enough accretion rate the inner region of the disc will be quenched. The logarithmic dependence on r_B/r_A means that in the Toomre regime of SF, depletion of a fixed gas supply $\dot{M}(r_B)$ is self-similar.

More explicitly, the gas supply is exhausted when $\dot{M}_{\text{supply}}/\dot{M}_{\text{SF}} = 1$, which occurs for

$$r_A = r_B \exp \left(-0.24 v_{220}^{-2} \epsilon_{0.01}^{-1} \frac{1.54}{f_R + \mu} \frac{\dot{M}_1}{(1.5 \dot{M}_1^{2/3} + 1)^{1/2}} \right), \quad (37)$$

where we have neglected the additional scalings with $\sqrt{1 + Q_g/Q_*}$ and we have introduced a few scaled parameters, $v_{220} = v_{\text{circ}}/(220 \text{ km s}^{-1})$, $\epsilon_{0.01} = \epsilon_{\text{ff}}/0.01$ and $\dot{M}_1 = \dot{M}(r_B)/(1 M_{\odot} \text{ yr}^{-1})$.

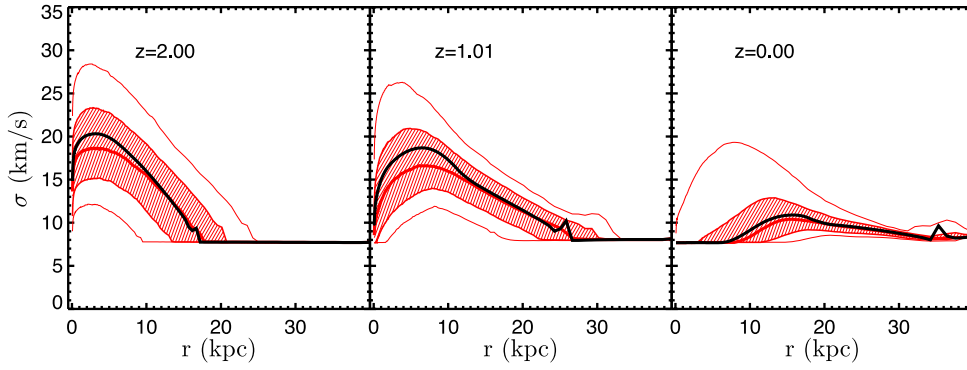


Figure 13. The velocity dispersion distribution for our models as a function of radius and time. The median, central 68 per cent (shaded) and 95 per cent for the stochastic accretion models are shown in red, with the smooth accretion model in black for reference. The high velocity dispersions at high redshift characterize galaxies undergoing ‘violent disc instability’, which manifests itself most dramatically with giant clumps. At low redshift, the turbulent velocity dispersions driven by GI are much lower and occur further out in the disc, largely as a result of the falling accretion rate. Galaxies undergo this transition, from violent, dynamical evolution to a ‘secular’ evolution, smoothly.

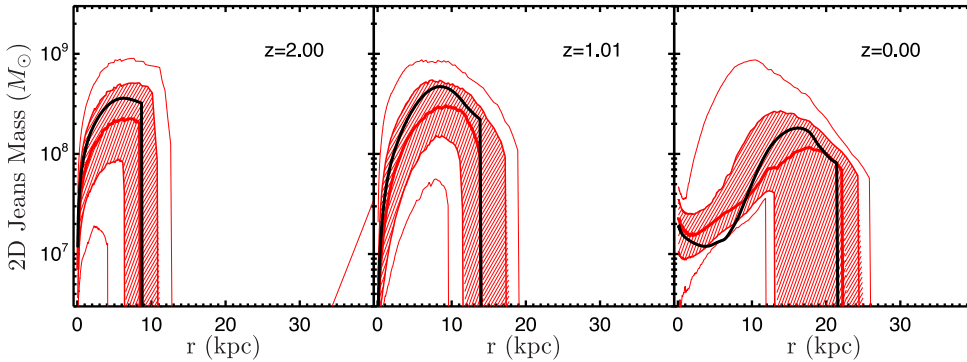


Figure 14. The characteristic size of clumps in the star-forming disc. Here we show the distribution of the 2D Jeans mass in regions where the molecular fraction is larger than $f_{\text{H}_2, \text{min}}$. The typical mass of gravitationally bound clumps decreases with time, and the peak moves outwards in radius. The median, central 68 per cent (shaded) and 95 per cent of the values at each radius and time for the ensemble of stochastically accreting galaxies are shown in red, along with the fiducial model in black.

We caution that this formula is for illustrative purposes only, since v_ϕ and σ are unlikely to be constant. For these values, it turns out that the exponent is fairly close to zero and so relatively insensitive to the exact values. The exponential evaluates to 0.86, 0.64 and 0.43 for $\dot{M}_1 = 1, 4$ and 10.

This is actually somewhat surprising, since in our fiducial model Fig. 6 shows that the mass flux at a few kpc is near $3 M_\odot \text{ yr}^{-1}$, yet the gas reaches the inner edge of the computational domain at $r = 80 \text{ pc}$ easily and GI transport is not shut off until much later. This illustrates the dramatic effect of the rotation curve. The essence of the effect is visible even in equation (37), namely by the time we reach radii well within the turnover in the rotation curve at $r_b = 3 \text{ kpc}$, v_ϕ is appreciably smaller than 220 km s^{-1} , meaning r_A/r_B should be much smaller. The two powers of v_ϕ come from (i) the dynamical time’s proportionality to the SF time – stars form more slowly if the freefall time $\propto r/v_\phi$ is longer, and (ii) the requirement that $Q = Q_{\text{GI}}$, which implies $\Sigma \approx \Sigma_{\text{GI}} \propto v_\phi$ – lower velocities and hence smaller shear mean less gas is required to destabilize the disc. Thus, lowering v_ϕ decreases both the surface density and the SFR for a fixed surface density.

Our simulations use a fixed rotation curve which increases as a power law with index $\beta_0 = 0.5$ near the centre, but galaxies with prominent bulges have what we would term negative values of β_0 , i.e. their rotation curves fall with radius near their centres (see e.g. Dutton 2009). As gas approaches the centre, it would see

higher rotation velocities, which, just the opposite of above, would increase the gas surface density required to maintain GI transport and speed up SF for fixed gas surface density, hence increasing the GI quenching radius r_A . We suggest that this may be a specific physical mechanism for morphological quenching (Martig et al. 2009). In our estimation, the formation of a bulge acts to quench the innermost regions of the galaxy by shutting off GI transport through the increase in v_ϕ , but other factors contribute, namely the available supply of gas \dot{M} and the radius at which stars begin to form efficiently in a galaxy, r_B . We also note that in our model this quenching is not caused by an increase in Q – the increase in Q and the decrease in SFR are both caused by the shutdown of GI transport.

4.3 The growth of bulges

Disc instabilities have of late been invoked to explain the growth of spheroids and AGN activity (Dekel et al. 2009b, 2013; Bournaud et al. 2011). Our fiducial choice of parameters certainly funnels gas to the very centres of our model galaxies at a rate of the order of solar masses per year until $z \sim 0.5$. We caution though that these results depend on our choice of rotation curve, and in particular the rotation curve at the very centre of the galaxy. Nonetheless, we can measure the growth of bulges in our simulations.

There are a number of components which we include in the bulge mass,

$$M_B(t) = \int_0^t \left(\dot{M}_*(r=r_0) + \frac{f_R}{\mu + f_R} (\dot{M}(r=r_0) + \int_0^{r_0} 2\pi r \dot{\Sigma}_{\cos} dr) \right) dt + \int_{r_0}^{r_g} 2\pi r (\Sigma_* - \Sigma_{*,\text{exp}}) dr. \quad (38)$$

Starting from $M_B(z = z_{\text{start}}) = 0$, mass enters the bulge a number of ways. First, there is the mass of stars which migrate off the inner boundary of the computational domain r_0 . Secondly, there is the gas which does the same, which we assume will quickly form stars. Thirdly, there is gas which, according to our cosmological accretion profile, would accrete within the inner boundary. Lastly, there are stars that are still within the computational domain, but which are in excess of an exponential stellar surface density profile extrapolated inwards from larger radii. We sum all of these components, reducing the gaseous terms by $f_R/(f_R + \mu)$ to account for the fact that for every unit mass of stars formed, only f_R will remain in remnants and $f_R + \mu$ will be lost from the gas supply. The exponential fit $\Sigma_{*,\text{exp}}$ is found by

$$\log \Sigma_{*,\text{exp}} = \log \Sigma_*(r_g) + r \frac{m \log(\Sigma_*(r_g)/\Sigma_*(r_g - \Delta r))}{\Delta r}, \quad (39)$$

where $r_g = 1.5r_{\text{acc}}(z)$ is the location at which we will fit the local exponential slope, Δr is the width of one cell and m is initially unity. The value of m is gradually reduced until $\Sigma_{*,\text{exp}} < \Sigma_*$ at every radius interior to r_g (typically $m = 1$ satisfies this condition immediately). This method may overestimate the bulge-to-total (BT) ratio if the stellar profile increases slower than an exponential towards the centre, while if the profile is rising faster than an exponential near r_g , the contribution to the bulge may be underestimated. The stellar surface density profiles are, however, quite exponential within the star-forming region and far from the bulge, likely owing to the mechanism proposed by Lin & Pringle (1987), so this is a reasonable if imperfect estimate. In practice, the flow of gas across the inner boundary, $\dot{M}(r = r_0)$, is the largest of the four terms by a factor of a few, followed by the excess above the exponential.

The growth of bulges measured by the BT ratio, with the bulge mass estimated by equation (38), is shown in Fig. 15. Although GI funnels gas to the centres of these galaxies, our simulations have SF efficient enough and a mass loading factor large enough, that the BT ratios tend to lie near 1/3, a fairly reasonable value for MW-mass galaxies. The trend with redshift seems to be a steep rise between $z = 2$ and 1, followed by a very gradual decrease from $z = 1$ to 0. This may be attributable to the efficient action of GI at high redshift and its subsequent quenching at lower redshift. Moreover, it is clear that galaxies for which GI transport is important at $z \sim 2$ need not end up as bulge-dominated galaxies at $z = 0$. These specific numbers are sensitive to both the angular momentum distribution of infalling gas and the parameters which influence SF, and hence GI quenching, near the centre of the galaxy. The galaxies in our sample all have the same accretion scalelength at $z = 0$, but if we include a 0.4 dex scatter in this parameter, comparable to the scatter in spin parameters observed for dark matter haloes in N -body simulations (Bullock et al. 2001), the central 95 per cent of $z = 0$ BT ratios for those galaxies stretches from 0.05 to 0.72.

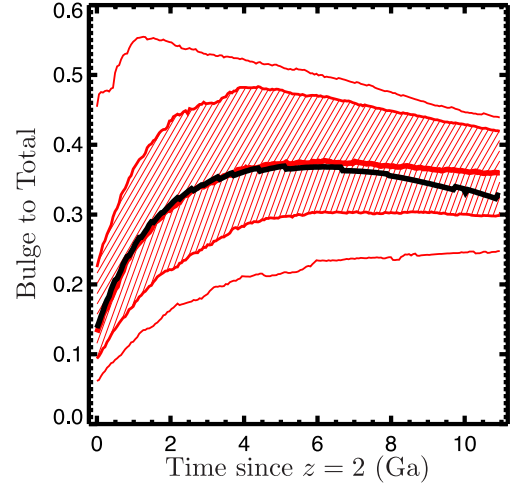


Figure 15. Estimated BT ratio of the stellar profile. Both the fiducial model (black) and the stochastic ensemble (red) follow similar trends, growing their bulges through GI transport at high redshift, then forming stars preferentially in the disc since $z = 1$.

4.4 Equilibrium between accretion and SF

The SF law has two regimes, so naturally there are two profiles where $\dot{\Sigma}_{\cos} = (f_R + \mu)\dot{\Sigma}_*^{\text{SF}}$. The simplest case is the single-cloud regime, defined by a constant molecular depletion time ~ 2 Gyr. In this regime,

$$\Sigma = \dot{\Sigma}_{\cos}(f_R + \mu)^{-1}\epsilon_{\text{ff}}^{-1}f_{\text{H}_2}^{-1}t_{\text{SC}}. \quad (40)$$

This equation is typically not applicable, however, since the outer regions of the disc in the single-cloud regime tend to still be gravitationally unstable even at $z = 0$.

Where SF tends to make a large impact is in the centre of the galaxy. In particular, once SF exhausts the mass flux from GI transport (see the previous section), the supply of gas quickly forms stars until SF equals the local rate of accretion. This equilibrium is local, in that it occurs independently at each radius, since gas is not being transported between radii. The equilibrium picks out a specific value of Σ , such that $\dot{\Sigma}_{\cos}$ (imposed externally) is roughly equal to $\dot{\Sigma}^{\text{SF}}$ (largely determined by Σ). By the time the disc reaches low redshift, we can assume $Q_g \gg 1$ in this region, so if $f_{\text{H}_2} \sim 1$ we can calculate that in the central regions of these galaxies,

$$\Sigma_{\text{eq}} = \left(\frac{3\pi\dot{\Sigma}_{\cos}^2 Q_{\text{lim}}\sigma_{\text{sf}}}{32\epsilon_{\text{ff}}^2 f_{\text{H}_2}^2 \kappa G(f_R + \mu)^2} \right)^{1/3} \approx 16 \frac{\text{M}_{\odot}}{\text{pc}^2} \frac{\dot{\Sigma}_{\cos,0.01}^{2/3}}{(f_R + \mu)^{2/3} f_{\text{H}_2}^{2/3}} r_1^{1/3} v_{220}^{-1/3} \sigma_{\text{th},8}^{1/3} \epsilon_{0.01}^{-2/3}. \quad (41)$$

We have used typical values of $\dot{\Sigma}_{\cos,0.01} = \dot{\Sigma}_{\cos}/(0.01 \text{ M}_{\odot} \text{ kpc}^{-2} \text{ yr}^{-1})$, $r_1 = r/(1 \text{ kpc})$ and $\sigma_{\text{th},8} = \sigma_{\text{sf}}/(8 \text{ km s}^{-1})$. Note that other sources of gas may be added to $\dot{\Sigma}_{\cos}$, although if they depend on the SFR (e.g., for a galactic fountain), the form of the solution will be a bit different. The assumed accretion rate corresponds closely to the redshift zero value for the smooth accretion history model, and the numerical value of Σ_{eq} , despite the approximations made, agrees quite well with the simulation. We see that as long as $\dot{\Sigma}_{\cos}$ is sufficiently flat, as is the case for an exponential on radial scales much less than the scalelength, the value of Σ_{eq} will have a moderate increase with radius. This relation will break down if radial transport of gas is operating, and if f_{H_2} is appreciably smaller than unity, there will be an implicit

dependence on Σ_{eq} on the right-hand side, since f_{H_2} is a function of Σ (and Z).

We saw in Section 3.1.3 that in our smooth accretion model, the inward mass flux from GI transport is exhausted beginning around $z = 0.5$, after which the central gas surface density is rapidly depleted by SF. We refer to this process as ‘GI quenching’. When GI transport is active, it essentially collects cosmological infall from all radii and sends most of that gas inwards and some outwards. This can concentrate most of the SF in the centre of the disc, i.e. gas does not form stars at the location it arrives, but in the centre of the galaxy. When GI transport is shut off, the centre of the galaxy loses this vast supply of gas virtually instantaneously. The surface density falls from $\Sigma \approx \Sigma_{\text{GI}} \propto 1/r$ to $\Sigma \approx \Sigma_{\text{eq}} \propto r^{1/3} \dot{\Sigma}_{\text{cos}}^{2/3}$ in a few depletion times, which may be significantly faster than 1 Gyr (Fig. 4).

We have found that even for large values of an α viscosity (see Appendix C), and even for a rotation curve quite favourable for transporting gas to the central regions of galaxies, the supply of gas to the central regions of galaxies at $z = 0$ via transport through the disc is negligible for a large fraction of the galaxy population. Moreover, gas within this region is unable to move any significant distance radially via these mechanisms. Therefore, the equilibrium which develops there is a balance between the *local* SF in some annulus and the *local* gas supply. In our model this comes from cosmological infall, but it could in principle also come from SN-induced accretion (Marinacci et al. 2010; Hobbs et al. 2013) or gas recycling from old stellar populations (Leitner & Kravtsov 2011). Therefore, we suggest that measuring the SFR and profile in the centres of local galaxies with low SFRs should directly determine the rate and profile with which those particular regions (regardless of the rest of the galaxy) are being supplied with cold gas.

5 SUMMARY

We have explored the evolution of an ensemble of typical disc galaxies with MW-like masses over the past 10 Gyr of cosmic history, with the aim of understanding what sets their surface density profiles.

In our model, discs begin their life at high redshift as exponential and gravitationally unstable in the vicinity of the initial exponential scalelength. This is a somewhat artificial initial condition, but by $z = 2$ (the simulations are started at $z = 2.5$), the gas has had sufficient time to migrate inwards and the discs become gravitationally unstable interior to the accretion scalelength. As more gas is added, the gravitationally unstable region spreads outwards. In this gravitationally unstable state, accreted gas at fairly large radii (of the order of the accretion scalelength) is funnelled towards the centre of the disc where the high surface densities and short dynamical times allow for efficient SF. Eventually, the cosmological accretion rate falls off and the supply of inflowing gas can be consumed by SF before the gas reaches the centre of the galaxy. At this point, the gas transport is shut off and the region of the galaxy interior to this point is quenched, with SF balancing only the local supply of gas. The main lessons we can draw from these results are as follows:

(i) The surface density at every radius is set by a slowly evolving equilibrium. In general, this is a balance among the three terms in the continuity equation: cosmological accretion, SF and GI transport. In this paper, we have described the properties of the disc when each pair of those terms is in balance.

(ii) At a given time, a galaxy will tend to have the following progression of regions, from outside inwards. First there is an out-

of-equilibrium, low column density region, where gas is building up from cosmological accretion but is not yet gravitationally unstable. Next, the galaxy is in equilibrium between infalling material and GI transport. Further in, SF takes up an increasing share of the responsibility for balancing incoming accretion – at this point all three terms in the continuity equation are important. Eventually, SF is so efficient that it outstrips the direct supply of gas and can only be balanced by GI transport from larger radii. Finally, if SF can use up the entire supply of GI-transported gas, there is a quenched region at the centre of the galaxy where SF balances only the direct accretion on to that radius.

(iii) If a region is gravitationally unstable, its gas kinetic energy will equilibrate on a dynamical time-scale, with local heating by GI-driven torques balancing cooling by turbulent dissipation. In a high surface density region where SF is efficient because of the high molecular fraction and short freefall times, SF can equilibrate with its gas supply within a few Gyr. The centres of galaxies, where both GI transport operates (at least at high redshift) and stars form efficiently, will therefore generically equilibrate first. Thus, galaxies equilibrate from the inside out.

(iv) In equilibrium, new accretion must be balanced by the available sinks: SF (plus galactic winds) and transport through the disc. Even at radii where SF is inefficient, GI transport alone is sufficient to balance accretion. GI transport operates through torques which redistribute angular momentum, allowing gas to be removed from where it accretes. To balance the accretion rate, the gas must lose angular momentum in proportion to the accretion rate, so in a steady state the accretion rate specifies the torque. The heating caused by these torques is balanced by turbulent dissipation. The turbulent dissipation rate is proportional to the kinetic energy in the gas, so this balance picks out a velocity dispersion. In summary, the mass flux sets the torque and hence a dissipation rate, which in turn sets the velocity dispersion, so the cosmological accretion rate sets the velocity dispersion.

(v) In general, both the inner and outer boundaries of the gravitationally unstable region move outwards in time. The inner edge moves outwards through a process we call GI quenching. As the cosmological infall rate drops, SF near the centre of the galaxy becomes capable of consuming all of the mass moving inwards via GI transport. If all of the gas is consumed on its way towards the centre of the disc, any part of the disc at smaller radii will be deprived of this large supply of gas. This picture is supported by a number of observational studies, including the depletion of gas near the centres of green valley galaxies (Fang et al. 2012), the link between quenching and a large inner surface density of stars (Cheung et al. 2012; Fang et al. 2013), and the rings of SF and centrally peaked Q observed in gas-rich high-redshift discs (Genzel et al. 2013). SF at a particular radius in this quenched region can only be supplied by whatever cold gas is arriving at that particular radius. In our model this is exclusively from direct accretion from the IGM, but there are other plausible sources.

(vi) The process of GI quenching becomes more effective at higher rotational velocities, which increase the SFR. Massive bulges increase the rotational velocity near the centre of a galaxy, so we propose that morphological quenching may occur through the following physical channel: GI transport moves gas to the centre of a galaxy forming a bulge, the central concentration of matter increases the rotational velocity, GI transport is quenched by the increased SFR (and the decreasing cosmological accretion rate) and so the SF in the central region drops dramatically as its gas supply is removed. The value of Q and Q_{gas} will rise as the gas surface density drops to its new, much lower, equilibrium value. This is distinct from the

mechanism proposed by Martig et al. (2009), wherein they claim that the formation of a spheroid removes the stellar disc and causes the gas disc to stabilize and hence SF to cease. In our model, the self-gravity of the stars has very little effect on the gas because the stars are assumed to be separately self-regulated to a fixed Q_* . Both models predict a rise in Q and a drop in the SFR; in our model, both of these are effects of the shutoff in GI transport (which may be hurried by an increased circular velocity from the formation of a bulge), whereas in Martig et al. (2009), Q increases through the removal of the stars' contribution to the self-gravity to the disc, which then causes the SFR to drop.

(vii) The growth of bulges in our simulations occurs primarily through GI transport of gas from the scale on which it is accreted to the centres of galaxies where it forms stars efficiently. Our galaxies all have the same $z = 0$ halo mass and accretion scalelength, and we recover a relatively narrow range of BT ratios around 0.3–0.4. If we use a more realistic scatter in accretion scalelength of 0.4 dex, the variety of BT ratios increases dramatically.

(viii) Our simulations show that at $z = 0$, some galaxies will be gravitationally unstable at radii $\lesssim 3$ kpc, while others will have undergone GI quenching. The surface density at small radii can therefore vary by an order of magnitude from galaxy to galaxy. This variability at small radii is in fact observed in the neutral gas profiles of nearby galaxies studied by Bigiel & Blitz (2012). Fundamentally, we predict that this variability is the result of variance in the cosmological accretion rate from galaxy to galaxy, which in turn determines whether the galaxy has undergone GI quenching. Another consequence of this variability is that some galaxies – those which have undergone GI quenching – will have a peak in their SF surface density in a ring. This may explain so-called ring galaxies without invoking a recent merger or bar-induced transport.

(ix) The outer edge of the gravitationally unstable region expands as more mass falls on to the galaxy. This is because some fraction of the accreted material will move to larger radii until it runs into the edge of the gravitationally unstable region, where it piles up until the disc at that radius also becomes gravitationally unstable.

(x) Although we have emphasized the equilibration of galaxies, we also observe situations where some region of the galaxy is out of equilibrium (meaning that the surface density is changing at a rate $\gtrsim 5$ per cent of the instantaneous accretion rate), even in the gravitationally unstable region, and even if the accretion history is perfectly smooth. This occurs primarily in the outer regions of the galaxy where the depletion time and even the dynamical time can be long enough for the cosmological accretion rate to change significantly – in other words the sinks for gas are in equilibrium with a past accretion rate. Equilibrium also breaks down near the moving boundaries between gravitationally stable and unstable regions – for instance when a new region of the galaxy has just lost its gas supply via GI quenching, it takes a depletion time to burn through the (now stationary) gas and reach a new equilibrium with direct accretion. As the surface density of gas, and hence the molecular fraction, declines with time, even the inner parts of the disc may experience depletion times much longer than 2 Gyr, and they too may drop out of equilibrium.

(xi) The turbulent velocity dispersion of gas in the galaxy falls over time, and the region of the disc subject to GI-driven transport and turbulence moves outwards. This may be interpreted as a smooth transition from violent to secular instability. The high velocity dispersions and shorter dynamical times of gas at small radii and high redshift lead to giant clumps (since the Jeans mass $\propto \sigma^4/\Sigma$) evolving rapidly, while at low redshift the gravitationally unstable region has a much longer dynamical time and is characterized by

lower clump masses. As predicted in the simpler models of Cacciato et al. (2012), the violent disc instability which operates at $z = 2$ no longer operates today in most MW-mass galaxies, but we show that the transition is gradual and the outskirts of the disc remain unstable even at $z = 0$.

We conclude that GI transport is an important driver of disc galaxy evolution. It provides a natural link between MW-like galaxies at the present day and their high-redshift progenitors, and plays a crucial role in determining the structure of disc galaxies.

ACKNOWLEDGEMENTS

The authors would like to thank Eyal Neistein for helpful discussions. JCF is supported by the National Science Foundation Graduate Research Fellowship under Grant Nos DGE0809125 and DGE1339067. MRK acknowledges support from the Alfred P. Sloan Foundation, from the NSF through CAREER grant AST-0955300, and by NASA through a Chandra Space Telescope Grant and through ATP grant NNX13AB84G. AB thanks the UCSC astronomy and astrophysics department for its hospitality during his summer visits. AD and AB acknowledge support from GIF grant G-1052-104.7/2009. AD also acknowledges support by ISF grant 24/12 and by NSF grant AST-1010033.

REFERENCES

- Agertz O., Kravtsov A. V., Leitner S. N., Gnedin N. Y., 2013, *ApJ*, 770, 25
- Balbus S. A., Hawley J. F., 1998, *Rev. Mod. Phys.*, 70, 1
- Bigiel F., Blitz L., 2012, *ApJ*, 756, 183
- Bigiel F., Walter F., Blitz L., Brinks E., de Blok W. J. G., Madore B., 2010, *AJ*, 140, 1194
- Bigiel F. et al., 2011, *MNRAS*, 730, L13
- Birnboim Y., Dekel A., 2003, *MNRAS*, 345, 349
- Bouché N. et al., 2010, *MNRAS*, 718, 1001
- Bournaud F., Elmegreen B. G., Martig M., 2009, *MNRAS*, 707, L1
- Bournaud F., Elmegreen B. G., Teyssier R., Block D. L., Puerari I., 2010, *MNRAS*, 409, 1088
- Bournaud F., Dekel A., Teyssier R., Cacciato M., Daddi E., Juneau S., Shankar F., 2011, *ApJ*, 741, L33
- Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, *ApJ*, 555, 240
- Burkert A. et al., 2010, *MNRAS*, 725, 2324
- Cacciato M., Dekel A., Genel S., 2012, *MNRAS*, 421, 818
- Caldú-Primo A., Schruha A., Walter F., Leroy A., Sandstrom K., de Blok W. J. G., Ianjamasimanana R., Mogotsi K. M., 2013, *AJ*, 146, 150
- Carlberg R. G., Sellwood J. A., 1985, *ApJ*, 292, 79
- Ceverino D., Dekel A., Bournaud F., 2010, *MNRAS*, 404, 2151
- Cheung E. et al., 2012, *ApJ*, 760, 131
- Cresci G. et al., 2009, *MNRAS*, 697, 115
- Danovich M., Dekel A., Hahn O., Teyssier R., 2012, *MNRAS*, 422, 1732
- Davé R., Oppenheimer B. D., Finlator K., 2011, *MNRAS*, 415, 11
- Dekel A., Birnboim Y., 2006, *MNRAS*, 368, 2
- Dekel A. et al., 2009a, *Nature*, 457, 451
- Dekel A., Sari R., Ceverino D., 2009b, *MNRAS*, 703, 785
- Dekel A., Zolotov A., Tweed D., Cacciato M., Ceverino D., Primack J. R., 2013, *MNRAS*, 435, 999
- Diemer B., More S., Kravtsov A. V., 2013, *ApJ*, 766, 25
- Dutton A. A., 2009, *MNRAS*, 396, 121
- Dutton A. A., van den Bosch F. C., 2009, *MNRAS*, 396, 141
- Dutton A. A., van den Bosch F. C., Dekel A., Courteau S., 2007, *ApJ*, 654, 27
- Elmegreen B. G., 2011, *MNRAS*, 737, 10
- Elmegreen B. G., Burkert A., 2010, *MNRAS*, 712, 294
- Elmegreen D. M., Elmegreen B. G., Hirst A. C., 2004, *ApJ*, 604, L21

Elmegreen D. M., Elmegreen B. G., Rubin D. S., Schaffer M. A., 2005, *ApJ*, 631, 85

Fang J. J., Faber S. M., Salim S., Graves G. J., Rich R. M., 2012, *ApJ*, 761, 23

Fang J. J., Faber S. M., Koo D. C., Dekel A., 2013, *ApJ*, 776, 63

Faucher-Giguere C., Keres D., Ma C., 2011, *MNRAS*, 417, 2982

Forbes J., Krumholz M., Burkert A., 2012, *ApJ*, 754, 48 (F12)

Förster Schreiber N. M. et al., 2009, *MNRAS*, 396, 1364

Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, *MNRAS*, 409, 515

Fu J. et al., 2013, *MNRAS*, 434, 1531

Gabor J. M., Bournaud F., 2013, preprint ([arXiv:1310.1923](https://arxiv.org/abs/1310.1923))

Gammie C. F., 2001, *ApJ*, 553, 174

Genel S., Dekel A., Cacciato M., 2012, *MNRAS*, 425, 788

Genzel R. et al., 2011, *ApJ*, 733, 101

Genzel R. et al., 2013, preprint ([arXiv:1310.3838](https://arxiv.org/abs/1310.3838))

Guedes J., Callegari S., Madau P., Mayer L., 2011, *ApJ*, 742, 76

Hobbs A., Read J., Power C., Cole D., 2013, *MNRAS*, 434, 1849

Hoffmann V., Romeo A. B., 2012, *MNRAS*, 425, 1511

Holmberg J., Nordström B., Andersen J., 2009, *A&A*, 501, 941

Hopkins P. F., Kereš D., Murray N., 2013, *MNRAS*, 432, 2639

Joung M. R., Mac Low M. M., Bryan G. L., 2009, *MNRAS*, 394, 137

Joung M. R., Putman M. E., Bryan G. L., Fernandez X., Peek J. E. G., 2012, *ApJ*, 759, 137

Kassin S. A., Devriendt J., Fall S. M., de Jong R. S., Allgood B., Primack J. R., 2012, *MNRAS*, 424, 502

Kaviraj S. et al., 2013, *MNRAS*, 429, L40

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, *MNRAS*, 363, 2

Komatsu E. et al., 2009, *ApJS*, 180, 330

Krumholz M. R., Burkert A., 2010, *ApJ*, 724, 895

Krumholz M. R., Dekel A., 2012, *ApJ*, 753, 16

Krumholz M. R., Tan J. C., 2007, *ApJ*, 654, 304

Krumholz M. R., Thompson T. A., 2012, *ApJ*, 760, 155

Krumholz M. R., Thompson T. A., 2013, *MNRAS*, 434, 2329

Krumholz M. R., McKee C. F., Tumlinson J., 2009, *MNRAS*, 393, 216

Krumholz M. R., Dekel A., McKee C. F., 2012, *ApJ*, 745, 69

Leitner S. N., Kravtsov A. V., 2011, *ApJ*, 734, 48

Leroy A. K. et al., 2009, *AJ*, 137, 4670

Lin D. N. C., Pringle J. E., 1987, *ApJ*, 320, L87

Mac Low M. M., Klessen R. S., Burkert A., Smith M. D., 1998, *Phys. Rev. Lett.*, 80, 2754

Mandelker N., Dekel A., Ceverino D., Tweed D., Moody C. E., Primack J., 2013, preprint ([arXiv:1311.0013](https://arxiv.org/abs/1311.0013))

Marinacci F., Binney J., Fraternali F., Nipoti C., Ciotti L., Londrillo P., 2010, *MNRAS*, 404, 1464

Martig M., Bournaud F., 2010, *ApJ*, 714, L275

Martig M., Bournaud F., Teyssier R., Dekel A., 2009, *ApJ*, 707, 250

Mo H. J., Mao S., White S. D. M., 1998, *MNRAS*, 295, 319

Neistein E., Dekel A., 2008, *MNRAS*, 383, 615

Neistein E., Maccio A. V., Dekel A., 2010, *MNRAS*, 403, 984

Nelson E. J. et al., 2013, *ApJ*, 763, L16

Noeske K. G. et al., 2007, *ApJ*, 660, L43

Ocvirk P., Pichon C., Teyssier R., 2008, *MNRAS*, 388, 1326

Oppenheimer B. D., Davé R., Kereš D., Fardal M., Katz N., Kollmeier J. A., Weinberg D. H., 2010, *MNRAS*, 406, 2325

Ostriker E. C., McKee C. F., Leroy A. K., 2010, *ApJ*, 721, 975

Rafikov R. R., 2001, *MNRAS*, 323, 445

Rodighiero G. et al., 2011, *ApJ*, 739, L40

Romeo A. B., 1992, *MNRAS*, 256, 307

Romeo A. B., 1994, *A&A*, 286, 799

Romeo A. B., Falstad N., 2013, *MNRAS*, 433, 1389

Romeo A. B., Wiegert J., 2011, *MNRAS*, 416, 1191

Scannapieco E., 2013, *ApJ*, 763, L31

Schruba A. et al., 2011, *AJ*, 142, 37

Sellwood J. A., Carlberg R. G., 1984, *ApJ*, 282, 61

Shakura N. I., Sunyaev R. A., 1973, *A&A*, 24, 337

Shu F. H. S., 1968, PhD thesis, Harvard University

Springel V. et al., 2005, *Nature*, 435, 629

Stone J. M., Ostriker E. C., Gammie C. F., 1998, *ApJ*, 508, L99

Tacconi L. J. et al., 2010, *Nature*, 463, 781

Tacconi L. J. et al., 2013, *ApJ*, 768, 74

Tinsley B. M., 1980, in *Fundamentals of Cosmic Physics*. Vol. 5, Gordon & Breach, New York, p. 287

Toomre A., 1964, *ApJ*, 139, 1217

van den Bosch F. C., 2002, *MNRAS*, 331, 98

van den Bosch F. C., Swaters R. A., 2001, *MNRAS*, 325, 1017

van Dokkum P. G. et al., 2013, *ApJ*, 771, L35

Walter F., Brinks E., de Blok W. J. G., Bigiel F., Kennicutt R. C., Jr, Thornley M. D., Leroy A., 2008, *MNRAS*, 386, 2563

Wang B., Silk J., 1994, *ApJ*, 427, 759

Wuyts S. et al., 2011, *ApJ*, 738, 106

Yang C.-C., Krumholz M., 2012, *ApJ*, 758, 48

APPENDIX A: CHANGES SINCE F12

Here we explicitly list the changes made to our simulation code. In addition to the items discussed below, our code here differs from that of F12 in our assumed rotation curve, assumed accretion rate, the metallicity evolution equation we use and the SF prescription we use. These changes are detailed in the main text.

A1 Finite volume/explicit mass conservation

The evolution equations for Σ and σ are written here in terms of \mathcal{T} , \dot{M} and $\partial\dot{M}/\partial r$, as opposed to \mathcal{T} , $\partial\mathcal{T}/\partial r$ and $\partial^2\mathcal{T}/\partial r^2$. The terms involving these quantities are mathematically identical, but this version is clearer physically. Moreover, when we solve these equations, we explicitly calculate the flux \dot{M} from cell $i + 1$ to i , via

$$\dot{M}_{i+1/2} = \frac{-1}{v_\phi(r_{i+1/2})(1 + \beta(r_{i+1/2}))} \frac{\mathcal{T}_{i+1} - \mathcal{T}_i}{r_{i+1} - r_i}, \quad (\text{A1})$$

where i 's indicate cell-centred quantities and $i + 1/2$'s are edge-centred. Using these fluxes, the change in surface density of cell i is then

$$\left(\frac{\partial\Sigma}{\partial t}\right)_{\text{transport}} = \frac{\dot{M}_{i+1/2} - \dot{M}_{i-1/2}}{2\pi r_i(r_{i+1/2} - r_{i-1/2})} \quad (\text{A2})$$

so that if mass is transported out of cell $i + 1$, it must reappear in cell i (or $i + 2$). Note that we are using a logarithmic grid, so $r_{i+1/2} = \sqrt{r_i r_{i+1}}$, and v_ϕ and β may be calculated at these values analytically because of our simple formula for the rotation curve.

The reason why this is an improvement is that, written in terms of \mathcal{T} and not \dot{M} , $\partial\Sigma/\partial t_{\text{transport}} \propto \partial^2\mathcal{T}/\partial r^2$. This derivative was computed using a minmod slope limiter, so for example if material attempted to enter or exit a cell from both directions (i.e. the value of $\dot{M}_{i+1/2}$ and $\dot{M}_{i-1/2}$ had opposite signs), $\partial\Sigma/\partial t_{\text{transport}} = 0$, and so the entering mass would be lost or the exiting mass would remain in the cell. For monotonic solutions of \mathcal{T} , this is a small effect, and so only became apparent when mass was added inside the computational domain instead of at its outer edge (see Appendix A3), which meant some regions would have mass flowing outwards.

A2 Treatment of stable regions where $Q > Q_{\text{GI}}$

In our previous work, when $Q > Q_{\text{GI}}$, we solved $f_{\text{transport}} = 0$. In contrast, in this work we simply set $\mathcal{T}_{\text{GI}} = 0$ in those regions. The difference between the two is somewhat subtle. The two treatments would be equivalent if the boundary conditions around the stable region were $\mathcal{T}_{\text{GI, boundary}} = 0$, but this will generally not be the case in our discs because the neighbouring unstable regions will have non-zero torques. Thus, our previous approach would lead to small but

non-zero mass fluxes in stable regions. Our new approach is more consistent with the physical picture we are presenting, namely that radial motion is caused by GI-induced turbulence.

A3 Accretion on to the disc instead of at the outer boundary

In our previous work, the accretion of gas on to the galaxy occurred only at the outer boundary of the galaxy, and the accretion rate was enforced by setting $(\partial \mathcal{T}_{GI}/\partial r)_{r=R} = \dot{M}_{\text{ext}}(t)v_{\text{circ}}$. We have abandoned this approach because it is inflexible and likely to be physically wrong. In particular, if all the gas comes in at $r = R$, then the value of R may strongly affect the results of the simulation, especially for discs where the accretion rate is not large enough to maintain a GI, e.g. low-mass galaxies or galaxies experiencing a lull in their accretion rate. For these galaxies, accretion at large radii leads to an unphysical pileup of gas in the outermost radial cell. Moreover, the hole in the gas distribution which we saw forming at the centre of our simulated galaxy is not a ubiquitous feature in real galaxies, suggesting that a more flexible accretion model might be necessary.

In this work, we still need to specify the boundary conditions at inner and outer edges of the computational domain. We opt for the simplest choice, $\mathcal{T}_{GI}(r = r_0) = \mathcal{T}_{GI}(r = R) = 0$, which should be reasonable so long as R is much larger than the radial scale of the accretion.

APPENDIX B: NEW STELLAR MIGRATION EQUATIONS

To derive the evolution of stars as a result of their migration through the disc, we will assume that stars obey $dQ_*/dt = \Delta Q_*/T_{\text{mig}}(2\pi\Omega)^{-1}$, i.e. that stars will exponentially ‘decay’ to a limiting value of Q_* above which they will be stable to GI, Q_{lim} , on some multiple of the local orbital time (Sellwood & Carlberg 1984; Carlberg & Sellwood 1985). As with the gas, we take the stars to be subject to gravitational torques which will lead to some velocity v_r^* of stars inwards or outwards at each radius such that Q_* approaches Q_{lim} . In analogy with the gas, we derive evolution equations for the stellar surface density and velocity dispersion which depend on this torque. To do so we begin with the continuity equation and the ϕ -component of the Jeans equations, both derived from the collisionless Boltzmann equation. The continuity equation is

$$\frac{\partial \rho_*}{\partial t} + \nabla \cdot (\rho_* \mathbf{v}^*) = 0. \quad (\text{B1})$$

The angled brackets define an average over the distribution function, namely $\langle v_i^* \rangle \equiv \rho_*^{-1} \int v_i^* f d^3 \mathbf{v}^*$. Note that for simplicity we have taken f to be the distribution function of mass rather than number of stars, where all stars are assumed to have the same mass. The ϕ -component of the Jeans equations is

$$\frac{\partial \rho_* \langle v_\phi^* \rangle}{\partial t} + \frac{\partial \rho_* \langle v_r^* v_\phi^* \rangle}{\partial r} + \frac{\partial \rho_* \langle v_\phi^* v_z^* \rangle}{\partial z} + \frac{2\rho_* \langle v_r^* v_\phi^* \rangle}{r} = 0. \quad (\text{B2})$$

As with the gas, we have assumed axisymmetry.

The evolution of the surface density follows almost immediately from integrating the continuity equation in z .

$$\frac{\partial}{\partial t} \int_{-\infty}^{\infty} \rho_* dz = -\frac{1}{r} \frac{\partial}{\partial r} \left(r \int_{-\infty}^{\infty} \rho_* \langle v_r^* \rangle dz \right) - \int_{-\infty}^{\infty} \frac{\partial}{\partial z} \rho_* \langle v_z^* \rangle dz. \quad (\text{B3})$$

We will assume that $\langle v_i^* \rangle$ does not vary much over the scaleheight of the disc, that the disc does not change orientation (so $\langle v_z^* \rangle = 0$), and

that $\rho_* \rightarrow 0$ for large and small values of z , so integrals over z of the z -derivative of a quantity weighted by ρ_* will vanish. Defining $\Sigma_* \equiv \int_{-\infty}^{\infty} \rho_* dz$, we have

$$\frac{\partial \Sigma_*}{\partial t} = -\frac{1}{r} \frac{\partial}{\partial r} (r \Sigma_* \langle v_r^* \rangle) = \frac{1}{2\pi r} \frac{\partial}{\partial r} \dot{M}_*, \quad (\text{B4})$$

where we have defined $\dot{M}_* \equiv -2\pi r \Sigma_* \langle v_r^* \rangle$ to be the inward mass flux of stars through the disc.

To relate this to the torque experienced by the stars, we integrate the ϕ -component of the Jeans equations in z ,

$$\frac{\partial}{\partial t} \Sigma_* \langle v_\phi^* \rangle + \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \int \rho_* \langle v_r^* v_\phi^* \rangle dz = 0. \quad (\text{B5})$$

We now define the quantity $\delta v_i \equiv v_i^* - \langle v_i^* \rangle$, the deviation of a particular velocity at a given point from the mean velocity at that point. As usual, we define $\langle \delta v_i \delta v_j \rangle \equiv \sigma_{ij}^2$, and so $\langle v_r^* v_\phi^* \rangle = \langle v_r^* \rangle \langle v_\phi^* \rangle + \sigma_{r\phi}^2$ (since by construction $\langle \delta v_i \rangle = 0$). Rearranging, we arrive at

$$\begin{aligned} \langle v_\phi^* \rangle \frac{\partial \Sigma_*}{\partial t} + \Sigma_* \frac{\partial \langle v_\phi^* \rangle}{\partial t} + \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \Sigma_* \langle v_r^* \rangle \langle v_\phi^* \rangle \\ = -\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \int \rho_* \sigma_{r\phi}^2 dz. \end{aligned} \quad (\text{B6})$$

Using the continuity equation and multiplying through by $2\pi r^2$ yields the evolution equation for specific angular momentum $j_* \equiv r \langle v_\phi^* \rangle$,

$$2\pi r \Sigma_* \frac{\partial j_*}{\partial t} + 2\pi r \Sigma_* \langle v_r^* \rangle \frac{\partial j_*}{\partial r} = \frac{\partial}{\partial r} \mathcal{T}_*, \quad (\text{B7})$$

where $\mathcal{T}_* \equiv -2\pi r^2 \int \rho_* \sigma_{r\phi}^2 dz$. As with the gas, we assume a slowly varying potential, in which case we have

$$-\dot{M}_* \frac{\partial j_*}{\partial r} = -\dot{M}_* v_\phi (1 + \beta) = \frac{\partial}{\partial r} \mathcal{T}_*, \quad (\text{B8})$$

so it is clear that the time derivative of Σ_* is proportional to the second derivative of the torque. At this point we have also assumed that $\langle v_\phi^* \rangle = v_\phi$, the circular velocity of the gas, so that here, as in e.g. equation (3), $\beta = \partial \ln v_\phi / \partial \ln r$.

To find the evolution of the stellar velocity dispersion, we begin with the collisionless Boltzmann equation,

$$\frac{\partial f}{\partial t} + v_i^* \frac{\partial f}{\partial x_i} - \frac{\partial \psi}{\partial x_i} \frac{\partial f}{\partial v_i^*} = 0. \quad (\text{B9})$$

Next, we multiply through by $v_j^* v_j^*$ and as usual integrate over $d^3 \mathbf{v}^*$. Since ψ , x_i and t are independent of \mathbf{v}^* , we have

$$\begin{aligned} \int v_j^* v_j^* f d^3 \mathbf{v}^* + \frac{\partial}{\partial x_i} \int v_i^* v_j^* v_j^* f d^3 \mathbf{v}^* \\ - \frac{\partial \psi}{\partial x_i} \int v_j^* v_j^* \frac{\partial f}{\partial v_i^*} d^3 \mathbf{v}^* = 0. \end{aligned} \quad (\text{B10})$$

The final term may be integrated by parts,

$$\begin{aligned} \frac{\partial \psi}{\partial x_i} \int v_j^* v_j^* \frac{\partial f}{\partial v_i^*} d^3 \mathbf{v}^* &= -\frac{\partial \psi}{\partial x_i} \int \frac{\partial v_j^* v_j^*}{\partial v_i^*} f d^3 \mathbf{v}^* \\ &= -2 \frac{\partial \psi}{\partial x_i} \int v_i^* f d^3 \mathbf{v}^*, \end{aligned} \quad (\text{B11})$$

while the second term may be expanded by again splitting up $v_k^* = \langle v_k^* \rangle + \delta v_k$, so that

$$\begin{aligned} \langle v_i^* v_j^* v_j^* \rangle &= \langle v_i^* \rangle \langle v_j^* \rangle \langle v_j^* \rangle + \langle v_i^* \rangle \sum_j \sigma_{jj}^2 \\ &\quad + 2 \langle v_j^* \rangle \sigma_{ij}^2 + \rho_*^{-1} \int f \delta v_i \delta v_j \delta v_j d^3 \mathbf{v}^*. \end{aligned} \quad (\text{B12})$$

For simplicity we drop the final term. This should be a reasonable approximation, since even though the δv_i are not necessarily small compared to $\langle v_i^* \rangle$, the integrand contains a quantity which averages to zero, δv_i , multiplied by a positive definite quantity $\delta v_j \delta v_j$. We are therefore reweighting an integral which would vanish for a constant weight and approximating it as zero.

With these two substitutions, we arrive at an equation for the evolution of the specific kinetic plus potential energy of the stars,

$$0 = \frac{\partial}{\partial t} \rho_* \left(\langle v^* \rangle^2 + \sum_i \sigma_{ii}^2 \right) + \nabla \cdot \rho \langle v^* \rangle \left(\langle v^* \rangle^2 + \sum_i \sigma_{ii}^2 \right) + \nabla \cdot (2\rho \langle v^* \rangle \cdot \sigma^2) + 2\rho_* \nabla \psi \cdot \langle v^* \rangle. \quad (\text{B13})$$

Here σ^2 is the tensor with components $\langle \delta v_i \delta v_j \rangle = \sigma_{ij}^2$. The gravitational work term may be replaced via the continuity equation, since

$$\begin{aligned} \nabla \cdot (\rho_* \langle v^* \rangle \psi) &= \psi \nabla \cdot (\rho_* \langle v^* \rangle) + \rho_* \langle v^* \rangle \cdot \nabla \psi \\ &= -\frac{\partial(\rho_* \psi)}{\partial t} + \rho_* \frac{\partial \psi}{\partial t} + \rho_* \langle v^* \rangle \cdot \nabla \psi. \end{aligned} \quad (\text{B14})$$

With this substitution, we can group the terms composing the specific energy together, so that if we define $\mathcal{A} = (\langle v^* \rangle^2 + \sum_i \sigma_{ii}^2 + 2\psi)$, we arrive at

$$\frac{\partial}{\partial t} \rho_* \mathcal{A} + \nabla \cdot \rho \langle v^* \rangle \mathcal{A} + \nabla \cdot (2\rho \langle v^* \rangle \cdot \sigma^2) - 2\rho_* \frac{\partial \psi}{\partial t} = 0. \quad (\text{B15})$$

Before integrating over z , we can use the continuity equation to make one more simplification,

$$\rho_* \frac{\partial}{\partial t} (\mathcal{A} - 2\psi) + \rho_* \langle v^* \rangle \cdot \nabla \mathcal{A} + \nabla \cdot (2\rho_* \langle v^* \rangle \cdot \sigma^2) = 0. \quad (\text{B16})$$

Next we approximate $\mathcal{A} \approx v_\phi^2 + \sum_i \sigma_{ii}^2 + 2\psi$, since the other components of $\langle v^* \rangle$ are small, $\langle v_r^* \rangle$, or zero, $\langle v_z^* \rangle$. We also approximate $\sigma_{\phi\phi}^2 \approx \sigma_{zz}^2$, in accordance with observations in the solar neighbourhood (e.g. Holmberg, Nordström & Andersen 2009). Finally, we will again assume that the potential changes slowly, so that $\partial v_\phi / \partial t = 0$. With these approximations, we have

$$\begin{aligned} 0 &= \rho_* \frac{\partial}{\partial t} (\sigma_{rr}^2 + 2\sigma_{zz}^2) + \rho_* \langle v_r^* \rangle \frac{\partial}{\partial r} (v_\phi^2 + \sigma_{rr}^2 + 2\sigma_{zz}^2 + 2\psi) \\ &\quad + \frac{2}{r} \frac{\partial}{\partial r} r \rho_* (\langle v_r^* \rangle \sigma_{rr}^2 + v_\phi \sigma_{\phi\phi}^2). \end{aligned} \quad (\text{B17})$$

Now we will integrate over z and assume, consistent with our approximation that $\int f \delta v_i \delta v_j \delta v_j d^3 v \approx 0$, that σ_{ii}^2 is roughly constant over a disc scaleheight. Employing the angular momentum conservation equation and assuming $\partial \sigma_{rr} / \partial t \approx 2\partial \sigma_{zz} / \partial t$ (Sellwood, private communication), we arrive at

$$\begin{aligned} \frac{\partial \sigma_{rr}}{\partial t} &= \frac{1}{2\pi r \Sigma_* (\sigma_{rr} + \sigma_{zz})} \left(\frac{v_\phi (\beta - 1)}{r^2} \mathcal{T}_* + \sigma_{rr}^2 \frac{\partial \dot{M}_*}{\partial r} \right. \\ &\quad \left. + \dot{M}_* \left(3\sigma_{rr} \frac{\partial \sigma_{rr}}{\partial r} + 2\sigma_{zz} \frac{\partial \sigma_{zz}}{\partial r} \right) \right). \end{aligned} \quad (\text{B18})$$

This is very similar to equation (3), since the procedures used to derive the two are quite similar. The primary distinction is that here we have split the velocity dispersion into a radial and non-radial component whereas for the gas they are assumed to be identical (a reasonable approximation since the gas is collisional). Besides that the only difference is in the numerical values of the coefficients.

With equation (B18), its counterpart for $\partial \sigma_{zz} / \partial t$, and the continuity equation (equation B4), we can follow a similar procedure as

detailed in Section 2.1 to solve for \mathcal{T}_* . In particular, we can again split the terms which appear in dQ_*/dt into those which contain \mathcal{T}_* and its radial derivatives, and those which do not,

$$\begin{aligned} \frac{dQ_*}{dt} &= \frac{\partial \Sigma_*}{\partial t} \frac{\partial Q_*}{\partial \Sigma_*} + \frac{\partial \sigma_{rr}}{\partial t} \frac{\partial Q_*}{\partial \sigma_{rr}} + \frac{\partial \sigma_{zz}}{\partial t} \frac{\partial Q_*}{\partial \sigma_{zz}} \\ &= f_{\text{transport}}^* \left(\Sigma_*, \sigma_{rr}, \sigma_{zz}, \mathcal{T}_*, \frac{\partial \mathcal{T}_*}{\partial r}, \frac{\partial^2 \mathcal{T}_*}{\partial r^2} \right) \\ &\quad + f_{\text{source}}^* (\Sigma, \sigma, Z, \Sigma_*, \sigma_{rr}, \sigma_{zz}). \end{aligned} \quad (\text{B19})$$

None of the equations used so far in this appendix contribute to f_{source}^* – the only way Q_* can change without transport is via SF, which increases Σ_* and typically reduces σ_{rr} and σ_{zz} , which in turn tends to lower Q_* . For the purposes of computing \mathcal{T}_* , we ignore f_{source}^* and simply solve $f_{\text{transport}}^* = \Delta Q_*/T_{\text{mig}}(2\pi\Omega)^{-1}$ when $Q_* < Q_{\text{lim}}$ and set $\mathcal{T}_* = 0$ otherwise. This allows Q_* to fall significantly below Q_{lim} , though in practice SF is typically slow enough that $Q_* \approx Q_{\text{lim}}$. As with the gas, if this equation yields a solution where $\mathcal{T}_* > 0$, we set $\mathcal{T}_* = 0$ in the offending cell.

APPENDIX C: SENSITIVITY TO PARAMETERS

Thus far we have used only the fiducial parameters, but each one is at least somewhat uncertain (e.g. ϵ_{ff} may vary by a factor of 3 in either direction), or may change for physical reasons (e.g., lower mass haloes will likely have smaller r_{acc} and v_{circ} , and larger μ). To explore the effects of each parameter, we have varied them one at a time from their fiducial values for the smooth accretion history. Fig. C1 shows the $z = 0$ surface density distribution when each parameter is varied. Essentially, all of the models are gravitationally unstable over a wide range of radii and parameter choices. We also show the metallicity distribution for all of these models in Fig. C2. The metallicities are in general hugely sensitive to the parameters, so much so that any attempt to draw a physical conclusion by fitting a metallicity gradient should be treated with extreme caution, since the same metallicity gradient can be produced by varying any number of parameters. We discuss each of the parameters in more detail in the following sections.

C1 Initial conditions – $\alpha_r, f_{g,0}, f_{\text{cool}}, \phi_0$

These parameters, the scaling of the accretion scalelength with halo mass, the initial gas fraction, the fraction of baryons which have cooled into a disc at $z = z_{\text{start}}$ and the initial ratio of stellar to gas velocity dispersion, are almost completely irrelevant for the $z = 0$ surface density distribution. In our framework, the surface density at each radius is set by an equilibrium relation, and so it is unsurprising that the initial conditions are washed out. The exception is that strong evolution of r_{acc} with halo mass, i.e. (probably unrealistically) high values of α_r , leads to smaller gravitationally unstable regions at $z = 0$, since so little mass was accreted directly at large radius.

C2 Rotation curve – $\beta_0, n, r_b, v_{\text{circ}}$

The shape of the rotation curve is controlled by these four parameters – inner power-law slope, the sharpness and location of the turnover from flat, and the overall normalization. The most dramatic effect of changing these parameters is in the inner region of the disc, where different values can change the surface density by an order of magnitude. This is because the rotation curve influences both the surface density in gravitationally unstable regions and the

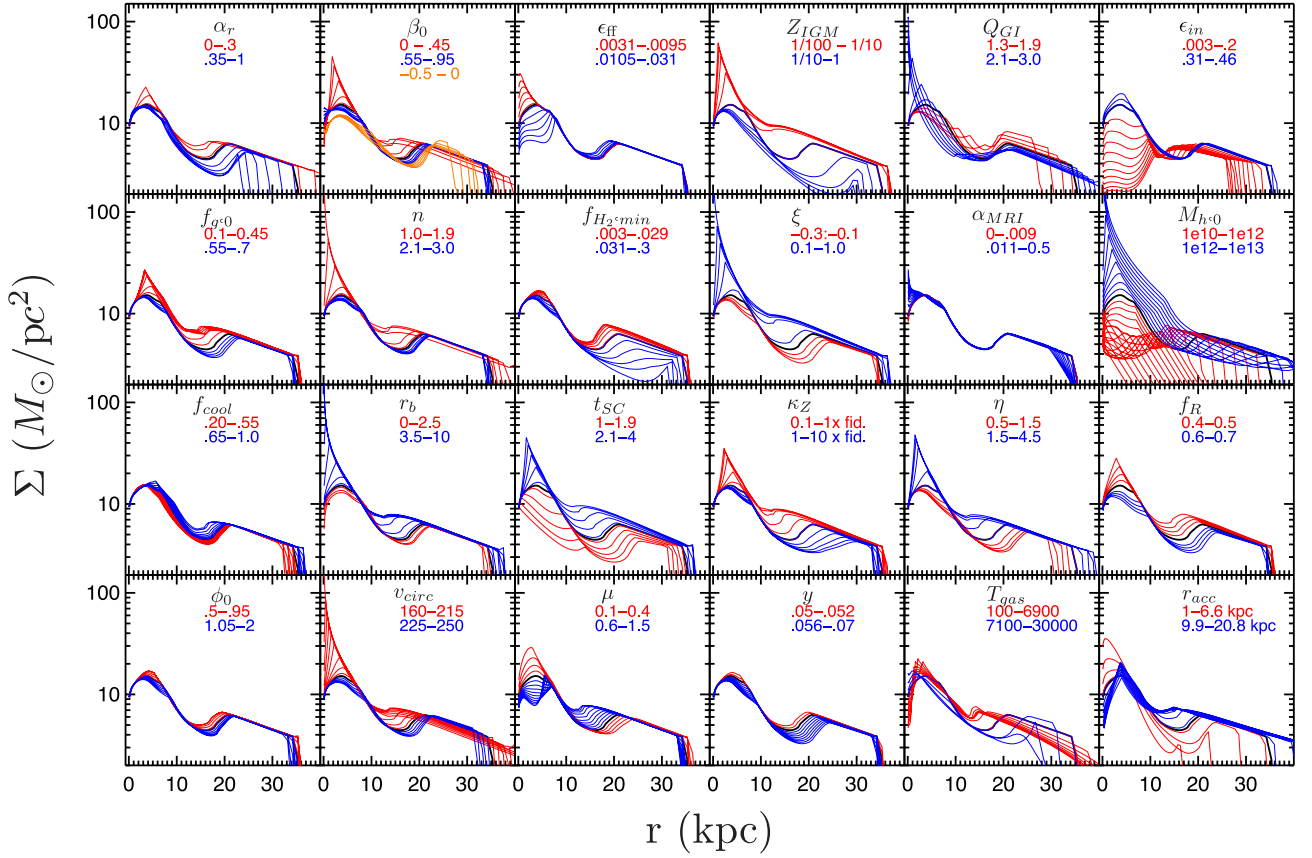


Figure C1. Surface density at $z = 0$. Each pane shows models where the given parameter is varied within the quoted range – red models have lower values of the parameter, blue higher. The models are arranged in six columns according to what the parameters are controlling – from left to right: initial conditions, rotation curve, SF, metallicity, gas transport and gas supply.

SFR in the central region, where $\dot{\Sigma}_*^{\text{SF}} \propto \kappa \propto v_\phi$, and hence has a strong influence on where exactly the disc is able to form stars fast enough to shut off GI transport to the innermost region. Negative values of β_0 have an effect at large radii too. In general, however, the qualitative behaviour of our models is largely insensitive to these parameters except near galactic centres.

C3 Star formation – ϵ_{ff} , $f_{\text{H}_2, \text{min}}$, t_{SC} , μ

Each of these parameters governs the rate at which gas is depleted from the galaxy, either into stars or galactic winds. As we would expect, ϵ_{ff} is important in the inner region of the disc where the SFR is in the Toomre regime, while $f_{\text{H}_2, \text{min}}$ is important in the outer disc where the gas is mostly atomic and hence $\text{SFR} \propto f_{\text{H}_2} = f_{\text{H}_2, \text{min}}$. Note however that the factor of 3 variation in ϵ_{ff} has a much larger effect than the order-of-magnitude variation in $f_{\text{H}_2, \text{min}}$. This is because the surface density is set by different equilibria – in the outer disc the surface density is mostly set by GI, whereas in the inner region the surface density is determined by whether SF has shut off GI transport to the central region or not, which in turn depends strongly on the SF law there. If GI transport has been shut off, then the surface density is set by cosmological infall balancing SF, so a change in the SF law given a fixed infall rate $\dot{\Sigma}_{\text{cos}}$ can have a large effect. The mass loading factor μ affects the rate of mass-loss everywhere in the disc, but again because of the different equilibria, it has a much stronger effect in the inner region. Again, however, we note that the qualitative results, as opposed to the precise numerical values of $\Sigma(r)$, are insensitive to these parameters.

C4 Metallicity – Z_{IGM} , ξ , κ_Z , y

The metallicity of the infalling and initial gas in the disc, the metal enhancement of galactic winds, the metal diffusion coefficient and the yield. The first three strongly influence the metallicity of the disc, as does the yield to a lesser extent. This in turn affects the H_2 fraction when the gas is near its transition surface density [higher (lower) surface densities will have $f_{\text{H}_2} \sim 1$ ($f_{\text{H}_2, \text{min}}$) respectively]. The H_2 fraction then has an effect on the SFR. In general, changes in the parameters which decrease the overall metallicity increase the surface density everywhere by decreasing the rate at which SF is depleting/ejecting the gas.

C5 Transport – Q_{GI} , α_{MRI} , η , T_{gas}

The parameters which control the radial transport of the gas have the potential to strongly affect the surface density, since much of the disc is gravitationally unstable. Q_{GI} and T_{gas} both directly affect Σ_{crit} , namely the minimum surface density for the gas to be gravitationally unstable. Meanwhile η only affects the energy balance in the disc. For all the parameters, the primary difference is in where the GI transport is shut off. Higher Q_{GI} and dissipation rate allow the gas to reach farther towards the centre before being consumed by SF. T_{gas} turns out not to matter all that much, primarily because in the limit of large accretion rates/low velocity dispersion floors, the energy balance is independent of T_{gas} . Perhaps the most dramatic parameter here is α_{MRI} , which has even less effect on the surface density distribution than the initial conditions.

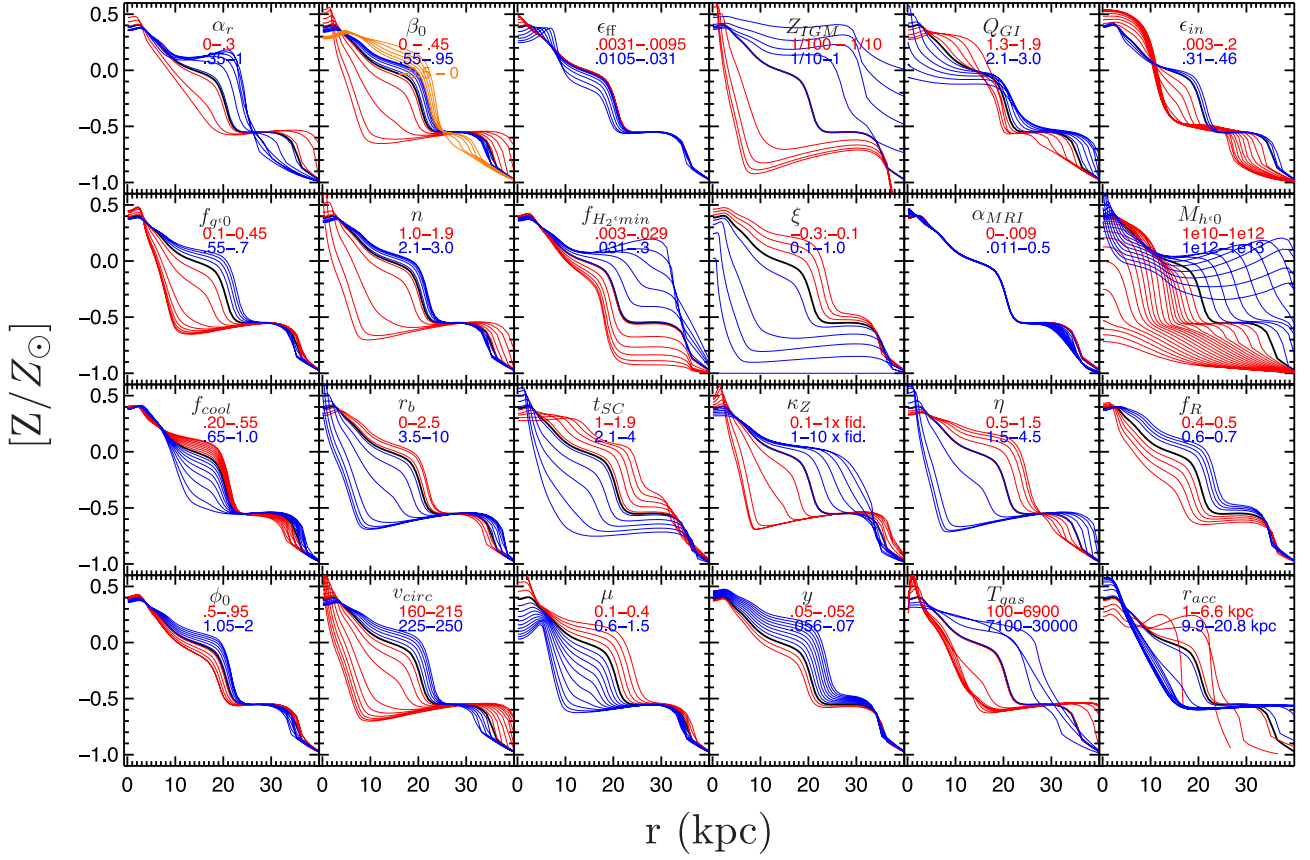


Figure C2. Metallicity for the same models as in Fig. C1.

C6 Gas supply – ϵ_{in} , $M_{\text{h},0}$, f_{R} , r_{acc}

Here we come to the most important parameters in setting the surface density – the quantity and distribution of the gas supply. These parameters, respectively, are the efficiency at $z = 0$ (assuming a fixed efficiency at $z = 2$), the halo mass at redshift zero (where we change only M_{h} and hence the accretion history, but no other parameters), the remnant fraction and finally the accretion scale-length. These models obey the trends one might expect. Less gas

means the region over which the gas is gravitationally unstable is smaller. Parts of the disc beyond this radius retain the exponential character of the accretion profile, and parts of the disc interior have their surface densities set by the balance between local accretion and local SF.

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