

A unified model for galactic discs: star formation, turbulence driving, and mass transport

Mark R. Krumholz,¹★ Blakesley Burkhart,² John C. Forbes² and Roland M. Crocker¹

¹*Research School of Astronomy & Astrophysics, Australian National University, Canberra, ACT 2611, Australia*

²*Harvard-Smithsonian Center for Astrophysics, 60 Garden St, Cambridge, MA 0213, USA*

Accepted 2018 March 28. Received 2018 March 26; in original form 2017 May 30

ABSTRACT

We introduce a new model for the structure and evolution of the gas in galactic discs. In the model the gas is in vertical pressure and energy balance. Star formation feedback injects energy and momentum, and non-axisymmetric torques prevent the gas from becoming more than marginally gravitationally unstable. From these assumptions we derive the relationship between galaxies’ bulk properties (gas surface density, stellar content, and rotation curve) and their star formation rates, gas velocity dispersions, and rates of radial inflow. We show that the turbulence in discs can be powered primarily by star formation feedback, radial transport, or a combination of the two. In contrast to models that omit either radial transport or star formation feedback, the predictions of this model yield excellent agreement with a wide range of observations, including the star formation law measured in both spatially resolved and unresolved data, the correlation between galaxies’ star formation rates and velocity dispersions, and observed rates of radial inflow. The agreement holds across a wide range of galaxy mass and type, from local dwarfs to extreme starbursts to high-redshift discs. We apply the model to galaxies on the star-forming main sequence, and show that it predicts a transition from mostly gravity-driven turbulence at high redshift to star-formation-driven turbulence at low redshift. This transition and the changes in mass transport rates that it produces naturally explain why galaxy bulges tend to form at high redshift and discs at lower redshift, and why galaxies tend to quench inside-out.

Key words: turbulence – stars: formation – ISM: kinematics and dynamics – galaxies: formation – galaxies: ISM – galaxies: star formation.

1 INTRODUCTION

1.1 Observational background

Despite their diversity in mass, spatial extent, and stellar and gas content, disc galaxies both in the local and distant Universe show a striking range of regularities. Perhaps the most famous of these is the Kennicutt–Schmidt relation (see reviews by Kennicutt 1998; Kennicutt & Evans 2012; Krumholz 2014), the observed correlation between the rate at which galaxies form stars and a combination of their gas content and their dynamical times. The rate of star formation implied by this relation is remarkably small: on average, galaxies turn only ~ 1 per cent of their gas into stars per dynamical time of the gas (Zuckerman & Evans 1974). This correlation between gas content and star formation, and the remarkably low efficiency of star formation that it implies, was first observed on galactic scales

and in the local Universe. However, subsequent work has shown that it continues to hold even at high redshift (e.g. Bouché et al. 2007; Daddi et al. 2008, 2010a,b; Genzel et al. 2010; Tacconi et al. 2013), and on ~ 1 kpc scales in the local Universe (Kennicutt et al. 2007; Bigiel et al. 2008; Leroy et al. 2008, 2013; Liu et al. 2011; Momose et al. 2013).

Indeed, the correlation and inefficiency extend down to even ~ 1 pc scales. There are a number of lines of evidence in favour of this conclusion, including direct star counts in star-forming clouds near the Sun (Heiderman et al. 2010; Lada, Lombardi & Alves 2010; Krumholz, Dekel & McKee 2012; Evans, Heiderman & Vutisalchavakul 2014; Salim, Federrath & Kewley 2015; Heyer et al. 2016), correlations between gas and indirect star formation tracers such as recombination lines to larger distances in the Milky Way (Vutisalchavakul, Evans & Heyer 2016), and correlations between star formation and tracers of dense gas in both Galactic and

★ E-mail: mark.krumholz@anu.edu.au

extragalactic systems (Krumholz & Tan 2007; García-Burillo et al. 2012; Usero et al. 2015).¹

A second regularity and noted galactic-scale correlation concerns the velocity dispersions of the gas in galaxies. In both local and high-redshift galaxies, this gas invariably displays superthermal linewidths indicative of transsonic or supersonic motion (Glazebrook 2013, and references therein). This is true regardless of whether these motions are traced using the 21 cm line of H I (van Zee & Bryant 1999; Petric & Rupen 2007; Tamburro et al. 2009; Burkhart et al. 2010; Ianjamasimanana et al. 2012, 2015; Stilp et al. 2013; Chepurnov et al. 2015), the low- J lines of CO (Caldú-Primo et al. 2013, 2015; Meidt et al. 2013; Pety et al. 2013), or the recombination lines of ionized gas (Cresci et al. 2009; Lehnert et al. 2009, 2013; Green et al. 2010, 2014; Le Tiran et al. 2011; Swinbank et al. 2012; Arribas et al. 2014; Genzel et al. 2014; Moiseev, Tikhonov & Klypin 2015). Observed linewidths are relatively independent of radius within a given galaxy, but vary significantly from galaxy to galaxy in a way that is well-correlated with galaxies' rates of star formation. Galaxies with star formation rates below $\sim 1 M_{\odot} \text{ yr}^{-1}$, typical of the local Universe (Kennicutt & Evans 2012), all have roughly the same velocity dispersion of $\approx 10 \text{ km s}^{-1}$. However, at the higher star formation rates found both in local starbursts and in main sequence star-forming galaxies at higher redshift, velocity dispersions increase roughly linearly, $\sigma \propto \dot{M}_*$ (Krumholz & Burkhart 2016), although with substantial scatter and subsidiary dependencies on quantities such as the galaxies' gas fractions, sizes, and rotational velocities.

These velocity dispersions feed naturally into a third observed correlation, which is that galaxy discs tend to be in a state of marginal gravitational stability. The gravitational stability of a disc can be characterized by the Toomre (1964) Q parameter, defined by $Q \approx \kappa \sigma / \pi G \Sigma$, where κ is the epicyclic frequency of the galaxy's rotation, σ is the velocity dispersion, and Σ is the surface mass density. Observed disc galaxies in both the local universe and at high redshift tend to have $Q \approx 1$ throughout their discs (e.g. Martin, Kobulnicky & Heckman 2002; Genzel et al. 2010; Meurer, Zheng & de Blok 2013; Romeo & Falstad 2013; Romeo & Mogotsi 2017).

A fourth and final observed correlation relates to the spatial distribution of gas and star formation in galaxy discs. Star formation correlates with molecular gas rather than total gas, and the H₂-rich regions of galaxies are preferentially located in their centres. Consequently, the scale length of the star formation is comparable to the stellar scale length, $\approx 2\text{--}4 \text{ kpc}$, and a factor of 2–3 smaller than the neutral gas scale length (Regan et al. 2001; Leroy et al. 2008; Schrubba et al. 2011; Bigiel & Blitz 2012). Within the molecule-dominated region, the gas depletion time is $\sim 1\text{--}2 \text{ Gyr}$ (Bigiel et al. 2008; Leroy et al. 2013), much less than a Hubble time. The fact that star formation has not ceased in the centres of all galaxy discs implies either that we live at a special time when all local disc centres are about to quench, or that there is an ongoing gas supply to fuel star formation. Direct accretion of cold gas from the intergalactic medium (e.g. Kereš et al. 2005; Dekel & Birnboim 2006; Dekel et al. 2009; Wetzel & Nagai 2015) and condensation from hot haloes in low redshift galaxies (Marinacci et al. 2010; Joungh, Bryan & Putman 2012; Fraternali et al. 2013; Hobbs et al. 2013), supplemented by mass returned by stellar evolution (Leitner & Kravtsov

2011), likely provide a sufficient mass supply for star formation. However, they do not naturally provide it at the small galactocentric radii where star formation takes place. Accretion from a hot corona is predicted to deliver most of its mass at radii of $\sim 3\text{--}4$ stellar scale lengths (e.g. Marasco, Fraternali & Binney 2012), and, at least at high redshift, cold accretion tends to join the disc at radii of $\sim 0.1\text{--}0.3$ virial radii, which is ~ 10 times the stellar scale length (Danovich et al. 2015), though there are exceptions associated with loss of angular momentum by counter-rotating streams and major mergers, which tend to trigger 'compaction' events (Zolotov et al. 2015; Tacchella et al. 2016a,b). Preventing quenching requires this gas then flow radially inward. Such flows have recently been detected directly in a number of nearby galaxy discs (Schmidt et al. 2016).

1.2 Theoretical background

Any successful theory of the structure and evolution of disc galaxies ought to be able to explain all of these observed regularities, but at present no such theory is available. This is at least in part because theoretical modelling has tended to focus on one or two of the observed correlations, without attempting to unify all of them into a single, coherent picture.

Several authors have attempted to develop theories that link the problems of velocity dispersion, marginal stability, and star formation fuelling (e.g. Bournaud, Elmegreen & Elmegreen 2007; Agertz et al. 2009b; Bournaud & Elmegreen 2009; Dekel, Sari & Ceverino 2009; Ceverino, Dekel & Bournaud 2010; Krumholz & Burkert 2010; Vollmer & Leroy 2011; Cacciato, Dekel & Genel 2012; Forbes, Krumholz & Burkert 2012; Forbes et al. 2014a; Goldbaum, Krumholz & Forbes 2015, 2016). The central premise in these models is that gravitational instability produces torques that both move mass inward and drive turbulence, simultaneously regulating galaxies to $Q \approx 1$, producing supersonic velocity dispersions, and fuelling star formation. Models in this class naturally explain why $Q \approx 1$, why star formation is not quenched in modern galaxy centres, and why high-redshift galaxies have high-velocity dispersions. If one couples them to an empirically determined star formation relation, they can also do a reasonable job of explaining both galaxy-scale star formation laws and the high star formation–velocity dispersion portion of the $\sigma\text{--}\dot{M}_*$ correlation (Zheng et al. 2013; Krumholz & Burkhart 2016; Wong et al. 2016).

However, these models do not naturally explain the minimum velocity dispersion to which galaxy discs seem to settle at $z \approx 0$. Even in quiescent galaxies similar to the Milky Way, observed ISM velocity dispersions are $\approx 10 \text{ km s}^{-1}$ (e.g. Ianjamasimanana et al. 2012), corresponding to bulk motions at a Mach number ~ 1 for gas at the typical warm neutral medium temperature of $\approx 7000 \text{ K}$ (Wolfire et al. 2003). Some energy input is required to maintain transsonic flows of this sort, and models based purely on gravitational instability-driven torques do not naturally produce such an input in quiescent discs. Because such models do not naturally make any predictions about star formation rates on either large or small scales, they also do not explain the physical origins of the star formation law. More generally these models usually do not include any specific treatment of star formation feedback or its coupling to the interstellar medium, an obvious omission.

Other authors have instead chosen to focus on the observed correlation between star formation and gas. Some authors have attempted to derive this correlation using a 'bottom up' approach, whereby one begins by attempting to explain the inefficiency of star formation on small scales, and then builds a galaxy-scale star formation

¹This conclusion has recently been questioned by Lee, Miville-Deschênes & Murray (2016), but we argue in this paper that this is likely an artefact of their methodology, which differs from that of all the other authors. See below for details.

relation as the sum of small-scale relations (Krumholz & McKee 2005; Krumholz, McKee & Tumlinson 2009b; Federrath & Klessen 2012, 2013; Padoan, Haugbølle & Nordlund 2012; Krumholz 2013; Federrath 2015; Burkhardt 2018). These small-scale relations, while theoretically motivated, can be checked directly against numerical simulations of self-gravitating turbulence, and the agreement is generally good (e.g. Padoan & Nordlund 2011; Federrath & Klessen 2012; Padoan et al. 2014). This approach allows one to explain the star formation rate on both small and large scales, and naturally incorporates star formation feedback on small scales. Furthermore, if these models are supplemented by chemical models that capture the transition between the warm, H I and cold, H₂ phases of the ISM (Krumholz et al. 2009b; Krumholz 2013), they also correctly capture the observed dependence of the star formation rate on the chemical phase and metallicity of the ISM (Bolatto et al. 2011; Wong et al. 2013; Shi et al. 2014; Filho et al. 2016; Jameson et al. 2016; Rafelski et al. 2016). On the other hand, these models are generally silent on the question of galaxies' velocity dispersions, gravitational stability, or long-term fuelling.

Conversely, some authors have attempted to derive the star formation rate and velocity dispersion using a 'top down' method, the fundamental assumption of which is that the star formation rate is set by considerations of force and energy balance within a galactic disc (e.g. Thompson, Quataert & Murray 2005; Ostriker, McKee & Leroy 2010; Hopkins, Quataert & Murray 2011; Ostriker & Shetty 2011; Faucher-Giguère, Quataert & Hopkins 2013; Hayward & Hopkins 2017). In these models, one considers a disc of a prescribed gas content and gravitational potential, and asks what star formation rate is required for star formation feedback to be vigorous enough to keep the disc in vertical pressure balance and energy balance. This approach has the advantage that it is rooted in simple physical considerations that must hold at some level, and it is the first step in the approach that we shall pursue in this paper. Moreover, it enables one to make predictions that link star formation, velocity dispersion, and Toomre stability, and thus unify three of the observed correlations discussed above.

However, top-down models that work solely based on the balance between feedback, vertical gravity, and dissipation have proven difficult to make work in practice. First of all, unless one posits a source of star formation feedback for which the momentum injected per star formed increases with gas surface density (e.g. as trapped infrared radiation pressure does in the model of Thompson et al. 2005), the natural prediction of these models is that the star formation rate per unit area should rise as the square of the gas surface density (e.g. equation 13 of Ostriker & Shetty 2011 or equation 18 of Faucher-Giguère et al. 2013). The predicted correlation $\dot{\Sigma}_* \propto \Sigma^p$ with $p \approx 2$ is substantially steeper than the observed correlation, which ranges between $p \approx 1$ in spatially resolved patches of local galaxies to $p \approx 1.5$ for rapidly star-forming galaxies as a whole.²

Secondly, because these models compute the star formation rate from the weight of the ISM, they naturally predict that the star formation rate at a given surface density is independent of the metallicity or chemistry of the ISM, since these factors do not alter

the weight. They can be reconciled with the strong observational evidence that metallicity and chemical phase do affect the rate of star formation only by positing that the efficiency of star formation feedback is metallicity-dependent. For example, the model of Ostriker et al. (2010) predicted that the regions of comparable gas surface density in the Small Magellanic Cloud and the Milky Way should form stars at nearly equal rates. Bolatto et al. (2011) found that this prediction was incorrect, and proposed a modification to the theory in which the efficiency of photoelectric heating scales inversely with metallicity, and thus stars pressurise the ISM more efficiently in low-metallicity galaxies. While this does fix agreement with the observations, Krumholz (2013) points out that the physical mechanism proposed by Bolatto et al. (2011) to produce the metallicity dependence is not correct, and, more generally, that there is no good reason to expect that feedback efficiencies will depend on metallicity in the ways required to explain the observations. In particular, supernovae are thought to be the dominant feedback mechanism in most galaxies, and supernova momentum injection is nearly independent of metallicity (e.g. Thornton et al. 1998; Martizzi, Faucher-Giguère & Quataert 2015; Gentry et al. 2017).

Thirdly, these models do not naturally predict either the sub-galactic star formation law or the gravitational stability parameter, forcing one to adopt one or the other based on empirical observations. If one adopts the observed sub-galactic star formation rate (e.g. Ostriker & Shetty 2011), then, as we shall show below, one predicts velocity dispersions and Toomre Q parameters sharply at odds with what is observed. Conversely, one can posit that star formation rates are very sensitive to the Toomre Q parameter, so that the star formation rate self-adjusts to maintain $Q \approx 1$ (e.g. Faucher-Giguère et al. 2013; Hayward & Hopkins 2017). By construction this produces the correct Toomre Q , but it still fails to reproduce the observed σ – \dot{M}_* correlation (because the predicted star formation law is too steep – Krumholz & Burkhardt 2016), and it also predicts that star formation on small scales is very efficient in high surface density galaxies, contrary to observations. Just to give one example of this difficulty: if star formation efficiencies on small scales were higher in high surface density galaxies, then the ratio of infrared emission (a star formation tracer) to HCN luminosity (a tracer of dense gas on small scales) should increase with star formation rate, whereas the observed trend is the opposite (García-Burillo et al. 2012; Usero et al. 2015). This approach also runs into observational difficulty with its central assumption that galaxies' star formation rates are very sensitive to the value of the Toomre Q parameter; observations strongly disfavour any such correlation (Leroy et al. 2008). Instead, both observations and simulations (Agertz, Teyssier & Moore 2009a; Goldbaum, Krumholz & Forbes 2015, 2016) seem to suggest that the response of a disc to a drop in Toomre Q is that the disc becomes non-axisymmetric and moves mass inwards, rather than that its star formation rate dramatically increases.

1.3 This work and its motivation

Our goal in this work is to unify models of galactic discs that focus on transport, star formation fuelling, and gravitational instability with those that focus on the energy and momentum balance of star formation feedback. We show below that this approach remedies many of the observational problems we have identified with the various theories that have been proposed to date. However, the need for such a synthesis can be driven home simply by more basic consideration of the observations and their energetic implications.

²Narayanan et al. (2012) and Faucher-Giguère et al. (2013) argue that one can steepen the relation and increase the value of p by adopting a CO to H₂ conversion factor that scales strongly with galaxy star formation rate. However, even adopting such a scaling, fits to the more recent and larger data sets favour $p \approx 1.7$ rather than 2 (c.f. fig. 3 of Thompson & Krumholz 2016), and recent dust-based measurements of gas content that are independent of CO suggest that even this is too steep (Genzel et al. 2015).

The turbulent energy per unit area contained in a galactic disc of gas surface density Σ_g and velocity dispersion σ_g is

$$\left(\frac{dE}{dA}\right)_{\text{turb}} \approx \frac{3}{2} \Sigma_g \sigma_g^2 = 3.1 \times 10^9 \Sigma_{g,10} \sigma_{g,10}^2 \text{ erg cm}^{-2}, \quad (1)$$

where $\Sigma_{g,10} = \Sigma_g / 10 \text{ M}_\odot \text{ pc}^{-2}$ and $\sigma_{g,10} = \sigma_g / 10 \text{ km s}^{-1}$; the scaling factors are typical values at the Solar Circle in the Milky Way. The energy should dissipate due to decay of turbulence over a time-scale comparable to the crossing time, but, in a disc with $Q \approx 1$, this time-scale is comparable to the galactic dynamical time $t_{\text{dyn}} = r/v_\phi$, where r is the galactocentric radius and v_ϕ is the rotation velocity. It is therefore of interest to consider possible sources of power that are capable of delivering this amount of energy per unit area over a time-scale t_{dyn} .

As noted above, Schmidt et al. (2016) directly detect flows of mass radially inwards through the discs of local spiral galaxies with mass fluxes $\dot{M}_{\text{in}} \sim 1 \text{ M}_\odot \text{ yr}^{-1}$. These observations are difficult due to the near cancellation of inflow and outflow rates around spiral arms and in outer regions where galaxies become significantly lopsided. At a minimum, the magnitude of the inflow should be regarded as significantly uncertain. However, we note that inflows rates of roughly this size must be ubiquitous to explain star formation fuelling. In a galaxy with a flat rotation curve, the amount of gravitational potential energy per unit area per unit time released by this flow of mass down the potential well is

$$\frac{d^2 E}{dt dA} \approx \frac{\dot{M}_{\text{in}} v_\phi^2}{2\pi r^2}, \quad (2)$$

so over a galactic dynamical time the flow delivers an energy per unit area

$$\left(\frac{dE}{dA}\right)_{\text{inflow}} \approx \frac{\dot{M}_{\text{in}} v_\phi}{2\pi r} = 6.5 \times 10^9 \dot{M}_{\text{in},1} v_{\phi,200} r_{10}^{-1} \text{ erg cm}^{-2}, \quad (3)$$

where $\dot{M}_{\text{in},1} = \dot{M}_{\text{in}} / 1 \text{ M}_\odot \text{ yr}^{-1}$, $v_{\phi,200} = v_\phi / 200 \text{ km s}^{-1}$, and $r_{10} = r / 10 \text{ kpc}$.

In comparison, star formation feedback is expected to inject energy at a rate per unit area

$$\frac{d^2 E}{dt dA} \approx \dot{\Sigma}_* \left\langle \frac{p_*}{m_*} \right\rangle \sigma_g, \quad (4)$$

where $\dot{\Sigma}_*$ is the star formation rate per unit area and p_*/m_* is the terminal momentum per unit mass delivered by star formation feedback. (We give a detailed explanation for the origin of this expression below, but intuitively it results simply from the assumption that motions driven by stellar feedback break up and add their energy to the turbulent background once their expansion velocities become comparable to the overall velocity dispersion, so the energy added per ‘injection event’ is of order the momentum injected times the velocity dispersion.) Simulations suggest the momentum per unit mass is $p_*/m_* \approx 3000 \text{ km s}^{-1}$ for single supernovae (Cioffi, McKee & Bertschinger 1988; Thornton et al. 1998; Kim & Ostriker 2015a; Martizzi et al. 2015; Walch & Naab 2015; Gentry et al. 2017; Kim, Ostriker & Raileanu 2017). Over a galactic dynamical time, and scaling to Solar Circle values again,

$$\begin{aligned} \left(\frac{dE}{dA}\right)_{\text{sf}} &\approx \dot{\Sigma}_* \left\langle \frac{p_*}{m_*} \right\rangle \sigma_g \frac{r}{v_\phi} \\ &= 3.1 \times 10^9 \dot{\Sigma}_{*,3} \sigma_{g,10} r_{10} v_{\phi,200}^{-1} \text{ erg cm}^{-2}, \end{aligned} \quad (5)$$

where $\dot{\Sigma}_{*,3} = \dot{\Sigma}_* / 10^{-3} \text{ M}_\odot \text{ pc}^{-2} \text{ Myr}^{-1}$.

The implication of this calculation is that, at least at the order of magnitude level, inflow and star formation feedback are comparably

important energetically in the Solar neighbourhood, and that both are capable of supplying enough energy to replenish the turbulence in the ISM over a galactic dynamical time. Moreover, if we were to repeat this calculation for other types of galaxies we might well get quite different results. The ratio of $(dE/dA)_{\text{inflow}}$ to $(dE/dA)_{\text{sf}}$ scales as $(\dot{M}_{\text{in}}/\dot{M}_*)(v_\phi^2/\sigma_g)$, where \dot{M}_* is the total star formation rate. We do not have direct measurements of \dot{M}_{in} except in local spirals, but assuming that $\dot{M}_{\text{in}}/\dot{M}_* \sim 1$, as would be required to explain star formation fuelling and as is observed locally, star formation should be energetically dominant in galaxies with smaller v_ϕ (for example local dwarfs), while inflow should dominate those with larger σ_g (for example high- z galaxies). Clearly it is not reasonable to ignore either star formation feedback or inflows in building a model of galaxy discs, as has been the practice for most work up to this point.

Below we build a minimal unified model that combines both of these processes. We show that, while simple, this model is far more successful than either feedback-only or inflow-only models at explaining the observed correlations obeyed by galaxy discs. We derive the model in Section 2, and compare it to a variety of observations in Section 3. We discuss the implications of our findings for galaxy formation in Section 4, and conclude in Section 5.

2 MODEL

In this section we develop a model for a galactic disc in both vertical hydrostatic and energy equilibrium, where the sources of energy input include both star formation feedback and gravitational potential energy released by inward flow of gas through the disc. A central premise of our model is that the gas is dynamically important and capable of adjusting its inflow rate to maintain marginal stability, rather than simply acting as a passive tracer whose transport rate is dictated by the stellar potential independent of the dynamical state of the gas. This premise likely fails in regions where the gas contributes a negligible mass fraction even at the mid-plane, for example, the central $\sim 3 \text{ kpc}$ of the Milky Way where the Galactic bar dominates the dynamics (e.g. Binney et al. 1991). We argue in Section 4.4 that the vast majority of the interstellar medium by both mass and star formation rate is not found in such regions, so that our model is applicable to the bulk of the ISM and star formation in the Universe. For now, however, we simply take as given that the transport rate is not dictated by a stellar bar or similar structures, but is able to self-adjust.

For convenience we summarize all the quantities used in our model in Table 1. We treat our model galaxy as a thin, axisymmetric disc characterized at every radius r by a total gas surface density Σ_g and 1D gas velocity dispersion σ_g . In addition to gas, the disc contains stars and dark matter. The dark matter has a density ρ_d , and we assume that its distribution is close to spherical. If there is a spheroidal stellar distribution, we also include its density in ρ_d . Other stars are in a disc, characterized by a surface density Σ_* and a 1D velocity dispersion σ_* . For simplicity we assume that both σ_g and σ_* are isotropic. In real stellar discs at $z \approx 0$ this assumption fails at the factor of ~ 2 level.

The gas and stars orbit within a steady gravitational potential, which we characterize by the velocity v_ϕ required for material in orbit to be in balance between centrifugal and gravitational forces in the corotating frame. The rotation curve has an index $\beta = d \ln v_\phi / d \ln r$, the angular velocity at radius r is $\Omega = v_\phi / r$, and the orbital period is $t_{\text{orb}} = 2\pi r / v_\phi$.

Table 1. Symbol definitions. The fiducial value listed is the value used in numerical evaluations and plots unless otherwise stated.

Symbol	Fiducial value	Meaning	Defining equation
Inputs to model			
Σ_g	–	Gas surface density	–
Σ_*	–	Stellar surface density	–
σ_g	–	Gas velocity dispersion (total thermal plus non-thermal)	–
σ_*	–	Stellar velocity dispersion	–
ρ_d	–	Dark matter density	–
v_φ	–	Galaxy rotation curve velocity	–
Ω	–	Galaxy angular velocity	–
t_{orb}	–	Galaxy orbital period, $t_{\text{orb}} = 2\pi/\Omega$	–
β	0	Rotation curve index, $\beta = d \ln v_\varphi / d \ln r$	–
$f_{g,Q}$	0.5	Fractional contribution of gas to Q	9
$f_{g,P}$	0.5	Fractional contribution of gas self-gravity to mid-plane pressure	20
f_{sf}	–	Fraction of ISM in star-forming molecular phase	30
Physics parameters			
Q_{min}	1	Minimum possible disc stability parameter	6
φ_{mp}	1.4	Ratio of total pressure to turbulent pressure at mid-plane	12
η	1.5	Scaling factor for turbulent dissipation rate	26
φ_Q	2	One plus ratio of gas to stellar Q	27
φ_{nt}	1	Fraction of velocity dispersion that is non-thermal	28
ϵ_{ff}	0.015	Star formation efficiency per free-fall time	30
$t_{\text{sf,max}}$	2 Gyr	Maximum star formation time-scale	32
φ_a	2	Offset between resolved and unresolved star formation law normalizations	58
Model outputs			
ρ_{min}	–	Minimum mid-plane density required to produce rotation curve	51
$t_{\text{orb,T}}$	–	Orbital period at which galaxies switch from GMC to Toomre regime	33
σ_{sf}	–	Gas velocity dispersion that can be sustained by star formation alone	39
Σ_{sf}	–	Gas surface density below which star formation alone can sustain turbulence	41
\dot{M}_{ss}	–	Steady-state mass inflow rate through the disc	49

We provide the source code to perform the computations involved in the model, and produce all the plots included in the paper, at <https://bitbucket.org/krumholz/kbfc17>.

2.1 Gravitational instability

A central ansatz of our model, following Krumholz & Burkert (2010), Cacciato et al. (2012), Forbes et al. (2012), and Forbes et al. (2014a) is that gravitational instability-driven transport will prevent the disc from ever becoming more than marginally gravitationally unstable. If the disc begins to become unstable, the instability will break axisymmetry and the subsequent torques will drive mass inward until marginal stability is restored. We therefore begin by expressing this condition. Modern treatments of gravitational instability include the effects of multiple stellar populations as well as gas, along with the effects of finite thickness and the dissipative nature of gas (Rafikov 2001; Romeo, Burkert & Agertz 2010; Elmegreen 2011; Romeo & Wiegert 2011; Hoffmann & Romeo 2012; Romeo & Falstad 2013). In this work we use the simple approximation given by Romeo & Falstad (2013),

$$Q \approx \left(Q_g^{-1} + \frac{2\sigma_g\sigma_*}{\sigma_g^2 + \sigma_*^2} Q_*^{-1} \right)^{-1}, \quad (6)$$

where

$$Q_g = \frac{\kappa \sigma_g}{\pi G \Sigma_g} \quad (7)$$

and similarly for Q_* . Here $\kappa = \sqrt{2(\beta+1)}\Omega$ is the epicyclic frequency. This expression is valid as long as $Q_g < Q_*$, the quasi-spherical dark matter halo contributes negligibly to the gravitational stability or instability of the system (i.e. $Q_d \gg Q_*$, where Q_d is the dark matter Q), and the ratio of vertical to radial velocity dispersions for the gas and stars is $\gtrsim 0.5$. The latter two conditions hold broadly across all the galaxies we shall consider; the first requires a bit more discussion, which we defer to the end of this section. For convenience, we can rewrite equation (6) as

$$Q = f_{g,Q} Q_g, \quad (8)$$

where

$$f_{g,Q} \equiv \frac{\Sigma_g}{\Sigma_g + [2\sigma_g^2/(\sigma_g^2 + \sigma_*^2)]\Sigma_*}. \quad (9)$$

The quantity $f_{g,Q}$ can be thought of as defining the effective gas fraction in the disc for the purposes of computing gravitational stability. It clearly behaves as we intuitively expect, in that $f_{g,Q} \rightarrow 1$ for $\Sigma_g \gg \Sigma_*$, and $f_{g,Q} \rightarrow 0$ for $\Sigma_g \ll \Sigma_*$. In the Solar neighbourhood, which has gas properties $\Sigma_g \approx 14 \text{ M}_\odot \text{ pc}^{-2}$ (McKee, Parravano & Hollenbach 2015) and $\sigma_g \approx 7 \text{ km s}^{-1}$ (Kalberla & Kerp 2009), and

stellar properties $\Sigma_* \approx 33 \text{ M}_\odot \text{ pc}^{-2}$ and $\sigma_* \approx 16 \text{ km s}^{-1}$ (McKee et al. 2015), we have $Q_* \approx Q_g \approx 1.5$ and $f_{g,Q} \approx 0.6$.

The condition for stability is that Q be larger than a value Q_{\min} of order unity that depends on the thickness of the disc (thicker discs can be stable at lower Q) and the gas equation of state (more dissipative equations require higher Q for stability). As a fiducial value we shall adopt $Q_{\min} = 1$, which is appropriate for discs that are relatively quiescent. There is some evidence from cosmological simulations that instability can set in at slightly higher $Q \sim 2$ –3 in the perturbed discs where a greater fraction of the turbulence is in compressive modes that do not support the gas (Inoue et al. 2016), but since this is only a factor of ~ 2 level effect and only then in some of the galaxies with which we are concerned, we will neglect this complication.

The case $Q_* < Q_g$, where stars rather than gas are the most unstable component, requires a bit more attention. Due to the fact that gas is dissipational and thus usually has a lower velocity dispersion than stars, it tends to be the most unstable component in any gas-rich system. Thus we expect $Q_* > Q_g$ to hold in local dwarfs and lower mass spirals, all star-forming galaxies at high redshift, and in all mergers and starbursts. However, massive local spirals like the Milky Way are sufficiently gas poor ($f_{\text{gas}} \sim 10$ –20 per cent – Saintonge et al. 2011) that for the most part they have $Q_* < Q_g$: Romeo & Mogotsi (2017) find $Q_g/Q_* \approx 0.5$ –10 for the HERACLES / THINGS sample, with the bulk of the data at $Q_g/Q_* \approx 3$. Our expression for Q (equation 6) assumes $Q_g < Q_*$, but the equivalent expression for $Q_g > Q_*$ (Romeo & Falstad 2013) differs only slightly when Q_g and Q_* are within a factor of a few of one another. Quantitatively, using the Solar neighbourhood velocity dispersions quoted above ($\sigma_g \approx 7 \text{ km s}^{-1}$, $\sigma_* \approx 16 \text{ km s}^{-1}$), the error produced by using equation (6) is 10 per cent for $Q_g/Q_* = 3$, and 17 per cent for $Q_g/Q_* = 10$. This is well below the factor of ≈ 2 uncertainty in Q_{\min} , so for simplicity we simply use equation (6) in all cases, rather than using a different form for large local spirals than for all the other types of galaxies we will consider. One might also worry that, in the $Q_* < Q_g$ regime, gravitational instabilities in the stars might not induce perturbations in the gas capable of driving transport. However, Romeo & Mogotsi (2017) find that the local spirals with $Q_* < Q_g$ are also in the regime where perturbations in the gas and stars are strongly coupled (e.g. see their fig. 5), so this is not a concern.

2.2 Vertical force balance

A second ansatz of our model, following a number of authors (e.g. Boulares & Cox 1990; Piontek & Ostriker 2007; Koyama & Ostriker 2009; Ostriker et al. 2010) is that the gas is in vertical hydrostatic equilibrium. The spatially averaged momentum equation for a time-steady isothermal gas reads (Krumholz 2017, equation 10.9; also see Kim & Ostriker 2015b)

$$\frac{\partial}{\partial z} \langle \rho_g (\sigma_{\text{th}}^2 + v_z^2 + v_A^2) \rangle - \frac{\partial}{\partial z} \left\langle \frac{B_z^2}{4\pi} \right\rangle - \langle \rho_g g_z \rangle = 0, \quad (10)$$

where ρ_g is the gas density, σ_{th} is the gas thermal velocity dispersion, v_z is the vertical velocity, v_A is the gas Alfvén speed, B_z is the z component of the magnetic field, g_z is the vertical gravitational acceleration, and we have oriented our coordinate system so the disc mid-plane lies in the xy plane; the angle brackets denote averaging over the area of the disc, where the area considered is small compared to the disc scale length, but large compared to the size of an individual molecular cloud of star-forming complex. The first term represents the force exerted by the gradient in thermal, turbu-

lent, and magnetic pressure, the second represents the force due to magnetic tension, and the third represents the force due to gravity. Magnetic tension tends to be subdominant except for unusual, artificially constructed magnetic field configurations, and thus we can generally drop the second term. This expression omits the contribution from cosmic ray pressure, but this is likely comparable to magnetic pressure in importance (e.g. Boulares & Cox 1990).

Integrating equation (10) from $z = 0$ to ∞ , and assuming that $\rho_g \rightarrow 0$ and the Alfvén speed remains finite as $z \rightarrow \infty$, we have

$$\rho_{g,\text{mp}} (\sigma_g^2 + v_A^2)_{\text{mp}} = - \int_0^\infty \langle \rho_g g_z \rangle dz, \quad (11)$$

where the subscript mp indicates that a quantity is to be evaluated at the disc mid-plane, and where we have dropped the angle brackets and implicitly understand that mid-plane terms represent area averages over the mid-plane; in writing this expression, we have relied on our assumption that the gas velocity dispersion is isotropic, so $\langle \rho_g v_z^2 \rangle = \rho_{g,\text{mp}} (\sigma_g^2 - \sigma_{\text{th}}^2)$. We write the left-hand side as

$$\rho_{g,\text{mp}} (\sigma_g^2 + v_A^2)_{\text{mp}} \equiv \phi_{\text{mp}} \rho_{g,\text{mp}} \sigma_g^2, \quad (12)$$

where ϕ_{mp} is a factor that represents the factor by which the mid-plane pressure exceeds that due to turbulent plus thermal pressure alone, due to magnetic and cosmic ray pressure. Equipartition between magnetic and kinetic degrees of freedom in the directions transverse to the field corresponds to an Alfvén Mach number of $2/3$, which is $\phi_{\text{mp}} = 1.4$ assuming that thermal pressure is unimportant compared to turbulent pressure. A cosmic ray pressure comparable to the magnetic pressure would increase this to $\phi_{\text{mp}} \approx 2$. On the other hand, if thermal pressure is non-negligible, for example in modern dwarf galaxies, then kinetic–magnetic equipartition implies ϕ_{mp} closer to unity, since the gas reaches equipartition only between the non-thermal motions and the magnetic field. The differences between $\phi_{\text{mp}} = 2$ and $\phi_{\text{mp}} = 1$ are small enough that we will not worry about it, and we will simply use $\phi_{\text{mp}} = 1.4$ as our fiducial value.

The term on the right-hand side of equation (11) depends on the distribution of gas, stars, and dark matter, since each of these components contributes to g_z . To parametrize this dependence, note that the potential ψ obeys the Poisson equation, which in cylindrical coordinates (assuming symmetry in the azimuthal direction) reads

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \psi}{\partial r} \right) + \frac{\partial^2 \psi}{\partial z^2} = 4\pi G \rho, \quad (13)$$

where ρ is the total density including all components. The radial gradient of ψ is related to the rotation curve by

$$\frac{v_\phi^2}{r} = \frac{\partial \psi}{\partial r}, \quad (14)$$

and using this in the Poisson equation we obtain

$$\frac{\partial g_z}{\partial z} = 4\pi G \rho - 2\beta \Omega^2, \quad (15)$$

where $g_z = \partial \psi / \partial z$ and $\beta = d \ln v_\phi / d \ln r$ is the rotation curve index. Integrating, we therefore have

$$g_z \approx \int_0^z (4\pi G \rho - 2\beta \Omega^2) dz'. \quad (16)$$

Note that, although it is tempting to approximate that $\beta \Omega^2$ is constant for small z , this approximation clearly fails for the common case of a flat rotation curve, $\beta = 0$, because $\beta = 0$ at the mid-plane but not above it – see appendix C of McKee et al. (2015) for

discussion. The weight is therefore

$$\int_0^\infty \langle \rho_g g_z \rangle dz = 2\pi G \int_0^\infty \rho_g \left[\Sigma(z) - \frac{1}{\pi G} \int_0^z \beta \Omega^2 dz' \right] dz, \quad (17)$$

where $\Sigma(z) = 2 \int_0^z \rho dz$ is the total column density of material at heights between $-z$ and z , and we assume symmetry about $z = 0$. If we write out the total column as the sum of the gas, stellar, and dark components, $\Sigma(z) = \Sigma_g(z) + \Sigma_*(z) + \Sigma_d(z)$, then we can integrate the gaseous part by the usual change of variables $d\Sigma_g = 2\rho_g dz$, yielding

$$\int_0^\infty \langle \rho_g g_z \rangle dz = \frac{\pi}{2} G \Sigma_g^2 + 4\pi G \int_0^\infty \rho_g \left(\Sigma_*(z) + \Sigma_d(z) - \frac{1}{2\pi G} \int_0^z \beta \Omega^2 dz' \right) dz. \quad (18)$$

The dark matter scale height is much larger than the gas scale height, so we can approximate $\Sigma_d(z) = 2\rho_d z$ in equation (18), where ρ_d is the dark matter density inside the plane. Similarly, the stellar scale height is at least as large as the gas scale height. We can therefore use the approximation suggested by Ostriker et al. (2010),

$$\int_0^\infty \langle \rho_g g_z \rangle dz \approx \frac{\pi}{2} G \Sigma_g^2 \cdot \left[1 + \frac{\zeta_d \rho_d + \zeta_* \rho_*}{\rho_{g,mp}} - \frac{4}{\pi G \Sigma_g^2} \int_0^\infty \rho_g \int_0^z \beta \Omega^2 dz' dz \right], \quad (19)$$

where $\rho_{g,mp}$ is the mid-plane gas density, $\rho_{*,mp}$ is the mid-plane stellar density, and ζ_d and ζ_* are numerical factors of order unity that depend on the gas density distribution and the relative scale heights of gas and stars.³ For the dark matter, which has a scale height much larger than the gas scale height, $\zeta_d \approx 1.33$. The stellar scale height can range from much larger than that of the gas, in which case $\zeta_* \approx 1.33$ as for the dark matter, to comparable to the gas, in which case $\zeta_* \approx 1$, with exact equality holding in the case where the gas and stars have identical vertical distributions.

We therefore define

$$f_{g,p} \equiv \left[1 + \frac{\zeta_d \rho_d + \zeta_* \rho_*}{\rho_{g,mp}} - \frac{4}{\pi G \Sigma_g^2} \int_0^\infty \rho_g \int_0^z \beta \Omega^2 dz' dz \right]^{-1} \quad (20)$$

so that

$$\int_0^\infty \langle \rho_g g_z \rangle dz = \frac{\pi}{2} G f_{g,p}^{-1} \Sigma_g^2. \quad (21)$$

The physical meaning of $f_{g,p}$ is that it is the fraction of the mid-plane pressure due to the local self-gravity of the gas (the unity term in equation 20), as opposed to local dark matter (as represented by the ρ_d term), local stars (as represented by the ρ_* term), or material of any type interior to the radius under consideration (as represented by the $\beta \Omega^2$ term). In the Solar neighbourhood, McKee et al. (2015) obtain estimates $\rho_{g,mp} = 0.041 M_\odot \text{pc}^{-3}$, $\rho_* = 0.043 M_\odot \text{pc}^{-3}$, and $\rho_d \ll \rho_*$. Using their equation 94, and adopting $\beta = 0$ at the mid-plane, gives $1/(\pi G) \int_0^z \beta \Omega^2 dz' = 0.01 M_\odot \text{pc}^{-2}$ at $z = 150 \text{pc}$, approximately the gas scale height. Using these values in equation (20), and adopting $\zeta_* = 1.33$ since the stellar scale height is

much larger than the gas scale height, gives $f_{g,p} \approx 0.4$ for the Solar neighbourhood, similar to $f_{g,Q}$.

Finally, inserting equations (12) and (21) into equation (11) gives

$$\rho_{g,mp} = \frac{\pi}{2\phi_{mp} f_{g,p}} G \left(\frac{\Sigma_g}{\sigma_g} \right)^2. \quad (22)$$

Rewriting in terms of Q , we arrive at our final expression for the mid-plane density,

$$\rho_{g,mp} = \frac{(1 + \beta) f_{g,Q}^2}{\pi Q^2 \phi_{mp} f_{g,p}} \left(\frac{\Omega^2}{G} \right). \quad (23)$$

2.3 Energy equilibrium

The third assumption of our model is that gas discs are in energy equilibrium, meaning that the rate at which energy is lost due to dissipation of turbulence (ultimately leading to radiative losses) balances the rate at which it is added due to star formation feedback and input of gravitational energy due to non-axisymmetric torques. We must therefore calculate each of these three rates.

2.3.1 Turbulent dissipation

Dissipation of supersonic turbulence has been subject to extensive study (Mac Low et al. 1998; Stone, Ostriker & Gammie 1998; Mac Low 1999; Lemaster & Stone 2009), and the consensus of this work is that the energy is lost to shocks (and, in weakly ionized plasmas, ion-neutral friction – Burkhart et al. 2015) in roughly a flow crossing time at the outer scale of the turbulence. Thus the dissipation rate per unit area should be the kinetic energy per unit area divided by the crossing time. To determine the crossing time, we approximate that the outer scale of the turbulence is of order the gas scale height, and following Forbes et al. (2012) we approximate this as

$$H_g \approx \frac{\sigma_g^2}{\pi G [\Sigma_g + (\sigma_g/\sigma_*) \Sigma_*]}, \quad (24)$$

where the factor σ_g/σ_* in the denominator has been chosen to interpolate between the two extreme cases where $\sigma_g/\sigma_* \ll 1$ and $\sigma_g/\sigma_* = 1$. In the former case, the gas is so much thinner than the stars that the stellar distribution contributes negligibly to the vertical gravity of the gas, while in the latter case the two components have approximately the same vertical distribution. With this approximation, we can write the loss rate as

$$\mathcal{L} = \eta \frac{\Sigma_g (\sigma_g^2 - \sigma_{th}^2)}{H_g / \sqrt{\sigma_g^2 - \sigma_{th}^2}} \quad (25)$$

$$= \frac{2(1 + \beta)}{\pi G Q^2} \eta \phi_Q \phi_{nt}^{3/2} f_{g,Q}^2 \Omega^2 \sigma_g^3. \quad (26)$$

In equation (25), the numerator is the kinetic energy per unit area, the denominator is the scale height crossing time, and σ_{th} is the purely thermal portion of the gas velocity dispersion, which is not subject to radiative loss because the gas temperature is assumed to be set by radiative equilibrium. The quantity η is a factor of order unity that defines the exact loss rate, with $\eta = 3/2$ corresponding to all the energy being radiated in a single scale height-crossing time; we adopt this as our fiducial value. The factors

$$\phi_Q \equiv 1 + \frac{Q_g}{Q_*} \quad (27)$$

³Note that Ostriker & Shetty (2011)'s equation 2 is a special case of equation (19); one can derive their equation by adopting $\beta = 1$, and assuming that the angular velocity Ω arises purely from a spherical matter distribution. Also note that our ζ_d and ζ_* differ from theirs by a factor of 4. We choose our normalization so that $\zeta \rightarrow 1$ exactly in the limiting case where the gas and stars have the same vertical distribution.

and

$$\phi_{\text{nt}} \equiv 1 - \frac{\sigma_{\text{th}}^2}{\sigma_{\text{g}}^2} \quad (28)$$

are both close to unity for most galaxies. We have $\varphi_Q = 2$ if $Q_{\text{g}} \approx Q_*$, and we adopt this as a fiducial value. Values of φ_Q significantly greater than unity are possible only if $Q_* < Q_{\text{g}}$. Similarly, the quantity φ_{nt} deviates significantly from unity only for gas velocity dispersions so small that they approach the thermal velocity dispersion, which is $\approx 5 \text{ km s}^{-1}$ in H I-dominated galaxies, and ≈ 0.2 – 0.5 km s^{-1} in H₂-dominated ones. For most purposes we will use $\varphi_{\text{nt}} = 1$ as a fiducial value, corresponding to $\sigma_{\text{g}} \gg \sigma_{\text{th}}$, but where necessary we will evaluate φ_{nt} numerically.

2.3.2 Driving by star formation

Following a number of authors (Matzner 2002; Krumholz, Matzner & McKee 2006; Goldbaum et al. 2011; Faucher-Giguère et al. 2013; Krumholz, Kruijssen & Crocker 2017), we approximate that the rate at which star formation adds energy to the gas is determined by the asymptotic momentum of shells of gas driven by supernovae or other forms of stellar feedback. Specifically, if an energetic feedback event (such as a supernova) occurs, it will sweep up a bubble of interstellar gas that will, after all the thermal energy injected by the event has been radiated, contain asymptotic radial momentum p . We approximate that this event adds an amount of energy $\approx p\sigma_{\text{g}}$ to the gas when the shell breaks up and merges with the turbulence. Thus if the star formation rate per unit area is $\dot{\Sigma}_*$, and the mean momentum injected per unit mass of stars formed is p_*/m_* , the rate of energy gain per unit area from star formation is

$$\mathcal{G} = \left\langle \frac{p_*}{m_*} \right\rangle \sigma_{\text{g}} \dot{\Sigma}_*. \quad (29)$$

As discussed above, for single supernovae $p_*/m_* \approx 3000 \text{ km s}^{-1}$ (Cioffi et al. 1988; Thornton et al. 1998; Kim & Ostriker 2015a; Martizzi et al. 2015; Walch & Naab 2015). The momentum injected may be somewhat enhanced by clustering, though probably by at most a factor of ~ 4 when averaging over a realistic cluster mass function (Sharma et al. 2014; Gentry et al. 2017, 2018; Kim et al. 2017). For simplicity we will ignore this effect and adopt the single supernova value $p_*/m_* \approx 3000 \text{ km s}^{-1}$ as our fiducial choice.

It is convenient to express the rate of star formation as

$$\dot{\Sigma}_* = \epsilon_{\text{ff}} f_{\text{sf}} \frac{\Sigma_{\text{g}}}{t_{\text{ff}}}. \quad (30)$$

Here f_{sf} is the fraction of the gas that is in a star-forming molecular phase rather than a warm atomic phase, and t_{ff} and ϵ_{ff} are the free-fall time and star formation rate per free-fall time in this gas. As noted above, there is extensive observational evidence that $\epsilon_{\text{ff}} \approx 0.01$ over a very wide range of star-forming environments (Krumholz & Tan 2007; García-Burillo et al. 2012; Krumholz, Dekel & McKee 2012; Evans, Heiderman & Vutsalchavakul 2014; Salim et al. 2015; Usero et al. 2015; Heyer et al. 2016; Vutsalchavakul, Evans & Heyer 2016; Leroy et al. 2017; Onus, Krumholz & Federrath 2018). We adopt $\epsilon_{\text{ff}} = 0.015$, the best fit from Krumholz et al. (2012), as our fiducial choice.

We pause here to note that, in contrast to the other studies cited, Lee et al. (2016), building on the work of Murray (2011), report the existence of a population of clouds with very high star formation efficiencies, $\epsilon_{\text{ff}} \approx 1$. If this result were correct, it would have profound implications for models such as the one we propose. However, it is hard to reconcile this observation with the results of

the numerous other studies cited above, which have failed to detect the purported high efficiency cloud population. We argue that the likely explanation for this discrepancy is a methodological bias. Lee et al. (2016) compute their efficiencies based on the ratio of ionizing luminosity to instantaneous gas mass. The difficulty with this technique is that the ionizing luminosity is a measure of stars formed ~ 3 – 5 Myr ago, rather than the instantaneous rate at which the gas that is currently present is forming stars. The high efficiency regions that Lee et al. (2016) identify are those associated with the largest and most luminous H II regions in the Milky Way, all of which have substantially disrupted their environments. Lee et al.’s method assumes that it is possible to map these giant bubbles one-to-one on to still-extant molecular clouds, neglecting the possibility that their present masses are not reflective of the mass of gas that went into making the ionizing stars. Such a discrepancy in mass could occur because the parent clouds have been disrupted into multiple pieces by stellar feedback, or because there have been substantial flows of mass in (ongoing accretion) or out (mass loss via feedback – Feldmann & Gnedin 2011) of the star-forming region.

In contrast, no studies that measure star formation rates using indicators other than ionizing luminosity, or that target embedded sources for which the cloud identification is much less uncertain, find a population of high efficiency clouds. Indeed, even using ionizing luminosity as a star formation tracer, but in external galaxies where there is no line of sight confusion and thus it is not necessary to try to assign individual H II regions to individual molecular clouds, Leroy et al. (2017) find $\epsilon_{\text{ff}} \lesssim 1$ per cent, and with a much smaller dispersion than Lee et al. (2016). This finding strongly supports the hypothesis that Lee et al.’s cloud matching procedure is the source of the discrepancy between their results and the rest of the literature. For this reason, we use the value of ϵ_{ff} found by all other techniques.

There is some subtlety in choosing f_{sf} and t_{ff} . Some authors have simply set $f_{\text{sf}} \approx 1$ and evaluated t_{ff} using the mid-plane density, and this approach is reasonable for starburst galaxies where the entire ISM is continuous, molecular, star-forming medium. However, such an approach is clearly not reasonable for galaxies like the Milky Way, where the mean density at the mid-plane is $\approx 1 \text{ cm}^{-3}$, but star formation occurs exclusively in molecular clouds that constitute only $f_{\text{sf}} \approx 30$ per cent of the mass, but are a factor of $\gtrsim 100$ denser, giving $t_{\text{ff}} \approx t_{\text{ff,mp}}/10$. Indeed, such an assumption is even problematic for galaxies on the star-forming main sequence at $z \sim 2$, since for some of these galaxies the mid-plane density implied by equation (23) is $\lesssim 10 \text{ cm}^{-3}$. This clearly cannot all be star-forming molecular material.

In our model we follow the approach set out in Forbes et al. (2014a), who base their model on the observations compiled by Krumholz et al. (2012). In this model, stars are assumed to form in a continuous medium with a free-fall time determined from $\rho_{\text{g,mp}}$ as long as the resulting star formation time-scale,

$$t_{\text{sf,T}} \equiv \frac{t_{\text{ff}}}{\epsilon_{\text{ff}}} = \frac{\pi Q}{4 f_{\text{g,Q}} \epsilon_{\text{ff}}} \sqrt{\frac{3 f_{\text{g,P}} \phi_{\text{mp}}}{2(1 + \beta)}} \frac{1}{\Omega}, \quad (31)$$

is shorter than $t_{\text{sf,max}} \approx 2 \text{ Gyr}$, the value that appears to result in galaxies like the Milky Way where the gas breaks up into individual molecular clouds whose densities are decoupled from the mean mid-plane density (Bigiel et al. 2008; Leroy et al. 2008, 2013). Following the terminology of Krumholz et al. (2012), we refer to the former case as the ‘Toomre regime’ and the associated time-scale $t_{\text{sf,T}}$ defined by equation (31) as the Toomre star formation time-scale, since when it applies the density in star-forming regions

is set by Toomre stability of the entire disc. We refer to the latter case as the ‘GMC regime’, since it applies when star-forming regions have densities determined by local considerations rather than global disc stability. Thus, we take the star formation rate to be

$$\dot{\Sigma}_* = f_{\text{sf}} \Sigma_g \max(t_{\text{sf},\text{T}}^{-1}, t_{\text{sf},\text{max}}^{-1}), \quad (32)$$

where the first case is the Toomre regime and the second is the GMC regime. In terms of the galactic orbital period, the condition for being in the Toomre regime is

$$t_{\text{orb}} < t_{\text{orb},\text{T}} \equiv \frac{8\epsilon_{\text{ff}} f_{g,Q}}{Q} \sqrt{\frac{2(1+\beta)}{3f_{g,P}\phi_{\text{mp}}}} t_{\text{sf},\text{max}} \quad (33)$$

$$= 35 f_{g,Q,0.5} f_{g,P,0.5}^{-1/2} \text{ Myr}, \quad (34)$$

where $t_{\text{orb}} = 2\pi/\Omega$ is the galactic orbital period, $f_{g,Q,0.5} \equiv f_{g,Q}/0.5$ and similarly for $f_{g,P,0.5}$, and the numerical evaluation uses the fiducial values given in Table 1.

The value of f_{sf} can be computed from theoretical models (e.g. Krumholz, McKee & Tumlinson 2009a; Krumholz et al. 2009b; McKee & Krumholz 2010; Krumholz 2013). For galaxies in the Toomre regime, one usually has $f_{\text{sf}} \approx 1$, but this is not true for galaxies in the GMC regime. For now we choose to leave f_{sf} as a free parameter. Finally, using equation (8), we have

$$\mathcal{G} = f_{\text{sf}} \frac{\sqrt{2(1+\beta)}}{\pi G Q} f_{g,Q} \left\langle \frac{p_*}{m_*} \right\rangle \Omega \sigma_g^2 \times \max \left(\sqrt{\frac{32(1+\beta)}{3\pi^2 f_{g,P} Q \phi_{\text{mp}}}} \epsilon_{\text{ff}} \Omega, \frac{1}{t_{\text{sf},\text{max}}} \right). \quad (35)$$

Note that we are implicitly neglecting other possible energy injection mechanisms, such as magnetorotational or thermal instability.

2.4 Radial transport

2.4.1 The transport rate equation

In a standard ‘top down’ derivation of the star formation law, the next step would be to equate the rates of loss from turbulent dissipation and gain from star formation feedback \mathcal{G} . Since these have different scalings – $\mathcal{L} \propto \Omega^2 \sigma_g^3$ and $\mathcal{G} \propto \epsilon_{\text{ff}} \Omega^2 \sigma_g^2$ (in the Toomre regime) or $\mathcal{G} \propto t_{\text{sf},\text{max}}^{-1} \Omega \sigma_g^2$ (in the GMC regime) – such equality can hold everywhere within the disc only if σ_g takes on a particular, fixed value (and hence Q is non-constant), or if ϵ_{ff} is non-constant. For example, Ostriker & Shetty (2011) make the former choice, while Faucher-Giguère et al. (2013) and Hayward & Hopkins (2017) make the latter. Neither option provides a particularly good match to observations, for the reasons discussed in Section 1.

Our model is based on the realization that there is an alternative source of energy, radial transport. Such transport injects energy at scales comparable to the gas scale height, which then cascades down to become turbulent on smaller scales. Krumholz & Burkert (2010) show that the time evolution of the gas velocity dispersion obeys

$$\frac{\partial \sigma_g}{\partial t} = \frac{\mathcal{G} - \mathcal{L}}{3\sigma_g \Sigma_g} + \frac{\sigma_g}{6\pi r \Sigma_g} \frac{\partial \dot{M}}{\partial r} + \frac{5(\partial \sigma_g / \partial r)}{6\pi r \Sigma} \dot{M} - \frac{1-\beta}{6\pi r^2 \Sigma_g \sigma_g} \Omega \mathcal{T}, \quad (36)$$

where \mathcal{T} is the torque exerted by non-axisymmetric stresses, and

$$\dot{M} = -\frac{1}{v_\phi(1+\beta)} \frac{\partial \mathcal{T}}{\partial r} \quad (37)$$

is the rate of inward mass accretion through the disc. Note that \dot{M} here and throughout refers explicitly to mass accretion *through* the disc rather than *on to* the disc from outside, unless explicitly stated otherwise. There is clear physical interpretation for equation (36). The first term on the right hand side is the net effect of star formation driving (\mathcal{G}) and dissipation of turbulence (\mathcal{L}), the second and third represent advection of kinetic energy as gas moves through the disc, and the final term represents transfer of energy from the galactic gravitational potential to the gas.

We pause here to comment on the physical assumptions that lie behind equation (36). This equation is simply the time- and azimuthally averaged version of the equation of energy conservation for a thin disc with a time-steady rotation curve, and it holds regardless of the nature of the torque \mathcal{T} . Thus it can apply equally well to gas transport driven by transient or steady spiral waves (as in a modern galaxy) or transport coming from the mutual torquing of giant clumps (as in a high- z galaxy). However, equation (36) does not include another energy source that is at least in principle possible: transfer of energy from stars to gas without any transport of the gas itself, for example due to stellar spiral arms or bars directly driving turbulent gas motions. That is, it is possible to ‘pay’ for an increase in gas kinetic energy by having the stars decrease their energy by either flowing down the potential well or decreasing their velocity dispersion, and such transfer could take place even if gas does not flow down the potential well, or even flows up it. Such star-to-gas direct transfer probably is important in some regions, particularly those with little gas and strong bars, as discussed in Section 4.4. However, numerous numerical simulations of both local (Agertz et al. 2009; Goldbaum et al. 2015, 2016) and high- z (Bournaud et al. 2007; Bournaud & Elmegreen 2009; Ceverino et al. 2010) galaxies offer strong evidence that direct star-to-gas energy transfer cannot be a dominant source of gas kinetic energy. These simulations show that gas does flow inward at roughly the rate predicted by our model, even when a live stellar disc and its spiral waves are included in the simulations, and, conversely, that turbulence and inflow occur even in simulations that do not include a massive stellar disc. Neither of these findings is consistent with the hypothesis that stellar driving rather than gas transport dominates the energy budget in most galaxies.

If we search for solutions where that gas is in energy equilibrium, $\partial \sigma_g / \partial t = 0$, then equation (36) implies that

$$\frac{\sigma_g^2}{2\pi r} \frac{\partial \dot{M}}{\partial r} + \frac{5\sigma_g \dot{M}}{2\pi r} \frac{\partial \sigma_g}{\partial r} - \frac{1-\beta}{2\pi r^2} \Omega \mathcal{T} = \mathcal{L} - \mathcal{G}. \quad (38)$$

This is a second-order ordinary differential equation in \mathcal{T} (since $d\dot{M}/dr$ involves the second derivative of \mathcal{T}), with $\mathcal{L} - \mathcal{G}$ as a forcing term. Physically valid solutions to this equation are subject to the constraint $\mathcal{T} \rightarrow 0$ as $r \rightarrow 0$, so that no torques are exerted (and thus no energy is added) at $r = 0$.

2.4.2 The critical velocity dispersion

Following Forbes et al. (2014a), we note that the solutions to this equation are only consistent with thermodynamic constraints when $\mathcal{L} \geq \mathcal{G}$, i.e. when the dissipation of turbulence is stronger than driving, so the forcing term is positive. If this inequality holds, then gravitationally driven turbulence transports mass inward and converts gravitational potential energy into turbulent motion at the rate required to maintain the gas in a state of marginal stability. In the opposite case, however, gravitational instability would be required to convert energy from random motions into a net outward trans-

port of mass, which is unphysical on thermodynamic grounds – the turbulence is assumed to be randomly oriented, so there is plausible physical mechanism by which it could self-organize to generate a net outward mass transport. If $\mathcal{L} = \mathcal{G}$ exactly, then driving by star formation is by itself sufficient to offset the decay of turbulence, and there is no gravitational instability or radial transport.

The condition that $\mathcal{L} = \mathcal{G}$ for a marginally stable disc with $Q = Q_{\min}$ is satisfied if the gas velocity dispersion (total thermal plus non-thermal) is

$$\sigma_g = \sigma_{\text{sf}} \equiv \frac{4f_{\text{sf}}\epsilon_{\text{ff}}}{\sqrt{3f_{g,P}\pi\eta\phi_{\text{mp}}\phi_Q\phi_{\text{nt}}^{3/2}}} \left\langle \frac{p_*}{m_*} \right\rangle \times \max \left[1, \sqrt{\frac{3f_{g,P}}{8(1+\beta)}} \frac{Q_{\min}\phi_{\text{mp}}}{4f_{g,Q}\epsilon_{\text{ff}}} \frac{t_{\text{orb}}}{t_{\text{sf,max}}} \right]. \quad (39)$$

With this definition, we can rewrite the equation for energy equilibrium, equation (38), as

$$\frac{\sigma_g^2}{2\pi r} \frac{\partial \dot{M}}{\partial r} + \frac{5\sigma_g \dot{M}}{2\pi r} \frac{\partial \sigma_g}{\partial r} - \frac{1-\beta}{2\pi r^2} \Omega \mathcal{T} = \mathcal{L} \left(1 - \frac{\sigma_{\text{sf}}}{\sigma_g} \right). \quad (40)$$

With the energy equation written in this way, the physical meaning of σ_{sf} becomes clear. It is the velocity dispersion that star formation alone is capable of maintaining, without any additional energy input from mass transport. As the velocity dispersion of the ISM approaches this limit, the net rate of turbulent dissipation diminishes, and the amount of gravitational transport required to maintain marginal stability does as well. The fraction of the energy supplied by star formation is simply $\sigma_{\text{sf}}/\sigma_g$, while the fraction supplied by gravity is $1 - \sigma_{\text{sf}}/\sigma_g$. Once the galaxy reaches $\sigma_g = \sigma_{\text{sf}}$ exactly, the mass inflow rate drops to 0, and the galaxy is no longer constrained to have $Q = Q_{\min}$; it can instead take on any value of $Q \geq Q_{\min}$.

We can also express the condition that $\mathcal{L} = \mathcal{G}$, and thus gravitational power shut off, in terms of the surface density. Combining equations (6) and (39), the critical surface density at which this occurs is

$$\Sigma_{\text{sf}} = \frac{8\sqrt{2(1+\beta)}f_{\text{sf}}\epsilon_{\text{ff}}}{\sqrt{3\pi G Q \eta \phi_{\text{mp}} \phi_Q \phi_{\text{nt}}^{3/2}}} \left\langle \frac{p_*}{m_*} \right\rangle \frac{f_{g,Q}}{f_{g,P}^{1/2}} \frac{1}{t_{\text{orb}}} \times \max \left[1, \sqrt{\frac{3f_{g,P}}{2(1+\beta)}} \frac{Q\phi_{\text{mp}}}{8f_{g,Q}\epsilon_{\text{ff}}} \frac{t_{\text{orb}}}{t_{\text{sf,max}}} \right]. \quad (41)$$

Transport shuts off wherever Σ_g falls below Σ_{sf} . Note that higher values of Q imply lower values of Σ_{sf} , i.e. the more gravitationally stable the disc, the lower the total surface density that can be maintained by star formation alone. The maximum surface density that can be sustained by star formation alone in a marginally stable disc is given by Σ_{sf} evaluated with $Q = Q_{\min}$.

Numerical evaluation of equations (39) and (41) requires some care due to the ϕ_{nt} term in the denominator. Our fiducial choice for this term is $\phi_{\text{nt}} = 1$, appropriate for highly supersonic gas ($\sigma_{\text{sf}} \gg \sigma_{\text{th}}$). In most cases this choice is not problematic. However, one regime of interest for our theory is H I-dominated regions like the outer Milky Way or the majority of $z = 0$ dwarfs, which have $f_{\text{sf}} \ll 1$. Examination of equation (39) would seem to suggest that sufficiently small values of f_{sf} will produce correspondingly small values of σ_{sf} , in which case the approximation that $\sigma_{\text{sf}} \gg \sigma_{\text{th}}$, and thus $\phi_{\text{nt}} \approx 1$, is no longer valid; indeed, for $\sigma_{\text{sf}} \rightarrow \sigma_{\text{th}}$ we have $\phi_{\text{nt}} \rightarrow 0$. Thus we cannot simply assume $\phi_{\text{nt}} = 1$ when evaluating equation (39) for H I-dominated regions; a more sophisticated approach is required.

If one substitutes the full definition $\phi_{\text{nt}} = 1 - (\sigma_{\text{th}}/\sigma_g)^2$ into equation (39), the resulting equation is a cubic in σ_g^2 . While we

can solve this exactly, the solution is extremely cumbersome and unenlightening. It is more useful to obtain the solution in the two limiting cases $\sigma_g \gg \sigma_{\text{th}}$ and $\sigma_g \rightarrow \sigma_{\text{th}}$; numerical solution of the full cubic shows that σ_{sf} transitions smoothly between the two limits. The solution for $\sigma_g \gg \sigma_{\text{th}}$ is simply what we would have obtained by naively plugging in $\phi_{\text{nt}} = 1$, which is

$$\sigma_{\text{sf}} = 11 \text{ km s}^{-1} f_{\text{sf}} f_{g,P,0.5}^{-1/2} \cdot \max \left(1, 1.0 f_{g,P,0.5}^{1/2} f_{g,Q,0.5}^{-1} t_{\text{orb},100} \right), \quad (42)$$

$$\Sigma_{\text{sf}} = 36 M_{\odot} \text{ pc}^{-2} f_{\text{sf}} f_{g,Q,0.5} f_{g,P,0.5}^{-1/2} t_{\text{orb},100}^{-1} \cdot \max \left(1, 1.0 f_{g,P,0.5}^{1/2} f_{g,Q,0.5}^{-1} t_{\text{orb},100} \right), \quad (43)$$

where $t_{\text{orb},100} = t_{\text{orb}}/100 \text{ Myr}$, for Σ_{sf} we have used $Q = Q_{\min}$, and for all quantities we have used the fiducial parameter choices as given in Table 1.

We treat the $\sigma_{\text{sf}} \approx \sigma_{\text{th}}$ limit by defining the Mach number \mathcal{M}_{sf} corresponding to σ_{sf} by

$$\sigma_{\text{sf}} = \sigma_{\text{th}} \sqrt{1 + \mathcal{M}_{\text{sf}}^2} \quad (44)$$

(so that $\sigma_{\text{th}} = \sigma_{\text{sf}}$ corresponds to $\mathcal{M}_{\text{sf}} = 0$), and solve equation (39) to first order in \mathcal{M}_{sf} . This gives

$$\mathcal{M}_{\text{sf}} = \left\{ \frac{4f_{\text{sf}}\epsilon_{\text{ff}}}{\sqrt{3f_{g,P}\pi\eta\phi_{\text{mp}}\phi_Q}} \left\langle \frac{p_*}{m_*} \right\rangle \frac{1}{\sigma_{\text{th}}} \cdot \max \left[1, \sqrt{\frac{3f_{g,P}}{8(1+\beta)}} \frac{Q_{\min}\phi_{\text{mp}}}{4f_{g,Q}\epsilon_{\text{ff}}} \frac{t_{\text{orb}}}{t_{\text{sf,max}}} \right] \right\}^{1/3} \quad (45)$$

$$= 0.60 f_{\text{sf},0.1}^{1/3} f_{g,P,0.5}^{1/6} \sigma_{\text{th},5}^{-1/3} \cdot \max \left(1, 1.0 f_{g,P,0.5}^{1/6} f_{g,Q,0.5}^{-1/3} t_{\text{orb},100}^{1/3} \right), \quad (46)$$

$$\Sigma_{\text{sf}} = \frac{\sqrt{8(1+\beta)}f_{g,Q}\sigma_{\text{th}}}{G Q_{\min} t_{\text{orb}}} \quad (47)$$

$$= 16 M_{\odot} \text{ pc}^{-2} f_{g,Q,0.5} \sigma_{\text{th},5} t_{\text{orb},100}^{-1}, \quad (48)$$

where $\sigma_{\text{th},5} = \sigma_{\text{th}}/5 \text{ km s}^{-1}$. Thus we find that, for the relatively modest star-forming fractions typical of H I-dominated regions, the maximum Mach number that can be sustained by star formation is of the order of 0.5. Since $\sigma_{\text{th}} \approx 5 \text{ km s}^{-1}$ in the warm neutral medium, this in turn implies overall velocity dispersions of ≈ 6 – 8 km s^{-1} . Thus we find that, regardless of the value of f_{sf} or various other parameters, our model predicts that the maximum velocity dispersion that can be sustained by star formation alone is $\sigma_{\text{sf}} \approx 6$ – 10 km s^{-1} . A corollary of this statement is that, if we observe a galaxy's velocity dispersion to be close to σ_{sf} , we can conclude that the turbulence within it is primarily powered by star formation, whereas if we observe the velocity dispersion to be $\gg \sigma_{\text{sf}}$, we can conclude that the turbulence is primarily powered by gravity. We also note that our finding that star formation at a rate consistent with the observed Kennicutt–Schmidt relation is capable of powering a velocity dispersion of $\approx 10 \text{ km s}^{-1}$ and no more is not new; several numerical simulations of supernova-driven turbulence have reached the same conclusion from their numerical experiments (e.g. Jounge, Mac Low & Bryan 2009; Kim, Kim & Ostriker 2011; Kim & Ostriker 2015b).

2.4.3 The steady-state mass inflow rate

With σ_{sf} defined, we are now in a position to calculate the mass inflow rate for galaxies with $\sigma_g > \sigma_{\text{sf}}$ and $Q = Q_{\text{min}}$. Krumholz & Burkert (2010) obtained a transport equation analogous to equation (40) in the limit $\sigma_g \gg \sigma_{\text{sf}}$, and for constant β (i.e. fixed rotation curve index) showed that it admits an analytic steady state solution with σ_g and \dot{M} independent of radius. Numerical solution of the full time-dependent system (equation 36) shows that galaxies tend to approach this steady state (Forbes et al. 2012, 2014a), so motivated by this result we look for similar solutions (β, σ, \dot{M} all independent of r) for the more general case given by equation (40).⁴ A solution of this form must have $\mathcal{T} = -\dot{M}v_\phi r$, and inserting this into equation (40) we immediately obtain that the mass inflow rate must be

$$\dot{M} = \dot{M}_{\text{ss}} \equiv \frac{4(1+\beta)\eta\phi_Q\phi_{\text{nt}}^{3/2}}{(1-\beta)GQ_{\text{min}}^2} f_{g,Q}^2 \sigma_g^3 \left(1 - \frac{\sigma_{\text{sf}}}{\sigma_g}\right) \quad (49)$$

$$= 0.71 f_{g,Q,0.5}^2 \sigma_{g,10}^3 \text{M}_\odot \text{yr}^{-1} \cdot \left(1 - \frac{\sigma_{\text{th}}^2}{\sigma_g^2}\right)^{3/2} \left(1 - \frac{\sigma_{\text{sf}}}{\sigma_g}\right), \quad (50)$$

where $\sigma_{g,10} = \sigma_g/10 \text{ km s}^{-1}$; the numerical evaluation uses the fiducial values in Table 1, except that we have retained the explicit dependence on ϕ_{nt} because it is important in H I-dominated regions, as explained above. The quantity \dot{M}_{ss} is the steady-state mass inflow rate that is required to keep a galactic disc in energy equilibrium. Thus we expect galactic discs with $\sigma_g \approx 10 \text{ km s}^{-1}$, and thus slightly above σ_{sf} and σ_{th} to have mass inflow rates of order $1 \text{ M}_\odot \text{yr}^{-1}$. As σ_g decreases and approaches both σ_{sf} and σ_{th} , the inflow rate rapidly falls to zero, while as it increases the inflow rate rises as $\dot{M}_{\text{ss}} \propto \sigma_g^3$.

We show some example equilibrium solutions in Fig. 1; the examples are representative of the range of galaxies to which we can apply our model, including a local dwarf, a local spiral similar to the Milky Way, a local ULIRG, and a high redshift star-forming disc. The exact parameters for each model are given in Table 2. All models use $f_{g,Q} = f_{g,P} = 0.5$, and an inner radius of 0.1 kpc. We use a value of f_{sf} computed using the KMT+ model of Krumholz (2013) with a clumping factor $f_c = 1$, since the gas surface densities here are the true ones rather than a beam-diluted average. To apply this theory we require a value for the mid-plane stellar plus dark matter density. If the rotation curve index is independent of radius, and is dominated by stars and dark matter, then the minimum density at the mid-plane required to produce the rotation curve is

$$\rho_{\text{min}} = \frac{v_\phi^2}{4\pi G r^2} (2\beta + 1). \quad (51)$$

The true value is likely to be somewhat higher, since ρ_{min} applies for a spherical mass distribution, which we would expect if dark matter alone were dominating the rotation curve; we therefore adopt a stellar density $\rho_* = 2\rho_{\text{min}}$. We compute the thermal velocity dispersion σ_{th} as $\sigma_{\text{th}} = f_{\text{sf}}\sigma_{\text{th,mol}} + (1 - f_{\text{sf}})\sigma_{\text{th,WNM}}$, where $\sigma_{\text{th,mol}} = 0.2 \text{ km s}^{-1}$ (appropriate for molecular gas) and $\sigma_{\text{th,WNM}} = 5.4 \text{ km s}^{-1}$ appropriate for warm neutral gas. The results illustrate the qualitative

behaviour of the model: local spirals and dwarfs with modest velocity dispersions and modest star formation rates have $\sigma_g/\sigma_{\text{sf}} \approx 1$, and as a result also have low-mass inflow rates, $\sim 10^{-2} \text{ M}_\odot \text{yr}^{-1}$ for the dwarf and $\sim 1 \text{ M}_\odot \text{yr}^{-1}$ for the spiral. In contrast, rapidly star-forming ULIRGs and high-redshift galaxies have high $\sigma_g/\sigma_{\text{sf}}$ and high inflow rates. The turbulence in these galaxies is driven almost entirely by inflow.

2.5 Equilibria without transport or without feedback

It is worth considering the alternatives to our model that result from omitting either feedback or transport, in order to demonstrate why both are important. First consider omitting feedback, as in Krumholz & Burkert (2010). This amounts to setting $p_*/m_* = 0$, and thus all the relations we have derived continue to apply, but with $\sigma_{\text{sf}} = 0$ and $\Sigma_{\text{sf}} = 0$.

The other alternative is models without transport, which require that $\mathcal{G} = \mathcal{L}$. As noted above, this requirement can be satisfied in two ways. One is that we can keep the star formation law (equation 30) fixed. In the GMC regime we have $\mathcal{G} \propto \Sigma_g \sigma_g$ while $\mathcal{L} \propto \Sigma_g^2 \sigma_g$, and thus $\mathcal{G} = \mathcal{L}$ is possible only for a single value of Σ_g ; since real galaxies clearly do not all have a single surface density, we discount this solution and instead focus on the Toomre regime. In the Toomre regime we have $\mathcal{G} = \mathcal{L}$ whenever $\sigma_g = \sigma_{\text{sf}}$ (equation 39). This implies that

$$Q = f_{g,Q} \frac{\kappa \sigma_{\text{sf}}}{\pi G \Sigma_g} = \frac{8\sqrt{2(1+\beta)}}{\sqrt{3}\pi\eta\phi_{\text{mp}}\phi_Q\phi_{\text{nt}}^{3/2}} f_{\text{sf}} \epsilon_{\text{ff}} \left\langle \frac{p_*}{m_*} \right\rangle \frac{f_{g,Q}}{f_{g,P} G \Sigma_g t_{\text{orb}}} \quad (52)$$

$$= 3.6 f_{g,Q,0.5} f_{g,P,0.5}^{-1/2} t_{\text{orb},100}^{-1} \Sigma_{g,10}^{-1}, \quad (53)$$

where $\Sigma_{g,10} = \Sigma_g/10 \text{ M}_\odot \text{pc}^{-2}$. Thus if we do not include transport and keep the star formation law fixed, the model still predicts that $Q \approx 1$ for Solar Circle conditions ($\Sigma_{g,10} \approx 1$, $t_{\text{orb},100} \approx 2$). However, for conditions like those found in ULIRGs ($\Sigma_{g,10} \sim 100$, $t_{\text{orb},10} \sim 3$) or high- z star-forming discs ($\Sigma_{g,10} \sim 10$, $t_{\text{orb},100} \sim 1$), the predicted value of Q is much smaller than unity.

Conversely, we can hold Q fixed and treat the quantity $f_{\text{sf}}\epsilon_{\text{ff}}$ as a free parameter, and use the relation $\mathcal{G} = \mathcal{L}$ to solve for it. In this case only the Toomre regime exists, and it is characterized by a star formation efficiency per free-fall time

$$\epsilon_{\text{ff}} = \frac{\sqrt{3}\pi\eta\phi_{\text{mp}}\phi_Q\phi_{\text{nt}}^{3/2}}{4f_{\text{sf}}} \left\langle \frac{p_*}{m_*} \right\rangle^{-1} f_{g,P}^{-1/2} \sigma_g \quad (54)$$

$$= 0.027 f_{\text{sf}}^{-1} f_{g,P,0.5}^{-1/2} \sigma_{g,10}. \quad (55)$$

Thus ϵ_{ff} is ~ 1 per cent for $\sigma_g \approx 10 \text{ km s}^{-1}$, but rises to $\gtrsim 10$ per cent for the higher velocity dispersions typically seen in ULIRGs or high-redshift star-forming discs. Note that equation (54) is identical, up to factors of order unity, to equation 37 of Faucher-Giguère et al. (2013).

3 COMPARISON TO OBSERVATIONS

We can use our steady state model to calculate a wide range of observables, and in this section we compare the model predictions to observations. We also compare contrasting models without transport and without feedback, in order to highlight how including both

⁴An important subtlety: in writing equation (46) we evaluated ϕ_{nt} using $\sigma_g = \sigma_{\text{sf}}$. This is the correct approach to finding the value of σ_{sf} that can be sustained by star formation alone. However, in equation (40), σ_{sf} must be evaluated using the actual value of σ_g , which may be larger.

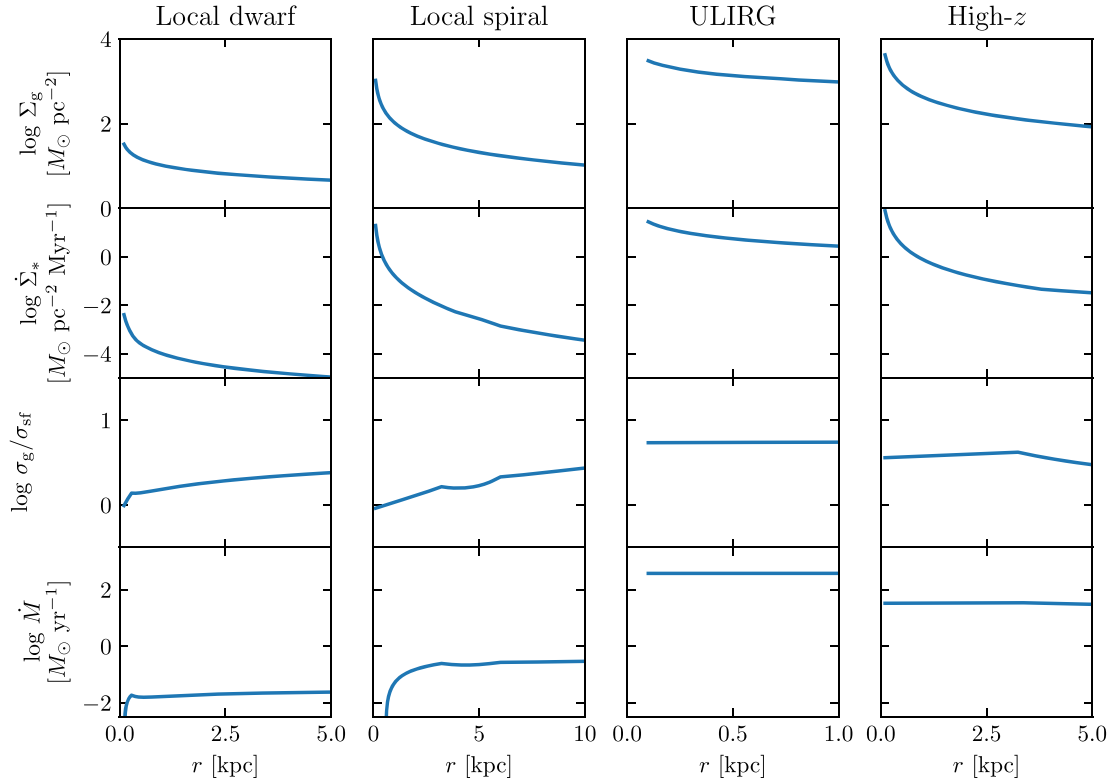


Figure 1. Example solutions for our fiducial model, using the parameters chosen for local dwarfs, local spirals, ULIRGs, and high-redshift star-forming galaxies given in Table 2. Note that different columns have different x -axis ranges. Rows show, from top to bottom, gas surface density Σ_g , star formation surface density Σ_* , ratio of gas velocity dispersion σ_g to dispersion provided by star formation σ_{sf} , and mass inflow rate \dot{M} .

Table 2. Parameters for example solutions. Note that r_{out} is the outermost radius at which we compute the solution, and Z' is the metallicity normalized to Solar used in the KMT+ model for f_{sf} (see the main text).

Parameter	Local dwarf	Local spiral	ULIRG	High- z
σ_g (km s $^{-1}$)	6	10	60	40
r_{out} (kpc)	5	10	1	5
v_ϕ at r_{out} (km s $^{-1}$)	60	200	250	200
β	0.5	0	0.5	0
Z'	0.2	1	1	1

mechanisms alters the results. Specifically, throughout this section we will consider four different models, to which we refer as follows:

Transport+feedback. This is our fiducial model. It has $\epsilon_{ff} = 0.015$ and two branches: $Q = Q_{min}$ with $\sigma_g > \sigma_{sf}$ (or equivalently $\Sigma_g > \Sigma_{sf}$), and $Q \geq Q_{min}$ with $\sigma_g = \sigma_{sf}$ (or $\Sigma_g \leq \Sigma_{sf}$).

No-feedback. This is identical to the transport+feedback model, except that $\sigma_{sf} = 0$ and $\Sigma_{sf} = 0$, so $Q = Q_{min}$ under all circumstances. This model is similar to the one proposed by Krumholz & Burkert (2010).

No-transport, fixed ϵ_{ff} . A model without transport, with $\epsilon_{ff} = 0.015$ fixed but Q allowed to vary freely. In this model the value of Q is given by equation (52). This model is similar to the one proposed by Ostriker & Shetty (2011).

No-transport, fixed Q . A model without transport, with $Q = Q_{min}$ fixed, but ϵ_{ff} allowed to vary freely. In this model, ϵ_{ff} takes on the value given by equation (54). This model is similar to the one proposed by Faucher-Giguère et al. (2013).

For each of these models we compute the star-forming fraction f_{sf} using the formalism of Krumholz (2013), with a clumping factor $f_c = 5$ (since we are now dealing with beam-diluted kpc-scale observations), Solar metallicity, and a stellar density equal to four times the minimum value given in equation (51).

3.1 The star formation law

A first test of any model of star formation is the prediction it makes for the star formation law, the relation between the gas content of galaxies and their star formation behaviour. Observationally, the star formation law can be expressed as a correlation between the surface density of star formation and either the gas surface density alone, or the gas surface density divided by the galactic orbital period. It can be measured averaged over entire galaxies, or measured in spatially resolved patches of galaxies. A successful model should be able to reproduce all these observed correlations.⁵

⁵One can also define a local star formation law, which relates the local rate of star formation within a given cloud to its volumetric properties (density, virial ratio, etc.). There are significant observational constraints on this relationship as well, as discussed in Section 2.3.2, but in this paper we have used these constraints as an input to the model, not an output, and thus our model cannot be said to predict this relation. However, the local volumetric star formation relation is distinct from the projected, area-averaged one, and it is perfectly possible to match observations of one without successfully reproducing the other. Indeed, in the following sections we will encounter a number of models that do exactly that. Thus the models we consider do constitute predictions for the areal star formation law.

3.1.1 Spatially resolved observations

First consider spatially resolved observations. For both the transport+feedback model and the no-feedback model, the star formation rate at each point in the disc is described by equation (32) with $\epsilon_{\text{ff}} = 0.015$. If we omit star formation feedback, only the $Q = Q_{\text{min}}$ solution branch exists, whereas in our fiducial transport+feedback model we can have $Q > Q_{\text{min}}$ for $\Sigma_g < \Sigma_{\text{sf}}$. (Recall that we are limiting our attention to discs in energy equilibrium without significant external energy input; external stimulation can produce $Q \gg Q_{\text{min}}$ – Inoue et al. 2016.) In practice, however, this makes relatively little difference in the star formation law unless we adopt $Q \gg Q_{\text{min}}$, though we shall see that it makes a considerable difference for other observables. Thus for simplicity we simply adopt $Q = Q_{\text{min}}$ everywhere, in which case the transport+feedback and no-feedback models are the same.

In the no-transport, fixed ϵ_{ff} model, the value of Q is given by equation (52). Substituting this into equation (32) (and recalling that the GMC regime does not exist in this case) gives a star formation law

$$\dot{\Sigma}_* = \pi G \eta \phi_{\text{mp}}^{1/2} \phi_Q \phi_{\text{nt}}^{3/2} \left\langle \frac{P_*}{m_*} \right\rangle^{-1} \Sigma_g^2. \quad (56)$$

This relation is identical up to factors of order unity to equation 10 of Ostriker & Shetty (2011), which is not surprising since it is based on the same physical assumptions.

In the no-transport, fixed Q model, we instead have $Q = Q_{\text{min}}$ and a value of ϵ_{ff} given by equation (54). Inserting this into equation (32) gives

$$\dot{\Sigma}_* = \pi^2 G \eta \phi_{\text{mp}}^{1/2} \phi_Q \phi_{\text{nt}}^{3/2} f_{g,P}^{-1} Q_{\text{min}} \left\langle \frac{P_*}{m_*} \right\rangle^{-1} \Sigma_g^2, \quad (57)$$

with no dependence on the orbital period. This equation is identical up to factors of order unity with equation 18 of Faucher-Giguère et al. (2013). It is also nearly identical to equation (56) – the scalings are the same, and the leading coefficients differ only by a factor of $\pi Q_{\text{min}}/f_{g,P} \sim 1$.⁶ Thus for the purposes of comparing to observation we need only consider one form of the no-transport model. An important point to note is that the factor f_{sf} vanishes in both equations (56) and (57), as it must, since in these models the star formation rate always self-adjusts to maintain force and energy balance without any help from transport.

We therefore have two prospective predictions of the star formation law to consider: our fiducial transport+feedback model [equation (32) evaluated with $Q = Q_{\text{min}}$], and a no-transport model (equation 56). We plot the model predictions together with resolved observations in Fig. 2. The fiducial model does a good job of

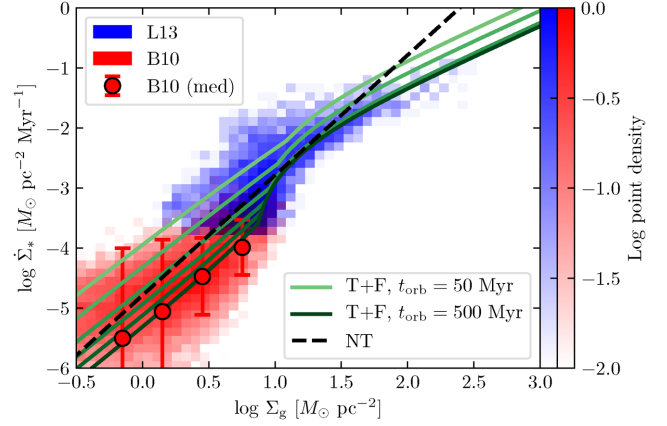


Figure 2. Comparison between theoretical model predictions of the star formation law and observation of nearby galaxies at ~ 1 kpc resolution. Lines represent models; solid green lines are the transport+feedback model (equation 32; T+F in the legend), evaluated for orbital times evenly spaced in logarithm from $t_{\text{orb}} = 50$ –500 Myr, with lighter colours (towards the top) corresponding to shorter orbital times. The dashed black line is the no-transport model (equation 56; NT in the legend), which has no dependence on orbital time. All models use the fiducial parameters given in Table 1, and we compute the star-forming molecular fraction f_{sf} from the KMT+ model (Krumholz 2013) as in Section 2.4.3, using a Solar-normalized metallicity $Z' = 1/3$, appropriate for dwarfs and outer discs. Coloured histograms show observations; colours indicate the distribution of individual pixels in the $\Sigma_g - \dot{\Sigma}_*$ plane for inner galaxies (blue; Leroy et al. 2013; L10 in the legend) and outer galaxies and dwarfs (red; Bigiel et al. 2010; B10 in the legend); red circles with error bars show the median and scatter of the outer galaxy data.

describing the data for plausible input values of t_{orb} – the range plotted is 50–500 Myr, which roughly covers the span of the data, which include regions from galactic centres to outskirts. In particular, the fiducial model properly captures the curvature seen in the data, where the slope of $\dot{\Sigma}_*$ versus Σ_g is clearly steeper in the range $\log(\Sigma_g/M_{\odot} \text{pc}^{-2}) \approx 0.5$ –1 than at either higher or lower surface density. In comparison, the no-transport model produces noticeably too steep a slope compared to the observations. The mismatch is most apparent at surface densities of $\sim 100 M_{\odot} \text{pc}^{-2}$, where a model without transport tends to overpredict the star formation rate by more than an order of magnitude. Moreover, the no-transport model is unable to reproduce the curvature of the data associated with the atomic- to molecular-dominated transition at $\approx 10 M_{\odot} \text{pc}^{-2}$, because the star formation rate is insensitive to the thermal or chemical state of the ISM in this case.

3.1.2 Unresolved observations

For unresolved observations, we have access only to the surface densities of gas and star formation averaged over the entire disc, and to the rotation period at the disc edge. To compare our model to such data, we must take care to average the model predictions in the same way. Doing so precisely requires knowing the radial variation of the gas surface density and all the other factors in equation (32), which is obviously not possible for unresolved observations. However, we can make a rough estimate for the effects of area-averaging by considering a disc with radially-constant values of the gas velocity dispersion σ_g , rotation curve index β , stability parameter Q , the various gas fractions $f_{g,Q}$ and $f_{g,P}$, and the star-forming fraction f_{sf} . From equation (6), we can see that such a disc

⁶Despite the fact that equations (56) and (57) make nearly identical predictions for the star formation law, the routes by which they arrive at these predictions are quite different. In deriving equation (56), one assumes that the star formation efficiency per free-fall time is constant. The scaling $\dot{\Sigma}_* \propto \Sigma_g^2$, implying a star formation time-scale that declines as $t_{\text{sf}} \propto 1/\Sigma_g$, arises because the gas velocity dispersion is constant, and this leads to a mid-plane density that increases as the square of Σ_g . This in turn leads to a free-fall time that scales as Σ_g^{-1} . In contrast, in deriving equation (57) one assumes that the mid-plane density is not varying, since it is fixed by the condition $Q = Q_{\text{min}}$. Instead, the efficiency of star formation is proportional to Σ_g . Thus equation (56) corresponds to a picture where the star formation process is not sensitive to the gas surface density in a galaxy, but the mid-plane density is, while equation (57) arises from a picture where the mid-plane density is independent of gas surface density, but the star formation process is not.

has a surface density that varies with radius as $\Sigma_g \propto v_\phi/r \propto r^{\beta-1}$. Thus if the disc extends from inner radius 0 to some finite outer edge, the area-averaged surface density is larger than the surface density at the edge by a factor of $2/(1+\beta)$.

The effects of area-averaging on the star formation rate depend on the star formation law. First consider our transport+feedback case or a case with no feedback, both of which follow equation (32). In discs where the majority of the star formation occurs in the GMC regime, where the star formation time-scale is constant, the area-averaged star formation rate is larger than the value at the outer edge by the same factor. However, in portions of the disc in the Toomre regime, equation (32) gives a star formation rate per unit area that varies as $\dot{\Sigma}_* \propto \Sigma_g \Omega \propto r^{2(\beta-1)}$. For $\beta \neq 1$, this gives an area-averaged star formation surface density that is larger than the value at the disc edge by a factor of $1/\beta$. Thus the area-averaged version of equation (32) can be written as

$$\langle \dot{\Sigma}_* \rangle \approx f_{\text{sf}} \langle \Sigma_g \rangle \phi_a \max \left[\frac{4\epsilon_{\text{ff}} f_{g,Q}}{\pi Q} \sqrt{\frac{2(1+\beta)}{3f_{g,P} \phi_{\text{mp}}}} \Omega_{\text{out}}, t_{\text{sf,max}}^{-1} \right], \quad (58)$$

where the angle brackets indicate area averages, and Ω_{out} is the angular velocity at the outer edge of the star-forming disc.

The factor ϕ_a represents the difference in the factors by which area-averaging enhances the star formation rate compared to the gas surface density. It is unity for discs in the GMC regime; in the Toomre regime it is $(1+\beta)/2\beta$ for $\beta \neq 0$. The case of a flat rotation curve, $\beta = 0$, requires special consideration, since in the Toomre regime such a disc has a total star formation rate that diverges logarithmically near the disc centre. As noted by Krumholz & Burkhardt (2016), this divergence is a result of the unphysical assumption that a flat rotation curve can continue all the way to $r = 0$; such a rotation curve has a divergent shear, which in turn makes the mid-plane density required to maintain constant Q , and thus the total star formation rate, diverge. If one instead considers the more realistic case of a rotation curve that is flat only to some finite inner radius r_0 , then the area-averaged star formation rate is larger than the value at the disc edge at radius r_1 by a factor of $2\ln(r_1/r_0)$, and thus $\phi_a = \ln(r_1/r_0)$. In practice this factor cannot be that large, because extended discs with flat rotation curves also tend to have much of their star formation in the GMC regime, where this extra enhancement does not occur. For this reason, we will adopt $\phi_a = 2$ as a fiducial value, recognizing that it can be somewhat larger or smaller depending on the rotation curve and how much of the disc is in the Toomre regime.

We can proceed analogously to derive the offsets between the local and disc-averaged star formation laws for the alternative no-transport models. In the no-transport, fixed Q model, the star formation law obeys $\dot{\Sigma}_* \propto \Sigma_g^2$ (equation 57), we again have $\dot{\Sigma}_* \propto r^{2(\beta-1)}$, and the factor ϕ_a is therefore the same as in the transport+feedback case. In the no-transport, fixed ϵ_{ff} model (equation 56), we cannot calculate the run of Σ_g versus radius from our assumptions, because the values of gas surface density Σ_g and velocity dispersion σ_g are independent of one another. Thus we cannot directly calculate ϕ_a without making an additional assumption about the radial variation of Σ_g . For simplicity, however, we will assume the same radial variation as in the $Q = Q_{\text{min}}$ models, and thus obtain the same ϕ_a . Thus the area-averaged versions of equation (56) or equation (57) are identical to the original versions, with an added factor of ϕ_a on the right-hand side.

We compare the model predictions to a sample of unresolved observations culled from the literature in Fig. 3. In plotting the data we use the CO-H₂ conversion factor α_{CO} recommended by Daddi

et al. (2010b), and we discuss this choice further in Appendix A. We see that the transport+feedback model agrees reasonably well with the data, while the no-transport model produces noticeably too steep a slope in both $\dot{\Sigma}_*$ versus Σ_g and $\dot{\Sigma}_*$ versus Σ_g/t_{orb} . The model including transport fares significantly better.

3.2 Gas velocity dispersions

A second observable that we can predict is the gas velocity dispersions in galaxies, and its correlation with star formation. Consider a galaxy with a constant gas velocity dispersion σ_g . Using the star formation relation equation (32) and our definition of Q (equation 6), we can write the star formation rate per unit area as

$$\dot{\Sigma}_* = f_{\text{sf}} \frac{\sqrt{8(1+\beta)} f_{g,Q}}{G Q} \frac{\sigma_g}{t_{\text{orb}}^2} \cdot \max \left[\frac{8\epsilon_{\text{ff}} f_{g,Q}}{Q} \sqrt{\frac{2(1+\beta)}{3f_{g,P} \phi_{\text{mp}}}}, \frac{t_{\text{orb}}}{t_{\text{sf,max}}} \right]. \quad (59)$$

As in Section 3.1.2, we can derive an unresolved version of this relation under the assumption that Q , f_{sf} , and β are constant with radius. Integrating over radius, we find that the total star formation rate is

$$\dot{M}_* = \sqrt{\frac{2}{1+\beta}} \frac{\phi_a f_{\text{sf}}}{\pi G Q} f_{g,Q} v_{\phi,\text{out}}^2 \sigma_g \cdot \max \left[\sqrt{\frac{2(1+\beta)}{3f_{g,P} \phi_{\text{mp}}}} \frac{8\epsilon_{\text{ff}} f_{g,Q}}{Q}, \frac{t_{\text{orb,out}}}{t_{\text{sf,max}}} \right], \quad (60)$$

where $v_{\phi,\text{out}}$ and $t_{\text{orb,out}}$ are the circular velocity and orbital period evaluated at the outer edge of the star-forming disc.

In our transport+feedback model, equations (59) and (60) are to be evaluated with $Q = Q_{\text{min}}$ if $\sigma_g > \sigma_{\text{sf}}$. If $\sigma_g = \sigma_{\text{sf}}$, then we can have any $Q \geq Q_{\text{min}}$. Finally, values of $\sigma_g < \sigma_{\text{sf}}$ are not possible in equilibrium.

Our alternative models have a variety of other behaviours. In the no-feedback model equations (59) and (60) are the same, but with $\sigma_{\text{sf}} = 0$, and thus $Q = Q_{\text{min}}$ for all σ_g , and all values of σ_g are allowed. Conversely, in the no-transport, fixed ϵ_{ff} model, σ_g can only take on the one value σ_{sf} ; no other values are allowed in equilibrium, and $\dot{\Sigma}_*$ are \dot{M}_* are independent of this. Finally, in the no-transport, fixed Q model, we have $Q = Q_{\text{min}}$, and we must use equation (54) for ϵ_{ff} . Substituting this value of ϵ_{ff} into equations (59) and (60) gives the relationships

$$\dot{\Sigma}_* = \frac{8(\beta+1)\pi\eta\sqrt{\phi_{\text{mp}}\phi_{\text{nt}}^3\phi_Q}}{G Q^2 \langle p_*/m_* \rangle f_{g,P}} \frac{\sigma_g^2}{t_{\text{orb}}^2} \quad (61)$$

$$\dot{M}_* = \frac{4\eta\sqrt{\phi_{\text{mp}}\phi_{\text{nt}}^3\phi_Q\phi_a}}{G Q^2 \langle p_*/m_* \rangle} \frac{f_{g,Q}^2}{f_{g,P}} v_{\phi,\text{out}}^2 \sigma_g^2. \quad (62)$$

Note that the transport+feedback and no-feedback models both predict $\dot{M}_* \propto \sigma_g$ for $\sigma_g > \sigma_{\text{sf}}$, while the two no-transport models predict very different scalings: no relationship between \dot{M}_* and σ_g for the no-transport, fixed ϵ_{ff} model, and a much stronger scaling, $\dot{M}_* \propto \sigma_g^2$, for the no-transport, fixed Q model. This difference, first pointed out by Krumholz & Burkhardt (2016), provides a very clear observational signatures that can be used to distinguish models with

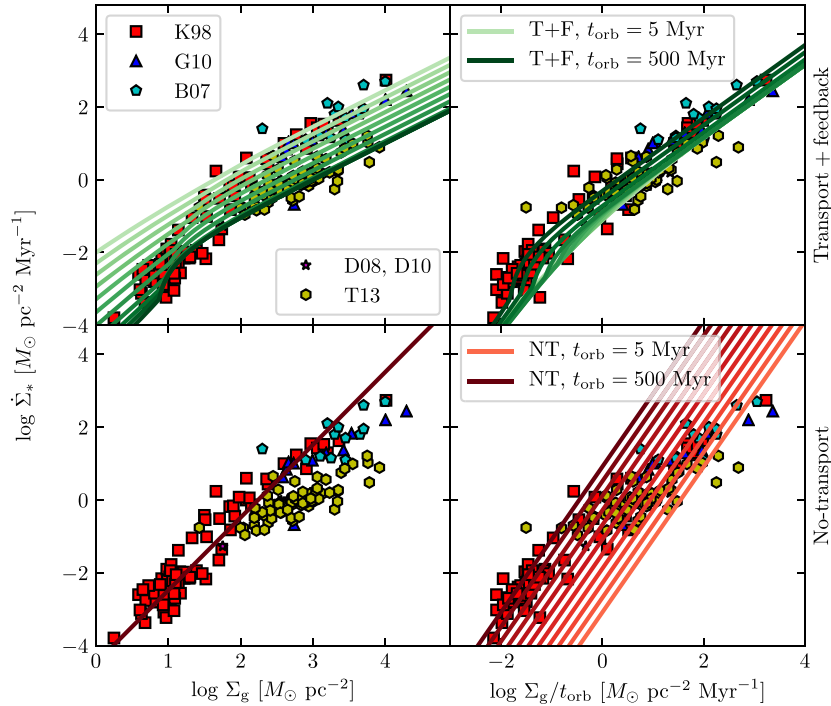


Figure 3. Comparison between theoretical model predictions of the star formation law and observations of marginally resolved galaxies, with one measurement per galaxy. In each panel, the y axis shows the star formation rate per unit area $\dot{\Sigma}_*$. In the left column the x-axis shows total gas surface density Σ , while in the right it shows Σ/t_{orb} . The top row shows our fiducial transport+feedback (T+F) model (equation 58), while the bottom row shows the no-transport (NT) model (equation 56). Coloured lines show the model predictions, evaluated using orbital periods from $t_{\text{orb}} = 5$ Myr (lighter colours) to 500 Myr (darker colours), with lines evenly spaced in $\log t_{\text{orb}}$; note that only one line appears in the lower left panel, because the relationship between Σ_g and $\dot{\Sigma}_*$ is independent of t_{orb} in the no-transport model. All plots have star-forming ISM fraction f_{sf} computed from the KMT+ model as in Section 3.1, and use the fiducial values given in Table 1, except that we use $\beta = 0.5$ rather than 0 because a substantial part of the sample consists of circumnuclear starbursts which are in regions with rising rather than flat rotation curves. Coloured points, which are the same in each panel, show data culled from the following sources: local galaxies from Kennicutt (1998, K98 in the legend), $z \sim 2$ sub-mm galaxies from Bouché et al. (2007, B07 in the legend), and galaxies on and somewhat above the star-forming main sequence at $z \sim 1-3$ from Daddi et al. (2008, 2010b, D08, D10 in the legend), Genzel et al. (2010, G10 in the legend), and Tacconi et al. (2013, T13 in the legend). The observations have been homogenized to a Chabrier (2005) IMF and the convention for α_{CO} suggested by Daddi et al. (2010a); see Appendix A and Krumholz et al. (2012) for details.

and without transport.⁷ The physical origin of this difference is easy to understand. The star formation rate is $\dot{\Sigma}_* = \epsilon_{\text{ff}} \Sigma_g / t_{\text{ff}}$. For a fixed rotation curve, orbital time, and gas fraction, the gas surface density scales as $\Sigma_g \propto \sigma_g / Q$, and the mid-plane density scales as $\rho_{\text{mp}} \propto Q^{-2}$, implying that the free-fall time-scales as $t_{\text{ff}} \propto Q$, with no explicit dependence on σ_g . The overall scaling is therefore $\dot{\Sigma}_* \propto \epsilon_{\text{ff}} \sigma_g / Q^2$. The difference between the transport+feedback and no-transport models then follows from their assumed variations in ϵ_{ff} and Q . Our fiducial transport+feedback model has ϵ_{ff} and Q both constant, so we obtain a linear scaling $\dot{\Sigma}_* \propto \sigma_g$. The no-transport, fixed ϵ_{ff} model has constant σ_g and varying Q , so it predicts no relationship between $\dot{\Sigma}_*$ and σ_g , with all the variations in star formation rate being driven by changes in Q . The no-transport, fixed Q model has $\epsilon_{\text{ff}} \propto \sigma_g$ (equation 54), so it predicts $\dot{\Sigma}_* \propto \sigma_g^2$.

To compare the various theoretical models to observations, since the results depend on f_{sf} , v_{ϕ} , t_{orb} , β , and the gas fraction, we must choose values for these parameters. For unresolved observations,

the data will also depend on φ_a . Following our approach in Section 2.4.3, we consider four different possibilities that should be broadly representative of the ranges these parameters can take. We label these cases local dwarf, local spiral, ULIRG, high- z , with the final case intended to be typical of the observed high-redshift star-forming discs. We summarize the chosen parameters in Table 3; all other parameters have their fiducial values as specified in Table 1, and we compute the ISM thermal velocity dispersion σ_{th} as in Section 2.4.3. Broadly speaking, the dwarf is characterized by a high gas fraction, a low star-forming fraction, and a low orbital velocity, and has $\varphi_a = 1$ because it is entirely in the GMC regime; the ULIRG has a high orbital velocity, a high gas fraction, and a short orbital period. It has a larger value of φ_a , since it is entirely in the Toomre regime. The local spiral and high- z star-forming disc have properties intermediate between these extremes, with the high- z system having a higher gas fraction, star-forming fraction, and φ_a . Finally, we note that each set of model parameters is found only in some finite range of star formation rates; for example, objects with $f_{\text{sf}} = 0.2$, as we adopt for our local dwarf case, do not generally produce star formation rates of $10 M_{\odot} \text{ yr}^{-1}$. For this reason we use each set of properties only up to some maximum or down to some minimum star formation rate. We give the limiting values $\dot{M}_{*,\text{min}}$ and $\dot{M}_{*,\text{max}}$ in Table 3 as well.

⁷The scaling between \dot{M}_* and gas fraction for the no-transport, fixed Q model that we obtain here is slightly different from that given in Krumholz & Burkhardt (2016), because here we have treated this model as having fixed total Q . In contrast, the Faucher-Giguère et al. (2013) model to which Krumholz & Burkhardt (2016) compare assumed fixed Q_g .

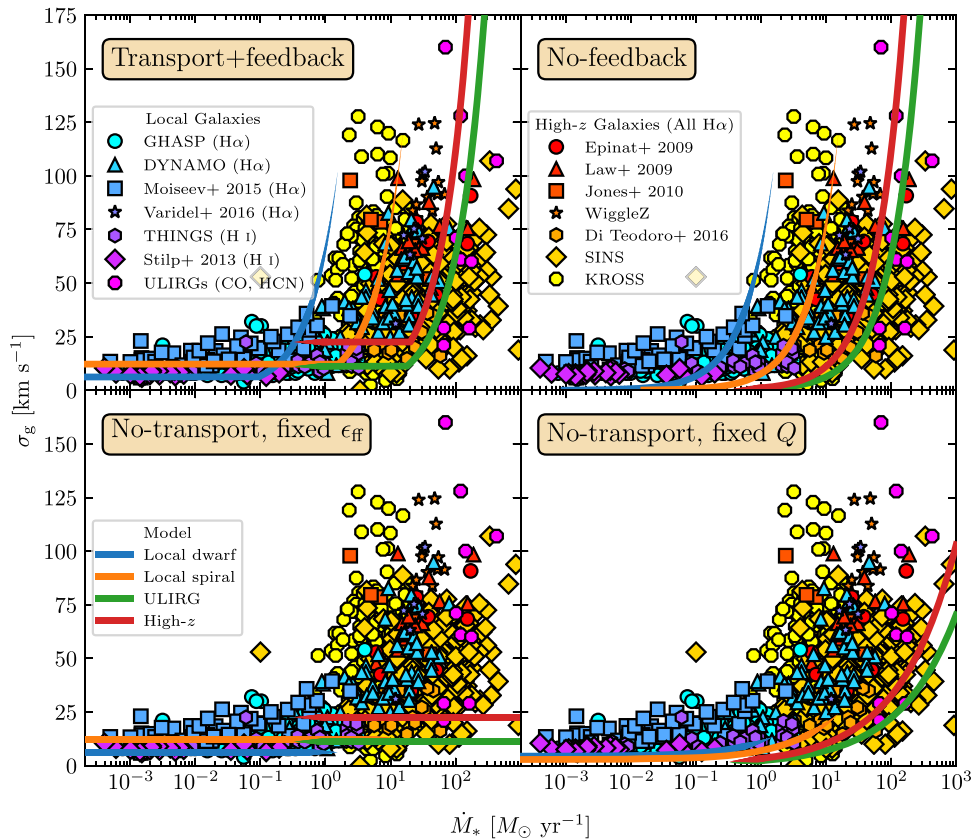


Figure 4. Comparison between the observed correlation between gas velocity dispersion and star formation rate and theoretical models. Solid lines represent theoretical models, with the model plotted indicated in each panel; clockwise from top left, these are the transport+feedback model, the no-feedback model, the no-transport, fixed Q model, and the no-transport, fixed ϵ_{ff} model. The lines shown are for four representative sets of parameters, corresponding roughly to those appropriate for local dwarf galaxies, local spiral galaxies, ULIRGs, and high- z star-forming discs; the lines fade outside the range of star formation rates for which they are applicable. See Table 3 and the main text for details. The coloured points represent observations, and are the same in every panel. Data shown include: $\text{H}\alpha$ observations of local galaxies from two surveys (GHASP, Epinat, Amram & Marcelin 2008, and DYNAMO, Green et al. 2014) as well as smaller studies (Moiseev et al. 2015; Varidel et al. 2016); H I observations of nearby galaxies from THINGS (Leroy et al. 2008; Walter et al. 2008; Ianjamasimanana et al. 2012) and from the survey of dwarfs by Stilp et al. (2013); a compilation of molecular line observations of nearby ULIRGs (Downes & Solomon 1998; Sanders et al. 2003; Veilleux et al. 2009; Scoville et al. 2015, 2017); $\text{H}\alpha$ observations of high-redshift galaxies from the samples of Epinat et al. (2009), Law et al. (2009), Lemoine-Busserolle et al. (2010), and the WiggleZ (Wisnioski et al. 2011) and SINS-KMOS-3D (Wisnioski et al. 2015; Wuyts et al. 2016) surveys at $z \sim 1$ –3; $\text{H}\alpha$ observations of lensed galaxies at $z \sim 2$ –3 from Jones et al. (2010); a sample of galaxies at $z \sim 1$ from the KMOS survey (Wisnioski et al. 2015) as analysed by Di Teodoro, Fraternali & Miller (2016), and a sample from the KROSS survey (Stott et al. 2016; Johnson et al. 2018). Full details on the data set are given in Appendix B.

Table 3. Parameter values used for the theoretical models shown in Fig. 4. See the main text for details.

Parameter	Local dwarf	Local spiral	ULIRG	High- z
f_{sf}	0.2	0.5	1.0	1.0
v_{ϕ} (km s $^{-1}$)	100	220	300	200
t_{orb} (Myr)	100	200	5	200
β	0.5	0.0	0.5	0.0
$f_{g,Q} = f_{g,P}$	0.9	0.5	1.0	0.7
φ_a	1	1	2	3
$\dot{M}_{*,\text{min}}$ (M $_{\odot}$ yr $^{-1}$)	–	–	1	1
$\dot{M}_{*,\text{max}}$ (M $_{\odot}$ yr $^{-1}$)	0.5	5	–	–

We compare the predictions of our transport+feedback model and the three alternatives to observations in Fig. 4. Details on the observations and our processing of them are given in Appendix B. We see that the transport+feedback model is in generally good agreement

with the observations at both low and high star formation rates. In particular, it captures the behaviour that the velocity dispersion reaches a floor of ≈ 10 km s $^{-1}$ at low star formation rates,⁸ while increasing rapidly with star formation rate at star formation rates above a few M $_{\odot}$ yr $^{-1}$. (The dwarf case rises to high velocity dispersions at star formation rates that are too low, but this is an artefact of choosing to fix f_{sf} , when in fact no galaxy with a star formation rate of $\gtrsim 1$ M $_{\odot}$ yr $^{-1}$ has $f_{\text{sf}} = 0.2$ and $v_{\phi} = 100$ km s $^{-1}$, as we have adopted in the dwarf case.) In contrast, the alternative models all have an obvious failing. The no-feedback model does well at high star formation rates, but fails to capture the floor imposed by star formation at low star formation rates, instead predicting that the velocity dispersion should fall to very small values. Conversely,

⁸Some of the data, particularly the GHASP and Moiseev et al. (2015) samples, have $\sigma_g \approx 20$ km s $^{-1}$ at low star formation rates, but this is likely an artefact of using $\text{H}\alpha$ -estimates of σ_g , as the other tracers are all systematically lower. See Appendix B for further discussion.

the no-transport, fixed ϵ_{ff} model correctly captures the behaviour at low star formation rates, but fails to reproduce the observed increase in velocity dispersion at higher star formation rates. Finally, the no-transport, fixed Q model has qualitatively correct behaviour, but seriously underpredicts the velocity dispersion at all star formation rates. This failure is a direct result of having too steep a relationship between star formation and gas surface density, as seen in Section 3.1.

3.3 Mass transport

A final observable, or at least potential observable, that we can predict is the correlation between mass inflow rate and physical properties of the star-forming disc. As discussed in Section 1, at present we have direct detections of inflow rates only for a handful of nearby galaxies, but we can compare our model to these, and predict the results of future observations and simulations.

To make predictions for this correlation using our transport+feedback model, for any choice of σ_g and ancillary parameters (gas fraction, rotation curve, etc.), we can use equation (49) to compute the mass inflow rate, and equation (60) to compute the corresponding star formation rate. When $\sigma_g \gg \sigma_{\text{sf}}$, this leads to a predicted scaling between inflow and star formation $\dot{M} \propto \dot{M}_*^3 / (v_\phi^6 f_{\text{sf}}^3)$, with the coefficient depending on the gas fraction, rotation curve index, and whether the galaxy is in the Toomre regime. We can use the same method for the no-feedback model simply by setting $\sigma_{\text{sf}} = 0$, but the results are only slightly different, so we refrain from showing them. Models without transport, depending on one's perspective, either predict that the inflow rate should be zero or make no predictions at all regarding its value.

We show predicted mass inflow rates for the same four example cases used in Section 3.2 in Fig. 5. Given the extremely strong scaling of the inflow rate with v_ϕ and f_{sf} , it is not surprising that the example cases cover a very wide range of possible inflow rates for a given star formation rate. Thus the model is consistent with the data, in that the data lie near the ‘local spiral’ parameter choices, where we expect them, but this is a relatively weak statement.

A more interesting test is to normalize out the dependence on the rotation curve velocity and star forming fraction. Our model predicts that $\dot{M}_* \propto f_{\text{sf}} v_\phi^2$, whereas \dot{M} is independent of these two parameters (except very close to $\sigma_g = \sigma_{\text{sf}}$), and thus our model makes a much stronger prediction for the correlation of \dot{M} with $\dot{M}_* / (f_{\text{sf}} v_\phi^2)$. For the observations, we take v_ϕ from Table 2 of Leroy et al. (2013); as a proxy for the star-forming fraction f_{sf} , we use $f_{\text{sf}} = \dot{\Sigma}_* / (\Sigma_g / 2 \text{ Gyr})$, where the values of Σ_g and $\dot{\Sigma}_*$ are taken from the same table.

We plot the correlation of \dot{M} with $\dot{M}_* / (f_{\text{sf}} v_\phi^2)$ in Fig. 6. We see that both the data and the models cluster much more tightly than in the plot of \dot{M} versus \dot{M}_* (note the difference in x-axis range in Fig. 5 versus Fig. 6), and that the data remain quite close to the model lines. The remaining difference between the theoretical model results for dwarfs and spirals are due to the differences in gas fraction and rotation curve index between the two, and the observations, with one possible exception, are well within the space of models that are plausibly spanned by the gas fraction and rotation curve index range of nearby galaxies. The outlier at the left side of the plot is NGC 2903. (In Fig. 5 this is the point to the left at the lowest inflow rate and second lowest star formation rate.) For this galaxy, Schmidt et al. (2016) state that their fits are likely unreliable due to complex kinematics driven by a strong bar.

While this comparison is encouraging, the observations from Schmidt et al. (2016) cover only a very narrow range of galaxy

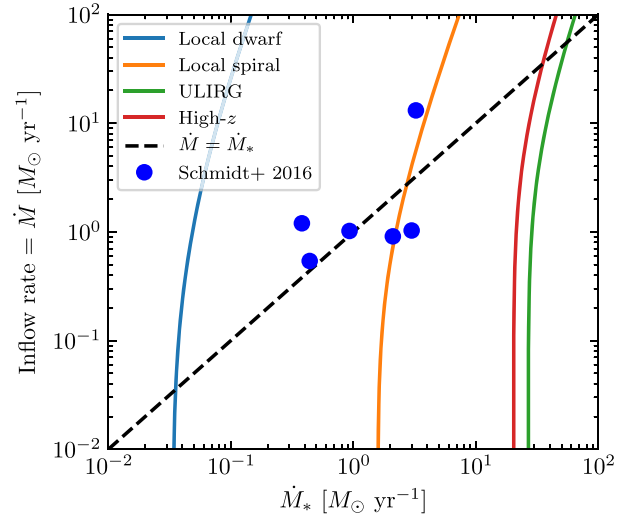


Figure 5. Inflow rate through this disc versus star formation rate. Blue points show the observations of Schmidt et al. (2016), while lines show the predictions of our transport+feedback model for the four example cases whose parameters are given in Table 3.

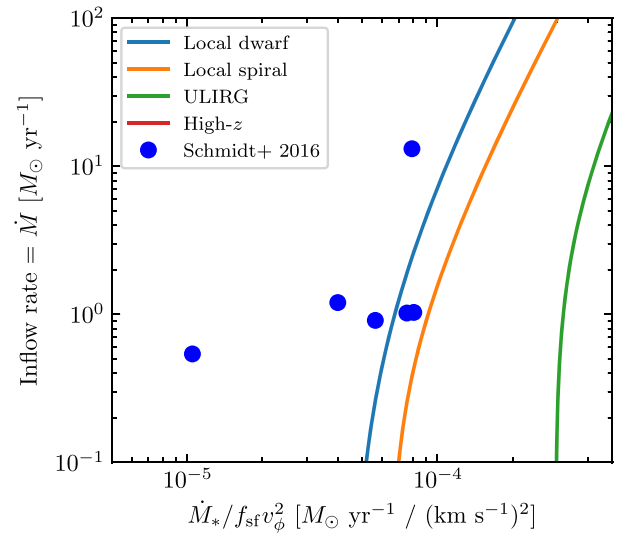


Figure 6. Same as Fig. 5, but with the star formation rate normalized by $f_{\text{sf}} v_\phi^2$. Note that, compared to Fig. 5, the x-axis range has been compressed from 4 to 2 dex.

properties. Stronger tests are clearly warranted. The most obvious target for such comparisons are nearby starburst galaxies, for which our transport+feedback model predicts large inflow rates. These galaxies are nearby enough that one can make high resolution CO or HCN maps from which kinematic information can be extracted. While the kinematics are likely to be complex and thus analysis will be more difficult than it is for quiescent spirals, the predicted signal is also much larger.

4 IMPLICATIONS FOR GALAXY FORMATION

4.1 Equilibrium inflow and star formation

The model we present here has important implications for the formation of disc galaxies. To explore these further, we begin by chang-

ing our perspective. Thus far we have been developing a theory that takes as input the gas content and other ancillary properties of a galaxy, and returns as output the inflow rate \dot{M} that is required to maintain pressure and energy equilibrium in a galaxy of given physical parameters, and the star formation rate \dot{M}_* that accompanies this equilibrium configuration. This framing of the problem is appropriate if we are interested in behaviours on time-scales short compared to the gas consumption or flow time-scales. However, over cosmological time-scales it is more natural to think of the inflow rate as given. Gas will fall on to the central galaxy at a rate dictated by cosmological structure formation, recycling of gas ejected at earlier epochs, and processing through the gaseous halo. At least for large galaxies where the gas consumption time-scale is shorter than the Hubble time, the galaxy will adapt its structure to be in equilibrium given this inflow rate, a point previously made by Dekel et al. (2009).

We can use this picture to calculate the evolution of galaxies' velocity dispersions. Tidal torque theory suggests that the specific angular momentum of infalling gas will increase with halo mass and cosmic time, and thus gas accreting on to galaxies tends to arrive at their outskirts, where orbital times, and thus star formation time-scales, are relatively long. Some of this gas will form stars before gravitational instability moves it inward, but some will be forced to flow inward towards the galactic bulge. This is particularly true at high redshift, when galaxies accrete quickly. We can therefore approximate that the inflow rate must be comparable to the infall rate (including in this rate recycling of gas ejected from the galaxy but not the halo), which in turn dictates the velocity dispersion in the galaxy, as suggested by Genel, Dekel & Cacciato (2012). In practice, we can calculate this velocity dispersion from equation (49), setting $\dot{M} = \dot{M}_{g,acc}$, where $\dot{M}_{g,acc}$ is the gas accretion rate on to the galaxy.

However, from the velocity dispersion and the galaxy rotation curve we can in turn compute the run of gas surface density, and thence the star formation rate. For a simple case of radially-constant β , f_{sf} , and gas fractions, we can compute this analytically using equation (60). A more realistic calculation would consider the time and radial variation of the gas fraction and star-forming ISM fraction, and such models have had significant success (e.g. Forbes et al. 2012, 2014a; Tonini et al. 2016), but this level of complexity requires semi-analytic solution, and our goal here is qualitative insight. For this reason, we choose to neglect these complications and simply ignore radial- and time-variation. Doing so allows us to compute an equilibrium star formation rate in the disc of the galaxy as a function of a galaxy's inflow rate and rotation curve, which in turn are functions of the halo mass and redshift. We emphasize that this is the star formation rate in the disc; in a simple equilibrium picture, the total star formation rate must equal the accretion rate minus the rate of mass ejection by star formation or black hole feedback. The procedure we have just outlined provides a means of computing the disc star formation rate from the inflow rate and the rotation curve, and with the remaining mass balance coming from a combination of star formation in a bulge and ejection of gas through galactic winds.

4.2 Cosmological halo evolution

To make use of the simple picture outlined above, we must have methods to calculate the halo accretion rate, circular velocity, and orbital period as a function of redshift. First consider the accretion rate. In the interest of simplicity we neglect the contribution from gas recycling, and just attempt to calculate the infall rate of pristine gas.

To first order this is determined by the dark matter accretion rate, which in the context of a Λ CDM cosmology⁹ can be calculated with the extended Press–Schechter (EPS) formalism, with some additional calibration from simulations. Following Krumholz & Dekel (2012) and Forbes et al. (2014a), we adopt the approximate dark matter accretion rates found by Neistein & Dekel (2008) and Bouché et al. (2010):

$$\dot{M}_{h,12} \approx -\alpha M_{h,12}^{1+\beta} \dot{\omega}, \quad (63)$$

where $M_{h,12}$ and $\dot{M}_{h,12}$ are the halo mass and accretion rate normalized to $10^{12} M_\odot$, and ω is the self-similar time variable of the EPS formalism (i.e. ω is time measured in units of the linear growth time for structure), whose time derivative is well fitted by

$$\dot{\omega} \approx -0.0476[1+z+0.093(1+z)^{-1.22}]^{2.5} \text{Gyr}^{-1}. \quad (64)$$

The functional form of equation (63) follows from the EPS formalism, and the value of β follows from the power spectrum of density fluctuations, while α is a free parameter to be calibrated from full dark matter simulations. The Neistein & Dekel (2008) fit, updated to current cosmological parameters, gives $\alpha = 0.628$ and $\beta = 0.14$. With these parameters, the accretion rate evaluates numerically to approximately

$$\dot{M}_h \approx 39 M_{h,12}^{1.1} (1+z)^{2.2} M_\odot \text{yr}^{-1}, \quad (65)$$

at $z < 1$, with a slightly steeper slope with $1+z$ at higher z . In practice, however, rather than use this approximate expression we generate histories of mass versus redshift by direct numerical integration of equation (63).

The next step in estimating the baryonic accretion rate is to correct for the efficiency with which baryons penetrate the hot halo of the galaxy. Following Forbes et al. (2014a), we compute the gas accretion rate using the model of Faucher-Giguère, Kereš & Ma (2011),

$$\dot{M}_{g,acc} = \epsilon_{in} f_b \dot{M}_h \quad (66)$$

$$\epsilon_{in} = \min \left[\epsilon_0 M_{h,12}^{\beta_{M_h}} (1+z)^{\beta_z}, \epsilon_{max} \right], \quad (67)$$

where $f_b \approx 0.17$ is the universal baryon fraction, and the parameters $(\epsilon_0, \beta_{M_h}, \beta_z, \epsilon_{max}) = (0.31, -0.25, 0.38, 1)$ are the results of fits to cosmological simulations. This combined with equation (63) enables us to compute the mass accretion rate for an arbitrary halo. The formula for ϵ_{in} is the result of a fit to the results of a series of SPH simulations run by Faucher-Giguère et al. (2011), and it is calibrated to be most accurate for $z > 2$. Despite this we will continue to use it at lower z ; examination of Faucher-Giguère et al.'s fig. 9 suggests that it remains reasonably accurate down to $z \approx 0$ for haloes below $10^{12} M_\odot$, but that it overestimates the accretion rates for more massive haloes at low redshift by a factor of a few.

In addition to the accretion rate, we need the circular velocity and orbital period, or, equivalently, the circular velocity and disc radius. We also follow Forbes et al. (2014a) in computing the characteristic

⁹Throughout this section we assume a cosmology with $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $h = 0.71$, and $\sigma_8 = 0.81$. We use this rather than more recent *Planck* cosmological parameters because we do not have calibrations of the accretion formulae we adopt below for the more recent cosmological parameters. In practice this will make little difference, since our goal in this section is to develop a rough intuitive model rather than perform precision calculations.

radius of the disc as

$$r_d \approx 0.035 r_{\text{vir}} = 0.035 \times 163 M_{h,12}^{1/3} (1+z)^{-1} \text{ kpc}, \quad (68)$$

where r_{vir} is the virial radius, the numerical value of 163 kpc is for a halo overdensity of 200, and the coefficient of 0.035 is roughly consistent with the findings by Kravtsov (2013) and Somerville et al. (2018) that galaxies have ratios of half-mass to virial radius $r_h/r_{\text{vir}} \simeq 0.015\text{--}0.018$.¹⁰ For the circular velocity, we note that a Navarro, Frenk & White (1997) profile has a maximum circular velocity (Mo, van den Bosch & White 2010)

$$v_{c,\text{max}} \approx 0.465 \sqrt{\frac{c}{\ln(1+c) - c/(1+c)}} v_{\text{vir}} \quad (69)$$

$$v_{\text{vir}} \approx 117 M_{h,12}^{1/3} (1+z)^{1/2} \text{ km s}^{-1}, \quad (70)$$

where c is the halo concentration; the numerical coefficients are again for a halo overdensity of 200, and we adopt $c = 10$ as a fiducial value. The true circular velocity will be somewhat larger than this because in the star-forming parts of galaxies the baryons contribute non-negligibly to the gravitational potential. We very roughly adopt a relation

$$v_\phi = \phi_v v_{c,\text{max}} \quad (71)$$

with $\phi_v = 1.4$, which gives $v_\phi = 200 \text{ km s}^{-1}$ for a $10^{12} M_\odot$ halo at $z = 0$. This is a crude approximation, but, as noted above, our goal here is a qualitative toy model, not a precise calculation. The orbital time follows immediately from r_d and v_ϕ .

4.3 Model results and interpretation

We now use the formalism of Section 4.2 to compute the mass and accretion histories of a range of haloes with present-day masses of $M_{h,0} = 10^{12}\text{--}10^{13} M_\odot$. For each one we compute the velocity dispersion σ_g , the velocity dispersion that can be supported by star formation σ_{sf} , and the disc star formation rate $\dot{M}_{*,\text{disc}}$ as outlined in Section 4.1.

We restrict our attention to haloes in this mass range for three reasons. First, haloes substantially smaller than this are not observable beyond the local Universe, while those substantially larger host clusters rather than single galaxies. Thus the observational sample beyond $z = 0$ is mostly limited to this mass range. Secondly, it is not clear if the equilibrium assumption is applicable in haloes smaller than $\sim 10^{12} M_\odot$. These host dwarf galaxies, and the gas consumption time-scale in modern-day dwarfs is generally comparable to or longer than the Hubble time (e.g. Bolatto et al. 2011; Hunter et al. 2012; Jameson et al. 2016). This might also be the case at higher redshift, but this is uncertain because the gas consumption time-scale depends on both the star formation rate and the rate at which star formation drives gas out via galactic winds, and the latter is highly uncertain for low-mass galaxies beyond the local universe (Forbes et al. 2014b). Limiting our attention to haloes above $10^{12} M_\odot$ avoids this issue. A related point is that the mass loading factor for these haloes is unlikely to be $\gtrsim 1$, so we need not adopt a complex model to treat this phenomenon either. Third and finally, to evaluate our models we require values of f_{sf} and σ_{th} , and these depend on the molecular fraction in the ISM. The dependence of

¹⁰To be precise, since our equilibrium models have gas surface density profiles $\Sigma_g \propto 1/r$ for flat rotation curves, implying a mass that scales as r , the outer radius should be exactly twice the half-mass radius. Thus our coefficient of 0.035 corresponds to $r_h/r_{\text{vir}} = 0.0175$.

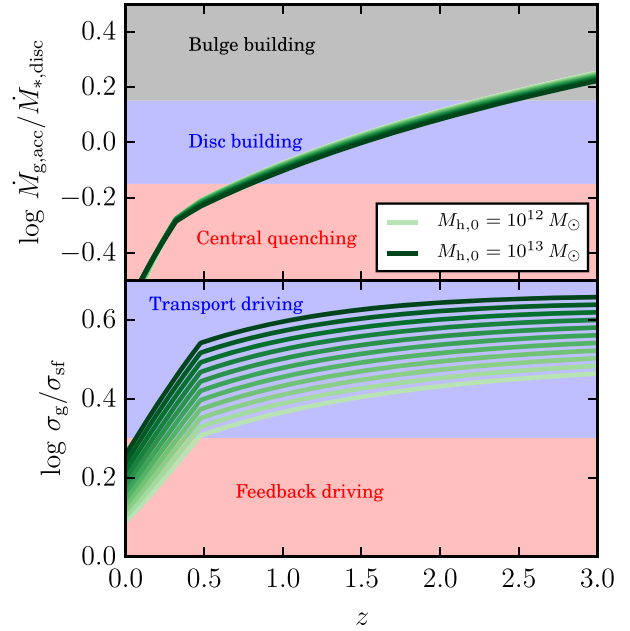


Figure 7. Evolution of the ratio of gas accretion rate (equation 66) to disc star formation rate (equation 60; top panel) and gas velocity dispersion (equation 49) to star formation-supported velocity dispersion (equation 39; bottom panel) as a function of redshift. Each line represents the evolutionary path of a particular halo, with the lightest colour (bottom lines in the lower panel) corresponding to a halo with a present-day mass of $M_{h,0} = 10^{12} M_\odot$, and the darkest (top lines in the lower panel) to a halo with a present-day mass of $M_{h,0} = 10^{13} M_\odot$. Intermediate lines are uniformly spaced by 0.1 dex in $M_{h,0}$. The inflection points visible at $z \approx 0.5$ corresponds to where haloes switch from star formation occurring mainly in the Toomre regime (at higher z) to occurring mainly in the GMC regime (at lower z). Shaded regions in the upper panel indicate regimes of bulge building, disc building, and central quenching, and in the lower panel indicate regions of transport-driven versus feedback-driven turbulence; see the main text for details.

this fraction on halo mass and redshift is highly complex and substantially uncertain (e.g. Obreschkow et al. 2009; Fu et al. 2010; Lagos et al. 2011; Krumholz & Dekel 2012; Forbes et al. 2014a). However, we expect that the molecular fraction will be smallest in small haloes at low redshift, since these combine low metallicity and low gas surface density. At $z = 0$, we observe that haloes of mass $10^{12} M_\odot$ (Milky Way-sized) host galaxies with $f_{\text{H}_2} \sim 0.5$ within the scale radius. By restricting our attention to haloes with present-day masses above this limit, we stay in the part of parameter space where the ISM is at least marginally molecule-dominated, and thus we can adopt $f_{\text{sf}} \approx 1$ and $\sigma_{\text{th}} \approx 0.2 \text{ km s}^{-1}$ without making too large an error. In contrast, $10^{11} M_\odot$ haloes (LMC-sized) have present-day molecular fractions ~ 0.1 (e.g. Jameson et al. 2016). Treating them would require that we adopt a model for the time evolution of the molecule fraction, which is beyond the scope of this paper.

4.3.1 Star formation

We are interested in two diagnostic ratios from these models, which we plot in the upper and lower panels of Fig. 7. The first of these is $\dot{M}_{g,\text{acc}}/\dot{M}_{*,\text{disc}}$, the ratio between the rate of accretion on to and then through the disc and the rate at which that accretion flow should convert into stars as it moves through the disc; Dekel &

Burkert (2014) describe this quantity as the ‘wetness factor’. If $\dot{M}_{g,acc} \gg \dot{M}_{*,disc}$, then the rate of mass flow on to the disc from the IGM, and through it towards the galactic centre, greatly exceeds the rate at which we expect that flow to convert into stars. Consequently, the majority of the flow will not be converted to stars before it reaches the galactic centre. It may still be ejected in outflows, but unless all of it is lost in this fashion, a substantial mass flux will reach the bulge region. Consequently, an era when $\dot{M}_{g,acc} \gg \dot{M}_{*,disc}$ should correspond to an era when galaxies are building up their bulges.

Now suppose the reverse holds, $\dot{M}_{g,acc} \ll \dot{M}_{*,disc}$. Taken literally this would mean that the star formation rate exceeds the gas accretion rate. Of course such a configuration cannot represent a steady state in which the rate of gas flow through the galaxy matches the rate of gas accretion into it, which violates the assumption we made in deriving $\dot{M}_{*,disc}$. To understand what occurs in this regime, it is helpful to consider what happens as a galaxy evolves, with our intuition guided by the results of more detailed time-dependent models (e.g. Forbes et al. 2014a). From Fig. 7 we see that at early times haloes have $\dot{M}_{g,acc}/\dot{M}_{*,disc} \gg 1$, and that this ratio gradually decreases to ~ 1 . When this ratio is ~ 1 , gas only barely reaches the galactic centre before the last of it is consumed and turned into stars; all star formation therefore occurs in the disc. As the gas supply tapers off over cosmic time and the ratio tries to drop even further, the gas supply is insufficient to keep up with the rate of consumption into stars. The equilibrium between supply and consumption is easiest to maintain in the outer parts of galaxies, both because this is where the majority of the gas lands, and because this is where the star formation rate per unit area is smallest. Thus the failure of equilibrium is likely to occur inside out: less and less of the gas that is accreting on to the galaxy will be able to reach the centre before transforming into stars. This reduction in central gas surface density in turn reduces $\dot{M}_{*,disc}$ compared to the value we have computed under the assumption of constant radial mass flux, thereby maintaining $\dot{M}_{g,acc} \sim \dot{M}_{*,disc}$. The price of this balance is that the centre of the galaxy ceases star formation, and thus quenches.

Examining the upper panels of Fig. 7, we see haloes show a clear progression from bulge building at high-redshift to disc building at intermediate redshift, to central quenching near the present day. We should not put too much weight on the exact redshift range over which this transition occurs, particularly since our fitting formula for gas accretion on to haloes likely overestimates the rate of cold gas flow at lower redshift. None the less, qualitatively this progression is a natural explanation for a commonly observed phenomenon: galaxies that transition from the blue, star-forming cloud to the quenched, red sequence do so by ceasing their star formation from the inside out, after a central stellar bulge builds up (e.g. Cheung et al. 2012; Fang et al. 2012, 2013; Genzel et al. 2014; Nelson et al. 2016; Belfiore et al. 2017). Models that include radial transport of gas via gravitational instability are able to reproduce this qualitative behaviour (e.g. Cacciato et al. 2012; Forbes et al. 2012; Forbes et al. 2014a; Stevens, Croton & Mutch 2016; Tonini et al. 2016), but the analytic model we develop here allows a particularly simple and straightforward explanation for both the inside-out quenching phenomenon and the redshift at which it occurs.

However, we end this section by cautioning that this simple steady state, quasi-equilibrium picture almost certainly misses some of the complications that occur in real cosmological galaxy formation. At least for haloes at and slightly above the more massive end of the range we consider, which tend to quench at $z \sim 2$, both observations (Barro et al. 2013, 2016, 2017; Tacchella et al. 2015, 2018) and simulations (Zolotov et al. 2015; Tacchella et al. 2016a,b)

suggest that galaxies pass through a phase of ‘compaction’, where the central gas surface density is driven to very high values, before finally quenching. Such compaction events are not captured in our simple steady state model, which may be more applicable to galaxies in less massive haloes such as those that host the Milky Way.

4.3.2 Turbulence driving

The other diagnostic ratio of interest is σ_g/σ_{sf} . Recall that $1 - \sigma_{sf}/\sigma_g$ is the fraction of the energy required to maintain the turbulence that comes from star formation feedback (equation 39). Thus $\sigma_g = 2\sigma_{sf}$ corresponds to star formation feedback and transport (gravity) contributing equally to the turbulence. In the lower panel of Fig. 7, we show the time evolution of σ_g/σ_{sf} , with the regions where transport driving and feedback driving dominate the energy budget highlighted. Note that, under our simplifying assumption that the inflow rate is always non-zero, we cannot ever reach $\sigma_g = \sigma_{sf}$, corresponding to the point where transport driving ceases completely.

The most obvious trend in Fig. 7 is that more massive haloes are further into the transport-driving regime, and spend more of their evolutionary history there. Massive galaxies, by virtue of their large accretion rates, tend to have high gas surface densities that require high velocity dispersions to maintain. Such velocity dispersions can only be maintained by gravitational power. The second clear trend is that σ_g/σ_{sf} drops as we approach $z = 0$. This is driven partly by a drop off in cosmological accretion rates, which produce less gas-rich discs that require lower values of σ_g to remain marginally stable. It is also partly by galaxies transitioning from the Toomre regime to the GMC regime in their star formation, which puts a ceiling on the depletion time and thus a floor on σ_{sf} . This effect is visible as the downturn in σ_g/σ_{sf} below $z \approx 0.5$.

Thus the qualitative picture to which we come is that the transition from transport-driven turbulence to feedback-driven turbulence depends primarily on galaxy mass, and secondarily on redshift, with transport-driving dominating at high mass and high- z .

4.4 Limits of model applicability

We conclude this discussion by considering to what types of galaxies, or within which parts of galaxies, our model applies. As noted in Section 2, our model assumes that the gas inflow rate is able to self-adjust to maintain marginal gravitational stability. The numerous simulations discussed in Section 1.2 leave little doubt that this self-adjustment works in one direction: if the gas becomes gravitationally unstable, it will develop non-axisymmetric structures that exert torques and drive a net inflow, thereby raising the velocity dispersion and pushing the system back towards stability. However, our model also assumes the converse, that in galaxies that are gravitationally stable there will not be net radial transport. There are clearly locations where this is not true, such as the inner few kpc of the Milky Way. In this region the gas fraction is so small that the gas effectively acts like a passive tracer moving in the fixed stellar potential. Both the gas and the combined gas-star disc are Toomre-stable, but the stars are arranged in a bar that, depending on the galactocentric radius, either forces gas to shock by preventing it from flowing on non-intersecting orbits (e.g. Binney et al. 1991; Sormani, Binney & Magorrian 2015) or drives acoustic instabilities that are unrelated to self-gravity (e.g. Montenegro, Yuan & Elmegreen 1999; Krumholz & Kruijssen 2015; Krumholz, Kruijssen & Crocker 2017). Moreover, these effects can drive gas outward as well as in. In any region where such effects are dominant, our model is not applicable.

However, this is not a major limitation because regions such as the Milky Way centre contribute very little to the total gas mass or star formation budget of the Universe. The Milky Way's centre is gas-depleted to an extent that is unusual even among local spirals (Bigiel & Blitz 2012), and is likely related to it being a 'green valley' galaxy on the verge of quenching (Bland-Hawthorn & Gerhard 2016). Even including the central molecular zone, the central few kpc of the Galaxy contain ~ 10 per cent of its star formation or ISM mass (e.g. Kruijssen et al. 2014). Indeed, this gas-poverty is likely the reason that there is a strong bar, since simulations show that bar formation does not take place until the gas fraction drops to well under 10 per cent (Athanasoulas, Machado & Rodionov 2013).

Beyond the Milky Way, the majority of gas in nearby spirals lies in regions where the gas and stellar surface densities and velocity dispersions are such that gas and stars are strongly coupled (Romeo & Mogotsi 2017, their fig. 5). Estimates of the fraction of galaxies that contain stellar bars at all vary from ~ 20 per cent (e.g. Melvin et al. 2014; Cervantes Sodi 2017) to ~ 60 per cent (e.g. Erwin 2017), with the higher figures largely coming from surveys capable of detecting bars below ~ 1 kpc in size. Thus large bars that could conceivably affect the dynamics of the majority of the ISM mass or star formation in a galaxy appear to be rare even at $z = 0$, as one might expect since even a strongly barred galaxy like the Milky Way has little of its gas or star formation in the region where the bar dominates the dynamics. Though there is significant debate in the literature about whether the bar fraction declines with redshift (e.g. Melvin et al. 2014 versus Erwin 2017), there are strong theoretical reasons to believe that bars were less prominent in the past, both due to the higher gas fractions found at $z > 0$ (e.g. Tacconi et al. 2013) and the fact that bars take time to grow. Thus if barred regions do not dominate the ISM mass budget at $z = 0$, they should be even less important in the early Universe. In summary, our model should apply to most of the ISM and most star-forming regions over most of cosmic time. We do, however, caution that it should not be applied to extremely gas-poor, bar-dominated regions like the Milky Way centre. These are better described by models that treat the stellar potential as decoupled from the gas (e.g. Binney et al. 1991; Sormani et al. 2015; Krumholz et al. 2017).

5 SUMMARY

We present a new model for the gas in galactic discs, based on a few simple physical premises. We propose that galactic discs maintain a state of vertical hydrostatic equilibrium, marginal gravitational stability, and balance between dissipation of turbulence and injection of turbulent energy by star formation feedback and radial transport. The inclusion of both radial transport and feedback as potential energy sources is the primary new feature of our model, and despite the apparent simplicity of this addition, it yields a dramatic improvement in both predictive power and agreement with observation compared to simpler equilibrium models.

We find that star formation alone is able to maintain a velocity dispersion of $\sigma_{\text{sf}} \approx 6\text{--}10 \text{ km s}^{-1}$, with the exact value depending on the gas fraction, the thermal velocity dispersion, and the fraction of the interstellar medium in the star-forming molecular phase. In galaxies where the gas surface density is low, this is sufficient to maintain energy and hydrostatic balance, and there is no net radial flow. However, in many observed galaxies this velocity dispersion is insufficient to keep the gas in a state of marginal gravitational stability. In this case, the instability produces spiral structures and clumps that exert non-axisymmetric torques, leading to a net mass flow inward. The inflow releases gravitational potential energy, which

manifests as non-circular, turbulent motions in the transported gas. The fraction of the turbulent power that originates from this process rather than from star formation feedback is $1 - \sigma_{\text{sf}}/\sigma_{\text{g}}$, where σ_{g} is the gas velocity dispersion. This fraction is small in quiescent star-forming galaxies at the present cosmic epoch, but is larger in both starbursts and high-redshift galaxies.

The model we derive from this simple picture shows excellent agreement with a range of observational diagnostics, including the star formation law for both resolved and unresolved systems and the relationship between galaxies' star formation rates and velocity dispersions. It is also consistent with the limited data available on the observed rates of radial inflow in nearby galaxies. The agreement holds across a wide range of galaxy types and masses, from nearby dwarfs with star formation rates $\lesssim 0.1 \text{ M}_{\odot} \text{ yr}^{-1}$, to starbursts and high-redshift discs with star formation rates $\gtrsim 100 \text{ M}_{\odot} \text{ yr}^{-1}$. We also predict that high gas inflow rates should be measurable in nearby starburst galaxies, whose kinematics have yet to be analysed for inflow. In contrast, we show that models that neglect either radial transport or star formation feedback fail at either high or low star formation rate, or in some cases both.

Our model provides a natural explanation for the cosmic epochs at which galaxies build up bulges and discs, and at which they quench. At high redshift, galaxies' mass transport rates naturally exceed their star formation rates; this is a natural consequence of the high velocity dispersions found in high-redshift galaxies, and the stronger scaling between velocity dispersion and transport rate ($\dot{M} \propto \sigma_{\text{g}}^3$) than between velocity dispersion and star formation rate ($\dot{M}_{\star} \propto \sigma_{\text{g}}$). As a result, they tend to move mass inward towards a bulge. As accretion rates decline as the density of the universe diminishes, transport rates decline as well, and do so faster than star formation rates. This leads to a configuration where most star formation occurs in galaxies' discs. Finally, once the star formation rate is smaller than the mass transport rate, gas does not reach galaxy centres at all, and the centres quench, explaining the common observation that quenching tends to occur inside out.

In future work we plan to apply this model to radially-dependent models of galaxy formation, such as those of Forbes et al. (2012) and Forbes et al. (2014a). Such an application promises to yield new insights into the origin of the radial structure of galactic discs, and the evolution of this structure over cosmic time. We also plan to test the model against cosmological simulations, where inflow rates are determined directly from the hydrodynamics (Burkhart et al. in preparation).

ACKNOWLEDGEMENTS

MRK acknowledges support from the Australian Research Council's *Discovery Projects* funding scheme (project DP160100695). BB is supported by the NASA Einstein Postdoctoral Fellowship and an Institute for Theory and Computation Fellowship. JCF is supported by an Institute for Theory and Computation Fellowship. We thank S. Oh for helpful discussions of bar surveys, and the referee, A. Dekel, for helpful comments.

REFERENCES

- Agertz O., Teyssier R., Moore B., 2009a, *MNRAS*, 397, L64
- Agertz O., Lake G., Teyssier R., Moore B., Mayer L., Romeo A. B., 2009b, *MNRAS*, 392, 294
- Arribas S., Colina L., Bellocchi E., Maiolino R., Villar-Martín M., 2014, *A&A*, 568, A14

- Athanassoula E., Machado R. E. G., Rodionov S. A., 2013, *MNRAS*, 429, 1949
- Barro G. et al., 2013, *ApJ*, 765, 104
- Barro G. et al., 2016, *ApJ*, 820, 120
- Barro G. et al., 2017, *ApJ*, 851, L40
- Belfiore F. et al., 2017, *MNRAS*, 466, 2570
- Bigiel F., Blitz L., 2012, *ApJ*, 756, 183
- Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, *AJ*, 136, 2846
- Bigiel F., Leroy A., Walter F., Blitz L., Brinks E., de Blok W. J. G., Madore B., 2010, *AJ*, 140, 1194
- Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, *MNRAS*, 252, 210
- Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529
- Bolatto A. D. et al., 2011, *ApJ*, 741, 12
- Bolatto A. D., Wolfire M., Leroy A. K., 2013, *ARA&A*, 51, 207
- Bouché N. et al., 2007, *ApJ*, 671, 303
- Bouché N. et al., 2010, *ApJ*, 718, 1001
- Boulares A., Cox D. P., 1990, *ApJ*, 365, 544
- Bournaud F., Elmegreen B. G., 2009, *ApJ*, 694, L158
- Bournaud F., Elmegreen B. G., Elmegreen D. M., 2007, *ApJ*, 670, 237
- Burkhart B., 2018, *ApJ*, preprint ([arXiv:1801.05428](https://arxiv.org/abs/1801.05428))
- Burkhart B., Stanimirovic S., Lazarian A., Kowal G., 2010, *ApJ*, 708, 1204
- Burkhart B., Lazarian A., Balsara D., Meyer C., Cho J., 2015, *ApJ*, 805, 118
- 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
- Caldú-Primo A., Schruha A., Walter F., Leroy A., Bolatto A. D., Vogel S., 2015, *AJ*, 149, 76
- Cervantes Sodi B., 2017, *ApJ*, 835, 80
- Ceverino D., Dekel A., Bournaud F., 2010, *MNRAS*, 404, 2151
- Chabrier G., 2005, in Corbelli E., Palla F., Zinnecker H., eds, *Astrophysics and Space Science Library*, Vol. 327, *The Initial Mass Function 50 Years Later*. Springer, Dordrecht, p. 41
- Chepurnov A., Burkhard B., Lazarian A., Stanimirovic S., 2015, *ApJ*, 810, 33
- Cheung E. et al., 2012, *ApJ*, 760, 131
- Cioffi D. F., McKee C. F., Bertschinger E., 1988, *ApJ*, 334, 252
- Cresci G. et al., 2009, *ApJ*, 697, 115
- Daddi E., Dannerbauer H., Elbaz D., Dickinson M., Morrison G., Stern D., Ravindranath S., 2008, *ApJ*, 673, L21
- Daddi E. et al., 2010a, *ApJ*, 713, 686
- Daddi E. et al., 2010b, *ApJ*, 714, L118
- Danovich M., Dekel A., Hahn O., Ceverino D., Primack J., 2015, *MNRAS*, 449, 2087
- Dekel A., Birnboim Y., 2006, *MNRAS*, 368, 2
- Dekel A., Burkert A., 2014, *MNRAS*, 438, 1870
- Dekel A., Sari R., Ceverino D., 2009, *ApJ*, 703, 785
- Dekel A. et al., 2009, *Nature*, 457, 451
- Di Teodoro E. M., Fraternali F., Miller S. H., 2016, *A&A*, 594, A77
- Downes D., Solomon P. M., 1998, *ApJ*, 507, 615
- Elmegreen B. G., 2011, *ApJ*, 737, 10
- Epinat B., Amram P., Marcelin M., 2008, *MNRAS*, 390, 466
- Epinat B. et al., 2009, *A&A*, 504, 789
- Erwin P., 2017, *MNRAS*, 474, 5372
- Evans N. J., II, Heiderman A., Vutisalchavakul N., 2014, *ApJ*, 782, 114
- 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-Giguère C.-A., Kereš D., Ma C.-P., 2011, *MNRAS*, 417, 2982
- Faucher-Giguère C.-A., Quataert E., Hopkins P. F., 2013, *MNRAS*, 433, 1970
- Federrath C., 2015, *MNRAS*, 450, 4035
- Federrath C., Klessen R. S., 2012, *ApJ*, 761, 156
- Federrath C., Klessen R. S., 2013, *ApJ*, 763, 51
- Feldmann R., Gnedin N. Y., 2011, *ApJ*, 727, L12+
- Filho M. E., Sánchez Almeida J., Amorín R., Muñoz-Tuñón C., Elmegreen B. G., Elmegreen D. M., 2016, *ApJ*, 820, 109
- Forbes J., Krumholz M., Burkert A., 2012, *ApJ*, 754, 48
- Forbes J. C., Krumholz M. R., Burkert A., Dekel A., 2014a, *MNRAS*, 438, 1552
- Forbes J. C., Krumholz M. R., Burkert A., Dekel A., 2014b, *MNRAS*, 443, 168
- Fraternali F., Marasco A., Marinacci F., Binney J., 2013, *ApJ*, 764, L21
- Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, *MNRAS*, 409, 515
- García-Burillo S., Usero A., Alonso-Herrero A., Graciá-Carpio J., Pereira-Santaella M., Colina L., Planesas P., Arribas S., 2012, *A&A*, 539, A8
- Genel S., Dekel A., Cacciato M., 2012, *MNRAS*, 425, 788
- Gentry E. S., Krumholz M. R., Dekel A., Madau P., 2017, *MNRAS*, 465, 2471
- Gentry E. S., Krumholz M. R., Madau P., Lupi A., 2018, *MNRAS*, preprint ([arXiv:1802.06860](https://arxiv.org/abs/1802.06860))
- Genzel R. et al., 2010, *MNRAS*, 407, 2091
- Genzel R. et al., 2014, *ApJ*, 785, 75
- Genzel R. et al., 2015, *ApJ*, 800, 20
- Glazebrook K., 2013, *Proc. Astron. Soc. Aust.*, 30, e056
- Goldbaum N. J., Krumholz M. R., Matzner C. D., McKee C. F., 2011, *ApJ*, 738, 101
- Goldbaum N. J., Krumholz M. R., Forbes J. C., 2015, *ApJ*, 814, 131
- Goldbaum N. J., Krumholz M. R., Forbes J. C., 2016, *ApJ*, 827, 28
- Green A. W. et al., 2010, *Nature*, 467, 684
- Green A. W. et al., 2014, *MNRAS*, 437, 1070
- Hayward C. C., Hopkins P. F., 2017, *MNRAS*, 465, 1682
- Heiderman A., Evans N. J., II, Allen L. E., Huard T., Heyer M., 2010, *ApJ*, 723, 1019
- Heyer M., Gutermuth R., Urquhart J. S., Csengeri T., Wienen M., Leurini S., Menten K., Wyrowski F., 2016, *A&A*, 588, A29
- Hobbs A., Read J., Power C., Cole D., 2013, *MNRAS*, 434, 1849
- Hoffmann V., Romeo A. B., 2012, *MNRAS*, 425, 1511
- Hopkins P. F., Quataert E., Murray N., 2011, *MNRAS*, 417, 950
- Hunter D. A. et al., 2012, *AJ*, 144, 134
- Ianjamasimanana R., de Blok W. J. G., Walter F., Heald G. H., 2012, *AJ*, 144, 96
- Ianjamasimanana R., de Blok W. J. G., Walter F., Heald G. H., Caldú-Primo A., Jarrett T. H., 2015, *AJ*, 150, 47
- Inoue S., Dekel A., Mandelker N., Ceverino D., Bournaud F., Primack J., 2016, *MNRAS*, 456, 2052
- Jameson K. E. et al., 2016, *ApJ*, 825, 12
- Johnson H. L. et al., 2018, *MNRAS*, 474, 5076
- Jones T. A., Swinbank A. M., Ellis R. S., Richard J., Stark D. P., 2010, *MNRAS*, 404, 1247
- Joung M. R., Mac Low M., Bryan G. L., 2009, *ApJ*, 704, 137
- Joung M. R., Bryan G. L., Putman M. E., 2012, *ApJ*, 745, 148
- Kalberla P. M. W., Kerp J., 2009, *ARA&A*, 47, 27
- Kennicutt R. C., Evans N. J., 2012, *ARA&A*, 50, 531
- Kennicutt R. C., Jr, 1998, *ApJ*, 498, 541
- Kennicutt R. C., Jr et al., 2007, *ApJ*, 671, 333
- Kereš D., Katz N., Weinberg D. H., Davé R., 2005, *MNRAS*, 363, 2
- Kim C.-G., Ostriker E. C., 2015a, *ApJ*, 802, 99
- Kim C.-G., Ostriker E. C., 2015b, *ApJ*, 815, 67
- Kim C.-G., Kim W.-T., Ostriker E. C., 2011, *ApJ*, 743, 25
- Kim C.-G., Ostriker E. C., Raileanu R., 2017, *ApJ*, 834, 25
- Koyama H., Ostriker E. C., 2009, *ApJ*, 693, 1346
- Kravtsov A. V., 2013, *ApJ*, 764, L31
- Kruijssen J. M. D., Longmore S. N., Elmegreen B. G., Murray N., Bally J., Testi L., Kennicutt R. C., 2014, *MNRAS*, 440, 3370
- Krumholz M. R., 2013, *MNRAS*, 436, 2747
- Krumholz M. R., 2014, *Phys. Rep.*, 539, 49
- Krumholz M. R., 2017, *Star Formation*, World Scientific Series in Astrophysics. World Scientific Publishing, Singapore
- Krumholz M. R., Burkert A., 2010, *ApJ*, 724, 895
- Krumholz M. R., Burkert B., 2016, *MNRAS*, 458, 1671
- Krumholz M. R., Dekel A., 2012, *ApJ*, 753, 16
- Krumholz M. R., Kruijssen J. M. D., 2015, *MNRAS*, 453, 739
- Krumholz M. R., McKee C. F., 2005, *ApJ*, 630, 250
- Krumholz M. R., Tan J. C., 2007, *ApJ*, 654, 304

- Krumholz M. R., Matzner C. D., McKee C. F., 2006, *ApJ*, 653, 361
- Krumholz M. R., McKee C. F., Tumlinson J., 2009a, *ApJ*, 693, 216
- Krumholz M. R., McKee C. F., Tumlinson J., 2009b, *ApJ*, 699, 850
- Krumholz M. R., Dekel A., McKee C. F., 2012, *ApJ*, 745, 69
- Krumholz M. R., Kruijssen J. M. D., Crocker R. M., 2017, *MNRAS*, 466, 1213
- Lada C. J., Lombardi M., Alves J. F., 2010, *ApJ*, 724, 687
- Lagos C. D. P., Lacey C. G., Baugh C. M., Bower R. G., Benson A. J., 2011, *MNRAS*, 416, 1566
- Law D. R., Steidel C. C., Erb D. K., Larkin J. E., Pettini M., Shapley A. E., Wright S. A., 2009, *ApJ*, 697, 2057
- Le Tiran L., Lehnert M. D., van Driel W., Nesvadba N. P. H., Di Matteo P., 2011, *A&A*, 534, L4
- Lee E. J., Miville-Deschênes M.-A., Murray N. W., 2016, *ApJ*, 833, 229
- Lehnert M. D., Nesvadba N. P. H., Le Tiran L., Di Matteo P., van Driel W., Douglas L. S., Chemin L., Bounaud F., 2009, *ApJ*, 699, 1660
- Lehnert M. D., Le Tiran L., Nesvadba N. P. H., van Driel W., Boulanger F., Di Matteo P., 2013, *A&A*, 555, A72
- Leitner S. N., Kravtsov A. V., 2011, *ApJ*, 734, 48
- Lemaster M. N., Stone J. M., 2009, *ApJ*, 691, 1092
- Lemoine-Busserolle M., Bunker A., Lamareille F., Kissler-Patig M., 2010, *MNRAS*, 401, 1657
- Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, *AJ*, 136, 2782
- Leroy A. K. et al., 2013, *AJ*, 146, 19
- Leroy A. K. et al., 2017, *ApJ*, 846, 71
- Liu G., Koda J., Calzetti D., Fukuhara M., Momose R., 2011, *ApJ*, 735, 63
- Mac Low M., 1999, *ApJ*, 524, 169
- Mac Low M., Klessen R. S., Burkert A., Smith M. D., 1998, *Phys. Rev. Lett.*, 80, 2754
- Marasco A., Fraternali F., Binney J. J., 2012, *MNRAS*, 419, 1107
- Marinacci F., Binney J., Fraternali F., Nipoti C., Ciotti L., Londrillo P., 2010, *MNRAS*, 404, 1464
- Martin C. L., Kobulnicky H. A., Heckman T. M., 2002, *ApJ*, 574, 663
- Martizzi D., Faucher-Giguère C.-A., Quataert E., 2015, *MNRAS*, 450, 504
- Matzner C. D., 2002, *ApJ*, 566, 302
- McKee C. F., Krumholz M. R., 2010, *ApJ*, 709, 308
- McKee C. F., Parravano A., Hollenbach D. J., 2015, *ApJ*, 814, 13
- Meidt S. E. et al., 2013, *ApJ*, 779, 45
- Melvin T. et al., 2014, *MNRAS*, 438, 2882
- Meurer G. R., Zheng Z., de Blok W. J. G., 2013, *MNRAS*, 429, 2537
- Mo H., van den Bosch F. C., White S., 2010, *Galaxy Formation and Evolution*. Cambridge Univ. Press, Cambridge
- Moiseev A. V., Tikhonov A. V., Klypin A., 2015, *MNRAS*, 449, 3568
- Momose R. et al., 2013, *ApJ*, 772, L13
- Montenegro L. E., Yuan C., Elmegreen B. G., 1999, *ApJ*, 520, 592
- Murray N., 2011, *ApJ*, 729, 133
- Narayanan D., Krumholz M. R., Ostriker E. C., Hernquist L., 2012, *MNRAS*, 421, 3127
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Neistein E., Dekel A., 2008, *MNRAS*, 383, 615
- Nelson E. J. et al., 2016, *ApJ*, 828, 27
- Obreschkow D., Croton D., De Lucia G., Khochfar S., Rawlings S., 2009, *ApJ*, 698, 1467
- Onus A., Krumholz M. R., Federrath C., 2018, *MNRAS*, preprint ([arXiv:1801.09952](https://arxiv.org/abs/1801.09952))
- Ostriker E. C., Shetty R., 2011, *ApJ*, 731, 41
- Ostriker E. C., McKee C. F., Leroy A. K., 2010, *ApJ*, 721, 975
- Padoan P., Nordlund Å., 2011, *ApJ*, 730, 40
- Padoan P., Haugbølle T., Nordlund Å., 2012, *ApJ*, 759, L27
- Padoan P., Federrath C., Chabrier G., Evans N. J., II, Johnstone D., Jørgensen J. K., McKee C. F., Nordlund Å., 2014, *Protostars and Planets VI*. University of Arizona Press, Tucson, p. 77
- Petric A. O., Rupen M. P., 2007, *AJ*, 134, 1952
- Pety J. et al., 2013, *ApJ*, 779, 43
- Piontek R. A., Ostriker E. C., 2007, *ApJ*, 663, 183
- Rafelski M., Gardner J. P., Fumagalli M., Neeleman M., Teplitz H. I., Groggin N., Koekemoer A. M., Scarlata C., 2016, *ApJ*, 825, 87
- Rafikov R. R., 2001, *MNRAS*, 323, 445
- Regan M. W., Thornley M. D., Helfer T. T., Sheth K., Wong T., Vogel S. N., Blitz L., Bock D. C.-J., 2001, *ApJ*, 561, 218
- Romeo A. B., Falstad N., 2013, *MNRAS*, 433, 1389
- Romeo A. B., Mogotsi K. M., 2017, *MNRAS*, 469, 286
- Romeo A. B., Wiegert J., 2011, *MNRAS*, 416, 1191
- Romeo A. B., Burkert A., Agertz O., 2010, *MNRAS*, 407, 1223
- Saintonge A. et al., 2011, *MNRAS*, 415, 32
- Salim D. M., Federrath C., Kewley L. J., 2015, *ApJ*, 806, L36
- Sanders D. B., Mazzarella J. M., Kim D.-C., Surace J. A., Soifer B. T., 2003, *AJ*, 126, 1607
- Schmidt T. M., Bigiel F., Klessen R. S., de Blok W. J. G., 2016, *MNRAS*, 457, 2642
- Schruba A. et al., 2011, *AJ*, 142, 37
- Scoville N. et al., 2015, *ApJ*, 800, 70
- Scoville N. et al., 2017, *ApJ*, 836, 66
- Sharma P., Roy A., Nath B. B., Shchekinov Y., 2014, *MNRAS*, 443, 3463
- Shi Y., Armus L., Helou G., Stierwalt S., Gao Y., Wang J., Zhang Z.-Y., Gu Q., 2014, *Nature*, 514, 335
- Somerville R. S. et al., 2018, *MNRAS*, 473, 2714
- Sormani M. C., Binney J., Magorrian J., 2015, *MNRAS*, 449, 2421
- Stevens A. R. H., Croton D. J., Mutch S. J., 2016, *MNRAS*, 461, 859
- Stilp A. M., Dalcanton J. J., Skillman E., Warren S. R., Ott J., Koribalski B., 2013, *ApJ*, 773, 88
- Stone J. M., Ostriker E. C., Gammie C. F., 1998, *ApJ*, 508, L99
- Stott J. P. et al., 2016, *MNRAS*, 457, 1888
- Swinbank A. M., Smail I., Sobral D., Theuns T., Best P. N., Geach J. E., 2012, *ApJ*, 760, 130
- Tacchella S. et al., 2015, *Science*, 348, 314
- Tacchella S., Dekel A., Carollo C. M., Ceverino D., DeGraf C., Lapiner S., Mandelker N., Primack J. R., 2016a, *MNRAS*, 458, 242
- Tacchella S., Dekel A., Carollo C. M., Ceverino D., DeGraf C., Lapiner S., Mandelker N., Primack J. R., 2016b, *MNRAS*, 457, 2790
- Tacchella S. et al., 2018, *ApJ*, preprint ([arXiv:1704.00733](https://arxiv.org/abs/1704.00733))
- Tacconi L. J. et al., 2013, *ApJ*, 768, 74
- Tamburro D., Rix H.-W., Leroy A. K., Mac Low M.-M., Walter F., Kennicutt R. C., Brinks E., de Blok W. J. G., 2009, *AJ*, 137, 4424
- Thompson T. A., Krumholz M. R., 2016, *MNRAS*, 455, 334
- Thompson T. A., Quataert E., Murray N., 2005, *ApJ*, 630, 167
- Thornton K., Gaudlitz M., Janka H.-T., Steinmetz M., 1998, *ApJ*, 500, 95
- Tonini C., Mutch S. J., Croton D. J., Wyithe J. S. B., 2016, *MNRAS*, 459, 4109
- Toomre A., 1964, *ApJ*, 139, 1217
- Usero A. et al., 2015, *AJ*, 150, 115
- van Zee L., Bryant J., 1999, *AJ*, 118, 2172
- Varidel M., Pracy M., Croom S., Owers M. S., Sadler E., 2016, *Proc. Astron. Soc. Aust.*, 33, e006
- Veilleux S. et al., 2009, *ApJS*, 182, 628
- Vollmer B., Leroy A. K., 2011, *AJ*, 141, 24
- Vutisalchavakul N., Evans N. J., II, Heyer M., 2016, *ApJ*, 831, 73
- Walch S., Naab T., 2015, *MNRAS*, 451, 2757
- Walter F., Brinks E., de Blok W. J. G., Bigiel F., Kennicutt R. C., Jr, Thornley M. D., Leroy A., 2008, *AJ*, 136, 2563
- Wetzel A. R., Nagai D., 2015, *ApJ*, 808, 40
- Wisnioski E. et al., 2011, *MNRAS*, 417, 2601
- Wisnioski E. et al., 2015, *ApJ*, 799, 209
- Wolfe M. G., McKee C. F., Hollenbach D., Tielens A. G. G. M., 2003, *ApJ*, 587, 278
- Wong T. et al., 2013, *ApJ*, 777, L4
- Wong O. I., Meurer G. R., Zheng Z., Heckman T. M., Thilker D. A., Zwaan M. A., 2016, *MNRAS*, 460, 1106
- Wuyts S. et al., 2016, *ApJ*, 831, 149
- Zheng Z., Meurer G. R., Heckman T. M., Thilker D. A., Zwaan M. A., 2013, *MNRAS*, 434, 3389
- Zolotov A. et al., 2015, *MNRAS*, 450, 2327
- Zuckerman B., Evans N. J., 1974, *ApJ*, 192, L149

APPENDIX A: CHOICE OF α_{CO}

Our sample of galaxies includes starburst and high-redshift systems, and for these galaxies there is significant uncertainty about the choice of α_{CO} , the conversion factor between CO luminosity and gas surface density (Bolatto, Wolfire & Leroy 2013). Throughout this paper, we have chosen to adopt the convention for α_{CO} recommended by Daddi et al. (2010b): $\alpha_{\text{CO}} = 4.6 \text{ M}_{\odot} (\text{K km s}^{-1} \text{ pc}^{-2})$ for local spirals, $\alpha_{\text{CO}} = 3.6 \text{ M}_{\odot} (\text{K km s}^{-1} \text{ pc}^{-2})$ for high-redshift discs, and $\alpha_{\text{CO}} = 0.8 \text{ M}_{\odot} (\text{K km s}^{-1} \text{ pc}^{-2})$ for starbursts / ULIRGs / sub-mm galaxies at all redshifts.

We note that, in contrast, Faucher-Giguère et al. (2013) adopt a theoretical α_{CO} from Narayanan et al. (2012). The effects of this choice are modest for most galaxies, but are significant for high-redshift star-forming disc galaxies, e.g. the PHIBSS sample of Tacconi et al. (2013). Using the Narayanan et al. (2012) α_{CO} for these objects gives them a ULIRG-like α_{CO} rather than a Milky Way-like one favoured by the observers, and that we have adopted. This in turn shifts these objects to lower Σ_{g} by a factor of ~ 5 , and produces a steepening of the best-fitting slope to ≈ 1.7 , compared to ≈ 1.4 for the conventional α_{CO} ; we refrain from plotting the data and fit here, since such a plot is shown in fig. 3 of Thompson & Krumholz (2016).

The question of the correct α_{CO} for disc galaxies on the star-forming main sequence at high redshift has recently been reinvestigated by Genzel et al. (2015) using dust observations, which provide an independent means of measuring the gas mass. They find that the conventional α_{CO} (i.e. one similar to that of the Milky Way) is a better match to their data than a ULIRG-like α_{CO} . We therefore retain the conventional α_{CO} in this work.

APPENDIX B: VELOCITY DISPERSION DATA

Here we provide details on our handling of the observations of velocity dispersion σ versus star formation rate \dot{M}_{*} , as discussed in Section 3.2. A general issue that arises when combining multiple data sets is that we are interested in the velocity dispersion of the cool atomic or molecular ISM, but much of the data available (and all of the high-redshift data) are H α measurements, which are likely dominated by gas in H II regions. Such gas has a thermal velocity dispersion of $\approx 10 \text{ km s}^{-1}$, and cometary expansion of photoionized gas adds a comparable non-thermal component on top of this. The effects of this are evident if one compares the velocity dispersions obtained from H I observations to those obtained by H α measurements in galaxies of similar, low star formation rates; the H α velocity dispersions are a factor of ~ 2 larger. Since the H α excess is larger than σ_{sf} , we must remove it in order to make meaningful comparisons. Following Krumholz & Burkhardt (2016), we adopt a combined thermal plus non-thermal velocity dispersion of $\sigma_{\text{H II}} = 15 \text{ km s}^{-1}$ for H II region gas, and subtract this in quadrature where needed. However, even after this correction the H α velocity dispersions in low star formation rate galaxies remain systematically larger than the values obtained by H I or molecular lines. There is no comparable discrepancy between H α and molecules at high star formation rates, so the most likely explanation for the offset at low star formation rates is contamination by motions within H II regions, which is a more severe problem when the overall velocity dispersion is lower.

Details on individual data sets follow, and our data compilation is available from <https://bitbucket.org/krumholz/kbfc17>.

GHASP. The GHASP survey includes H α measurements for a range of local galaxies; values of σ and \dot{M}_{*} are reported in Epinat

et al. (2008); we retrieved the data from the VizieR data base entry associated with the paper. The velocity dispersions σ_{g} we plot are listed as σ_{res} in their tables. Despite the fact that these are H α measurements, the fitting method essentially measures the dispersion of velocity centroids inside each fitted ring, and thus is insensitive to the broadening of the line. For this reason, we do not subtract $\sigma_{\text{H II}}$ from these data points.

Epinat et al. (2009). This paper reports H α measurements for galaxies at $1.2 < z < 1.6$. The quantity we plot as σ_{g} is the local mean velocity dispersion σ_0 given in their table 5, with $\sigma_{\text{H II}}$ subtracted in quadrature.

Law et al. (2009). This paper reports H α and [O III] observations of galaxies at $z \sim 2\text{--}3.3$. For our star formation rate estimate, we use the values given in the paper based on nebular emission (their SFR_{neb}), and the velocity dispersion we use is their σ_{mean} , with $\sigma_{\text{H II}}$ subtracted in quadrature. While the correction factor should be somewhat smaller for [O III] than for H α due to the smaller thermal broadening of the heavy ion, the velocity dispersions reported are so large that the quadrature subtraction has not noticeable effect in any event. For galaxies where more than one value of σ_{mean} is given, we plot an average of the values listed.

Jones et al. (2010). This paper reports H α observations of gravitationally lensed galaxies at $z \sim 2\text{--}3$. We take our star formation rates and velocity dispersions directly from their tables, subtracting $\sigma_{\text{H II}}$ in quadrature.

Lemoine-Busserolle et al. (2010). This study reports H α star formation rates and velocity dispersions from galaxies at $z \sim 3$. The value we plot as σ_{g} is their σ_{mean} , with $\sigma_{\text{H II}}$ subtracted in quadrature.

WiggleZ. We plot a sample of H α observations of galaxies from the WiggleZ survey as reported by Wisnioski et al. (2011). We convert their reported H α luminosities to star formation rates using the conversion given by Kennicutt & Evans (2012), and we use their reported σ_{mean} as the gas velocity dispersion σ_{g} after subtracting $\sigma_{\text{H II}}$ in quadrature.

THINGS. We use the global H I velocity dispersions measured in local galaxies by the THINGS survey, as reported by Janjamasimanana et al. (2012). Our σ_{g} is their single Gaussian fit value. We take corresponding star formation rates from Leroy et al. (2008) if they are given there, or from the literature compilation of Walter et al. (2008) if not.

Stilp et al. (2013). This paper reports H I measurements of local dwarf galaxies. For σ_{g} we use their σ_{central} , the velocity dispersion of their central Gaussian component.

DYNAMO. The DYNAMO survey reports H α measurements of local galaxies selected to be have conditions similar to those commonly seen at high- z . Kinematic data are reported in Green et al. (2014), and we retrieve the quantities we plot from the VizieR entry associated with the paper. We use their H α -estimated star formation rates, and their estimated σ as the velocity dispersion, after subtracting $\sigma_{\text{H II}}$ in quadrature.

Moiseev et al. (2015). This paper reports a survey of H α emission from local dwarf galaxies. We derive star formation rates for this sample by converting their reported H α luminosities using the conversion of Kennicutt & Evans (2012). For the velocity dispersion, the values reported in the paper have already been corrected for an assumed thermal broadening of 9.1 km s^{-1} . For consistency with the other data sets, we remove this correction by adding 9.1 km s^{-1} in quadrature, then subtracting $\sigma_{\text{H II}}$ in quadrature.

Varidel et al. (2016). This is a sample of local analogues of high- z galaxies similar to DYNAMO. For these galaxies we use the star formation rates derived from line emission, and for velocity

Table B1. Compilation of ULIRG data. References are as follows: (1) Scoville et al. (2017); (2) Scoville et al. (2015); (3) Downes & Solomon (1998); (4) Veilleux et al. (2009); (5) Sanders et al. (2003). For galaxies where the star formation rates are from (4) or (5), we have converted the reported total IR luminosity to a star formation rate using the conversion of Kennicutt & Evans (2012). For IR data from Veilleux et al. (2009), we subtract off the active galactic nucleus (AGN) contribution using their estimates. For IR data from Sanders et al. (2003), we adopt an AGN fraction of 50 per cent.

Name	\dot{M}_* ($M_\odot \text{ yr}^{-1}$)	σ_g	σ_g Ref	\dot{M}_* Ref
Arp 220 West	120	128	(1)	(1)
Arp 220 East	120	61	(1)	(1)
NGC 6240	70	160	(2)	(2)
Mrk 231	176	60	(3)	(4)
VII Zw 31	66	21	(3)	(5)
IRAS 10565+2448	163	29	(3)	(4)
Arp 193	66	29	(3)	(5)
IRAS 17208–0014	428	107	(3)	(4)
IRAS 23365+3604	102	71	(3)	(4)

dispersion we use the flux-weighted mean (σ_m in Varidel et al.’s notation).

Di Teodoro et al. (2016). This is an analysis of $H\alpha$ observations of galaxies from the KMOS survey (Wisnioski et al. 2015) using a novel kinematic fitting method. We use the star formation rates and velocity dispersions taken from their table 1, subtracting σ_{HII} in quadrature.

SINS. The SINS-KMOS-3D survey (Wisnioski et al. 2015) includes $H\alpha$ measurements for 248 $z \sim 1$ –3 galaxies, as reported by Wuyts et al. (2016). The velocity dispersions and star formation rates were provided by R. Genzel (2017, private communication). The values we plot have σ_{HII} subtracted in quadrature.

KROSS. The KMOS Redshift One Spectroscopic Survey (KROSS; Stott et al. 2016) is a spectroscopic survey of redshift

one galaxies using KMOS on the VLT. Johnson et al. (2017) provides measurements of velocity dispersion based on the $H\alpha$ line in 472 galaxies from the survey, along with star formation rates for the same galaxies. The values we plot have σ_{HII} subtracted in quadrature.

ULIRGs. We have compiled a new sample of measured velocity dispersions in local ULIRGs, drawn from the literature. Velocity dispersions for all galaxies except NGC 6240 are based on CO measurements; NGC 6240 is based on HCN. The data are given in Table B1.

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