# A model for the origin of bursty star formation in galaxies

## Claude-André Faucher-Giguère[*]

*Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, IL 60647, USA*

## ABSTRACT

We propose a simple analytic model to understand when star formation is time steady versus bursty in galaxies. Recent models explain the observed Kennicutt–Schmidt relation between star formation rate and gas surface densities in galaxies as resulting from a balance between stellar feedback and gravity. We argue that bursty star formation occurs when such an equilibrium cannot be stably sustained, and identify two regimes in which galaxy-scale star formation should be bursty: (i) at high redshift ($z \gtrsim 1$) for galaxies of all masses, and (ii) at low masses (depending on gas fraction) for galaxies at any redshift. At high redshift, characteristic galactic dynamical time-scales become too short for supernova feedback to effectively respond to gravitational collapse in galactic discs (an effect recently identified for galactic nuclei), whereas in dwarf galaxies star formation occurs in too few bright star-forming regions to effectively average out. Burstiness is also enhanced at high redshift owing to elevated gas fractions in the early Universe. Our model can thus explain the bursty star formation rates predicted in these regimes by recent high-resolution galaxy formation simulations, as well as the bursty star formation histories observationally inferred in both local dwarf and high-redshift galaxies. In our model, bursty star formation is associated with particularly strong spatiotemporal clustering of supernovae. Such clustering can promote the formation of galactic winds and our model may thus also explain the much higher wind mass loading factors inferred in high-redshift massive galaxies relative to their $z \sim 0$ counterparts.

**Key words:** stars: formation – galaxies: dwarf – galaxies: formation – galaxies: high-redshift – galaxies: ISM – galaxies: starburst.

## 1 INTRODUCTION

Observations of relatively tight relationships between the star formation rates (SFRs) of galaxies and their stellar masses ($M_\star$; e.g. Noeske et al. 2007; Reddy et al. 2012; Rodighiero et al. 2014) and between their SFRs and gas surface densities (the Kennicutt–Schmidt, KS, relation; e.g. Kennicutt 1998; Genzel et al. 2010) indicate that star formation must proceed relatively smoothly in typical galaxies when the SFR is averaged over long time-scales ($\gtrsim 100$ Myr). In most galaxy formation models to date, including large-volume cosmological hydrodynamic simulations (e.g. Springel & Hernquist 2003; Vogelsberger et al. 2014; Schaye et al. 2015; Davé, Thompson & Hopkins 2016), semi-analytic models (e.g. Fu et al. 2013; Benson 2014; Henriques et al. 2015; Somerville, Popping & Trager 2015) and analytic 'equilibrium' models tied to gas accretion rates from the intergalactic medium (e.g. Bouché et al. 2010; Davé, Finlator & Oppenheimer 2012; Lilly et al. 2013), star formation in individual galaxies is in fact effectively assumed to proceed smoothly in intervals between disturbances like galaxy mergers.

However, several recent galaxy simulations with resolution sufficient to resolve the gravitational collapse of individual gravitationally bound clouds (GBCs) in the interstellar medium (ISM) and to model stellar feedback on the scale of individual star-forming regions predict much more variable star formation histories in some regimes. Bursty star formation[1] at high redshift and in dwarf galaxies is in particular a key prediction of the Feedback In Realistic Environments (FIRE) cosmological zoom-in simulations, which model stellar feedback in a spatially and temporally resolved manner (e.g. Hopkins et al. 2014; Faucher-Giguère et al. 2015; Muratov et al. 2015; Sparre et al. 2017). This is illustrated in Fig. 1, which shows star formation histories for five simulated galaxies from the FIRE project, ranging from dwarf galaxies to Milky Way-mass galaxies, normalized by the running mean SFR averaged over a time-scale of $\approx 300$ Myr. All FIRE galaxies are bursty at high

---

[1] In this paper, 'bursty' refers to galaxies in which a significant fraction of star formation occurs in recurrent bursts, even if the SFR is time steady when averaged over cosmological time-scales. This is in contrast, for example, to isolated bursts of star formation triggered by galaxy mergers.

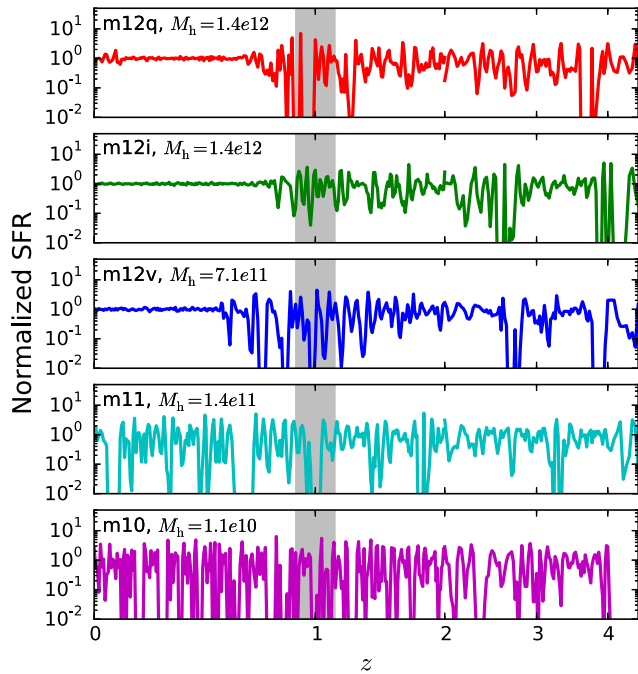[*] E-mail: cgiguere@northwestern.edu

**Figure 1.** Normalized SFR versus redshift for simulated galaxies from the FIRE project, in decreasing order of $z = 0$ halo mass (labelled at the top-left of each panel) from top to bottom. The instantaneous SFR is normalized by the running mean SFR, boxcar averaged over $\approx 300$ Myr. At high redshift ($z \gtrsim 1$), all simulated galaxies exhibit bursty star formation. The more massive galaxies settle into a more time-steady mode of star formation at lower redshifts but the dwarf galaxies sustain bursty star formation down to $z = 0$. Data from Muratov et al. (2015).

redshift ($z \gtrsim 1$, though with significant dispersion in the transition redshift). The more massive simulated galaxies (halo mass $M_h \sim 10^{12} \, M_\odot$) settle into a more time-steady mode of star formation at $z \lesssim 1$, while lower mass galaxies continue to be bursty all the way to the present time. In simulations of $\sim L^\star$ galaxies, the transition from bursty star formation at high redshift to more time-steady star formation at later times seems to be associated with the transition from highly dynamic and morphologically disturbed galaxies to more well-ordered discs familiar from observations of the nearby Universe (e.g. Hopkins et al. 2014; Agertz & Kravtsov 2015).

Our goal in this paper is to develop an analytic model to understand when and where galactic star formation is expected to be bursty versus time steady. We seek in particular to explain the results of simulations like the FIRE simulations. We however stress that similar bursty star formation is not limited to the FIRE simulations but is also seen in many other high-resolution simulations using different codes (e.g. Governato et al. 2012; Teyssier et al. 2013; Agertz & Kravtsov 2015; Domínguez et al. 2015).

Bursty star formation has particularly important implications for dwarf galaxies, as simulations indicate that the accompanying time-variable outflows can transfer energy to the central parts of dark matter haloes. This process produces cored dark matter halo profiles, resolving a primary tension between the predictions of pure cold dark matter simulations and observations of dwarf galaxies (e.g. Pontzen & Governato 2012; Madau, Shen & Governato 2014; Chan et al. 2015; Oñorbe et al. 2015; Wetzel et al. 2016; Fitts et al. 2017). Bursty star formation continuing to late times in dwarfs causes their outflows to recycle many times (Anglés-Alcázar et al. 2017), which likely plays an important role in maintaining a population of dwarf

galaxies blue to $z \sim 0$, a challenge in many galaxy formation models (e.g. Henriques et al. 2015).

The burstiness of star formation identified in recent galaxy formation simulations may explain several observational indications of time variable star formation on different time-scales. Observationally, the burstiness of star formation can be probed by comparing SFR measurements using indicators sensitive to different time-scales (Weisz et al. 2012; Domínguez et al. 2015; Sparre et al. 2017). Two of the most used indicators for this purpose are the H$\alpha$ nebular optical recombination line and the ultraviolet (UV) continuum. Whereas H$\alpha$ is excited by ionizing radiation from the most massive stars and is sensitive to SFR variations on time-scales $\lesssim 5$ Myr, the UV continuum is due to non-ionizing photospheric emission from stars with lifetimes up to $\sim 300$ Myr.

In the local Universe, observations show that the scatter in the H$\alpha$-to-UV ratio increases with decreasing galaxy mass (e.g. Weisz et al. 2012). By modelling the observations using toy star formation histories, Weisz et al. (2012) showed that the increased scatter towards low masses can be explained by increasing burstiness in dwarf galaxies.[2] Using a different observational approach combining the 4000 Å break and H$\delta_A$ stellar absorption line indices with SFR/$M_\star$ derived from emission line measurements, Kauffmann (2014) also found that the burstiness of recent star formation increases from $M_\star \sim 10^{10}$ to $\sim 10^8 \, M_\odot$. Interestingly, this technique probes variability on longer times scales, $\gtrsim 100$ Myr, indicating that SFRs can fluctuate on a broad range of time-scales. A similar conclusion was reached by Bauer et al. (2013) based on an analysis of the distribution of specific SFRs as a function of stellar mass at $z \lesssim 0.3$. Both Weisz et al. (2012) and Kauffmann (2014) find that in low-mass galaxies the amplitude of star formation bursts can be up to a factor of $\sim 30$. While some differences between SFRs inferred using different observational indicators can be explained in the context of steady star formation histories, such as due to incomplete sampling of the initial mass function (IMF), explanations based on constant SFRs tend to underestimate the magnitude of observed effects (e.g. Lee et al. 2009; Fumagalli, da Silva & Krumholz 2011).

At higher redshift, Guo et al. (2016) find that the observed H$\beta$-to-UV ratio increases with $M_\star$ at $0.4 < z < 1$.[3] By comparing with lower redshift measurements, Guo et al. (2016) also show that the H$\beta$-to-UV ratio decreases with increasing redshift. Guo et al. (2016) argue that their results are well explained by increasing star formation burstiness with both decreasing stellar mass and increasing redshift. van der Wel et al. (2011) identified an abundant population of $M_\star \sim 10^8 \, M_\odot$ 'extreme emission line' galaxies at $z \sim 1.7$, which they interpret as having recently experienced intense starbursts of duration $\sim 15$ Myr (for observations of extreme emission line galaxies at higher redshift, see Forrest et al. 2017). At $2.1 < z < 2.6$, Shivaei et al. (2015) measure larger scatter in the SFR–$M_\star$ relation for $M_\star \sim 10^{9.5}$–$10^{11.5} \, M_\odot$ galaxies when H$\alpha$ is used to measure the SFR relative to the UV continuum (see also Smit et al. 2016 for observational constraints at $z \sim 4$). While Shivaei et al. (2015) caution that uncertainties in dust attenuation and IMF variations preclude directly interpreting this measurement in terms of bursty star formation, Sparre et al. (2017) compared the difference in the scatter

---

[2] The mean H$\alpha$-to-UV ratio decreases with decreasing galaxy mass, which can also be interpreted as a signature of bursty star formation: if time intervals between bursts are sufficiently long, H$\alpha$ will be depressed relative to the UV continuum most of the time.

[3] Since H$\beta$ is also powered by ionizing radiation, it probes the same time-scale as H$\alpha$.

between the simulated H $\alpha$- and UV continuum-derived SFR–$M_\star$ relations predicted by the FIRE simulations to the observations of Shivaei et al. (2015) and showed that the observed scatter difference is consistent with the order-of-magnitude SFR variations predicted by the simulations on time-scales as short as a few Myr.

Our model for bursty star formation builds on previous work on understanding the regulation of star formation by stellar feedback. Several studies have shown that simple analytic models in which ISM pressure sustained by stellar feedback (e.g. via turbulence) balances the weight of disc gas can explain the observed KS relation (e.g. Thompson, Quataert & Murray 2005; Ostriker, McKee & Leroy 2010; Ostriker & Shetty 2011; Faucher-Giguère, Quataert & Hopkins 2013). Recently, Torrey et al. (2017) showed that such equilibrium models break down in galactic nuclei, where local dynamical time-scales are too short for stellar feedback to effectively respond and establish a steady balance between feedback and gravity. Torrey et al. (2017)'s simulations showed that the failure of stellar feedback to establish a steady equilibrium leads to bursty star formation in galactic nuclei. We analyse here other limits of feedback-regulated star formation models and argue that bursty star formation at high redshift and in dwarf galaxies can be understood as two different failures of stellar feedback to establish a steady equilibrium, the first due to the time-scale for stellar feedback and the second due to the discreteness (stochastic sampling) of star-forming regions. At high redshift, where gas fractions are elevated, the two effects act in concert. Our derivations deliberately involve a number of simplifications and are intended to explain in simple terms *why* star formation is bursty in some systems but time steady in others, rather than to be quantitatively exact. The fully dynamical numerical simulations referenced above are better suited for more detailed predictions.

We describe our star formation burstiness model and present our results in Section 2. Section 3 discusses our findings and concludes. Appendices summarize supporting data and calculations. Throughout, we assume a standard flat $\Lambda$ cold dark matter ($\Lambda$CDM) cosmology with parameters consistent with the latest constraints ($H_0 \approx 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 1 - \Omega_\Lambda \approx 0.27$ and $\Omega_b \approx 0.046$; Planck Collaboration XIII 2016).

## 2 ANALYTIC MODEL

### 2.1 Preliminaries

Our model is based on several ideas for how star formation is regulated in galaxies. There are different models for galactic star formation in the literature but there is as yet no generally agreed upon theory. We therefore begin by briefly reviewing the key assumptions necessary to understand our model. This is not intended to be a thorough review of the field (for reviews that also discuss alternate theories, see e.g. McKee & Ostriker 2007; Krumholz 2014) but rather to summarize the elements that we adopt in our modelling. We also note that our model for star formation burstiness is primarily based on recent simulations and analytic models of star formation regulation in galaxies. Although we review some observational constraints on star-forming regions for context below, we generally do not anchor our modelling to them; as we will explain, many results on individual star-forming regions are likely not representative of the physics that modulate galaxy-integrated SFRs.

*Most star formation occurs in Toomre-scale GBCs.* It is well known that stars form in molecular clouds (e.g. Myers et al. 1986; Mooney & Solomon 1988; Scoville & Good 1989; Williams & McKee 1997). In this paper, we are interested in variability in the integrated SFR of galaxies, which is a sum over the SFRs of individual star-forming clouds. To correctly capture the expected SFR variance, it is critical to carefully define what counts as independent star-forming regions in the sum.

The ISM is turbulent and highly inhomogeneous. As a result, clouds that undergo gravitational collapse generally have several different centres of collapse, each potentially corresponding to its own overdensity of molecular gas and nascent star cluster. These multiple centres of collapse and resulting star clusters are evident in the filamentary images of GBC simulations (e.g. Padoan, Haugbølle & Nordlund 2012; Lee, Chang & Murray 2015; Grudić et al. 2016; Raskutti, Ostriker & Skinner 2016). The same simulations show that when the parent cloud collapses, a fraction $\epsilon_{\text{GBC}}$ of the initial gas mass is converted into stars on a time-scale of just $\approx (1–3) t_{\text{ff}}^{\text{GBC}}$, where $t_{\text{ff}}^{\text{GBC}}$ is the gravitational free fall time of the cloud evaluated at its mean initial density. When stellar feedback is included, it usually truncates star formation and limits $\epsilon_{\text{GBC}}$. Since the different centres of collapse all form their stars during the collapse of the parent cloud, the parent cloud can be approximated as experiencing a single coherent burst on a time-scale $\approx (1–3) t_{\text{ff}}^{\text{GBC}}$.

If we are interested in galaxy-integrated burstiness, what distribution of independent star-forming clouds should we sum over? Since parent GBCs can each contains a large number of molecular overdensities and form a large number of different star clusters, the usual molecular cloud or star cluster mass functions measured in observations (e.g. Rosolowsky 2005; Fall & Chandar 2012; Rice et al. 2016; Miville-Deschênes, Murray & Lee 2017) do not count the relevant independent star formation units but are instead expected to include correlated centres of collapse. We therefore appeal to theory to guide us.

In general, galactic discs appear to be well described by a Toomre $Q$ parameter near unity (Toomre 1964; Goldreich & Lynden-Bell 1965), corresponding to a state of marginal gravitational stability. In a smooth disc with $Q = 1$, a single physical scale, the 'Toomre mass' $M_T \approx \pi h^2 \Sigma_g$ (where $h$ is the disc thickness and $\Sigma_g$ is its gas surface density) is subject to gravitational instability. Perturbations on smaller length scales are stabilized by pressure in the disc, while perturbations on larger scales are stabilized by rotation. Smooth discs are therefore expected to fragment into GBCs of a definite mass scale $\sim M_T$. During gravitational collapse, each Toomre-scale GBC would then hierarchically fragment into a large collection of smaller clouds, which together would correspond to the observed cloud mass function.

Real galactic discs are not smooth but turbulent and in reality turbulent density fluctuations make a finite range of physical scales unstable to gravitational collapse. Hopkins (2013) developed a general theory of gravitational fragmentation in turbulent media and addressed the 'clouds within clouds' problem using an excursion set formalism (a generalization of the methodology used in cosmological structure formation to distinguish locally bound structures from larger structures that contain them; e.g. Bond et al. 1991). In the excursion set terminology, the largest (parent) gravitationally bound structures correspond to 'first crossings' of the collapse barrier. When applied to rotating discs, Hopkins (2013) showed that the mass in first-crossing structures is strongly concentrated in objects near the Toomre scale (Hopkins 2013 's fig. 3).

While the full mass function of GBCs extends over many orders of magnitude, with a mass spectrum $dN/dm \sim m^{-\alpha}$ with $\alpha \sim 2$ in agreement with observational constraints on molecular cloud and star cluster mass functions, the first-crossing mass distribution decreases much more steeply with decreasing mass. The first-crossing mass distribution is instead strongly peaked near $M_T$ (i.e. with most

of the mass contained within just $\sim$1–3 orders of magnitude of $M_T$). These first-crossing, parent, Toomre-scale GBCs undergoing coherent gravitational collapse are the independent clouds that we should sum over to properly capture the variance expected from stochastic sampling of star-forming regions.

*Within an individual GBC, star formation occurs in a burst.* As mentioned above, dynamical simulations indicate that GBCs typically form stars in one main burst lasting $\approx$(1–3)$t_{\rm ff}^{\rm GBC}$. During this burst, the SFR in the GBC increases with time, until the collapse of the cloud is disrupted by stellar feedback. For example, the numerical simulations of Lee et al. (2015) indicate that SFR $\propto t$ in GBCs. At high redshift, where GBC masses tend to be large because of elevated gas fractions and where free fall times are shorter (see Sections 2.3–2.5 below), this can lead to quite intense bursts in individual GBCs:

$$\mathrm{SFR} \sim 10\,\mathrm{M_\odot\,yr^{-1}} \left(\frac{\epsilon_{\rm GBC}}{0.1}\right) \left(\frac{m}{10^9\,\mathrm{M_\odot}}\right) \left(\frac{t_{\rm ff}^{\rm GBC}}{10\,\mathrm{Myr}}\right). \tag{1}$$

We must address here observations that appear to contradict rapid, time dependent in star formation in molecular clouds. Indeed, a large number of measurements indicate low star formation efficiencies per free fall time in molecular gas, $\epsilon_{\rm ff}^{\rm mol} \sim 0.01$, across large cloud samples and on different scales (e.g. Krumholz & Tan 2007; García-Burillo et al. 2012; Krumholz, Dekel & McKee 2012; Evans, Heiderman & Vutisalchavakul 2014; Heyer et al. 2016). Unfortunately, because the relevant units of coherent star formation in our model are Toomre-scale GBCs, it is generally not possible to directly compare star formation efficiencies measured for individual molecular clouds to our assumption of dynamic star formation in GBCs. This is because most observed star-forming clouds are not representative of Toomre-scale GBCs. Indeed, as explained above, most molecular clouds identified in observations are likely subunits of more massive GBCs. Furthermore, most observed molecular clouds are not actually gravitationally bound. In the large molecular cloud catalogue of Miville-Deschênes et al. (2017), for example, only some of the most massive clouds are bound according to their measured virial parameter; by number most molecular clouds are unbound.

Nevertheless, observations do provide some support for dynamic star formation. Simulations with continuously driven turbulence predict that the star formation efficiency per free fall time is not universal but rather a function of the virial parameter of the cloud, $\alpha_{\rm vir}$, which measures the degree of gravitational boundedness via the ratio of kinetic energy to gravitational binding energy (Bertoldi & McKee 1992). Dynamic star formation, on average, predicts an increasing star formation efficiency with decreasing virial parameter (e.g. Padoan et al. 2012; Lee et al. 2015). A strong trend in qualitative agreement with this prediction has recently been observed in M51(Leroy et al. 2017). In the Milky Way, attempts to identify a similar relationship with the virial parameter have however not revealed a clear trend (e.g. Vutisalchavakul et al. 2016). This suggests that the observational results may be sensitive to choices of observational tracers or definitions of the virial proxy or star formation efficiency (e.g. Leroy et al. studied regions of fixed size, whereas Vutisalchavakul et al. analysed clouds of varying size). Following a different approach, Lee, Miville-Deschênes & Murray (2016) analysed the dispersion of observationally inferred $\epsilon_{\rm ff}^{\rm mol}$ in Milky Way molecular clouds and argued that the large dispersion is inconsistent with a time-independent $\epsilon_{\rm ff}^{\rm mol}$ and instead favours a time-variable efficiency.

Although clearly more work is needed test dynamic star formation observationally, we take the fact that dynamic star formation appears to be a generic prediction of simulations with self-gravity as our main motivation for assuming a time-dependent SFR in GBCs.

*On galactic scales, star formation is regulated by a balance between gravity and stellar feedback.* Averaged over entire galaxies, the star formation efficiency per free fall time, $\epsilon_{\rm ff}^{\rm gal}$, has low mean value $\sim$0.02 and relatively small dispersion (e.g. Kennicutt 1998; Genzel et al. 2010; Krumholz et al. 2012). Our view in this paper is that the low star formation efficiency per free fall time on galactic scales is set by a global balance between the ISM pressure excited by stellar feedback and gravity, with only a weak (or no) dependence on the local star formation efficiency within individual GBCs. In this picture, $\epsilon_{\rm ff}^{\rm gal}$ is low because only a relatively small SFR is needed for stellar feedback to support the ISM against runaway gravitational collapse. Several analytic models have been formulated based on this ansatz and appear broadly consistent with observed star formation efficiencies (Thompson et al. 2005; Ostriker & Shetty 2011; Faucher-Giguère et al. 2013). Moreover, a number of hydrodynamic simulations including stellar feedback indicate that feedback can maintain the galactic ISM in rough vertical hydrostatic balance and/or that the star formation efficiency in a feedback-regulated ISM is not sensitive to the small-scale (sub-GBC) star formation prescription (e.g. Hopkins, Quataert & Murray 2011; Shetty & Ostriker 2012; Kim & Ostriker 2015; Orr et al. 2017; Torrey et al. 2017).

Our picture of feedback-regulated star formation is distinct from another popular class of models in which it is assumed that the star formation efficiency in molecular clouds is universally low owing to the properties of supersonic turbulence in media with virial parameter of order unity (e.g. Krumholz & McKee 2005; Krumholz, McKee & Tumlinson 2009; Federrath 2013). In those models, the low star formation efficiency on galactic scales is inherited from the low star formation efficiency $\epsilon_{\rm ff}^{\rm mol} \sim 0.02$ in molecular clouds. In future work, it would be interesting to also investigate the galaxy-scale star formation variability predicted by universal star formation efficiency models, as this could provide a new test of the physics of star formation regulation.

## 2.2 The FG13 feedback equilibrium model

The starting point for our analysis is the analytic feedback-regulated model of Faucher-Giguère et al. (2013, hereafter FG13). The FG13 model predicts a KS relation that emerges from a balance between gravity and supernova (SN) feedback in galactic discs, and also shows how the KS relation imposes certain consistency requirements on the number of massive star-forming GBCs active at any time. We refer to that paper for more details and summarize here only the essential elements.

We approximate galaxies with a two-zone model consisting of a volume-filling ISM and a collection of Toomre-scale GBCs in which star formation is confined. The ISM is modelled as a thin disc with radius-dependent mean gas density $\bar\rho$ and is assumed to be in vertical hydrostatic balance, supported by turbulence excited by SN feedback. For simplicity we assume flat rotation curves, which we model using an isothermal potential with velocity dispersion $\sigma$ (see also Thompson et al. 2005). The circular velocity is then $v_{\rm c} = \sqrt{2}\sigma$.

Throughout this paper, we assume that galactic discs tend to self-regulate to a Toomre parameter:

$$Q = \frac{\kappa c_{\rm T}}{\pi G \Sigma_{\rm g}} = \frac{2\sigma c_{\rm T}}{\pi G \Sigma_{\rm g} r} \approx 1, \tag{2}$$

where $\kappa \equiv \sqrt{4\Omega^2 + \mathrm{d}\Omega^2/\mathrm{d}\ln r} = 2\sigma/r$ is the epicyclic frequency and $c_{\rm T}$ is the velocity dispersion of the turbulence (e.g. Quirk 1972;

Kennicutt 1989; Martin & Kennicutt 2001). We expect the ISM to self-regulate to this value, corresponding to marginal gravitational stability, because turbulence dissipation tends to decrease $Q$ below unity. When $Q < 1$, gravitational fragmentation increases the SFR and the associated heating of the disc by stellar feedback pushes the disc back to $Q \approx 1$. Using $\Sigma_g \equiv 2h\bar{\rho}$, where $h$ is the gaseous disc scale height, we can solve for the free fall time in the ISM at the half-mass radius $r_{1/2}$,

$$t_{\rm ff}^{\rm disc}(r_{1/2}) = \left( \frac{3\pi}{32\,G\,\bar{\rho}(r_{1/2})} \right)^{1/2}$$
$$= \left( \frac{3Q}{64 \times 2^{1/2}} \right)^{1/2} t_{\rm orb}(r_{1/2}), \tag{3}$$

where $t_{\rm orb}(r) \equiv 2\pi r/v_c$ is the orbital time. For $Q = 1$, the free fall time reduces to a constant fraction of the orbital time:

$$t_{\rm ff}^{\rm disc}(r_{1/2}) \approx 0.2 t_{\rm orb}(r_{1/2}). \tag{4}$$

By balancing the turbulent pressure with the weight of the gaseous disc normal to the disc plane, FG13 derived the following expression for the KS relation:

$$\dot{\Sigma}_\star = \frac{2\sqrt{2}\pi G Q}{\mathcal{F}} \left( \frac{P_\star}{m_\star} \right)^{-1} \Sigma_g^2. \tag{5}$$

Here, $\dot{\Sigma}_\star$ is the star formation rate surface density, $P_\star/m_\star$ is the momentum injected in turbulence by SN feedback per stellar mass formed and $\mathcal{F}$ encapsulates uncertain factors of order unity. FG13 showed that $\mathcal{F} = 2$ provides a good fit to observations for $P_\star/m_\star \approx 3000\,{\rm km\,s^{-1}}$, appropriate for SN feedback and a standard Kroupa IMF (e.g. Cioffi, McKee & Bertschinger 1988; Ostriker & Shetty 2011; Martizzi, Faucher-Giguère & Quataert 2015). FG13 also showed that the prediction of this model can be expressed in terms of the dimensionless star formation efficiency per free fall time on galactic scales (such that $\dot{\Sigma}_\star \equiv \epsilon_{\rm ff}^{\rm gal} \Sigma_g / t_{\rm ff}^{\rm disc}$) as

$$\epsilon_{\rm ff}^{\rm gal} \approx \frac{\sqrt{3}\pi}{2^{7/4}\mathcal{F}} \frac{f_g v_c}{P_\star/m_\star}, \tag{6}$$

where $f_g$ is the gas mass fraction. For most observed galaxies, this galaxy-scale star formation efficiency is small, of order one to a few per cent (e.g. Genzel et al. 2010; Krumholz et al. 2012). Starbursts can be induced in galaxy mergers as a result of strong gravitational torques that efficiently funnel large amounts of gas into the nucleus of the merger remnant (e.g. Toomre & Toomre 1972; Barnes & Hernquist 1991). In this case, the burst can occur on a short time-scale because the high-density gas concentration in the nucleus can reach short dynamical times $\lesssim 1\,{\rm Myr}$ in the inner 100 pc. However, galaxy mergers are too rare to explain the frequent SFR fluctuations seen in Fig. 1.

In the following sections, we describe two regimes in which ordinary (non-merging) galaxies can experience large deviations from the KS relation. These deviations and the accompanying bursts of star formation can occur either spontaneously due to instabilities in the disc or as a result more minor external perturbations, such as smooth inflows from the intergalactic medium.

## 2.3 Breakdown of equilibrium due to the supernova feedback time-scale

Implicit in FG13's equilibrium model is the assumption that SN feedback can respond sufficiently rapidly to gravitational collapse of the disc to establish hydrostatic balance. Torrey et al. (2017) noted that this assumption breaks down in galactic nuclei, where
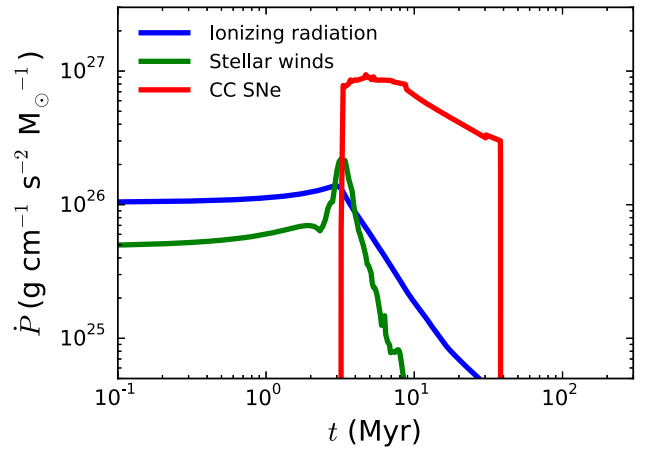


**Figure 2.** Momentum flux output by the main stellar feedback processes (ionizing radiation, stellar winds and core-collapse SNe) per stellar mass formed as a function of time since a burst of star formation (see Section 2.3 for details). Prompt feedback from ionizing radiation and stellar winds can disrupt stellar birth clouds but the integrated momentum output in the ISM is dominated by SNe.

local dynamical time-scales are shorter than a 'stellar feedback' time-scale. We argue here that a similar effect is in part responsible for bursty star formation in high-redshift galaxies (as a whole). In the early Universe the characteristic dynamical time-scales of galaxies were shorter, but stellar evolution proceeded at a constant rate. As the disc free fall time becomes smaller than the stellar feedback time-scale, it becomes impossible for the bulk of galaxies to reach a tight balance between feedback and gravity. In this limit, the instantaneous star formation efficiency in the disc can deviate from the median KS relation by a large factor.

Fig. 2 shows the momentum flux output by the main stellar feedback processes (ionizing radiation, stellar winds and core-collapse SNe) per stellar mass formed as a function of time since a burst of star formation. The plot, produced using STARBURST99 v7.0.1 (Leitherer et al. 1999), assumes a Kroupa (2001) IMF and solar metallicity. The SN energy rate returned by STARBURST99 is for the 'prompt' kinetic energy of SN ejecta. As SNe expand into the ISM, the radial momentum of SN remnants is enhanced by about an order of magnitude during the Sedov–Taylor (energy-conserving) phase (e.g. Cioffi et al. 1988; Blondin et al. 1998; Martizzi et al. 2015). This 'boosted' momentum is the momentum available to drive interstellar turbulence and galactic winds. We therefore plot in the figure the boosted SN momentum calculated using fits to the evolution of SNRs in an inhomogeneous ISM from the simulations of Martizzi et al. (2015).[4] As the figure shows, SNe dominate the momentum input in the ISM by a large factor relative to radiation and stellar winds.

SN feedback has two characteristic time-scales. The first is the time-scale ($t_{\rm 1st} \approx 3\,{\rm Myr}$) for the first SNe to explode following a burst of star formation. Before this time, SN feedback cannot oppose the gravitational collapse converting gas into stars. The second is the time-scale for most SNe to explode. This is roughly the time-scale over which SN feedback acts. For a Kroupa IMF, $\approx 60$ per cent of the cumulative energy and momentum from core-collapse SNe have been returned in the ISM by $\approx 20\,{\rm Myr}$ following a burst of

---

[4] Specifically, we use the fit appropriate for an inhomogeneous ISM with mean density $\langle n_{\rm H} \rangle = 1\,{\rm cm^{-3}}$, Mach number $\mathcal{M} = 30$ and solar metallicity. Each SN is assumed to have an energy of $10^{51}\,{\rm erg}$.

star formation (insensitive to metallicity). We thus assume a SN feedback time-scale $t_{SN} \approx 20\,\mathrm{Myr}$ for the numerical estimates in this paper.

The characteristic time-scales of SN feedback introduce two different effects. In a medium with $t_{ff} \ll t_{1st}$, most of the gas can be converted into stars in a rapid burst before SN feedback can disperse the birth cloud. As we will show below, typical free fall times are particularly short in the early Universe so this limit could be relevant for explaining the early formation of bound star clusters that are observed as old globular clusters today (which requires a high star formation efficiency). Properly understanding this limit would require modelling the effects of radiative feedback that acts before $t_{1st}$, which is beyond the scope of this paper. We focus instead on effects introduced by the longer time-scale $t_{SN}$. Since SNe dominate the total momentum output into the ISM, we assume that SN feedback is the primary source of turbulent pressure in the ISM.[5]

When $t_{ff}^{disc} \lesssim t_{SN}$, a feedback-supported galactic disc will be susceptible to large deviations from the KS relation.

Initially, this is because gravitational collapse on a time-scale $\approx t_{ff}^{disc}$ can form stars with an elevated efficiency before SN feedback can respond (on a time-scale $\approx t_{SN}$) and push the disc back towards vertical hydrostatic balance with $Q \approx 1$. Subsequently, an overshoot effect can occur. This is because stellar feedback continues to operate for a fixed period of time (set by stellar evolution) after star formation ends locally, regardless of the state of the surrounding gas, which can lead to strong gas blowouts. Such bursts of outflowing gas are seen in the galactic nucleus simulations of Torrey et al. (2017). They are also seen as 'gusty' galactic winds following star formation bursts in cosmological simulations such as the FIRE simulations shown in Fig. 1. The near evacuation of the ISM by strong outflows explains the periods of highly suppressed star formation apparent in Fig. 1, where blowouts can suppress the instantaneous SFR by up to $\gtrsim 2$ orders of magnitude.

We show below that the limit of unstable feedback regulation described above, leading to star formation burst-outflow-suppressed star formation cycles, is typical of high-redshift ($z \gtrsim 1$) galaxies of all masses and of gas-rich dwarf galaxies all the way to the present time.

For the free fall time, we use a simple model for the cosmological evolution of galactic discs, inspired by classic models in which disc sizes can be predicted by assuming that baryons have the same specific angular momentum as their parent dark matter haloes (e.g. Fall & Efstathiou 1980; Dalcanton, Spergel & Summers 1997; Mo, Mao & White 1998). Specifically, we assume that at any redshift the half-mass disc radius is a constant fraction $\approx 2$ per cent of the halo virial radius:

$$r_{1/2} \approx 0.02 R_{vir}. \tag{7}$$

Shibuya, Ouchi & Harikane (2015) show that this scaling holds in observations (on average to better than a factor of 2) from $z = 0$ to $z \sim 8$. In a study focused at low redshift, Kravtsov (2013) found that galaxies obey a consistent scaling over more than eight orders of magnitude in stellar mass. Here, $R_{vir}$ is the virial radius for haloes defined to have a redshift-dependent enclosed overdensity $\Delta_c(z)$ relative to the critical density, $\rho_c$, following Bryan &

Norman (1998). Section A1 in Appendix A provides more details on how we compute galaxy radii.

We note that cosmological hydrodynamic simulations clearly show that feedback (and not just angular momentum inherited from halo formation) is critical to produce realistic galaxy sizes (e.g. Sales et al. 2010; Crain et al. 2015; Agertz & Kravtsov 2016). We do not attempt here to model in detail what sets galaxy sizes, but rather simply use the above observationally supported scaling. We focus in this paper on late-type star-forming galaxies. Early-type galaxies, which we do not consider, may retain a smaller fraction of the specific angular momentum of their parent haloes and may therefore be somewhat more compact (e.g. Genel et al. 2015).

Next, we need a model for the circular velocities of galaxies. In Section A2 of Appendix A, we show that a model in which galaxy circular velocities are a constant factor of the maximum circular velocity of NFW (Navarro, Frenk & White 1997) haloes simultaneously matches the observed Tully–Fisher relation at $z = 0$ and its evolution to $z = 2$. In this model, the galaxy circular velocity is related to halo properties following:

$$v_c \approx 0.465 \left( \frac{c_{vir}}{A(c_{vir})} \right)^{1/2} \left( \frac{GM_{vir}}{R_{vir}} \right)^{1/2}, \tag{8}$$

where $c_{vir}$ is the halo concentration (see equation A9) and $A$ is a dimensionless function of concentration (see equation A8). We use the Moster, Naab & White (2013) abundance matching relation to convert between halo mass and galaxy stellar mass.

For a disc with $Q = 1$, we can use equation (4) to evaluate the free fall time in the ISM:

$$
\begin{aligned}
t_{ff}(r_{1/2}) &\approx 0.0264 \left( \frac{1}{G\rho_c(z)} \right)^{1/2} g(z) \\
&\approx \frac{0.0764}{H(z)} g(z),
\end{aligned} \tag{9}
$$

where

$$g(M_{vir}, z) \equiv \left( \frac{A(c_{vir})}{\Delta_c(z)c_{vir}} \right)^{1/2} \tag{10}$$

and $H$ is the Hubble parameter. In our numerical calculations, we evaluate the function $g$ using the full equations given in the Appendix but note that it is only weakly dependent on mass and redshift.

Fig. 3 compares $t_{ff}^{disc}(r_{1/2})$ for different fixed stellar masses and $t_{SN}$ versus redshift. For $z \lesssim 1.3$, $t_{SN} < t_{ff}^{disc}(r_{1/2})$ and SN feedback has enough time to establish an equilibrium and effectively regulate star formation in galaxies. At $z \gtrsim 1.3$, however, the feedback time-scale becomes longer than $t_{ff}^{disc}(r_{1/2})$ so galaxies can experience intense bursts of star formation to which SN feedback cannot respond sufficiently rapidly.[6] This result is nearly independent of galaxy mass and we propose that it explains in part why high-resolution cosmological simulations with time-resolved stellar feedback show bursty

---

[5] There may be exceptions, such as at very high gas surface densities where multiple scatterings of re-processed infrared radiation can enhance radiative feedback (Scoville 2003; Thompson et al. 2005) or in low surface density regions, where photoionization and photoheating can be important (e.g. Ostriker et al. 2010).
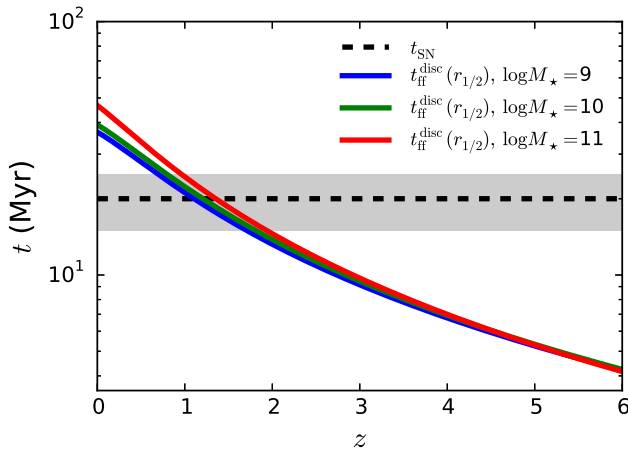
[6] Another relevant time-scale is the time necessary for pre-existing ISM turbulence to dissipate (absent a driving source) $t_{turb}$. If $t_{turb}$ were longer than $t_{ff}^{disc}$, turbulent pressure support would prevent the disc from collapsing on a free fall time and our argument would require modification. However, we can show that $t_{turb} \lesssim t_{ff}^{disc}$ in general. Assuming that the largest turbulent eddies have a size equal to the disc thickness $h$, $t_{turb} \approx h/c_T$ (this is true for both supersonic and subsonic turbulence, and also in the presence of magnetic fields; e.g. Stone, Ostriker & Gammie 1998; Mac Low 1999). Using $h/r \approx c_T/v_c$ (FG13) to eliminate $h$ and $c_T$ in favour of $r$ and $v_c$, we find that for a disc with $Q \approx 1$, $t_{turb} \approx 0.16 t_{orb} \lesssim t_{ff}^{disc}$ (see equation 4), independent of other disc properties.

**Figure 3.** Comparison of the free fall time at the half-mass radius of galactic discs, $t_{\rm ff}^{\rm disc}$, as a function of redshift for different stellar masses (solid coloured curves) to the SN feedback time-scale $t_{\rm SN} = 20\,{\rm Myr}$ (over which most of the SN feedback acts; dashed black) as a function of redshift. Galactic dynamical times are a constant fraction of the age of the Universe, so they are shorter at high redshift (only weakly dependent on galaxy mass), while the SN feedback time-scale is constant. At $z \gtrsim 1$, characteristic free fall times become shorter than the SN feedback time-scale and the feedback cannot effectively respond to gravitational collapse. Galactic star formation should be bursty in this regime The grey band covers $\pm 5\,{\rm Myr}$ around the fiducial feedback time-scale as an indication of the sensitivity of the results to the particular choice.

star formation histories at high redshift for all galaxies (e.g. Hopkins et al. 2014; Feldmann et al. 2017). Quantitatively, this result is in good agreement with the transition from bursty to time-steady star formation occurring around $z \sim 1$ for massive galaxies in the FIRE simulations (Fig. 1).

Since $g$ is only weakly dependent on $M_{\rm vir}$ and $z$, we can push our analytics further to derive an expression for the 'burstiness redshift' $z_{\rm burst}$ above which we expect bursty star formation. For this estimate, we use $g \approx 0.034$. Then,

$$t_{\rm ff}^{\rm disc}(r_{1/2}) \approx \frac{0.0026}{H(z)}. \tag{11}$$

This is a version applied to galactic discs of the well-known result that the characteristic dynamical time of haloes is a constant fraction of $1/H(z)$. Since the age of the Universe $\sim 1/H(z)$ at any redshift, this implies that the characteristic free fall times of galactic discs must become shorter than the constant $t_{\rm SN}$ at high redshift.

At $z \gtrsim 1$, where dark energy is negligible, $H(z) \approx H_0\sqrt{\Omega_{\rm m}(1+z)^3}$. We can then analytically solve for $z_{\rm burst}$ by setting $t_{\rm SN} = t_{\rm ff}^{\rm disc}(r_{1/2})$:

$$z_{\rm burst} \approx \left(\frac{0.0026}{H_0 t_{\rm SN}\sqrt{\Omega_{\rm m}}}\right)^{2/3} - 1$$

$$\approx 2.3 \left(\frac{t_{\rm SN}}{20\,{\rm Myr}}\right)^{-2/3} - 1. \tag{12}$$

This matches the nearly mass-independent crossing of $t_{\rm SN}$ and $t_{\rm ff}^{\rm disc}(r_{1/2})$ in Fig. 3. The transition from bursty to time steady will in practice be gradual around that redshift because feedback and gravitational collapse both operate continuously in time, as well as because galaxies have significant dispersion in their properties at any mass and redshift (which affect their internal free fall times). We indicate this in the figure with the horizontal grey band, which covers $\pm 5\,{\rm Myr}$ around the fiducial feedback time-scale as an

approximation of the sensitivity of our results to the choice of feedback time-scale.

### 2.4 Breakdown of equilibrium due to the discreteness of star formation

In the FIRE simulations shown in Fig. 1, massive galaxies settle into a more time-steady mode of star formation at low redshift but dwarf galaxies continue to experience bursty star formation down to $z \sim 0$. The burstiness of dwarfs at $z < 1$ cannot be explained by the previous galaxy-scale time-scale argument, but we propose that it can instead be explained by the discreteness of star-forming regions.

As discussed in Section 2.1, most of the star formation in galaxies at any given time occurs in Toomre-scale GBCs. In individual GBCs, the simulations summarized in Section 2.1 show that the SFR tends to be bursty as the cloud collapses. In the limit in which only a small number of GBCs contribute to the galactic SFR at any given time, the galactic SFR will inherit the time dependence of GBC-scale star formation. In nearby galaxies in which the KS relation has been studied at high spatial resolution, the scatter is measured to increase by a large factor when the relation is measured for galaxy patches of size $\lesssim 300\,{\rm pc}$–$1\,{\rm kpc}$ (Schruba et al. 2010; Liu et al. 2011) rather than averaging over the entire galaxy. These observations are consistent with stochastic sampling of bright star-forming regions causing departures from the median KS relation, since fewer star-forming regions are averaged over when measuring on small scales within galaxies (e.g. Orr et al. 2017; Torrey et al. 2017). In analogy with the averaging necessary for the KS relation to be tight, we expect that only galaxies in which a sufficiently large number of GBCs efficiently form stars at any time can produce time-steady galactic SFRs.

In what follows, we derive a scaling for how the number of Toomre-scale GBCs depends on galaxy properties. Since galaxies contain GBCs spanning a spectrum of masses but most star formation is expected to occur within coherently collapsing Toomre-scale clouds, we characterize galaxies by the parameter $N_{\rm U}$, defined as the number of GBCs that would be present if all of the self-gravitating mass were distributed in clouds of mass $M_{\rm U} \approx M_{\rm T}$, the maximum (upper) GBC mass. Appendix B shows how the minimum normalized SFR variance, $\sigma_{\rm SFR}/\langle {\rm SFR} \rangle$, scales with $N_{\rm U}$. Specifically, for individual GBCs with SFR $\propto t$,

$$\frac{\sigma_{\rm SFR}}{\langle {\rm SFR} \rangle} \approx \frac{1}{\sqrt{\gamma N_{\rm U}/0.5}}, \tag{13}$$

where $\gamma$ is the fraction of the time each GBC actually forms stars. As shown in the Appendix, the exact numerical value of the numerator in the above expression depends on the parameters describing the GBC mass function, but only weakly. This result shows that the SFR variance is roughly the Poisson variance expected for the number of Toomre-scale GBCs actively forming stars at any given time. Equation (13) is the minimum SFR variance expected because it only takes into account stochastic sampling of Toomre-scale regions. In particular, this estimate neglects the near-complete SFR suppression that can follow star formation bursts owing to evacuation of the gas reservoir (see Section 2.3), or temporary enhancements due to accretion of new gas (either smoothly from the intergalactic medium or in galaxy mergers). These effects can significantly increase the actual SFR variance that will be observed in bursty galaxies.

Assuming that all star formation proceeds in GBCs and that each GBC ultimately converts a fraction $\epsilon_{\rm int}^{\rm GBC}$ of its initial gas mass

into stars (determined by how stellar feedback operates in GBCs), the requirement that galaxies lie on the KS relation constrains the fraction $f_{GBC}$ of the total ISM mass found in GBCs at any time. Using $M_U \approx M_T \approx \pi h^2 \Sigma_g$, we find that

$$N_U = \frac{2(\pi r_{1/2}^2)\Sigma_g(r_{1/2})f_{GBC}}{M_U}$$

$$= 2\left(\frac{r_{1/2}}{h}\right)^2 f_{GBC}$$

$$= \frac{16}{Q^2 f_g^2} f_{GBC}. \tag{14}$$

The last expression uses the fact that for our disc model $h/r = Q f_g / 2^{3/2}$ (i.e. more gas-rich discs are thicker) to eliminate the disc thickness in favour of the gas fraction. Using this relation, we can also show that $M_U \propto f_g^3 M_\star$. We note that, because the SFR in individual GBCs is time dependent, only a fraction of all GBCs will be observed as luminous star-forming regions at any given time, consistent with the wide range of instantaneous star formation efficiencies inferred in Milky Way GBCs (e.g. Lee et al. 2016).

To make progress, we need an expression for $f_{GBC}$. FG13 showed that since we assume that all star formation occurs in GBCs, we can write

$$\epsilon_{ff}^{gal} = \frac{f_{GBC}}{\tilde{t}_{GBC}}\epsilon_{int}^{GBC}, \tag{15}$$

where

$$\tilde{t}_{GBC} \equiv \left(\frac{t_{GBC}}{t_{ff}^{disc}}\right) \tag{16}$$

and $t_{GBC}$ is the lifetime of GBCs. This is simply the consistency condition that the sum of star formation occurring in all GBCs must equal the total galactic star formation. In what follows, we assume $\tilde{t}_{GBC} \approx 1$. This is a good approximation to the GBC simulations discussed in Section 2.1, noting that GBCs assemble on a time-scale $t_{ff}^{disc}$, and models show that GBCs collapse and get dispersed by stellar feedback on a time-scale comparable to their free fall time (e.g. Grudić et al. 2016; Kim, Kim & Ostriker 2016). Then

$$f_{GBC} \approx \frac{\epsilon_{ff}^{gal}}{\epsilon_{int}^{GBC}}, \tag{17}$$

and therefore, using equation (6) for $\epsilon_{ff}^{gal}$,

$$N_U \approx \frac{8\sqrt{3}\pi v_c}{2^{3/4}\mathcal{F}Q^2 f_g(P_\star/m_\star)\epsilon_{int}^{GBC}}. \tag{18}$$

The model thus predicts decreasing $N_U$, and therefore increasing SFR variability, with increasing gas fraction (at fixed $v_c$).

Defining a critical minimum number of Toomre-scale GBCs necessary to exhibit steady star formation, $N_U^{crit}$, we can derive a minimum galaxy circular velocity necessary for steady star formation:

$$v_{c,burst} \approx \frac{2^{3/4}\mathcal{F}Q^2(P_\star/m_\star)\epsilon_{int}^{GBC}f_g}{8\sqrt{3}\pi}N_U^{crit}$$

$$\approx 46\,\mathrm{km\,s^{-1}}\left(\frac{f_g}{0.2}\right)\left(\frac{P_\star/m_\star}{3000\,\mathrm{km\,s^{-1}}}\right)$$

$$\times \left(\frac{\epsilon_{int}^{GBC}}{0.1}\right)\left(\frac{N_U^{crit}}{10}\right), \tag{19}$$

where the last numerical expression assumes $Q = 1$, $\mathcal{F} = 2$ and $P_\star/m_\star = 3000\,\mathrm{km\,s^{-1}}$ as before. Fig. 4 plots $N_U$ versus $v_c$ for different gas fractions. We note that in reality there is no single critical
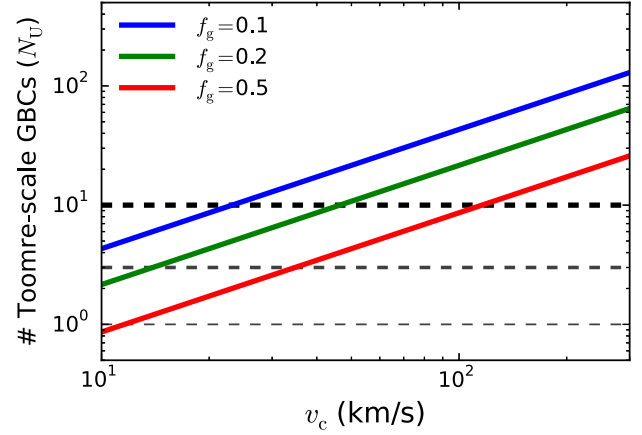


**Figure 4.** Number of Toomre-scale GBCs as a function of galaxy circular velocity, for different gas mass fractions indicated by the different colours (the curves do not depend on redshift). Within a single GBC, star formation can be highly time dependent as gravitational collapse proceeds, so a large number of GBCs must contribute in order for the galactic average to be time steady. The horizontal dashed lines indicate different values $N_U = 1$, 3 and 10 below which we expect increased SFR variability (Section 2.4 quantifies the minimum SFR variance expected versus $N_U$). Burstiness is predicted to increase in dwarf galaxies with low $v_c$. Burstiness should also increase with increasing gas fraction at fixed stellar mass because the Toomre mass increases strongly with $f_g$, so that the total number of Toomre-scale GBCs decreases.

threshold below which galaxies will be bursty; rather that there will be a gradual transition of increasing burstiness with decreasing $N_U$, with minimum variance quantified by equation (13) or its more general version in Appendix B.

If we crudely assume that a galaxy must have at least $N_U^{crit} \approx 10$ Toomre-scale GBCs in order to exhibit time-steady star formation, then this implies that for a fiducial gas fraction $f_g = 0.2$ galaxies with circular velocity $v_c \lesssim 46\,\mathrm{km\,s^{-1}}$ should be bursty, regardless of redshift. For gas fractions of $f_g \approx 0.5$, which are common at high redshift, all galaxies with $v_c \lesssim 116\,\mathrm{km\,s^{-1}}$ should be bursty. Indeed, the observations compiled in Section A3 of Appendix A show that gas fractions can be near unity at high redshift, in which case the star-forming discreteness effect can induce burstiness up to circular velocities $v_c \sim 230\,\mathrm{km\,s^{-1}}$ corresponding to massive galaxies. We stress, however, that our model also predicts a smooth transition from time-steady to bursty star formation with decreasing $v_c$, which is consistent with the trends found in observations (e.g. Weisz et al. 2012; Kauffmann 2014).

We can also express our predictions for $N_U$ in terms of stellar mass using the model for the $M_\star$–$v_c$ relation summarized in the previous section. This is done in Fig. 5. The curves in this figure depend on redshift because the conversion between $M_\star$ and $v_c$ depends on redshift. To connect more directly with observations, the solid grey curves in each panel show our model predictions for the typical gas fractions as a function galaxy mass and redshift evaluated using the scaling relations summarized in Appendix A.

## 2.5 Feedback time-scale and discreteness effects acting together: the effects of gas fractions

One effect apparent in Fig. 5 is that high-redshift galaxies (here with emphasis on $z = 2$) can be bursty owing to both because their discs collapsing too rapidly for feedback to respond effectively *and* because they contain few Toomre-scale GBCs at any given time.
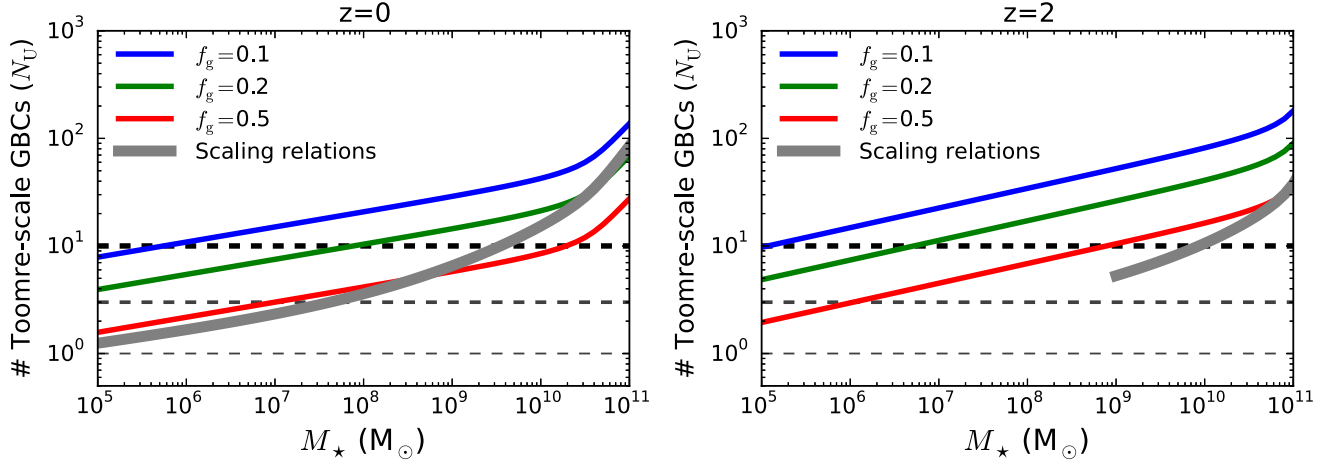
**Figure 5.** Number of Toomre-scale GBCs (as in Fig. 4) as a function of galaxy stellar mass. These curves depend on redshift because the $M_\star$–$v_c$ relation is redshift dependent. In each panel, the solid coloured curves show predictions for fixed gas fractions and the thick grey curve shows the prediction for the typical gas fraction for the given galaxy mass and redshift using the scaling relations summarized in Appendix A. The horizontal dashed lines indicate different values $N_U = 1$, 3 and 10 below which we expect increased SFR variability (Section 2.4 quantifies the minimum SFR variance expected versus $N_U$). At $z = 2$, galaxies of a given stellar mass are predicted to be more bursty (fewer massive GBCs) because of their enhanced gas fractions. This effect is in addition to the burstiness expected at high redshift from the SN feedback time-scale being longer than the characteristic ISM free fall times (Fig. 3).

As mentioned above, this is because high-redshift galaxies have high typical gas fractions, $f_g \gtrsim 0.5$ being common (e.g. Daddi et al. 2010; Tacconi et al. 2010, 2013). This result is in agreement with rest-UV observations of massive high-redshift galaxies, which show a relatively small number of very massive 'giant' star-forming clumps (the analogues of local GMCs; e.g. Elmegreen et al. 2007; Genzel et al. 2011; Wuyts et al. 2012). As a consistency check, present-day galaxies like the Milky Way should exhibit time-steady star formation according to our model, and this is in agreement with high-resolution cosmological simulations that predict that such galaxies settle into a more steady mode of star formation at $z \lesssim 1$ (Hopkins et al. 2014; Muratov et al. 2015; Sparre et al. 2017).

Fig. 6 summarizes the parameter space in which we expect burstiness either due to short galactic dynamical time-scales or to the small number of Toomre-scale GBCs. As the figure shows, our model predicts that only relatively massive galaxies at low redshift can sustain time-steady star formation.

In a different analytic analysis also expanding on the FG13 KS relation model, Hayward & Hopkins (2017) noted a different effect of gas fractions. At high gas fractions – corresponding to thick discs with high turbulent velocities – stellar feedback is predicted to be more efficient at driving galactic winds. Hayward & Hopkins (2017) argue that this effect explains why strong galactic winds are prevalent at high redshift but become weaker at late times in massive galaxies. When strong galactic winds are present, the fallback on to galaxies of outflows as they recycle (e.g. Oppenheimer et al. 2010; Anglés-Alcázar et al. 2017) can enhance the time variability of star formation by sustaining a recurrent series of 'inflow–star formation–outflow' cycles, on top of the gas accretion (and galaxy mergers) expected from the development of large-scale structure.

## 3 DISCUSSION AND CONCLUSIONS

By examining two limits in which a feedback-regulated model for the origin of the KS relation (FG13) breaks down, we have shown that the SFRs of galaxies are expected to be bursty at $z \gtrsim 1$ (for galaxies of all masses) and in dwarf galaxies (at all redshifts). At $z \gtrsim 1$, the characteristic free fall time in galactic discs is generically
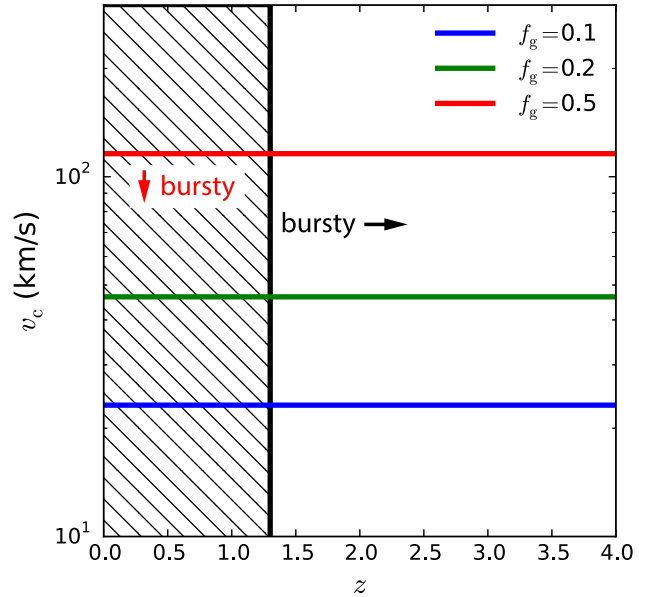


**Figure 6.** Summary of where bursty star formation is expected in the space of galaxy circular velocity versus redshift. All galaxies at $z \gtrsim 1$ are expected to be bursty (the model predicts a transition at $z = 1.3$). Below $z \sim 1$, low circular velocity galaxies are also bursty (depending on gas fraction). In this figure, the horizontal lines for different gas fractions correspond to $N_U = 10$, but in reality a continuum of burstiness is expected, increasing with decreasing $N_U$. The only galaxies in which stellar feedback can sustain steady star formation are relatively massive galaxies at $z \lesssim 1$.

shorter than the time-scale $t_{SN} \approx 20$ Myr necessary for SN feedback to output most of its energy. Therefore, SN feedback cannot establish a stable balance with gravity (Section 2.3). Below $z \approx 1$, galaxies in which star formation is confined to just a small number Toomre-scale GBCs will inherit a time-dependent SFR from their GBCs. Only in galaxies in which a reasonably large number of Toomre-scale GBCs (quantified by the parameter $N_U$) contribute to the galactic SFR can the total SFR become time steady by

averaging (Section 2.4). Our model predicts that $N_U$ decreases with galaxy mass (or, equivalently, galaxy circular velocity) and can thus explain why dwarf galaxies remain bursty all the way to $z = 0$. The effect is enhanced by the high gas fractions in dwarf galaxies, which make individual Toomre-scale GBCs relatively massive compared to the stellar mass of the galaxy.

At high redshift, including around the peak of the cosmic star formation history at $z \sim 2$ that is the focus of a number of large-scale observational efforts (e.g. Grogin et al. 2011; Steidel et al. 2014; Kriek et al. 2015; Wisnioski et al. 2015), both the time-scale and discreteness effects that we have identified should simultaneously occur owing to elevated gas fractions. In other words, high-redshift galaxies are expected to exhibit bursty star formation both because stellar feedback cannot respond sufficiently rapidly to gravitational collapse of the disc *and* because star formation occurs in a small number of massive clumps.

Overall, our model predicts that galaxies with time-steady SFRs are the exception rather than the norm, with only fairly massive galaxies at $z \lesssim 1$ (such as the Milky Way) being capable of sustaining time-steady star formation. We stress that on cosmological time-scales, galactic SFRs are sustained by gas accretion from the intergalactic medium (e.g. Kereš et al. 2005; Dekel et al. 2009; Faucher-Giguère, Kereš & Ma 2011; Davé, Finlator & Oppenheimer 2012). As a result, galaxy-averaged SFRs can evolve smoothly on cosmological time-scales even in the regimes in which our model predicts variability on shorter time-scales (the short time-scale variability must average out to satisfy constraints on the evolution of the SFR on cosmological time-scales imposed by the external gas supply). The amplitude of the SFR variability is increased by gas blowouts that follow star formation bursts (which can completely suppress star formation for a time) and gas fallback (which can fuel enhanced star formation). We also note that the causes of SFR variability identified in this paper are in addition to other known sources of star formation burstiness, such as galaxy mergers (e.g. Barnes & Hernquist 1991; Hopkins et al. 2010; Sparre & Springel 2016).

The mechanisms identified in this paper can induce star formation variability on a broad range of time-scales. The shorter time-scale variability should be apparent in indicators sensitive to star formation on $\lesssim 5$ Myr time-scales, such as H $\alpha$, but would be missed when SFRs are measured using indicators sensitive only to longer time-scales, such as the UV continuum or its dust-processed counterparts in the infrared or millimetre ranges (e.g. Hayward et al. 2014). As shown in previous studies, the intrinsic SFR burstiness in nearby dwarf galaxies owing to the discreteness of star formation units (sometimes phrased in terms of stochastic sampling of star clusters) likely plays an important role in explaining the observed distribution of H $\alpha$-to-UV ratios (Fumagalli et al. 2011; Weisz et al. 2012), in addition to stochastic sampling of the IMF. Our model predicts that galaxies, including more massive ones, are also bursty at high redshift. The intrinsic SFR variability should thus contribute to differences in SFRs inferred with different indicators at high redshift as well. One intriguing possibility is that recurrent star formation bursts in ordinary galaxies could explain substantial populations of extreme emission line galaxies observed at high redshift (e.g. van der Wel et al. 2011; Forrest et al. 2017).

A testable prediction, for which there is already some support, is that the scatter in the SFR–$M_\star$ relation should be larger when measured using H $\alpha$ than with longer time-scale SFR indicators at high redshift and in dwarfs (see Sparre et al. 2017). The transitions from bursty to time steady at late time and in massive galaxies predicted by the model are also found in recent cosmological hydrodynamic simulations that model stellar feedback in a spatially and time re-

solved manner (e.g. Hopkins et al. 2014; Agertz & Kravtsov 2015). It is in fact a generic prediction of such simulations that the SFRs of relatively massive galaxies like the Milky Way only become time steady at $z \lesssim 1$. Lower resolution simulations that do not explicitly resolve star formation into GBC units or do not model the time correct dependence of stellar feedback processes cannot capture the effects identified in this work. This is typically the case even in the latest, state-of-the-art large-volume cosmological simulations, which still have relatively coarse resolution in the ISM (e.g. Vogelsberger et al. 2014; Schaye et al. 2015; Davé et al. 2016). Because of this, current large-volume simulations likely predict star formation histories that are artificially smooth.

In this paper, we assumed that FG13's KS relation model could be applied to all galaxies, but that model was developed for high-$\Sigma_g$ galaxies and assumes that the ISM is primarily supported by SN-driven turbulence. Our burstiness criteria could be quantitatively refined by considering modifications of the KS relation model involving other feedback processes and interstellar chemistry, which can be more important at lower $\Sigma_g$. We nevertheless expect the arguments presented herein to hold approximately outside the range of strict applicability of the FG13 model. This is because a KS relation is *observed* across essentially the entire galaxy population (e.g. Kennicutt 1998; Bigiel et al. 2008; Genzel et al. 2010) and previous theoretical studies have shown that the basic principles of star formation regulation by stellar feedback can be generalized to other feedback mechanisms, including stellar radiation (Thompson et al. 2005; Ostriker et al. 2010; Hayward & Hopkins 2017). Extensions of our feedback-regulated model that include additional processes such as radiation and ISM chemistry should be subject to similar failures of stable regulation as the ones identified for SN feedback in this paper, but may differ in their detailed predictions.

Finally, we note that observations suggest that galactic winds are most prevalent in the high-redshift and dwarf regimes in which our model predicts bursty star formation (e.g. Martin 2005; Steidel et al. 2010; Bordoloi et al. 2014) and that galaxy-scale outflows become significantly weaker at late times in large galaxies (e.g. Heckman 2002; Heckman et al. 2015). Similar trends are found in the wind mass loading factors predicted by recent cosmological simulations that resolve the generation of galaxy-scale outflows from the injection of feedback energy on the scale of star-forming regions (e.g. Muratov et al. 2015; Anglés-Alcázar et al. 2017). It would thus be very interesting to understand better the connection between SFR variability and galactic winds. One intriguing possibility is that powerful galactic winds are associated with galaxies in which SNe are strongly clustered in space and/or time (conditions associated with bursty star formation in our model). SN clustering can promote the formation of hot superbubbles that can more effectively break out of galactic discs than isolated SN remnants and this may be key to driving powerful outflows (e.g. Martizzi, Faucher-Giguère & Quataert 2015; Girichidis et al. 2016; Fielding et al. 2017; Li, Bryan & Ostriker 2017).

# REFERENCES

Agertz O., Kravtsov A. V., 2015, ApJ, 804, 18

Agertz O., Kravtsov A. V., 2016, ApJ, 824, 79

Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017, MNRAS, 470, 4698

Barnes J. E., Hernquist L. E., 1991, ApJ, 370, L65

Bauer A. E. et al., 2013, MNRAS, 434, 209

Benson A. J., 2014, MNRAS, 444, 2599

Bertoldi F., McKee C. F., 1992, ApJ, 395, 140

Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846

Blondin J. M., Wright E. B., Borkowski K. J., Reynolds S. P., 1998, ApJ, 500, 342

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Bordoloi R. et al., 2014, ApJ, 796, 136

Bouché N. et al., 2010, ApJ, 718, 1001

Bryan G. L., Norman M. L., 1998, ApJ, 495, 80

Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, ApJ, 555, 240

Chan T. K., Kereš D., Oñorbe J., Hopkins P. F., Muratov A. L., Faucher-Giguère C.-A., Quataert E., 2015, MNRAS, 454, 2981

Cioffi D. F., McKee C. F., Bertschinger E., 1988, ApJ, 334, 252

Crain R. A. et al., 2015, MNRAS, 450, 1937

Cresci G. et al., 2009, ApJ, 697, 115

Daddi E. et al., 2010, ApJ, 713, 686

Dalcanton J. J., Spergel D. N., Summers F. J., 1997, ApJ, 482, 659

Davé R., Finlator K., Oppenheimer B. D., 2012, MNRAS, 421, 98

Davé R., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265

Dekel A. et al., 2009, Nature, 457, 451

Domínguez A., Siana B., Brooks A. M., Christensen C. R., Bruzual G., Stark D. P., Alavi A., 2015, MNRAS, 451, 839

Dutton A. A., Conroy C., van den Bosch F. C., Prada F., More S., 2010, MNRAS, 407, 2

Elmegreen D. M., Elmegreen B. G., Ravindranath S., Coe D. A., 2007, ApJ, 658, 763

Evans N. J., II, Heiderman A., Vutisalchavakul N., 2014, ApJ, 782, 114

Fall S. M., Chandar R., 2012, ApJ, 752, 96

Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189

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 (FG13)

Faucher-Giguère C.-A., Hopkins P. F., Kereš D., Muratov A. L., Quataert E., Murray N., 2015, MNRAS, 449, 987

Federrath C., 2013, MNRAS, 436, 3167

Feldmann R., Quataert E., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2017, MNRAS, 470, 1050

Fielding D., Quataert E., Martizzi D., Faucher-Giguere C.-A., 2017, MNRAS, 470, L39

Fitts A. et al., 2017, MNRAS, 471, 3547

Forrest B. et al., 2017, ApJ, 838, L12

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

Fumagalli M., da Silva R. L., Krumholz M. R., 2011, ApJ, 741, L26

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., Fall S. M., Hernquist L., Vogelsberger M., Snyder G. F., Rodriguez-Gomez V., Sijacki D., Springel V., 2015, ApJ, 804, L40

Genzel R. et al., 2010, MNRAS, 407, 2091

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

Girichidis P. et al., 2016, MNRAS, 456, 3432

Goldreich P., Lynden-Bell D., 1965, MNRAS, 130, 97

Governato F. et al., 2012, MNRAS, 422, 1231

Grogin N. A. et al., 2011, ApJS, 197, 35

Grudić M. Y., Hopkins P. F., Faucher-Giguère C.-A., Quataert E., Murray N., Kereš D., 2016, preprint (arXiv:1612.05635)

Guo Y. et al., 2016, ApJ, 833, 37

Hayward C. C., Hopkins P. F., 2017, MNRAS, 465, 1682

Hayward C. C. et al., 2014, MNRAS, 445, 1598

Heckman T. M., 2002, in Mulchaey J. S., Stocke J. T., eds, ASP Conf. Ser. Vol. 254, Extragalactic Gas at Low Redshift. Astron. Soc. Pac., San Francisco, p. 292

Heckman T. M., Alexandroff R. M., Borthakur S., Overzier R., Leitherer C., 2015, ApJ, 809, 147

Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663

Heyer M., Gutermuth R., Urquhart J. S., Csengeri T., Wienen M., Leurini S., Menten K., Wyrowski F., 2016, A&A, 588, A29

Hopkins P. F., 2013, MNRAS, 430, 1653

Hopkins P. F., Hernquist L., Cox T. J., Keres D., Wuyts S., 2009, ApJ, 691, 1424

Hopkins P. F., Younger J. D., Hayward C. C., Narayanan D., Hernquist L., 2010, MNRAS, 402, 1693

Hopkins P. F., Quataert E., Murray N., 2011, MNRAS, 417, 950

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581

Kauffmann G., 2014, MNRAS, 441, 2717

Kennicutt R. C., Jr, 1989, ApJ, 344, 685

Kennicutt R. C., Jr, 1998, ApJ, 498, 541

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

Kim C.-G., Ostriker E. C., 2015, ApJ, 815, 67

Kim J.-G., Kim W.-T., Ostriker E. C., 2016, ApJ, 819, 137

Kravtsov A. V., 2013, ApJ, 764, L31

Kriek M. et al., 2015, ApJS, 218, 15

Kroupa P., 2001, MNRAS, 322, 231

Krumholz M. R., 2014, Phys. Rep., 539, 49

Krumholz M. R., McKee C. F., 2005, ApJ, 630, 250

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

Krumholz M. R., McKee C. F., Tumlinson J., 2009, ApJ, 699, 850

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

Lee J. C. et al., 2009, ApJ, 706, 599

Lee E. J., Chang P., Murray N., 2015, ApJ, 800, 49

Lee E. J., Miville-Deschênes M.-A., Murray N. W., 2016, ApJ, 833, 229

Leitherer C. et al., 1999, ApJS, 123, 3

Leroy A. K. et al., 2017, ApJ, 846, 71

Li M., Bryan G. L., Ostriker J. P., 2017, ApJ, 841, 101

Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, ApJ, 772, 119

Liu G., Koda J., Calzetti D., Fukuhara M., Momose R., 2011, ApJ, 735, 63

McGaugh S. S., 2005, ApJ, 632, 859

McKee C. F., Ostriker E. C., 2007, ARA&A, 45, 565

Mac Low M.-M., 1999, ApJ, 524, 169

Madau P., Shen S., Governato F., 2014, ApJ, 789, L17

Martin C. L., 2005, ApJ, 621, 227

Martin C. L., Kennicutt R. C., Jr, 2001, ApJ, 555, 301

Martizzi D., Faucher-Giguère C.-A., Quataert E., 2015, MNRAS, 450, 504

Miville-Deschênes M.-A., Murray N., Lee E. J., 2017, ApJ, 834, 57

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

Mooney T. J., Solomon P. M., 1988, ApJ, 334, L51

Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, MNRAS, 454, 2691

Myers P. C., Dame T. M., Thaddeus P., Cohen R. S., Silverberg R. F., Dwek E., Hauser M. G., 1986, ApJ, 301, 398

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

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

Oñorbe J., Boylan-Kolchin M., Bullock J. S., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., Quataert E., Murray N., 2015, MNRAS, 454, 2092

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

Orr M. et al., 2017, MNRAS, preprint (arXiv:1701.01788)

Ostriker E. C., Shetty R., 2011, ApJ, 731, 41

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

Padoan P., Haugbølle T., Nordlund Å, 2012, ApJ, 759, L27

Planck Collaboration XIII, 2016, A&A, 594, A13

Pontzen A., Governato F., 2012, MNRAS, 421, 3464

Quirk W. J., 1972, ApJ, 176, L9

Raskutti S., Ostriker E. C., Skinner M. A., 2016, ApJ, 829, 130

Reddy N. A., Pettini M., Steidel C. C., Shapley A. E., Erb D. K., Law D. R.,
    2012, ApJ, 754, 25

Rice T. S., Goodman A. A., Bergin E. A., Beaumont C., Dame T. M., 2016,
    ApJ, 822, 52

Rodighiero G. et al., 2014, MNRAS, 443, 19

Rosolowsky E., 2005, PASP, 117, 1403

Sales L. V., Navarro J. F., Schaye J., Dalla Vecchia C., Springel V., Booth
    C. M., 2010, MNRAS, 409, 1541

Schaye J. et al., 2015, MNRAS, 446, 521

Schruba A., Leroy A. K., Walter F., Sandstrom K., Rosolowsky E., 2010,
    ApJ, 722, 1699

Scoville N., 2003, J. Korean Astron. Soc., 36, 167

Scoville N. Z., Good J. C., 1989, ApJ, 339, 149

Seljak U., 2000, MNRAS, 318, 203

Shetty R., Ostriker E. C., 2012, ApJ, 754, 2

Shibuya T., Ouchi M., Harikane Y., 2015, ApJS, 219, 15

Shivaei I. et al., 2015, ApJ, 815, 98

Smit R., Bouwens R. J., Labbé I., Franx M., Wilkins S. M., Oesch P. A.,
    2016, ApJ, 833, 254

Somerville R. S., Popping G., Trager S. C., 2015, MNRAS, 453, 4337

Sparre M., Springel V., 2016, MNRAS, 462, 2418

Sparre M., Hayward C. C., Feldmann R., Faucher-Giguère C.-A., Muratov
    A. L., Kereš D., Hopkins P. F., 2017, MNRAS, 466, 88

Springel V., Hernquist L., 2003, MNRAS, 339, 312

Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N., Bogosavljević
    M., Rudie G. C., Rakic O., 2010, ApJ, 717, 289

Steidel C. C. et al., 2014, ApJ, 795, 165

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

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

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

Teyssier R., Pontzen A., Dubois Y., Read J. I., 2013, MNRAS, 429, 3068

Thompson T. A., Quataert E., Murray N., 2005, ApJ, 630, 167

Toomre A., 1964, ApJ, 139, 1217

Toomre A., Toomre J., 1972, ApJ, 178, 623

Torrey P., Hopkins P. F., Faucher-Giguère C.-A., Vogelsberger M., Quataert
    E., Kereš D., Murray N., 2017, MNRAS, 467, 2301

van der Wel A. et al., 2011, ApJ, 742, 111

Vogelsberger M. et al., 2014, MNRAS, 444, 1518

Vutisalchavakul N., Evans N. J., II, Heyer M., 2016, ApJ, 831, 73

Weisz D. R. et al., 2012, ApJ, 744, 44

Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D.,
    Quataert E., 2016, ApJ, 827, L23

Williams J. P., McKee C. F., 1997, ApJ, 476, 166

Wisnioski E. et al., 2015, ApJ, 799, 209

Wuyts S. et al., 2012, ApJ, 753, 114

# APPENDIX A: GALAXY SCALING RELATIONS

To predict in which galaxies star formation is expected to be bursty, we need to know how galaxy properties evolve with mass and redshift. This appendix summarizes galaxy scaling relations that we use for this purpose.

## A1  Disc radii

For the disc radii, we follow the observational study of Shibuya et al. (2015) and assume that the half-mass radius of a galaxy[7] at any redshift is a constant fraction of the halo virial radius,

[7] In Shibuya et al. (2015), this result is demonstrated empirically for the effective radii of galaxies; here we identify the effective radius with the half-mass radius.
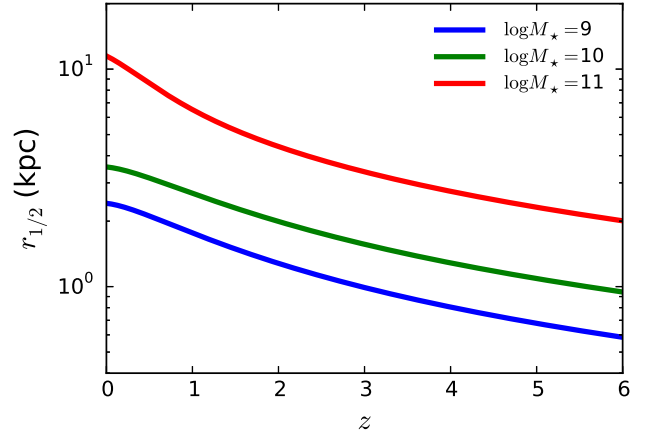


**Figure A1.** Model half-mass radius of galaxy discs $r_{1/2} \approx 0.02 R_{\rm vir}$ as a function of redshift, for different fixed stellar masses. At fixed mass, galaxy discs are smaller at high redshift.

$r_{1/2} \approx 0.02 R_{\rm vir}$. We use the Bryan & Norman (1998) halo definition, in which the mean enclosed overdensity depends on redshift:

$$\Delta_{\rm c} = 18\pi^2 + 82d - 39d^2, \tag{A1}$$

where, in a flat Universe with $\Omega_{\rm m} + \Omega_{\Lambda} = 1$,

$$d \equiv \Omega_{\rm m}^z - 1 \tag{A2}$$

and

$$\Omega_{\rm m}^z = \frac{\Omega_{\rm m}(1+z)^3}{\Omega_{\rm m}(1+z)^3 + \Omega_{\Lambda}}. \tag{A3}$$

The halo virial mass and radius are related by

$$M_{\rm vir} = \frac{4\pi}{3} R_{\rm vir}^3 \Delta_{\rm c}(z) \rho_{\rm c}(z), \tag{A4}$$

where

$$\rho_{\rm c} = \frac{3H^2}{8\pi G} \tag{A5}$$

is the critical density. At high redshift, $\Delta_{\rm c} \to 178$ but $\Delta_{\rm c}(z=0) \approx 97$. For any halo, we define the virial velocity as

$$V_{\rm vir} = \left( \frac{G M_{\rm vir}}{R_{\rm vir}} \right)^{1/2}. \tag{A6}$$

To connect to observations, we also require a conversion between halo mass and the stellar mass of the central galaxy. For this, we use the fitting formula as a function of both mass and redshift from the abundance matching analysis of Moster et al. (2013) for $M_\star$–$M_{\rm vir}$. Fig. A1 shows the predicted evolution of galaxy discs sizes as a function of redshift, for different fixed stellar masses. Galaxies of a given mass are smaller at high redshift because $R_{\rm vir} \propto (1+z)^{-1}$ as $z \to \infty$ (at fixed halo mass).

## A2  Disc circular velocities

To obtain more accurate results, it is useful to use the galaxy rotation velocity rather than the virial velocity of the parent halo in our calculations (although the two are similar to order unity). To do this, we derive a simple model calibrated to observations of the Tully–Fisher relation to capture the mass and redshift evolution of the relation between galaxy rotation velocity, $v_{\rm c}$, and stellar mass. Our model is based on the analysis of the evolution of haloes in cold dark matter cosmologies by Bullock et al. (2001).
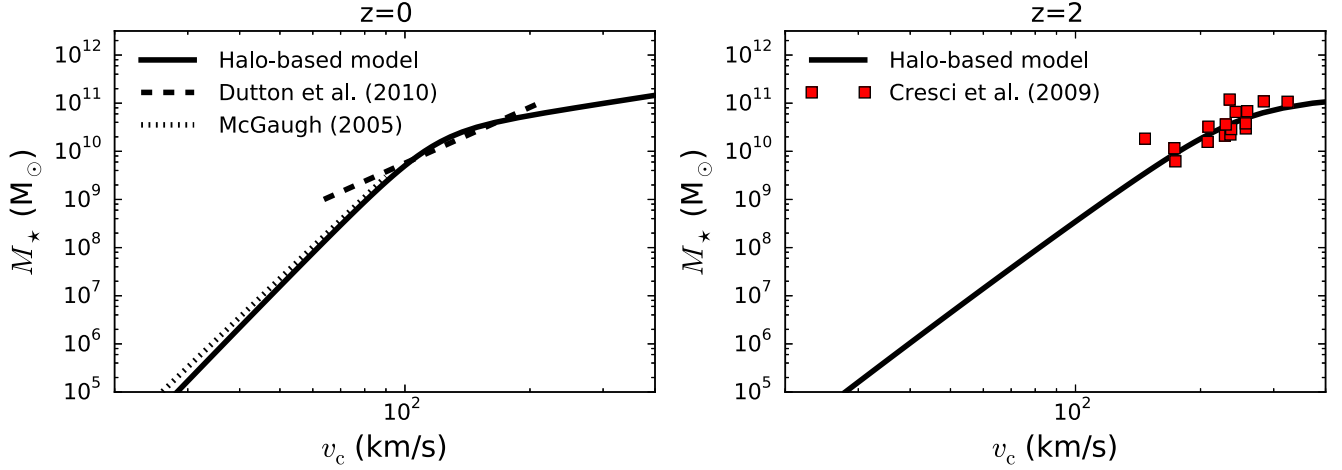
**Figure A2.** Stellar mass versus galaxy circular velocity. The solid curves show the halo-based model used in this paper. At $z = 0$ (left), the model is compared to observational constraints on the Tully–Fisher relations from McGaugh (2005) and Dutton et al. (2010). At $z = 2$, the model is compared to observations of massive galaxies at $1.5 \lesssim z \lesssim 2.5$ from Cresci et al. (2009). The model accurately reproduces the observed $z = 0$ Tully–Fisher relation over a large dynamic range of galaxy masses and correctly predicts its evolution to $z = 2$ (at least at the massive end).

For haloes that follow an NFW profile (Navarro et al. 1997), the maximum circular velocity in the halo, $v_{\mathrm{max}}$, is related to the halo virial velocity through

$$\frac{V_{\mathrm{max}}^2}{V_{\mathrm{vir}}^2} \approx 0.216 \frac{c_{\mathrm{vir}}}{A(c_{\mathrm{vir}})}, \qquad (A7)$$

where $c_{\mathrm{vir}}$ is the halo concentration and

$$A(c_{\mathrm{vir}}) \equiv \ln(1 + c_{\mathrm{vir}}) - \frac{c_{\mathrm{vir}}}{1 + c_{\mathrm{vir}}}. \qquad (A8)$$

For the concentration, we use

$$c_{\mathrm{vir}} \approx 15 \left(\frac{M_{\mathrm{vir}}}{10^{12}\,\mathrm{M}_\odot}\right)^{-0.2} (1 + z)^{-1}. \qquad (A9)$$

At $z = 0$, this matches the mass dependence from Seljak (2000) and at higher redshift this follows the scaling $c_{\mathrm{vir}} \propto (1 + z)^{-1}$ proposed by Bullock et al. (2001). We assume that the galaxy circular velocity at $r_{1/2}$ ($v_{\mathrm{c}}$) is a constant fraction of $v_{\mathrm{max}}$,

$$v_{\mathrm{c}} = \beta V_{\mathrm{max}}, \qquad (A10)$$

and calibrate $\beta$ to observations of the Tully–Fisher relation. As we show below, $\beta = 0.9$ provides a good fit to both $z \approx 0$ and $z \approx 2$ observational constraints, and so we adopt this value.

With these ingredients, we can compute the galaxy circular velocity for any combination of stellar mass and redshift. In Fig. A2, we plot $M_\star$ versus $v_{\mathrm{c}}$ for our model at $z = 0$ and $z = 2$. At $z = 0$, we also show the best-fitting Tully–Fisher relation from Dutton et al. (2010) for massive galaxies and an approximate fit to lower mass galaxies to the observations compiled by McGaugh (2005),

$$M_\star \approx 3 \times 10^9\,\mathrm{M}_\odot \left(\frac{v_{\mathrm{c}}}{90\,\mathrm{km\,s^{-1}}}\right)^{8.3} \ (v_{\mathrm{c}} < 90\,\mathrm{km\,s^{-1}}). \qquad (A11)$$

This fit to the McGaugh (2005) data is only approximate because many of the low-mass galaxies in the sample only have an upper limit on their stellar mass. At $z = 2$, we show stellar masses and circular velocities measured using integral field observations of massive galaxies from the Spectroscopic Imaging survey in the Near-infrared with SINFONI (SINS) survey by Cresci et al. (2009). As the figure shows, our simple model for the $M_\star$–$v_{\mathrm{c}}$ relation provides a good fit to both the $z = 0$ and $z = 2$ observations.
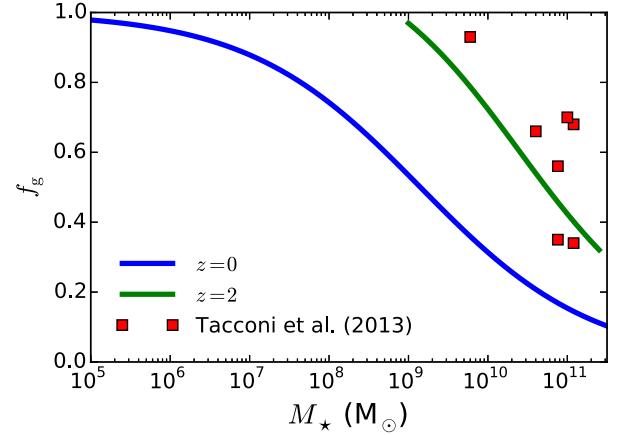


**Figure A3.** Gas mass fraction as a function of galaxy stellar mass used to evaluate our model predictions at $z = 0$ and $z = 2$ (solid curves). The red squares show gas fractions derived from observations of disc galaxies in Tacconi et al. (2013) in the redshift interval $1.5 < z < 2.5$.

### A3 Gas fractions

For the gas fractions, we start with the fitting formulae from Hopkins et al. (2009):

$$f_{\mathrm{g},0} = \left[1 + \left(\frac{M_\star}{10^{9.15}\,\mathrm{M}_\odot}\right)^{0.4}\right]^{-1},$$

$$f_{\mathrm{g}}(M_\star, z) = f_{\mathrm{g},0}\left[1 - \tau(z)(1 - f_0^{3/2})\right]^{-2/3}, \qquad (A12)$$

where $\tau(z)$ is the fractional lookback time to redshift $z$ (such that $\tau(z = 0) = 0$ and $\tau(z \to \infty) = 1$). At $z = 2$, we find that multiplying the $f_{\mathrm{g}}$ implied by these expressions by 1.2 provides a better fit to the more recent gas fraction measurements compiled in Tacconi et al. (2013). We therefore multiply $f_{\mathrm{g}}$ by this factor for our $z = 2$ predictions (but not for $z = 0$). At this redshift, we only consider relatively massive galaxies for which there are good gas fraction measurements ($M_\star \sim 10^9$–$10^{11}\,\mathrm{M}_\odot$). Fig. A3 shows the gas fractions used to evaluate our model predictions as a function of stellar mass, at $z = 0$ and $z = 2$. In this figure, the data points

from Tacconi et al. (2013) are restricted to galaxies in the redshift interval $1.5 < z < 2.5$ and that are classified as discs (we exclude galaxies classified as mergers or dispersion dominated). Two trends are noteworthy: gas fractions increase with decreasing stellar mass and with increasing redshift. These gas fraction trends drive the burstiness due to star formation discreteness in dwarf galaxies and in high-redshift galaxies predicted in Section 2.4.

## APPENDIX B: SFR VARIANCE VERSUS $N_{\mathrm{U}}$

In Section 2.4, we related the burstiness of star formation to the expected number of most massive (Toomre-scale) GBCs, $N_{\mathrm{U}}$, in a galaxy. In this section, we show that the fractional variance of the instantaneous galaxy SFR is primarily determined by $N_{\mathrm{U}}$ under general assumptions consistent with the star formation model on which this paper is based (see Section 2.1). Specifically, we derive an explicit expression for

$$\frac{\sigma_{\mathrm{SFR}}}{\langle \mathrm{SFR} \rangle} \tag{B1}$$

in terms of $N_{\mathrm{U}}$, where $\langle \mathrm{SFR} \rangle$ is the mean SFR in the galaxy (averaged over a long time-scale $\gg 100$ Myr) and $\sigma_{\mathrm{SFR}}$ is the standard deviation of the instantaneous SFR.

We start with a general function describing the time evolution of the SFR within an individual GBC as a function of mass $m$ and time $t$ since GBC formation, $\mathrm{SFR}_i(m, t)$. We will later evaluate our final result for different variations of $\mathrm{SFR}_i$. The GBC mass function,

$$\frac{\mathrm{d}N}{\mathrm{d}m} = \begin{cases} A m^{-\alpha} & M_{\mathrm{L}} \leq m \leq M_{\mathrm{U}}, \\ 0 & \text{otherwise}, \end{cases} \tag{B2}$$

describes the distribution of GBCs present in the galaxy at any given time. At any time, each such GMC will be 'caught' at a random time $0 \leq t < t_{\mathrm{GBC}}$ during its evolution, where $t_{\mathrm{GBC}}$ is the GBC lifetime. The total galactic SFR is, by definition, the sum over the star formation rates of GBCs in the galaxy:

$$\mathrm{SFR} = \Sigma_{i=1}^{N_{\mathrm{tot}}} \mathrm{SFR}_i. \tag{B3}$$

We treat $\mathrm{SFR}_i$ as a function of two random variables, $m$ and $t$. For each $i$, $m$ and $t$ are drawn from a joint distribution

$$f_{m,t}(m, t) = f_m(m) f_t(t), \tag{B4}$$

which factors into marginal probability density functions for $m$ and $t$ since these are independent random variables.

The probability density function for $m$ is proportional to the GMC mass function, but normalized such that $\int \mathrm{d}m f_m(m) = 1$:

$$f_m(m) = \begin{cases} \tilde{A} m^{-\alpha} & M_{\mathrm{L}} < m \leq M_{\mathrm{U}}, \\ 0 & \text{otherwise}, \end{cases} \tag{B5}$$

where

$$\tilde{A} = \frac{(\alpha - 1)(M_{\mathrm{L}} M_{\mathrm{U}})^{\alpha - 1}}{M_{\mathrm{U}}^{\alpha - 1} - M_{\mathrm{L}}^{\alpha - 1}} \tag{B6}$$

$$= \frac{\alpha - 1}{(\beta^{1-\alpha} - 1)} M_{\mathrm{U}}^{\alpha - 1}. \tag{B7}$$

In the last expression, we defined $\beta \equiv M_{\mathrm{L}}/M_{\mathrm{U}}$, which will simplify later results. The probability density function for $t$ is simply uniform over the GBC lifetime:

$$f_t(t) = \begin{cases} 1/t_{\mathrm{GBC}} & 0 \leq t \leq t_{\mathrm{GBC}}, \\ 0 & \text{otherwise}. \end{cases} \tag{B8}$$

We are now ready to evaluate $\langle \mathrm{SFR} \rangle$ and $\sigma_{\mathrm{SFR}}$:

$$\langle \mathrm{SFR} \rangle = N_{\mathrm{tot}} \langle \mathrm{SFR}_i \rangle, \tag{B9}$$

where

$$N_{\mathrm{tot}} = \int_{M_{\mathrm{L}}}^{M_{\mathrm{U}}} \mathrm{d}m \frac{\mathrm{d}N}{\mathrm{d}m} \tag{B10}$$

$$= A \frac{(\beta^{1-\alpha} - 1)}{\alpha - 1} M_{\mathrm{U}}^{1-\alpha}. \tag{B11}$$

For the variance, we start with the general result

$$\sigma_{\mathrm{SFR}}^2 = \langle \mathrm{SFR}^2 \rangle - \langle \mathrm{SFR} \rangle^2, \tag{B12}$$

and proceed to evaluate $\langle \mathrm{SFR}^2 \rangle$. We have

$$\mathrm{SFR}^2 = \Sigma_i^{N_{\mathrm{tot}}} \mathrm{SFR}_i^2 + \Sigma_{i \neq j} \mathrm{SFR}_i \mathrm{SFR}_j, \tag{B13}$$

and therefore, since $\mathrm{SFR}_i$ and $\mathrm{SFR}_j$ are independent but identically distributed for $i \neq j$,

$$\langle \mathrm{SFR}^2 \rangle = N_{\mathrm{tot}} \langle \mathrm{SFR}_i^2 \rangle + N_{\mathrm{tot}}(N_{\mathrm{tot}} - 1) \langle \mathrm{SFR}_i \rangle^2. \tag{B14}$$

As before, the problem reduces to evaluating moments of $\mathrm{SFR}_i$, the SFR evolution of an individual GBC.

Using equations (B9), (B12) and (B14), we find that

$$\frac{\sigma_{\mathrm{SFR}}^2}{\langle \mathrm{SFR}^2 \rangle} = \frac{1}{N_{\mathrm{tot}}} \frac{\langle \mathrm{SFR}_i^2 \rangle}{\langle \mathrm{SFR}_i \rangle^2} + \frac{N_{\mathrm{tot}}(N_{\mathrm{tot}} - 1)}{N_{\mathrm{tot}}^2} - 1 \tag{B15}$$

$$\approx \frac{1}{N_{\mathrm{tot}}} \frac{\langle \mathrm{SFR}_i^2 \rangle}{\langle \mathrm{SFR}_i \rangle^2}. \tag{B16}$$

The last approximation holds so long as $N_{\mathrm{tot}} \gg 1$. To connect with the results of Section 2.4, we must express our results in terms of $N_{\mathrm{U}}$ rather than $N_{\mathrm{tot}}$. By definition,

$$N_{\mathrm{U}} = \frac{\int_{M_{\mathrm{L}}}^{M_{\mathrm{U}}} \mathrm{d}m(\mathrm{d}N/\mathrm{d}m)m}{M_{\mathrm{U}}} \tag{B17}$$

$$= A \frac{(1 - \beta^{2-\alpha})}{2 - \alpha} M_{\mathrm{U}}^{1-\alpha}, \tag{B18}$$

and therefore, using equation (B11),

$$N_{\mathrm{tot}} = \frac{(2 - \alpha)}{(\alpha - 1)} \frac{(\beta^{1-\alpha} - 1)}{(1 - \beta^{2-\alpha})} N_{\mathrm{U}}. \tag{B19}$$

Equation (B16) is convenient because it allows us to calculate how the fractional variance of the galactic SFR depends on parameters of the GBC mass function and on the function $\mathrm{SFR}_i$ describing the time evolution of individual GBCs.

Consider a specific, but rather general parametrization of GBC evolution:

$$\mathrm{SFR}_i(m, t) = \begin{cases} 0 & -(1 - \gamma)t_{\mathrm{GBC}} \leq t < 0, \\ B m t^\delta & 0 \leq t \leq \gamma t_{\mathrm{GBC}}. \end{cases} \tag{B20}$$

In this expression, $t_{\mathrm{GBC}}$ is the total GBC lifetime. The parameter $\gamma$ defines the fraction of $t_{\mathrm{GBC}}$ during which the GBC forms stars at a significant rate. This parameter allows us to model a possible early phase during which, for example, strong turbulence generated during the assembly of the GBC prevents star formation from occurring by maintaining a high virial parameter (e.g. Padoan et al. 2012). The proportionality to $m$ captures the fact that, on average, we expect the SFR of an individual GBC to scale with its mass. This is true, for example, if $\mathrm{SFR}_i \sim m/t_{\mathrm{GBC}}$. Finally, the power-law index $\delta$ characterizes the SFR evolution as the cloud collapses. Recent

simulations suggest that to a good approximation the SFR within GBCs increases linearly with time ($\delta \approx 1$) after the first stars form (Lee et al. 2015; Grudić et al. 2016; Raskutti et al. 2016). Real GBC are likely disrupted more gradually than the abrupt cut-off at the end implied by the simple model in equation (B20). For simplicity, we do not model this explicitly here but note that the sensitivity of our main result below on different $\mathrm{SFR}_i$ parametrizations can be gauged from the dependence on the parameters $\gamma$ and $\delta$. The constant pre-factor $B$ defines the absolute normalization of the SFR; its value has no impact on the following result as it cancels out exactly in equation (B12).

For the $\mathrm{SFR}_i$ in equation (B20), we find

$$\frac{\sigma_{\mathrm{SFR}}^2}{\langle \mathrm{SFR} \rangle^2} = \frac{1}{\gamma N_{\mathrm{U}}} \frac{(1+\delta)^2}{(1+2\delta)} \left( \frac{2-\alpha}{3-\alpha} \right) \left( \frac{1-\beta^{3-\alpha}}{1-\beta^{2-\alpha}} \right). \qquad (B21)$$

This expression involves several parameters but its numerical value is not very sensitive to most. For example,

$$\frac{\sigma_{\mathrm{SFR}}}{\langle \mathrm{SFR} \rangle} = \frac{S}{\sqrt{\gamma N_{\mathrm{U}}/0.5}}, \qquad (B22)$$

where $S \approx 1.1, 1.1, 1.0, 0.9, 0.8$ and $1.2$ for $(\alpha, \beta, \delta) = (1.5, 0.1, 1)$, $(1.8, 0.1, 1)$, $(2, 0.1, 1)$, $(1.8, 0.01, 1)$, $(1.8, 0.001, 1)$ and $(1.8, 0.1, 2)$.

This result has a simple interpretation. In our model, the total SFR is dominated by Toomre-scale GBCs. Thus, at any given time the galactic SFR is roughly a sum over $\sim \gamma N_{\mathrm{U}}$ Toomre-scale GBCs, each caught at a random time during its evolution after star formation has begun. Since in individual GBCs the SFR is assumed to increase with time (before disruption by feedback), the total SFR can be approximated as a sum over some GBCs that are effectively in a 'low SFR' state and others that are in a 'high SFR' state. The number of 'high SFR' GBCs varies with time due to stochastic effects, with a Poisson fractional standard deviation $\propto 1/\sqrt{\gamma N_{\mathrm{U}}}$. This quantitatively illustrates our claim in the main text that the variability of the galactic SFR increases with decreasing $N_{\mathrm{U}}$.

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