Title

Numerical Experiments in Galactic Disks: Gravitational Instability, Stochastic Accretion, and Galactic Winds

Author

Forbes, John C.

Publication Date

2016

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

NUMERICAL EXPERIMENTS IN GALACTIC DISKS: GRAVITATIONAL INSTABILITY, STOCHASTIC ACCRETION, AND GALACTIC WINDS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

ASTRONOMY & ASTROPHYSICS

by

John C. Forbes

June 2016

The Dissertation of John C. Forbes
is approved:

Professor Mark R. Krumholz, Chair

Professor Douglas N. C. Lin

Professor Jason X. Prochaska

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Numerical Experiments in Galactic Disks:

Gravitational Instability, Stochastic Accretion, and Galactic Winds

by

John C. Forbes

Using 0D, 1D, and 3D models of galaxies, I explore different problems in galaxy evolution most suited to each technique. In the simplest case, a galaxy is described by a few numbers integrated via coupled ordinary differential equations. By allowing the galaxies to respond to a stochastic accretion rate, I show a natural way of generating the finite scatter observed in several galaxy scaling relations: the correlation between a galaxy's stellar mass and its star formation rate or metallicity. By comparing this simple model to observations, we constrain the process by which gas accretes onto galaxies, which must occur, but is essentially impossible to observe directly. Adding an additional dimension to the models, we explore the structure of galactic disks as a function of radius. We find that turbulence driven by gravitational instability in the disks and the resulting migration of gas can explain a wide variety of phenomena, including the age-velocity dispersion correlation of stars in the solar neighborhood, the central quenching star formation in disk galaxies, rings of star formation, and the observed radial profile of gas column densities. Finally, we run a set of fully three-dimensional galaxy simulations to try to understand what physics is responsible for basic properties of galaxies, including the rate at which they form stars, and the rate at which they eject mass in large-scale winds. We find that supernovae are capable of driving moderate metal-enhanced winds, but surprisingly they have very little effect on the star formation rates of dwarf galaxies. Instead, ordinary photoelectric heating dominates the star formation law in low-mass galaxies.

Acknowledgments

This thesis would not have been possible with the support of a great many people. Mark Krumholz was my primary adviser throughout my time as a graduate student. He is known worldwide as an excellent scientist, and that reputation is well-deserved. He also possesses an excellent instinct for management and keeping projects moving in the right direction which is much less common but incredibly valuable for his students. Doug Lin, my secondary adviser, has a remarkable ability to enthusiastically apply the same physics to a shocking variety of astrophysical scenarios and scales, and it has been a lot of fun working with him. The final member of my committee, J. Xavier Prochaska, has provided excellent advice at crucial moments in my graduate career, and served as a connection to what the universe is really doing outside of theoretical models.

Beyond my committee, I would also like to thank my significant other, Erin Zampaglione, whose support has been beyond measure. I would also like to thank my collaborators, Nathan Goldbaum, Avishai Dekel, Andreas Burkert, Antoine Petit, and Yicheng Guo for working with me on much of the science presented in this thesis. The Krumholz research group, which has over time included Anna Rosen, Eric Gentry, Chalence Safranek-Schrader, Evan Carter, Ji-hoon Kim, Chao-Chin Yang, Yi Feng, Yusuke Fujimoto, Jonathan Craig, and Donald Dukes (plus Nathan and Antoine) has always been helpful and supportive in diagnosing problems and inspiring in their own accomplishments. Much of the work in this thesis had a large computational component, much of which was executed on the Hyades machine, operated, assembled, maintained, and supported by Shawfeng Dong; thank you for your help over the years Shaw.

Before even getting to graduate school, a number of scientists - Jason Melbourne, Chris Hirata, Oliver Gutsche, Maury Goodman, and Don Lincoln - took a chance by working with me and giving me the opportunity to learn how to do research. My parents, Tom and Dorothy Forbes, also did a huge amount to get me where I am. I am incredibly grateful and fortunate. I would also like to thank the administrators in the department, Maria Sliwinski, Jenna Scarpelli, and Patti Schell, each of whom has been unfailingly helpful and kind, and the department chairs over my time here, Sandy Faber, Greg Laughlin, and Enrico Ramirez-Ruiz, each of whom had an outsized impact on my way of thinking. Finally, I would like to thank each of the many graduate students in the astronomy department with whom I have overlapped, including of course my cohort Hassen Yesuf, Anna Rosen, Morgan MacLeod, Elizabeth Lovegrove, Caroline Morley, Jenn Burt, and Maria Uribe. Everyone here does a great deal to make this department a friendly and supportive place through all the ups and downs of graduate school.

Chapter 2 is based on Forbes, J.C., Krumholz, M.R., Burkert, A. and Dekel, A. (2014) MNRAS, 443 84 . Financial support for each of these authors was provided via the following sources. JCF was supported by the National Science Foundation Graduate Research Fellowship under Grant Nos. DGE0809125 and DGE1339067. MRK acknowledges support from the NSF through CAREER grant AST-0955300, and by NASA through ATFP grant NNX13AB84G. AB acknowledges support from the Cluster of Excellence “Origin and Structure of the Universe”. AD was supported by ISF grant 24/12, by GIF grant G-1052-104.7/2009, by a DIP grant, and by the I-CORE Program

of the PBC and The ISF grant 1829/12. The monte carlo simulations were carried out on the UCSC supercomputer Hyades, which is supported by NSF grant AST-1229745. This work also benefited from helpful conversations with Erica Nelson, Pieter van Dokkum, Peter Behroozi, and Eyal Neistein.

Chapter 3 is based on Forbes, J., Krumholz, M., and Burkert A. (2012) ApJ 754 48. For this work, we thank M. Cacciato and A. Dekel for stimulating conversations and the anonymous referee for a thorough and helpful report. J.F. was supported by a Graduate Research Fellowship from the National Science Foundation. M.R.K. acknowledges support from the Alfred P. Sloan Foundation, from the NSF through the grant CAREER-0955300, and NASA through the Astrophysics Theory and the Fundamental Physics Grant NNX09AK31G and a Chandra Space Telescope Grant. A.B. thanks his colleagues at the astronomy department at UCSC for their hospitality and support.

Chapter 4 is based on Forbes, J.C., Krumholz, M.R., Burkert, A. and Dekel, A. (2014) MNRAS, 438 1552. For this work, the authors would like to thank Eyal Neistein for helpful discussions. JCF was supported by the National Science Foundation Graduate Research Fellowship under Grant Nos DGE0809125 and DGE1339067. MRK acknowledges support from the Alfred P. Sloan Foundation, from the NSF through CAREER grant AST-0955300, and by NASA through a Chandra Space Telescope Grant and through ATFP grant NNX13AB84G. AB thanks the UCSC astronomy and astrophysics department for its hospitality during his summer visits. AD and AB acknowledge support from GIF grant G-1052-104.7/2009. AD also acknowledges support by

ISF grant 24/12 and by NSF grant AST-1010033.

Chapter 5 is based on a paper submitted to Nature in October 2015 and recently accepted, co-authored by myself, Mark Krumholz, Nathan Goldbaum, and Avishai Dekel. JCF and MRK acknowledge support from Hubble Archival Research grant HST-AR-13909. This work was also supported by NSF grants AST-09553300 and AST-1405962, NASA ATP grant NNX13AB84G, and NASA TCAN grant NNX14AB52G (JCF, MRK, and NJG) and by Australian Research Council grant DP160100695. AD acknowledges support from the grants ISF 124/12, I-CORE Program of the PBC/ISF 1829/12, BSF 2014-273, and NSF AST-1405962. Simulations were carried out on NASA Pleiades and the UCSC supercomputer Hyades, supported by NSF grant AST-1229745.

The first author of these publications wrote the text, performed and analyzed the simulations, and produced the figures. The other authors contributed to supervision of the projects, editing the manuscripts, and for NJG, aided in the development of new functionality in the publicly-available enzo hydrodynamics code. Each coauthor has given explicit permission for their co-authored work to be included in this dissertation.

Chapter 1

Introduction

All models are wrong but some are useful. - George Box

Galaxies can be viewed from many angles by those with different inclinations. To the cosmologist they are mere tracers of the underlying, dominant dark matter structure of the universe. To those concerned with supernovae, stars, and black holes, galaxies are just "the host" in which the interesting phenomena occur. However, the dynamics, history, and governing physics of galaxies themselves plays a critical role for both sets of people. Individual galaxies are stunning in their beauty, complexity, and diversity. Each has a unique history extending through most of the age of the universe, and each experiences a huge variety of physical processes at a variety of scales. As in many areas of astrophysics, observers face a difficult challenge in understanding individual galaxies because they can only be viewed at a single instant in their history. To bring galaxies to life as dynamical objects requires modelling and simulation.

Models of galaxies can range from nearly trivially simple - a handful of coupled differential equations - to hydrodynamic simulations run with codes developed over decades. Each is extremely powerful. Simple models build intuition and cost relatively little in terms of computation, both in terms of CPU and human time. Simulations require much more of both, but have the power to predict unforeseen phenomena. This thesis is an eclectic collection of work using models of both types, each developed to answer a different question about the evolution of galaxies.

A common theme is the "equilibrium model," which posits that the timescale on which galaxies equilibrate between inflows, outlfows, and star formation, is short enough that the gas mass and metal content of galaxies remain roughly constant in time. These quantities evolve as a galaxy accretes new dark matter mass, or as the universe expands thereby reducing the cosmological gas supply. However, the instantaneous gas content, star formation rate, and metallicity of a galaxy are set only by its current mass and accretion rate. In other words, the galaxy's gas forgets its history. The projects detailed in this thesis perturb the equilibrium model, test its assumptions, and try to find where it breaks down.

1.1 A brief overview of galaxy evolution theory

Our modern theoretical understanding of the evolution of galaxies is built on a skeleton of dark matter. It has been known observationally, though not without a great deal of skepticism, that matter besides stars and stellar remnants is necessary to explain the kinematics of galaxies, both in terms of their dynamics within clusters (Zwicky, 1933), and in terms of their rotational velocities at large galactocentric radius (e.g. Rubin et al., 1980). With compelling observational evidence for the gravitational dominance of dark matter, a theoretical picture emerged wherein dark matter structures collapsed gravitationally into bound objects, while gas dissipated its energy via radiative cooling and condensed in their centers (Binney, 1977; Rees & Ostriker, 1977; Silk, 1977; White & Rees, 1978; Fall & Efstathiou, 1980; Blumenthal et al., 1984; White & Frenk, 1991).

Following the development of analytic models for the formation of galaxies, ever-increasing computational power allowed the development of a wide array of numerical methods. Semi-analytic models (White & Frenk, 1991; Somerville & Primack, 1999) integrate the evolution equations for large collections of individual galaxies to attempt to reproduce the statistical properties of observed galaxies. Dark matter-only cosmological N-body simulations (Bullock et al., 2001; Springel et al., 2005) allow the detailed study of the underlying dark matter, and are generally used as inputs in semi-analytic models. N-body plus hydrodynamic cosmological simulations (Katz, 1992; Davé et al., 1997), and idealized hydrodynamic simulations of individual galaxies (Mac Low & Ferrara, 1999) attempt to capture consequences of baryonic physics, i.e. gas cooling, star formation, and various forms of feedback.

An unavoidable conclusion in a universe where galaxies form in halos of cold dark matter is that individual galaxies will merge not infrequently over the course of their lifetimes. Much effort was devoted to understanding merger trees, how galaxies behaved during mergers (Hopkins et al., 2008; Cox et al., 2006) and developing empirical measures of the merger rate at a variety of redshifts (Lotz et al., 2008). Despite the interest in mergers, data from large surveys at a variety of redshifts showed that star-forming galaxies lie on narrow scaling relations with only a small fraction of the cosmic star formation rate occurring in starbursts (Rodighiero et al., 2011).

This fact likely contributes to the success of a variety of simple models wherein the basic quantities of a galaxy are integrated forward in time using an average accretion history of dark and baryonic matter (Bouché et al., 2010; Cacciato et al., 2012; Dekel et al., 2013). Taking this simplicity one step further, Davé et al. (2012) proposed that the properties of a galaxy can be understood without integrating equations over cosmological times, but rather that they are set instantaneously by an equilibrium between inflows, outflows, and star formation. In essence, the evolutionary equations describing a galaxy may be replaced with algebraic equations by setting the time derivatives to zero. The fact that such a model could have any success is initially surprising, but may be easily understood.

1.2 The equilibrium model

The gas content of a galaxy is set by the continuity equation, which is a simple accounting of the source and sinks of gas mass,

$$\dot{M}_g = \dot{M}_{\rm acc} - f_R \dot{M}_{\rm SFR} - \dot{M}_{\rm out} \tag{1.1}$$

The rate of change in gas mass $\dot{M}_g$ is the combined effect of the gas accretion rate $\dot{M}_{\rm acc}$, the outflow rate $\dot{M}_{\rm out}$, and gas which is locked in long-lived stellar remnants, $f_R \dot{M}_{\rm SFR}$, where $\dot{M}_{\rm SFR}$ is the star formation rate, and f_R is the remnant fraction. This equation may be cast in a different form by defining the mass loading factor $\eta = \dot{M}_{\rm out}/\dot{M}_{\rm SFR}$, and the gas depletion time $t_{\rm dep} = M_g/\dot{M}_{\rm SFR}$,

$$\dot{M}_g = \dot{M}_{\rm acc} - (f_R + \eta)M_g/t_{\rm dep} \tag{1.2}$$

Note that these two equations are identical, and no assumptions have been made yet.

If we now make the assumption that $\dot{M}_{\rm acc}$, f_R, η, and $t_{\rm dep}$ are all constant, or at least varying more slowly than any other timescale in the system, then the equation becomes conceptually simple, and in fact has a simple analytic solution, namely

$$M_g(t - t_0) = M_g(t_0)e^{-(t-t_0)/t_{\rm loss}} + \dot{M}_{\rm acc}t_{\rm loss}\left(1 - e^{-(t-t_0)/t_{\rm loss}}\right) \tag{1.3}$$

where t_0 is some arbitrary time at which M_g is known, and $t_{\rm loss} = t_{\rm dep}/(\eta + f_R)$. In this equation M_g forgets its initial value and approaches its equilibrium value, namely $\dot{M}_{\rm acc}t_{\rm loss}$ exponentially on a timescale $t_{\rm loss}$. This means that if $t_{\rm loss}$ is short in comparison to cosmological timescales, the galaxy's internal state will rapidly reach an equilbrium where $\dot{M}_g \approx 0$.

The fact that galaxies follow narrow scaling relations is quite congruous with the equilbrium model. In the equilibrium model essentially all of a galaxy's properties are set only by its current dark matter mass, which in turn determines the galaxy's mass accretion rate and values of the parameters in the continuity equation, namely η and $t_{\rm dep}$. If a galaxy's properties are simply set by its mass, then narrow scaling relations are a natural result. The accretion rate's dependence on mass is quite well-understood as a generic consequence of structure formation in a universe dominated by cold dark matter – the specific dark matter accretion rate is nearly independent of halo mass. The scaling of the depletion time and mass loading factor with mass are much less well-understood, but might well be expected to obey physics that in some way depends on a galaxy's mass.

The equilibrium model is a fascinating framework for understanding the fundamental nature of galaxies, and why they obey the scaling relations that they do. The eclectic results included in this thesis may be viewed as addressing different open questions about the equilibrium model and its implications.

- First, we note that although galaxies follow narrow scaling relations, these scaling relations have a finite width. In the next chapter, we detail a version of the equilibrium model that relaxes the assumption that $\dot{M}_{\rm acc}$ is slowly changing with time, and instead allow it to vary randomly with a lognormal distribution and some timescale on which it remains coherent. In other words, we 'kick' the equilibrium model to see how it responds. With this minimalistic model, we can produce the finite scatter observed in several scaling relations followed by star-forming galaxies, and place surprising constraints on several parameters of the model by comparing to observations.
- In the following two chapters, we construct a much more sophisticated model of star-forming galaxies, essentially modelling them as accretion disks subject to large-scale gravitational instability and hence torques which transfer mass and angular momentum. We examine the resulting radial structure of star-forming disks (especially Chapter 4), and its consequences for the stellar populations formed in them (Chapter 3). We find

that just as global equilibrium is important in setting the global properties of a galaxy, a local equilibrium among terms in the continuity equation for the gas column density determines the structure of star-forming disks.

- Finally, in Part II we move from simple models to full-fledged hydrodynamical simulations of dwarf galaxies. The equilibrium model applies when $t_{\rm loss}$ is extremely short. In dwarf galaxies where stars form extremely slowly, $t_{\rm dep}$ often exceeds the age of the universe, and it is not clear *a priori* whether the equilibrium model applies. In many cosmological simulations that have had their feedback parameters adjusted to reproduce the galaxy scaling relations, the mass loading factors are set so large for dwarfs that $t_{\rm loss} = t_{\rm dep}/(f_R + \eta)$ becomes short, i.e. the long depletion times are offset by the extremely strong winds. We set out to measure η directly in high-resolution simulations to see whether this scenario is borne out by the physics of supernova feedback (Chapter 6), and along the way we gain insight into the cause of the observed extreme depletion times in dwarfs (Chapter 5).

Part I

Simple Models

Chapter 2

The Consequences of Stochastic Accretion

2.1 Introduction

Large galaxy surveys in the past decade have taught us that star-forming galaxies fall on a main sequence, a tight correlation between their stellar mass and star formation rates. (Daddi et al., 2007; Noeske et al., 2007; Elbaz et al., 2007). This correlation contains remarkably little scatter - the star formation rate varies by about ±0.34 dex at fixed stellar mass (Whitaker et al., 2012; Guo et al., 2013).

The existence of the star-forming main sequence (MS) and its small scatter have substantially affected our understanding of galaxy evolution. The cosmological paradigm of the past few decades, ΛCDM, predicts that dark matter halos, and hence the galaxies occupying them, form hierarchically – large halos are built from mergers of smaller halos. The expectation was therefore that mergers between galaxies would be a major driver of their evolution over cosmological times, frequently triggering starbursts. The small scatter in the main sequence at multiple redshifts has shown, however, that most galaxies are not in fact experiencing any dramatic effects of major mergers (Rodighiero et al., 2011), and most stars form in 'normal' galaxies lying along this relation.

In addition to the main sequence, galaxies follow other scaling relations. Practically this means that at a given point in cosmic history many properties of galaxies are set by a single parameter associated with the galaxy's mass. This fact has led to the development of a series of models dubbed "equilibrium", "regulator", or "bathtub" models (Dekel et al., 2009a; Bouché et al., 2010; Davé et al., 2012; Cacciato et al., 2012; Genel et al., 2012; Lilly et al., 2013; Feldmann, 2013), in which the properties of a galaxy are self-regulated near a stable equilibrium between accretion of gas, star formation, and outflows. The fundamental ingredients are a star formation rate that increases with increasing gas mass in the galaxy, and a timescale on which the gas reservoir returns to its equilibrium value much shorter than the timescale on which bulk properties of the galaxy (mass, accretion rate, outflow efficiency, star formation timescale, etc.) vary.

Equilibrium models have enjoyed success in heuristically reproducing the average trends in galaxy evolution. However, such models are inherently incapable of quantifying higher-order effects, including scatter in individual scaling relations and various fundamental metallicity relations (FMRs), found by numerous groups (Mannucci et al., 2010, 2011; Lara-López et al., 2013; Bothwell et al., 2013; Stott et al., 2013), typically parameterized as a quadratic function $Z(M_*,\mathrm{SFR})$.

Many theoretical studies are focussed on reproducing first-order relations, which is sufficiently difficult in and of itself that second-order relations are often neglected (however, for an exception see Dutton et al., 2010). In principle, however, to fully understand galaxy evolution we should be able to understand not only average or median galaxy properties, but their full distribution. A significant advantage of the equilibrium models is that fitting them to any first order relation is trivial, so extending them to understand the higher-order relations is easier than with any other method.

In this work we take one step beyond equilibrium models, allowing the mass accretion rate to vary with a fixed log-normal scatter. Such a scatter is expected based on N-body simulations (e.g. Neistein & Dekel, 2008) and hydrodynamic simulations (e.g. Dekel et al., 2013). Depending on the timescale on which the accretion rate varies, a population of these galaxies may never be in equilibrium. That is, their masses and metallicities may change substantially. However they may still reach a statistical equilibrium, in which the full joint distribution of all galaxy properties becomes time-invariant.

In section 2.2 we introduce the basic formulation of this model and explore the implications for the origin of the width of the star-forming main sequence. We add metallicity to the model in section 2.3, allowing us to examine the FMR and the scatter in the mass-metallicity relation. The model we construct in these sections is independent of the first-order effects considered by most equilibrium models, which allows us to avoid uncertainty in the numerical values of numerous important parameters in galaxy evolution (e.g. the mass loading factor). In section 2.4 we re-dimensionalize our model, taking a guess at the correct scalings and normalizations to use, and demonstrate the ability of this model to understand unknown physics based solely on the scatter in galaxy scaling relations. We discuss the limitations of our model, quantitative constraints it can place on the accretion process, and alternative explanations for the intrinsic width of these relations in section 2.5, and conclude in section 2.6.

2.2 A very simple model

To examine the origin of and scatter within the MS and FMR, we have constructed a minimal model which contains enough physics to produce such features. We do not aim to fully reproduce galaxy properties, tune parameters to match observations, or build upon dark matter merger trees. Instead we aim for simplicity and intuition. In our model we describe the state of a galaxy by two numbers, its cold gas mass M_g and its metalicity $Z = M_Z/M_g$, where M_Z is the mass in metals. The gas mass evolves according to

$$\frac{dM_g}{dt} = \dot{M}_{\mathrm{ext}}(t) - \frac{M_g}{t_{\mathrm{loss}}}, \tag{2.1}$$

where t_{loss} is a the characteristic time over which gas is lost to the system (through the formation of stars and the launching of galactic winds). Here we are implicitly assuming that a given galaxy ejects gas in galactic winds at a rate directly proportional to the star formation rate, the constant of proportionality being defined as the mass loading factor η. We also assume that all stellar evolution happens instantaneously so that a fixed proportion of all mass which forms stars is immediately returned to the gas reservoir of the galaxy, and a fixed fraction f_R is permanently locked in stellar remnants. We explicitly relate the loss rate with these quantities as follows

$$\frac{M_g}{t_{\rm loss}} = (f_R + \eta)\dot{M}_{\rm SF} = (f_R + \eta)\frac{M_g}{t_{\rm dep}}, \tag{2.2}$$

where $\dot{M}_{\rm SF}$ is the star formation rate, and $t_{\rm dep}$ is the depletion time of all gas in the galaxy. This equation demonstrates that a great deal of (poorly-constrained) physics is hidden in $t_{\rm loss}$. In this and the following section, however, we will simply scale the time coordinate to $t_{\rm loss}$ and consider ensembles of galaxies with fixed $t_{\rm loss}$. This is a powerful technique, since our conclusions in these sections will be independent of the values and scaling relations of η, $t_{\rm dep}$, and so forth.

The external accretion rate is $\dot{M}_{\rm ext}$, which we parameterize as a lognormal distribution with fixed median and scatter,

$$\dot{M}_{\rm ext}(t) = \exp(\mu + \sigma x(t)), \tag{2.3}$$

where $x(t)$ is a random variable distributed as a standard normal (zero mean, unit variance), with a new value drawn at a fixed time interval $t_{\rm coherence}$. A more realistic model might have a spectrum of timescales over which x would vary, but for simplicity, clarity, and analytical tractability we will use a single “coherence time”.

Equation 2.1 is simple enough that it may be solved analytically given a sequence of random numbers, i.e. $x(t)$. It is convenient to first non-dimensionalize by scaling the mass loss rate to the median accretion rate,

$$\Psi = \frac{M_g}{t_{\text{loss}}} e^{-\mu}, \tag{2.4}$$

and the time to t_{loss} via $t = \tau t_{\text{loss}}$. Thus τ represents time in units of mass loss timescales, and one may think of Ψ loosely as the star formation rate or the gas mass, though strictly speaking it is the mass loss rate per median accretion rate. The evolution equation then becomes

$$\frac{d\Psi}{d\tau} = -\Psi + e^{\sigma x(t)} \tag{2.5}$$

The value of μ has entirely dropped out, so the full distribution of Ψ is determined solely by the inherent scatter in the accretion rate, σ, and the number of mass loss times over which the accretion rate remains constant, $\tau_c \equiv t_{\text{coherence}}/t_{\text{loss}}$.

This equation has exactly the same structure as the radiative transfer equation where accretion acts as the source term, time in units of mass loss timescales is similar to the optical depth, and the instantaneous value of Ψ is analogous to the intensity of radiation.

The mass loss rate as a function of time can now be solved analytically if we are given a sequence of standard normals, x_k, $k = 0, ...$, where each k corresponds to a new draw from the lognormal accretion rate distribution. In particular, suppose we know Ψ_k, the value of Ψ at the time that a new accretion rate is drawn. The value of Ψ from then up to the subsequent draw is given by

$$\Psi(\Delta) = \Psi_k e^{-\Delta} + e^{\sigma x_k}(1 - e^{-\Delta}) \text{ for } 0 \leq \Delta \leq \tau_c \tag{2.6}$$

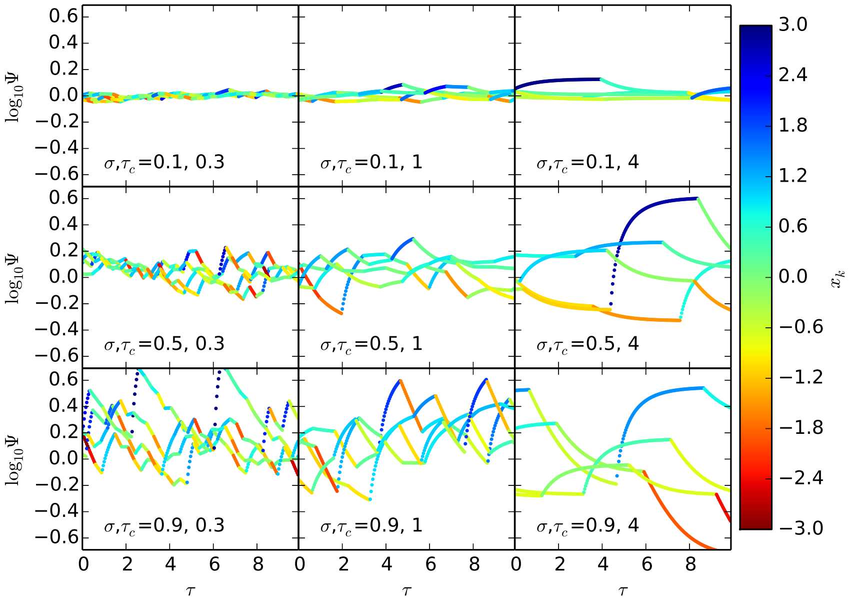

Figure 2.1: Ψ as a function of time. For each pair of σ (increasing top to bottom) and τ_c (increasing left to right), we show the trajectories of five random galaxies over the course of a randomly selected 10 star formation times and colored by the instantaneous value of $x(t)$ – bluer colors mean higher accretion rates. The galaxies exponentially approach $\Psi = e^{\sigma x_k}$. When τ_c is short, the many changes in the accretion rate never allow the galaxies to reach the accretion rate – instead they remain near the average value.

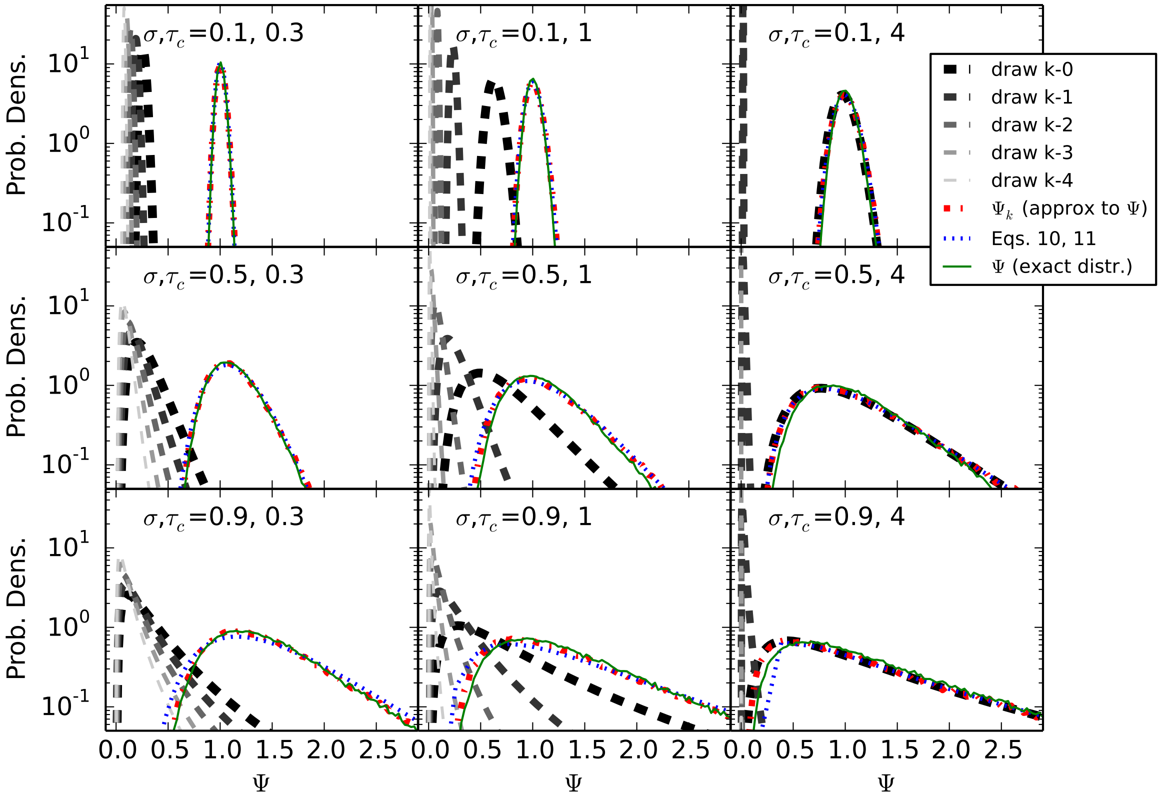

Figure 2.2: The construction of Ψ. The red, blue, and green lines show the probability density of a galaxy with a given combination of σ and τ_c having, at a randomly selected time, a given value of Ψ. shows the full distribution, red shows the distribution of Ψ_k, namely Ψ at a switch in the accretion rate, and blue shows the analytic approximation to Ψ_k. These approximations are remarkably good, at least for this selection of σ and τ_c. The dashed gray lines show the lognormal distributions from which numbers are drawn and added together to compute Ψ_k. As τ_c increases, fewer draws from the accretion distribution contribute to Ψ, until eventually only the last draw matters and all the others are exponentially suppressed (rightmost panel). As the intrinsic width of the accretion distribution increase (top to bottom), the accretion rate distributions increasingly overlap, increasing the chances that the most recent accretion rate does not dominate in determining the current value of Ψ.

where we have defined $\Delta \equiv \tau - k\tau_c$, the time since the most recent switch in the accretion rate. This is simply a relaxation equation. We can see this more explicitly by defining $r_k = \Psi_k / e^{\sigma x_k}$, the mass loss rate at the time of the switch in accretion rate relative to the newly-chosen accretion rate – this is a measure of how far from equilibrium the system is immediately after a switch. Large values mean that the galaxy is losing much more mass than it is accreting, values less than unity mean gas is building up, and $r_k = 1$ means that the galaxy has equilibrated to its current accretion rate. Note that $r_k > 0$. Using r_k, we see that

$$\Psi(\Delta)/e^{\sigma x_k} = 1 + e^{-\Delta}(r_k - 1) \text{ for } 0 \leq \Delta \leq \tau_c \tag{2.7}$$

In other words, the loss rate approaches the accretion rate exponentially on a mass loss timescale, with the deviation determined by the deviation at the time the accretion rate switched ($\Delta = 0$).

Using equation 2.6, we can recursively compute Ψ_{k+1} from the previous Ψ_k by setting $\Delta = \tau_c$,

$$\Psi_{k+1} = \Psi_k e^{-\tau_c} + e^{\sigma x_k}(1 - e^{-\tau_c}) \tag{2.8}$$

The sequence of x_k's thereby determine a sequence of Ψ_k's, which are used as a scaffolding to construct the full solution $\Psi(\tau)$ for a given realization of the random variables. Examples of such realizations for various values of σ and τ_c are shown in figure 2.1.

We may also solve the recursion relation (equation 2.8) for Ψ_k explicitly, again given the sequence x_k.

$$\Psi_k = (1 - e^{-\tau_c}) \sum_{i=0}^{k-1} e^{-(k-i-1)\tau_c} e^{\sigma x_i} + \Psi_0 e^{-k\tau_c} \tag{2.9}$$

We see that the mass loss rate at a switch in the accretion rate is therefore simply a sum of lognormally-distributed random variables. Each draw from the distribution loses influence as it recedes into the past. In particular $(k-i)\tau_c$ is simply the number of loss times since x_i was drawn. The prefactor of $1 - e^{-\tau_c}$ accounts for the fact that each individual draw of the accretion history matters less as the coherence time gets shorter. For example, even if the accretion rate is very large, if the coherence time is very short, the galaxy will only experience that accretion rate for a very short time.

In the limit that $\tau_c \gg 1$, only the most recent draw from the distribution matters and all previous draws are exponentially suppressed. In the opposite limit, $\tau_c \ll 1$, the leading factor $1 - e^{-\tau_c} \to \tau_c \approx N_{\text{loss}}^{-1}$, where N_{loss} is the number of draws from the accretion rate distribution in a given mass loss time. For very short coherence times, Ψ_k becomes an average of the lognormal accretion rates over the most recent mass loss timescale, with more recent accretion rates weighted somewhat more heavily. The full probability distribution of Ψ and Ψ_k (which serves as an approximation to Ψ) for various values of τ_c and σ are shown in figure 2.2, along with the probability distributions of the 5 most recent draws from the appropriate lognormal distributions which are added together to give Ψ_k. As the coherence times grow longer (panels from left to right), fewer draws from the accretion rate contribute to the current mass loss rate. As the intrinsic scatter in the accretion rate increases (top to bottom), the probability density of the draws from the accretion rate overlap more, meaning that it is increasingly likely that the most recent accretion rate is not the largest contributor to the current mass loss rate.

The distribution of a random variable, i.e. Ψ_k which is the sum of lognormally distributed random variables, i.e. $(1-e^{-\tau_c})e^{-j\tau_c+\sigma x_j}$, may itself be approximated as a lognormal random variable (Fenton, 1960) with variance (in the log)

$$\sigma^2_{\Psi_k} \approx \ln\left[1 + \left(e^{\sigma^2}-1\right)\frac{1-e^{-\tau_c}}{1+e^{-\tau_c}}\right]. \tag{2.10}$$

and median

$$\mu_{\Psi_k} \approx \sigma^2/2 - \sigma^2_{\Psi_k}/2. \tag{2.11}$$

Equation 2.10 represents a good analytical guess for the width of the star-forming main sequence. In the limit of large τ_c, $\sigma_{\Phi_k} \to \sigma$ as expected – when the galaxies are able to equilibrate to their accretion rates, the width of the main sequence is simply the intrinsic width in the accretion rate, σ. In this limit $\mu_{\Phi_k} \to 0$, i.e. the median mass loss rate approaches the median accretion rate.

In the limit of short τ_c, the fraction $(1-e^{-\tau_c})/(1+e^{-\tau_c})$ reduces to $\tau_c/2$. Thus when σ is also sufficiently small (namely $\sigma^2 \ll 1$), the variance of Ψ_k reduces to $\sigma^2\tau_c/2$. Physically, when the accretion rate varies rapidly compared to the mass loss time, any individual draw from the accretion rate distribution becomes unimportant and all that matters is the long-term average. The exception is when σ is large, in which case extremely large accretion rates become common and σ_{Ψ_k} again approaches

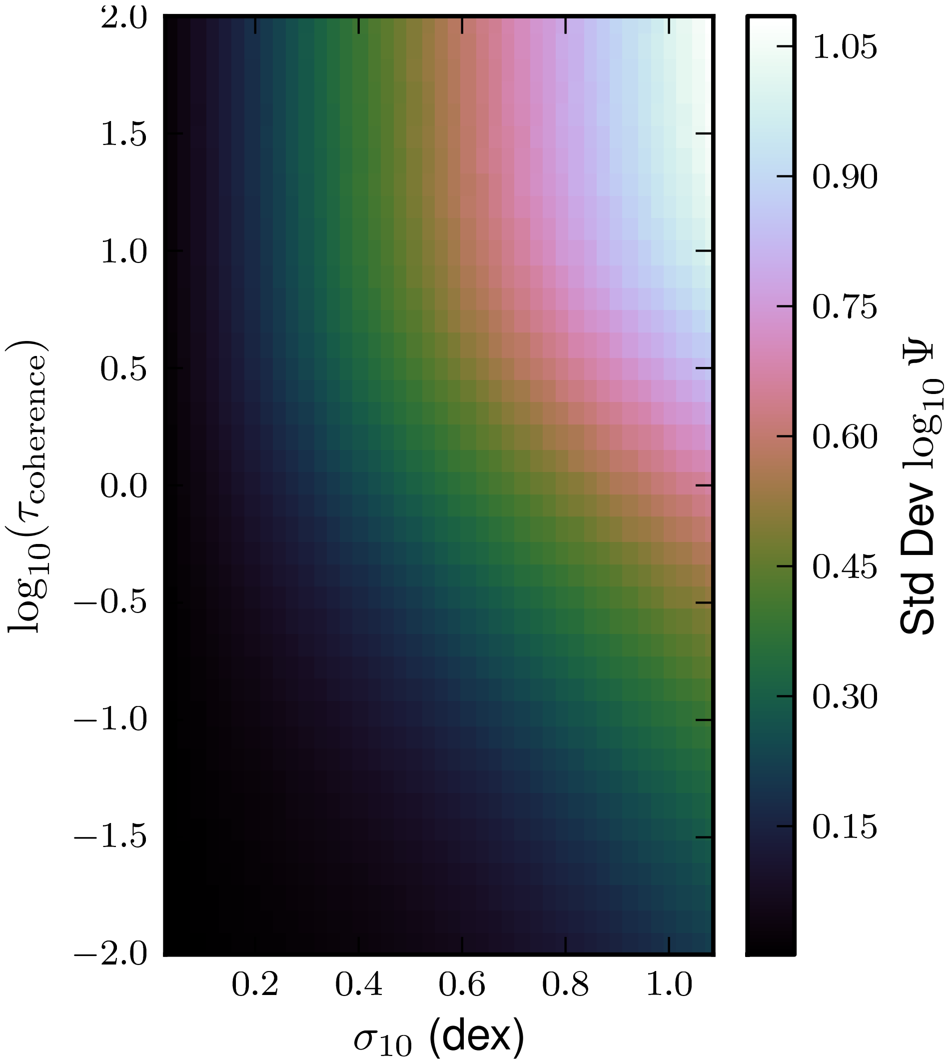

Figure 2.3: The width of the "main sequence". Each pixel represents an ensemble of galaxies with fixed τ_c and σ, wherein the mass loss rate Ψ was measured for each galaxy at a random time. The standard deviation of $\log\Psi$ in each bin is plotted above. Longer coherence times (upwards in the plot) allow the galaxies to equilibrate so that Ψ approaches the accretion rate, so the scatter in Ψ approaches σ. Coherence times shorter than the mass loss timescale, i.e. $\log_{10}\tau_c \lesssim 1$, lead to a reduction in the scatter roughly in proportion to τ_c – individual draws from the accretion rate matter increasingly less, and the galaxies do not have enough time for Ψ to approach the accretion rate before it changes.

σ, though we caution that the distribution becomes less log-normal as σ increases with small τ_c. Figure 2.3 shows σ_{Ψ_k} computed with a Monte Carlo simulation (for details see appendix 2.7).

From this discussion, we see that τ_c and σ compete in setting the width of the main sequence, even after σ_{Ψ_k} has been scaled by the intrinsic accretion rate. This suggests that σ may be interpreted as a third timescale in the problem, namely the number of star formation times necessary to forget a typical accretion event.

2.3 Including metallicity

The metal content of the galaxy is evolved according to the instantaneous recycling approximation, in the spirit of Tinsley (1980) and Maeder (1992),

$$\frac{dM_gZ}{dt} = Z_{IGM}\dot{M}_{\rm ext} + \dot{M}_{\rm SF}(f_Ry - f_RM_Z/M_g - \eta Z_w) \tag{2.12}$$

New metals are added along with accreting matter (the first term), and respectively produced by, locked up in the products of, and ejected from the galaxy by, star formation.

The star formation rate, taking into account galactic winds and stellar evolution (see equation 2.2), is $\dot{M}_{\rm SF} = (f_R + \eta)^{-1}M_g/t_{\rm loss}$. The yield y is defined as the mass of metals produced during the course of stellar evolution per unit mass of gas locked in stellar remnants – if 1000 $M_\odot$ of gas forms stars, a total of $yf_R1000M_\odot$ of metals will be produced by these stars and returned to the ISM. Following a procedure we will discuss further in Chapter 4, we parameterize the wind metallicity Z_w as

$$Z_w = Z + \xi y f_R / \max(\eta, 1 - f_R). \tag{2.13}$$

The usual assumption throughout the literature on chemical evolution is that $Z_w = Z$, i.e., before gas is ejected from a galaxy by stellar feedback, it is assumed to be perfectly well-mixed with the ambient ISM (for recent exceptions see Peeples & Shankar, 2011; Krumholz & Dekel, 2012; Vogelsberger et al., 2013). This is a strong assumption because galactic winds are likely to be preferentially metal-enriched. Physically this is because metals are produced in the same places, sometimes by the same events, which are likely to cause galactic-scale outflows, namely sites of recent star formation, where massive stars emit ionizing radiation and end their lives as supernovae. The assumption that $Z_w = Z$ corresponds in our model to $\xi = 0$ – the other extreme value is $\xi = 1$, corresponding to exactly no mixing between the metal-rich ejecta of stars and the ISM.

Non-dimensionalizing as in the previous section, we arrive at

$$\frac{d\Psi Z}{d\tau} = -\Psi Z + Z_{IGM} e^{\sigma x(t)} + q\Psi \tag{2.14}$$

where all of the factors associated with star formation and feedback can be collected into a single parameter

$$q \equiv \frac{y f_R}{f_R + \eta} \left(1 - \frac{\eta \xi}{\max(1 - f_R, \eta)} \right). \tag{2.15}$$

Much of the uncertainty in modeling the metallicity evolution of galaxies is encapsulated in q. We note that in the limit that η is large (probably the case for low-mass galaxies), $q \to y f_R(1-\xi)/\eta$, while for small values of η, $q \to y$. In this sense q may be considered an effective yield, and the more important galactic winds are, the more uncertain this parameter becomes.

Combining equations 2.14 and 2.1, we can obtain the evolution equation for metallicity,

$$\frac{dZ}{d\tau} = \frac{Z_{IGM} - Z}{\Psi/e^{\sigma x}} + q \tag{2.16}$$

This equation represents a competition between the effective yield of new metals formed during the course of stellar evolution and dilution of metals by accretion. In equilibrium, namely when the accretion rate is constant for many mass loss times, $\Psi/e^{\sigma x} \to 1$, so for $dZ/d\tau = 0$, we find that the equilibrium metallicity is

$$Z_{eq} = q + Z_{IGM}. \tag{2.17}$$

The corresponding quantity for the relative mass loss rate is $\Psi_{eq} = e^{\sigma x}$, where x is a standard normal. In this sense, metallicity is very different from Ψ – regardless of the accretion rate, the metallicity approaches a constant value, whereas the mass loss rate approaches whatever random value of the accretion rate it is being fed.

The factor of q can be factored out of the metallicity evolution equation by defining

$$Z^{\dagger} = \frac{Z - Z_{IGM}}{q}, \tag{2.18}$$

the metallicity scaled to its equilibrium value by the effective yield – Z_{eq} corresponds

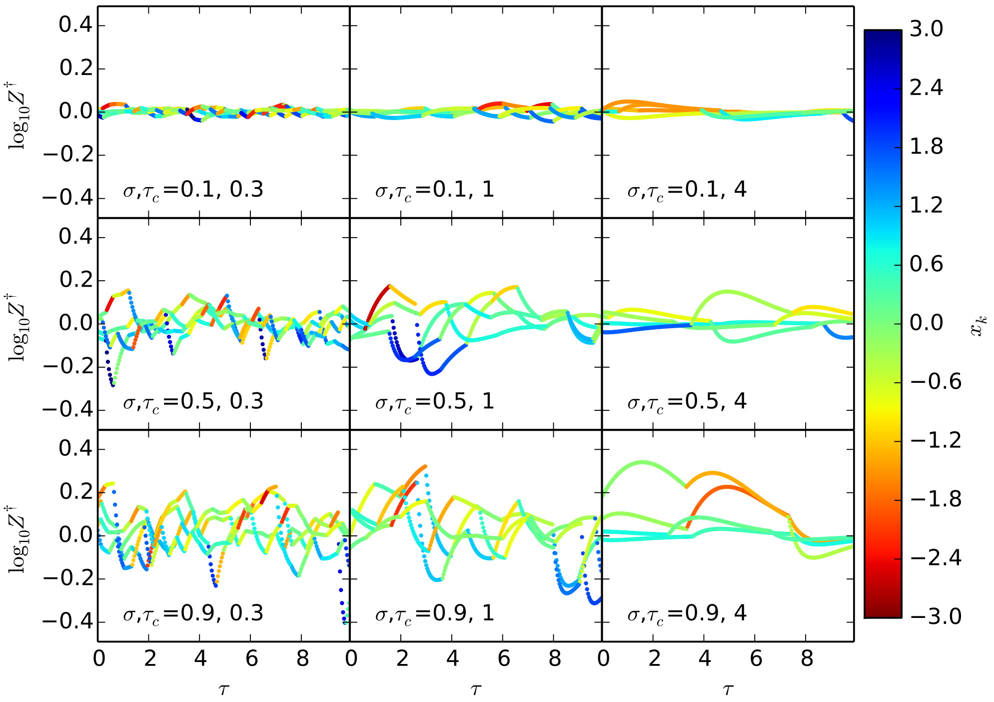

Figure 2.4: $Z^{\dagger}$ as a function of time. For each pair of σ (increasing top to bottom) and τ_c (increasing left to right), we show the trajectories of five random galaxies over the course of a randomly selected 10 star formation times and colored by the instantaneous value of $x(t)$ – bluer colors mean higher accretion rates. We can immediately see that high accretion rates tend to lead to low metallicities. We also see that for large coherence times (rightward) each galaxy reaches an extremum in metallicity before returning to its equilibrium value ($Z^{\dagger} = 1$)

to $Z^\dagger = 1$. The evolution equation for $Z^\dagger$ then becomes independent of q, yielding

$$\frac{dZ^\dagger}{d\tau} = \frac{-Z^\dagger}{1 + e^{-\tau}(r_k - 1)} + 1 \tag{2.19}$$

The denominator is recognizable as the mass loss rate normalized by the current accretion rate, $e^{\sigma x_k}$.

As with the star formation equation, we may solve equation 2.19 by elementary methods for a given fixed value of r_k, i.e. during the period when the accretion rate is constant. The solution is

$$Z^\dagger(\Delta) = \frac{Z^\dagger(0) r_k + e^\Delta - 1 + (r_k - 1)\Delta}{e^\Delta - 1 + r_k} \quad \text{for } 0 \leq \Delta \leq \tau_c \tag{2.20}$$

In the limit of large Δ, we recover that $Z^\dagger \to 1$, i.e. the metallicity returns to its equilibrium value, though depending on how different the initial mass loss rate is from the accretion rate, $\Psi_k / e^{\sigma x_k} = r_k$, the value of $Z^\dagger$ may change dramatically over the course of a star formation time ($\Delta \sim 1$) before exponentially returning to unity. Examples of such trajectories are shown in figure 2.4. Notice that in the right-most panels, despite the somewhat large value of τ_c, there are significant deviations from equilibrium immediately following a change in the accretion rate. For example, if the accretion rate increases, the ISM begins getting diluted immediately by the large accretion rate. The star formation rate, and hence the rate of metal production, takes some time to adjust to the new accretion rate while gas builds up in the ISM. It therefore takes some time before metal production can 'pollute' the large supply of low-metallicity infalling gas and return to the equilibrium $Z^\dagger = 1$, wherein for every unit of new clean gas added to the galaxy, star formation provides enough mass in metals to bring it to $Z = q + Z_{IGM}$. Similarly, if the accretion rate decreases suddenly, the rate at which metals are produced exceeds the rate necessary to pollute the new, smaller, supply of accreting gas, and the metallicity of the ISM increases until star formation and galactic winds can burn through the gas reservoir, allowing the galaxy to adjust to a lower rate of star formation (and hence metal production).

Immediately after a change in the accretion rate, i.e. in the limit that $\Delta \ll 1$,

$$Z^\dagger(\Delta) \approx Z^\dagger(0) + (1 - Z^\dagger(0)e^{\sigma x_k}/\Psi_k)\Delta \tag{2.21}$$

Thus we can see that immediately following a change in accretion rate, whether the metallicity will increase or decrease depends only on the ratio $Z^\dagger(0)e^{\sigma x_k}/\Psi_k$. When this number is greater than unity, because the metallicity is high and/or the new accretion rate is larger than the present mass-loss rate, the metals will be diluted, whereas if this number is less than one, e.g. in a low-metallicity galaxy and/or one facing a sudden drop in accretion rate, stars will be forming fast enough to pollute the gas reservoir. This relation demonstrates the basic physical mechanism which gives rise to the the mass-star formation-metallicity plane in our model, but also shows that this is a statistical relation only. A higher accretion rate (and hence an increasing star formation rate) is not guaranteed to produce a lower metallicity

We can solve for the maximum deviation from $Z^\dagger = 1$ if the galaxy begins in equilibrium, namely $Z^\dagger(0) = 1$. In this case,

$$Z^\dagger_{\text{extreme}} = \frac{(r_k - 1)(1 + W_0((r_k - 1)/e)}{r_k - 1 + \exp(1 + W_0((r_k - 1)/e))} \tag{2.22}$$

where $W_0(x)$ is the Lambert W function, namely the real solutions to $y = W_0(y)e^{W_0(y)}$.

As in the previous section, we can construct a recursion relation by setting $\Delta = \tau_c$ in equation 2.20. As with the mass loss rate, we see that the metallicity at the switches in accretion rate is given by a sum wherein the effects of long-past accretion are exponentially suppressed,

$$\begin{aligned} Z^\dagger_k &= Z^\dagger_0 \prod_{i=0}^{k-1} \frac{r_i}{r_i + e^{\tau_c} - 1} \\ &+ \sum_{i=0}^{k-1} \frac{e^{\tau_c} - 1 + \tau_c(r_i - 1)}{e^{\tau_c} - 1 + r_i} \prod_{j=i}^{k-1} \frac{r_j}{e^{\tau_c} - 1 + r_j} \end{aligned} \tag{2.23}$$

Since $r_k > 0$ and $e^{\tau_c} > 1$, each factor in both of the products is guaranteed to be between 0 and 1 – for a median value of $r_k = 1$ each factor becomes $e^{-\tau_c}$. In that sense this equation is very similar to equation 2.9 for Ψ_k. We show how $Z^\dagger$ is constructed in figure 2.5 – in addition to the full distributions of $Z^\dagger$ and $Z^\dagger_k$, we show the distributions for galaxies with Ψ in the top and bottom quartiles of galaxies with those values of σ and τ_c. As in figure 2.2, previous terms in the sum recede into irrelevancy over the course of a few star formation times. There are, however, crucial differences between the construction of $Z^\dagger_k$ and Ψ_k. To construct Ψ_k, independent lognormal variables were added together to get another quantity which was roughly lognormal. Here on the other hand, even though an individual r_i is indeed roughly lognormally distributed (since $r_k = \Psi_k / e^{\sigma x_k}$

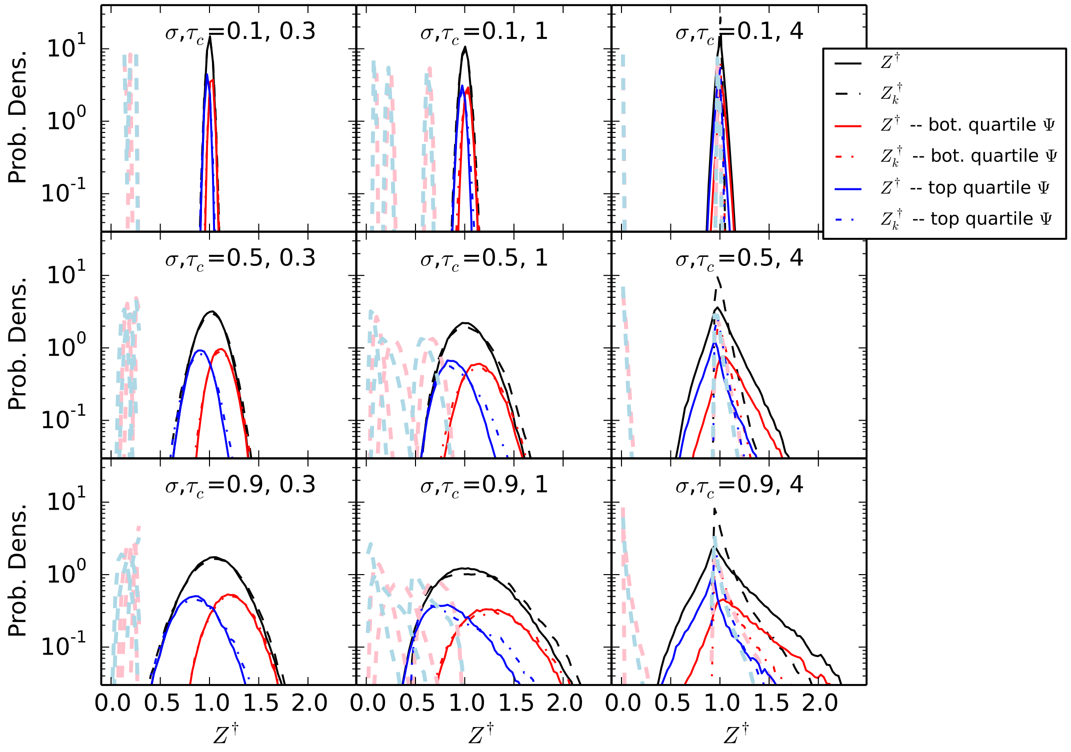

Figure 2.5: The construction of $Z^\dagger$. For the same parameters shown in the previous figures, we plot the PDF of $Z^\dagger$ (solid line), the approximation to it using $Z_k^\dagger$ (equation 2.23), and we also split the probability distribution into galaxies with high (low) star formation rates in blue (red). We can immediately see that high metallicity galaxies tend to have low star formation rates, though the probabilities do overlap substantially. Light blue and red lines show the PDF of the contributions to $Z_k^\dagger$, split into the top and bottom quartiles in star formation rate, from old terms in the summation of equation 2.23. As with Ψ, the probability distribution of $Z^\dagger$ may be thought of as the sum of a series of random draws wherein the influence of older draws is exponentially forgotten, although galaxies which end up in the blue (red) bin at the present time are likely to have had preferentially higher (lower) SFRs in the past, at least for $\tau_c \lesssim 1$. Unlike with Ψ, $Z^\dagger$ becomes noticeably non-log-normal for larger values of τ_c, since galaxies sampled at a random time are likely to be near the equilibrium value $Z^\dagger = 1$, regardless of the accretion rate.

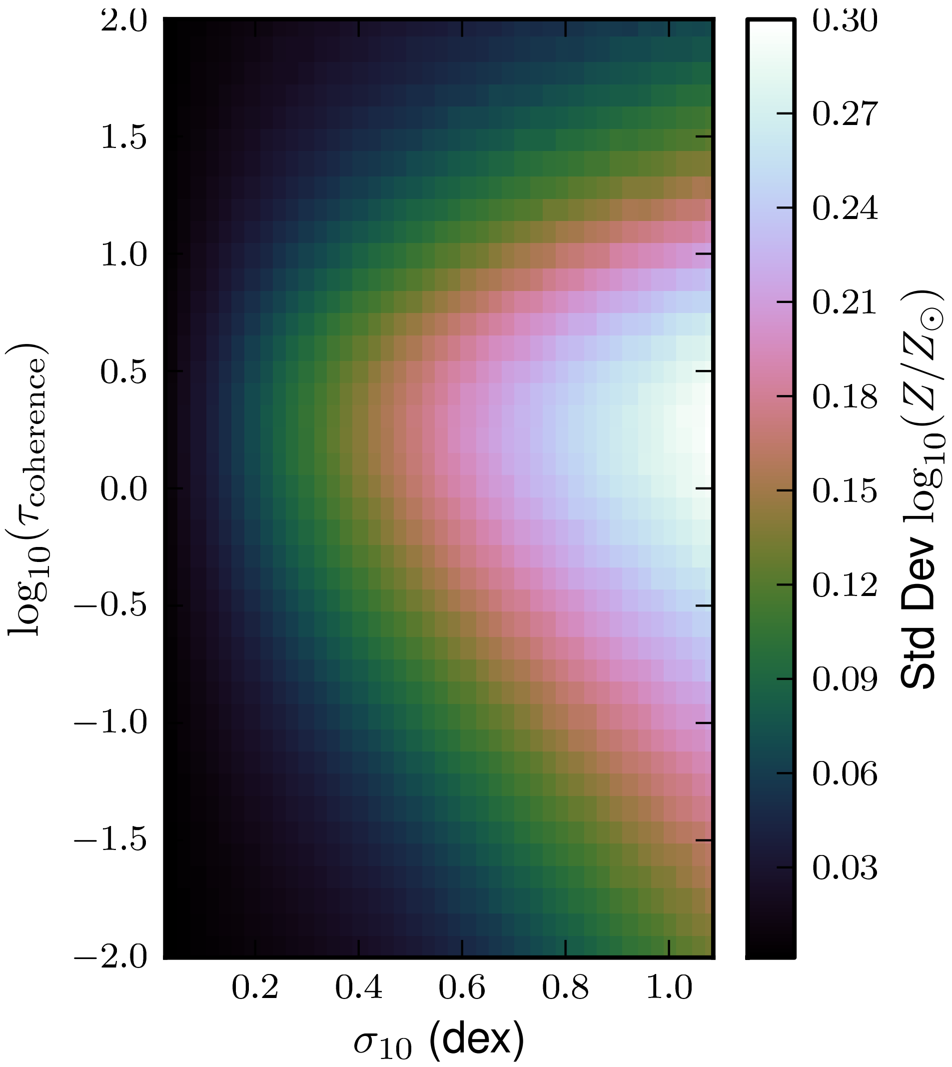

Figure 2.6: The width of the mass-metallicity relation. Each pixel represents an ensemble of galaxies with fixed τ_c and σ, wherein the metallicity Z was measured for each galaxy at a random time. The standard deviation of $\log Z/Z_\odot$ in each bin is plotted above. Longer coherence times (upwards in the plot) allow the galaxies to equilibrate so that Z approaches Z_{eq} (a constant value regardless of the accretion rate). Coherence times shorter than the mass loss timescale, i.e. $\log_{10}\tau_c \lesssim 1$, lead to a reduction in the scatter roughly in proportion to τ_c – individual draws from the accretion rate matter increasingly less. Larger intrinsic scatters (rightward in the plot) negate this effect by making large accretion events typical.

and Ψ_k is roughly lognormal), it is not the r_i's being summed, but rather a more complicated function of r_i and τ_c. Moreover, the r_i are not independent of each other. Perhaps worst of all, since $Z^\dagger$ does not monotonically approach its equilibrium value each time the accretion rate changes, but rather increases to $Z^\dagger_{\text{extreme}}$ before returning to $Z = Z_{eq}$, $Z^\dagger_k$ becomes a bad approximation of the full distribution of $Z^\dagger$ when $\tau_c \gtrsim 1$ (though eventually, for $\tau_c \gg 1$, the probability distributions of $Z^\dagger$ and $Z^\dagger_k$ both approach $\delta(Z^\dagger - 1)$).

Despite these difficulties, we can see that $p(Z^\dagger) \approx p(Z^\dagger_k)$ for shorter accretion times, and we can also begin to see that Ψ and $Z^\dagger$ are anti-correlated. Galaxies which have Ψ in the top quartile of their ensemble tend to have lower metallicities. This is a very simple physical effect, namely the competition between dilution of metals by new infall and pollution of metals by stellar evolution. If the galaxy is burning through excess gas from previous accretion events ($r_k > 1$), the metallicity will increase as the gas reservoir is polluted, whereas if the galaxy is accreting more gas than it is losing ($r_k < 1$), the star formation rate is slow enough that new gas is added faster than metals can be produced to pollute it. Using the same Monte Carlo simulations used to produce figure 2.3, we can explicitly show the correlation between $\log \Psi$ and $\log Z$ (figure 2.7), defined in general according to

$$Corr(X,Y) = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left(\sum_i (X_i - \bar{X})^2\right)\left(\sum_j (Y_j - \bar{Y})^2\right)}}, \tag{2.24}$$

where each sum is over all galaxies in the ensemble and $\bar{X}$ indicates an average over the

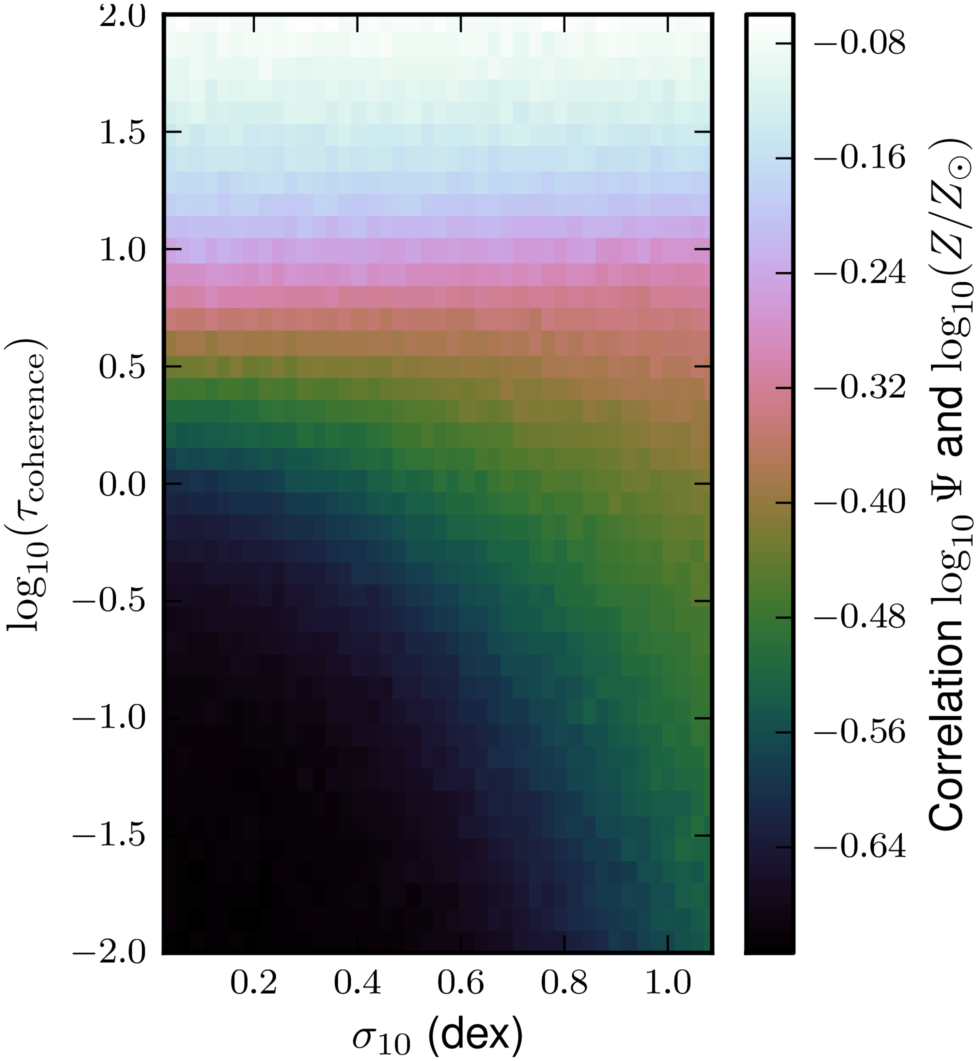

Figure 2.7: The correlation between Ψ and $Z/Z_\odot$. For small scatters and rapid variability in the accretion rate, the star formation rate and metallicity are substantially anti-correlated. Increasing the coherence time allows galaxies to return to their equilibrium Z regardless of the accretion rate, wiping out the anti-correlation.

ensemble. The correlation is strongest in the ‘linear regime’, namely small rapid perturbations. When the coherence time exceeds a few star formation times, the correlation disappears – $Z^\dagger$ is always close to unity regardless of the accretion rate. The correlation also weakens as the intrinsic scatter in the accretion rate increases.

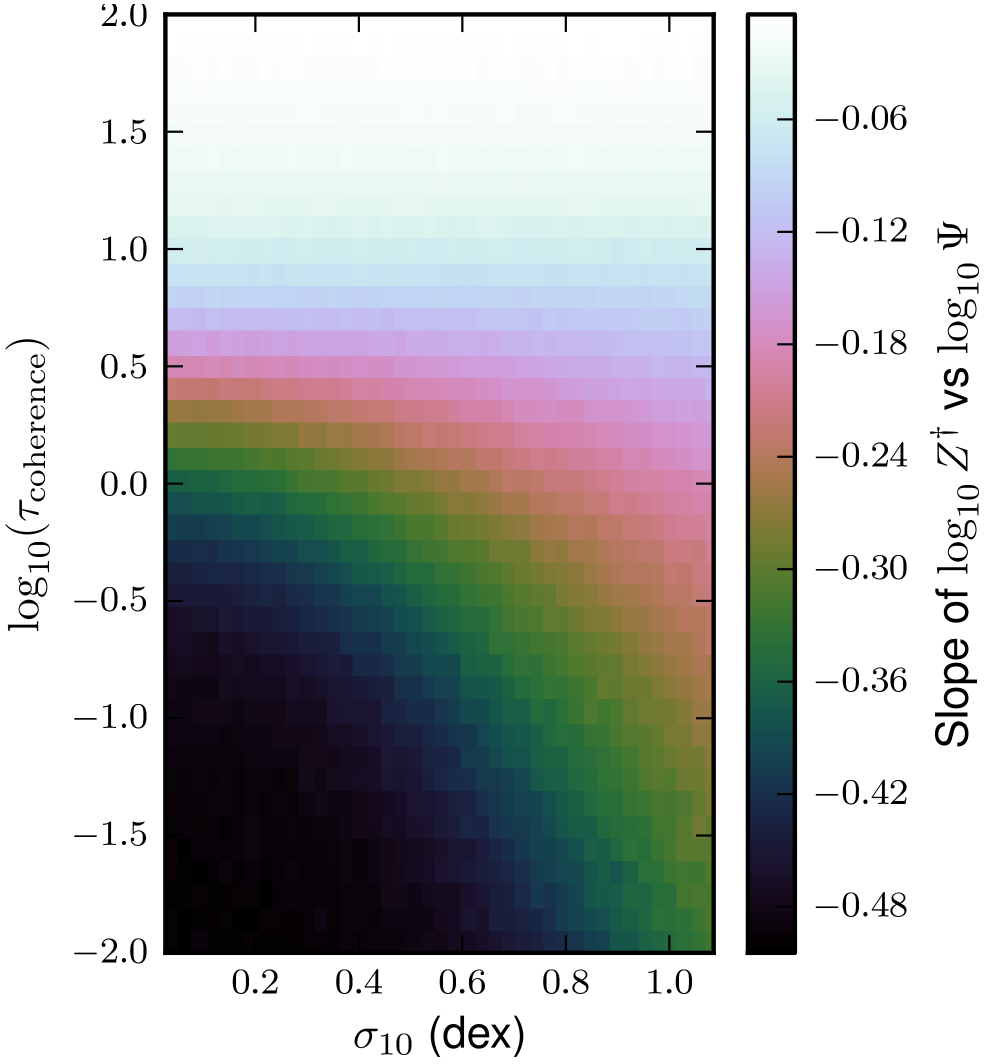

Figure 2.8: The "slope of the FMR". For each τ_c and σ we fit a linear model to the scatter plot of $\log_{10} \Psi$ vs $\log_{10}(Z/Z_{\odot})$ of all galaxies drawn from the distribution, and plot the resulting slope here. We find uniformly that the slope is negative, though for the same reasons mentioned in the plot of the correlation between these two variables, the slope flattens for large values of τ_c, and to a lesser degree, for large values of σ.

2.4 Construction of Galaxy Scaling Relations – an initial guess

The results we have derived thus far, namely the joint probability density of Ψ and $Z^{\dagger}$, are applicable only at fixed values of μ, σ, $t_{\mathrm{coherence}}$, t_{loss}, Z_{IGM}, and q. We refer to an ensemble of galaxies with fixed values of these parameters as a simple (or stationary) galactic population (SGP). As we saw in the previous sections, the only variables which affect the joint distribution of Ψ and $Z^{\dagger}$ are σ and $\tau_c = t_{\mathrm{coherence}}/t_{\mathrm{loss}}$. However, to compute physical quantities, i.e. the star formation rate, the metallicity, etc., one must specify the other variables.

The power of our approach using SGPs is that, to a reasonable approximation, the properties of a star-forming galaxy are set by a single parameter having to do with the size of the galaxy (e.g., stellar mass, halo mass, or K-band luminosity). One could therefore hope that, at a fixed stellar mass, the population of galaxies may be well-described by a single SGP. Essentially, all of the scalings of mass loading factor, accretion rate, etc. which account for the slope and zero-points of galaxy scaling relations would be taken out, leaving only the intrinsic scatter.

Here we make an educated guess as to how to map the dimensionless model presented in the first few sections of this paper to observable galaxy scaling relations. Although alternative assumptions may be preferred by other practitioners, we hope this exercise will be at least illustrative. In section 2.4.1 we use results from N-body-only dark matter simulations to make guesses for the parameters of the accretion process: μ, σ, and $t_{\mathrm{coherence}}$. Each of these parameters thereby has a predicted scaling with halo

Table 2.1: Important parameters used in Chapter 2

Parameter	Description
	The accretion process
μ	The log base e of the median baryonic accretion rate.
σ	The log-normal scatter in the (DM and baryonic) and accretion rate
$\Delta\omega$	The scale-free time step taken to generate accretion histories
$t_{\mathrm{coherence}}$	The time over which the accretion rate is constant before a new random value is drawn.
ϵ	The fraction of $f_b\dot{M}_{DM}$ which reaches the gas reservoir
f_ϵ	A fixed fraction by which the baryonic accretion rate may be reduced – fixed by 1st order scaling relations
Z_{IGM}	Metallicity of infalling baryons, fixed at 2×10^{-4}
	Star Formation
$\dot{M}_{\mathrm{SF}}$	The star formation rate
M_g	The gas mass available to form stars
η	The ratio of mass lost in galactic winds to stars formed – fixed by 1st order scaling relations
t_{dep}	The depletion time, $M_g/\dot{M}_{\mathrm{SF}}$ – fixed by 1st order scaling relations
f_R	The remnant fraction - fraction of mass not returned to the ISM from a simple stellar population, fixed at 0.54
t_{loss}	The loss time, i.e. the timescale on which the reservoir is depleted, $t_{dep}/(\eta+f_R)$
y	Mass of metals returned to the ISM per mass locked in stellar remnants, fixed at 0.054
ξ	Metallicity enhancement of galactic winds, fixed at 0 (perfect mixing)
q	A combination of y, f_R, ξ, and η we call the effective yield
	Stationary Galactic Population (SGP) Parameters
τ_c	$t_{\mathrm{coherence}}/t_{loss}$ – an input to the SGP
σ_{10}	This is the same σ as above, but divided by $\ln 10$ – an input to the SGP
Ψ	The ratio of the mass loss rate (M_g/t_{loss}) to the median accretion rate e^{μ} – an output of the SGP
$Z^{\dagger}$	The metallicity of the gas reservoir, subtracting out Z_{IGM} and normalizing to q – an output of the SGP
x_k	For a given galaxy, the kth draw from a standard gaussian, which sets the accretion rate.
	Other sources of scatter (see especially section 2.5.1)
$\sigma_{\log_{10} M_*}$	Log-normal scatter added to $M_*(M_h)$, assumed uncorrelated with anything else
$\sigma_{\log_{10} t_{dep}}$	Log-normal scatter added to the depletion time
$\sigma_{\log_{10}\eta}$	Log-normal scatter added to the mass loading factor η
σ_μ	Gaussian scatter in μ. This is like including a stochastic component whose coherence time $\gg t_{loss}$.

mass and redshift. In this procedure we leave one free parameter f_ϵ, a constant less than unity, to describe how much the accretion rate is reduced below this guess. We then adopt the assumptions that $y = 0.054$, $f_R = 0.54$ (appropriate for a Chabrier (2005) IMF with yields from solar metallicity stars Maeder (1992) – see appendix A of Krumholz & Dekel (2012)), $Z_{IGM} \ll 0.02$, and $\xi = 0$. To fully specify the SGP model, the only remaining parameters are η, t_{dep}, and f_ϵ. In section 2.4.2, we adopt a value of f_ϵ and powerlaw scalings of t_{dep} and η with halo mass such that we fit three galaxy scaling relations: M_* vs. SFR, Z, and M_g/M_*. To do so we need to assign a value of M_* to each halo mass, for which we take the Behroozi et al. (2013a,b) relations with a fixed log-normal scatter of 0.19, consistent with various observational constraints (Behroozi et al., 2013a; Reddick et al., 2013).

With all of these parameters specified as a function of halo mass and redshift, we can construct synthetic versions of these galaxy scaling relations, including their intrinsic scatter, and the higher-order fundamental metallicity relation, $Z(M_*, \mathrm{SFR})$. Section 2.4.3 describes how we take the synthetic relations and extract three higher-order quantities which we will use to constrain our model: the scatter in the MS, the scatter in the MZR, and the "slope" of the FMR, namely the logarithmic derivative of the metallicity with respect to the star formation rate at fixed stellar mass. These three quantities can then be compared directly to observations, which allows us to rule out our initial guess, and non-trivially constrain σ, $t_{\mathrm{coherence}}$, and other potential sources of scatter (section 2.5.1). The parameters used throughout the paper are summarized in table 4.1.

2.4.1 Baryonic Accretion

As is often the case for modellers of galaxies, we will begin with the dark matter. For the purposes of this simple model, we will rely on the EPS-like (Press & Schechter, 1974; Sheth & Tormen, 1999) formalism presented in Neistein & Dekel (2008) and Neistein et al. (2010). From this we will derive approximations for μ, σ, and $t_{\mathrm{coherence}}$. For a WMAP5 cosmology, Neistein et al. (2010) can fit the cumulative mass function of halos found in an N-body simulation if they construct halo accretion histories according to

$$\Delta S = \exp(\mu_p + x\sigma_p) \tag{2.25}$$

where x is a standard normal drawn at a fixed interval $\Delta\omega$, and

$$\mu_p = (0.132 \log_{10} S + 2.404) \log_{10}(\Delta\omega) + 0.585 \log_{10} S - 0.436 \tag{2.26}$$

$$\sigma_p = (-0.333 \log_{10} S - 0.321) \log_{10}(\Delta\omega) + 0.0807 \log_{10} S + 0.622 \tag{2.27}$$

In these equations, S is a measure of the amplitude of the dark matter power spectrum, and ω is a measure of time, similar to redshift.

Note that in general $\mu_p \neq \mu$ and $\sigma_p \neq \sigma$ because these refer to the mean and scatter in S-ω, rather than M-t space. To compute an estimate of the accretion rate, we take a grid of M_h and z – for each M_h we can compute S and draw a large number of ΔS values, from which we can compute ΔM_h, the change in halo mass. For each z we can compute Δt, the time between the current redshift and a time $\Delta\omega$ earlier. We then approximate each baryonic accretion rate as

$$\dot{M}_{\rm ext} \approx f_b \epsilon \frac{\Delta M_h}{\Delta t} \tag{2.28}$$

From our ensemble of ΔM_h, we can approximate μ and σ as the mean and standard deviation of $\ln \dot{M}_{\rm ext}$. We note that the distribution of $\dot{M}_{\rm ext}$ is not guaranteed to be lognormal, and this procedure produces distributions with non-zero skew and kurtosis in $\ln \dot{M}_{\rm ext}$. We neglect this non-gaussian component and approximate $\dot{M}_{\rm ext}$ as lognormal, keeping in mind that this is merely a guess at the true baryonic accretion rate. We take the efficiency factor to be

$$\epsilon = f_\epsilon \min(0.31(M_h/10^{12}M_\odot)^{-0.25}(1+z)^{0.38}, 1.0), \tag{2.29}$$

where we leave $f_\epsilon < 1$ a free parameter to be fit in the next section. The remaining factors come from the fitting formula of Faucher-Giguère et al. (2011), which accounts for the suppression of accretion in high-mass halos presumably due to hot virialized gas.

We estimate $t_{\rm coherence}$ by the fact that in the dark matter simulations, the merger trees become non-Markov for $\Delta\omega \lesssim 0.5$, indicating that the accretion rates over time intervals shorter than that are correlated (Neistein & Dekel, 2008). We therefore use

$$t_{\rm coherence} \sim \left| \left(\frac{d\omega}{dz}\right)^{-1} \frac{dt}{dz} \Delta\omega \right|, \tag{2.30}$$

with $\Delta\omega = 0.5$. Obviously this is only a rough estimate, since a real accretion history is likely to have more structure in Fourier space than the single period we assume here.

We encourage those with cosmological simulations to measure this quantity, both in dark-matter only and baryonic simulations. Another plausible value for $t_{\mathrm{coherence}}$ might be the dynamical time of the halo, or some other timescale related to the baryon cycle. This choice might be appropriate if the primary supply of gas is re-accreting winds.

2.4.2 Fitting the first-order relations

To compute τ_c and re-dimensionalize the SGP, we still need to know the mass-loss timescale, i.e. the depletion time and η, as well as ξ. These values are sufficiently uncertain that it is worth digressing to discuss how they may be chosen to fit the first-order galaxy scaling relations, namely the star-forming main sequence, the mass-metallicity relation, and the stellar mass-gas mass relation.

To compare our model with many galaxy scaling relations, computed as a function of stellar mass M_*, we must pick an M_*. Since our model is purely equilibrium-based, we have no way to compute integrated quantities like M_* besides appealing to other empirical relations. We employ the Behroozi et al. (2013a,b) model of the relation between stellar mass and halo mass, including a scatter in stellar mass at fixed halo mass of 0.19 dex (Reddick et al., 2013). Another common approach is to simply use M_* rather than M_h (Lilly et al., 2013) as the parameter by which to scale the SGP parameters (η, ξ, etc.)

In a standard equilibrium model (Davé et al., 2012), the 'center' of a galaxy scaling relation may be determined by setting time-derivatives to zero. For instance, setting $dM_g/dt = 0$ yields an equilibrium star formation rate

$$\dot{M}_{\mathrm{SF},eq} = \dot{M}_{\mathrm{ext}}/(\eta + f_R). \tag{2.31}$$

Immediately we can see the critical importance of two unknown pieces of physics – anything which displaces the baryonic accretion rate $\dot{M}_{\mathrm{ext}}$ away from the naive estimate $f_b\dot{M}_{DM}$ (e.g. preheating, halo quenching, AGN heating), and anything which removes gas from the star-forming gas reservoir of the galaxy (supernovae, radiation pressure, cosmic rays, or direct AGN). Typically the former are invoked at high mass and the latter at low mass.

Interestingly, the depletion timescale does not enter into $\dot{M}_{\mathrm{SF},eq}$. This is because the gas mass is free to adjust to whatever it needs to be so that inflowing gas is balanced by sinks for the gas – old stellar remnants and galactic outflows. For instance, a longer t_{dep} would simply mean a larger gas mass would be required to reach the same equilibrium between inflows and sinks.

$$M_{g,eq} = \dot{M}_{\mathrm{SF},eq} t_{dep} = \dot{M}_{\mathrm{ext}} t_{\mathrm{loss}} \tag{2.32}$$

Thus we can see that star formation rates can only be affected by physics which alters the star formation timescale (e.g. H_2 regulation Krumholz & Dekel, 2012) to the degree that the galaxies are out of equilibrium. These physical considerations do, however, affect the equilibrium size of the gas reservoir.

If we make the additional restriction that $dZ/dt = 0$, we arrive at the equilibrium metallicity

$$Z_{eq} = Z_{IGM} + q = Z_{IGM} + \frac{y f_R}{f_R + \eta}\left(1 - \frac{\eta\xi}{\max(1 - f_R, \eta)}\right). \tag{2.33}$$

The metallicity is a particularly powerful probe of feedback physics because Z_{eq} is independent of the accretion rate, meaning that "preventive feedback" or any other considerations which affect $\dot{M}_{\rm ext}$ do not affect the first-order mass-metallicity relation (MZR) – essentially all that matters are η, the mass loading factor, and ξ or some other measure of the mixing between ejecta and the ISM. The drawback of using the MZR is that it is bedeviled by large systematic uncertainties in converting characteristics of metal emission lines into actual gas-phase metal abundances (Kewley & Ellison, 2008).

Given these equilibrium relations, we attempt to find powerlaw scalings of η and t_{dep}, and a constant value of f_ϵ, which will roughly fit the observed galaxy scaling relations. This leads us to take $f_\epsilon = 0.5$, $\eta = (M_h/10^{12}M_\odot)^{-2/3}$, and $t_{dep} = 3$ Gyr $(M_h/10^{12}M_\odot)^{-1/2}$, as described below. These fits were done by hand, which is reasonable given the uncertainties in the mean relations.

To approximately fit the centers of the observed $z = 0$ main sequence, we had to use a steep scaling of the mass loading factor with halo mass, namely $\eta = (M_h/10^{12}M_\odot)^{-2/3}$. This is a result of the short cooling times and/or cold streams in low-mass halos which efficiently supply cold gas at a rate proportional to the dark matter accretion rate, which is much larger than the observed star formation rate in these galaxies. That is, ϵ is large for low-mass galaxies in our initial guess. If low-mass galaxies are in statistical equilibrium, these large values of η are necessary. It is possible that some other physical process is decreasing ϵ at these masses, or that the galaxies are not in equilibrium (Krumholz & Dekel, 2012; Kuhlen et al., 2012, 2013). In the latter scenario, η may vary quite weakly, leaving the mass loss timescales at low masses to be comparable to the depletion times, themselves comparable to or much longer than the age of the universe. We discuss this possibility further in section 2.5.2, but for now we adopt $\eta \propto M_h^{-2/3}$.

We do not fit the MZR particularly well in this simple model (see figure 2.9). That is, most calibrations of the MZR are not well-described by the powerlaw in M_h we assume here. This is not extremely concerning given the systematic uncertainties in the observations. The scaling of η needed to fit the star-forming main sequence does leave the MZR in the right neighborhood, without adjusting the fiducial values of $Z_{IGM} = 2 \times 10^{-4} \approx Z_\odot/100$, $\xi = 0$, or f_ϵ and η.

In contrast, the gas fraction data are fit remarkably well by our fiducial scalings. Given the reasonable fit of the SFR-M_* relation, we were left with one parameter to vary to fit $M_g/M_* - M_*$, namely t_{dep}. Despite the (lack of) trends from the GASS (Schiminovich et al., 2010) and COLD GASS (Saintonge et al., 2011) surveys in the total gas depletion time with mass, we find that we need t_{dep} =3 Gyr $(M_h/10^{12} M_\odot)^{-1/2}$. Since η scales even more steeply, the mass loss timescale actually shortens with decreasing halo mass in this scenario, scaling as $t_{\rm loss} \propto M_h^{1/6}$.

The result of applying these scalings is shown in figure 2.9. For each M_h below $10^{12.3} M_\odot$ provided in the Behroozi et al. (2013a,b) $M_*(M_h)$ relation, we apply the scalings above to compute the necessary physical parameters to specify the SGP, draw

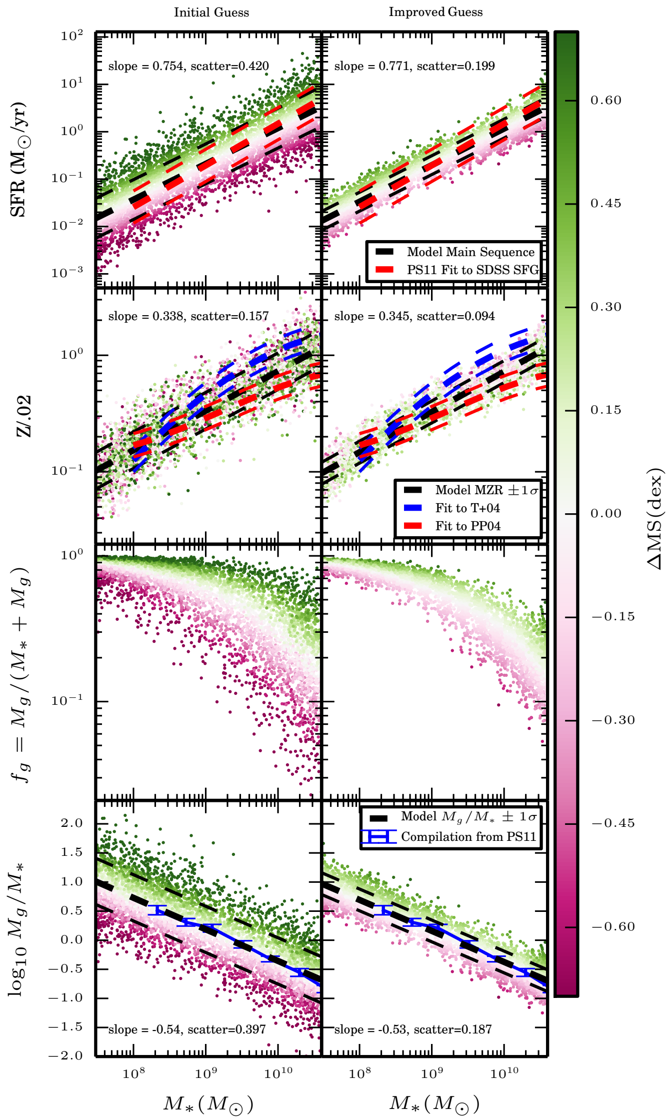

Figure 2.9: Scaling relations. We compare observed galaxy scaling relations and their scatters (colored lines) with scaling relations (black lines) calculated by drawing galaxies (colored points) from a sequence of SGPs using reasonable guesses of how the SGP parameters scale with halo mass. The left panel is our initial guess, while the right panel shows an improved guess which better fits the observational constraints on the scatter in these relations. In particular, the initial guess yields a model MS with intrinsic width wider than the observed scatter, while the improved guess has a much small intrinsic scatter. The colors indicate offset from the Main Sequence, namely the black dashed line fit in the top panel. PS11 refers to Peeples & Shankar (2011), and T+04 and PP04 refer to metallicity callibrations used by Tremonti et al. (2004) and Pettini & Pagel (2004). The M_g/M_* data come from McGaugh (2005); Leroy et al. (2008); Garcia-Appadoo et al. (2009).
.

Table 2.2: The two reference models we use to generate synthetic galaxy scaling relations and FMRs

	Initial Guess	Improved Guess
σ (dex)	0.45	0.225
$\Delta\omega$	0.5	0.2
$\sigma_{\log_{10} M_*}$ (dex)	0.19	0.07
$\sigma_{\log_{10} t_{dep}}$ (dex)	0	0
$\sigma_{\log_{10} \eta}$ (dex)	0	0
σ_{μ} (dex)	0	0.08

300 samples from the SGP, and assign each one a stellar mass according to $M_*(M_h)$, plus a fixed scatter, $\sigma_{\log_{10} M_*} = 0.19$ dex. The resultant population of galaxies, as designed, fits various observational constraints at $z = 0$, represented in each figure by the various colored lines (which include representative scatters) from Peeples & Shankar (2011). The black lines show power law fits and the computed $\pm 1\sigma$ scatter of the population. The observational fits are not necessarily precise, and in particular the mass-metallicity relation famously has many different fits depending on which calibration is used (Kewley & Ellison, 2008). The two columns in figure 2.9 have the same scalings of $t_{\rm dep}$, η, and q, but different values of σ, $\Delta\omega$, and $\sigma_{\log_{10} M_*}$. The left column shows the initial guess discussed here and in the previous section, while the right column has values of the accretion process parameters more in line with observations (the "Improved Guess" model – see table 2.2 and sections 2.4.3 and 2.5.1).

It is worth emphasizing that fitting or not fitting the observed relations should be construed neither as success nor failure for our model – the equilibrium relations (equations 2.31, 2.32, 2.33) have enough free parameters that fitting the observations is not challenging. However it is encouraging that relatively little tuning was required for reasonable fits, and that we end up with models with a reasonable physical basis (energy-driven winds for $\eta \propto M_h^{-2/3}$, long depletion times at low masses owing to low H_2 fractions, η of order unity and t_{dep} of order 3 Gyr at high masses).

2.4.3 Information in the scatter

Now that we have a reasonable fit for the redshift zero median relations, we can return to our main goal – understanding the higher-order relationships in the data. In particular, we would like to understand the scatter in the MS and the MZR, and the (negative) slope implied by the FMR of metallicity with respect to star formation rate at fixed stellar mass. Now that we have synthetic data, we can follow a simple procedure to fit a synthetic MS, MZR, and FMR. For the former two, we simply fit a linear model with least-squares regression to (the log base 10) of SFR vs M_* and Z vs M_*. For each synthetic galaxy, we can then subtract off the linear fit for SFR or Z at that galaxy's M_* to find its residual. Finally, the scatter is calculated as the sample standard deviation of the residuals.

Meanwhile, we fit a synthetic FMR both with a linear model

$$\log_{10} Z = b_0 + b_1 \log_{10} M_* + b_2 \log_{10} \dot{M}_{\rm SF}. \tag{2.34}$$

and, as is common practice, a quadratic model[1]

$$\log_{10} Z = a_0 + a_1 m + a_2 s + a_3 m^2 + a_4 ms + a_5 s^2 \tag{2.35}$$

---

[1] Note that the quadratic fits reported in the literature are typically fits to $12 + \log_{10}(O/\mathrm{H})$, whereas here we are fitting to the metallicity. This only makes a difference in the a_0 term, and not to the slope with respect to the star formation rate at fixed stellar mass, on which we will concentrate.

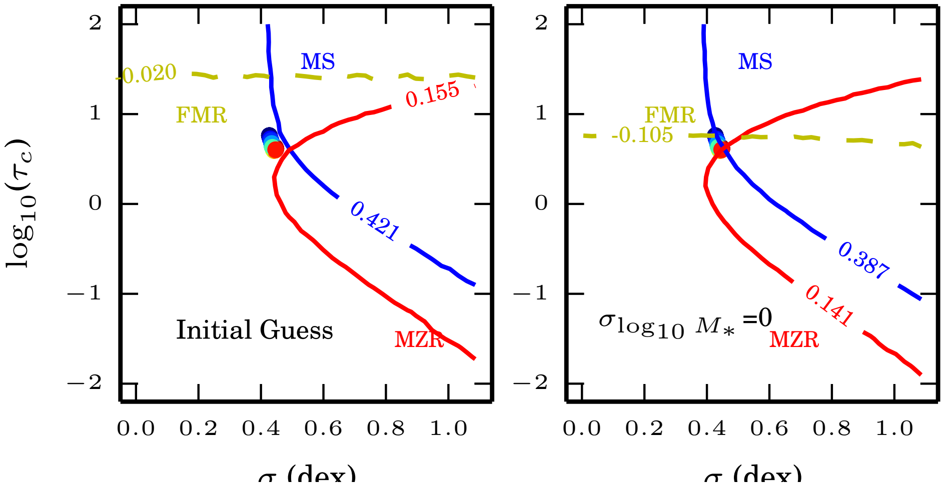

Figure 2.10: Information in the scatter of the MS and MZR. Here we show contours extracted from the SGP predictions, namely the values of τ_c and σ that we would have predicted given the synthetic scatters and FMR slope. The input values are shown as colored points near (.45,.8), where colors represent a wide range of halo mass. If we artificially reduce the scatter in M_* at fixed halo mass (right panel), the SGP predictions recover the correct input values, but including the scatter (left panel) increases the synthetic scatter and flattens the FMR slope.

Here $m = \log_{10} M_*/M_\odot - 10$ and $s = \log_{10} \dot{M}_{\rm SF}/(M_\odot\ {\rm yr}^{-1})$. From the linear fit, we can read off a value for $\partial \log_{10} Z/\partial \log_{10} \dot{M}_{\rm SF}$ at fixed M_*, namely b_2. For the quadratic fit, the slope is a function of both mass and $\dot{M}_{\rm SF}$.

For all three of these quantities – the two scatters and the (logarithmic) slope of Z vs $\dot{M}_{\rm SF}$ at fixed stellar mass, we can compare both to observations and to predictions from the full joint distribution of Ψ and $Z^\dagger$ by a single SGP (figures 2.3, 2.6, and 2.8). The scatters in the MS, MZR, and FMR are not identical to that in a SGP because the relationship between stellar and halo mass also has some scatter. Thus a population of galaxies at fixed M_* represents a weighted sum of SGPs.

Thus there will be additional scatter in these synthetic observations, as compared to the SGP predictions. This is illustrated in figure 2.10 – we compare the input values of τ_c and σ (the colored points, nearly on top of each other) to contours extracted from the heatmaps. In particular, given the synthetic scatters and FMR slope, we can read off from the SGP figures values of τ_c and σ which are consistent with these synthetically-observed values. When we artificially set $\sigma_{\log_{10} M_*} = 0$, these contours converge at the input values, but including the scatter in M_* flattens the FMR slope and increases the scatter in the MS and MZR, leaving a general region in τ_c - σ space where the contours are close to each other, but no trivial way to recover the input values.

This at once shows both the promise and the difficulty of using observed second order relations to predict these parameters. The constraints are largely independent of each other, so one could hope to nontrivially constrain the acceptable values of σ and τ_c. However, additional sources of scatter not included in the simple dimensionless SGP predictions can make it difficult to recover the values of these parameters simply by reading off where the contours intersect in this diagram.

Nonetheless, this diagram (figure 2.10) and the associated SGP predictions (figures 2.3, 2.6, and 2.8) provide some heuristic guidance. We see that in this parameter space, the input values of σ and τ_c are nearly independent of halo mass. This means that the fiducial model would predict no change in the scatters of the MS or MZR, nor any change in the slope of the FMR, which is roughly consistent with observations. We also see that to reduce the synthetic scatter to below the observed scatter, $\lesssim \pm 0.34$ dex (Whitaker et al., 2012) for the MS and $\lesssim \pm 0.1$ dex (Kewley & Ellison, 2008) (both of which may be regarded as upper limits on the intrinsic scatters), we could reduce the input accretion scatter, σ, or dramatically reduce the coherence time (and therefore τ_c). Our "Improved Guess" model adjusts the initial guess to reduce the two scatters and steepen the FMR slope. In particular we reduce σ by a factor of two and $\Delta\omega$ in our estimate of $t_{\text{coherence}}$ to 0.25 (see next section).

2.5 Discussion

In the previous section we set up a fiducial set of assumptions to map the stationary galactic populations of sections 2.2 and 2.3 into observable parameters. We were easily able to match the first-order relations, but our first guess produced scatters in our synthetic MS and MZR that were too large. In this section we examine in more detail the full range of SGP-based models that are consistent with the observed constraints on the scatters in the MS and MZR and the slope of the FMR, and the range of halo masses and redshifts over which SGP-based models are valid in principle.

2.5.1 A more general model – do all galaxies at a fixed M_h correspond to one SGP?

The analysis of SGPs in sections 2.2 and 2.3 explicitly assumes that a given galaxy has had the same values of μ, σ, t_{loss}, $t_{\text{coherence}}$,Z_{IGM}, and q for eternity. This is clearly false – galaxies increase their mass over time, moving them along any presumed scaling relations in e.g. t_{dep} or η, while other quantities likely depend explicitly on time, e.g. μ and $t_{\text{coherence}}$. The statistical equilibrium model we have proposed here, and other simpler models, may still be successful in describing galaxies because these quantities plausibly vary slowly relative to the internal timescales of the galaxy, i.e. the loss time. Whereas the typical equilibrium model assumes this of the accretion rate, our model relaxes that particular assumption and allows the accretion rate to vary, possibly very quickly, relative to other timescales.

Our model was constructed with the goal of understanding the scatter in galaxy scaling relations by examining the role of a known (and significant) scatter in dark matter accretion rates among galaxies at a given mass. However, it is also plausible that the mass loading factor, the depletion time, or some other quantity may vary between galaxies near a fixed mass, or within a given galaxy on relatively short timescales. The former situation may be handled by our model by having multiple SGPs with different values of e.g. η at the same mass. The latter situation cannot be handled by SGPs as we have formulated them.

The scenario in which scatter in η, t_{dep}, or μ is responsible for the scatter in galaxy scaling relations has several distinct predictions compared to the stochastic accretion model we have presented. In particular, the scatter, rather than being stochastic, would be constructed from several nearly parallel, slightly offset, equilibrium relations. One could likely find acceptable values for the scatter in the SGP parameters which reproduced the observed scatters, since each equilibrium relation depends on a different combination of the SGP parameters. Thus one might expect to be able to have e^{μ} vary by ~ 0.34 dex, and Z_{IGM} to vary by ~ 0.1 dex.

This model would indeed produce, more or less, the observed scatters in the MS and MZR, but it would not account for the decreasing metallicity with increasing

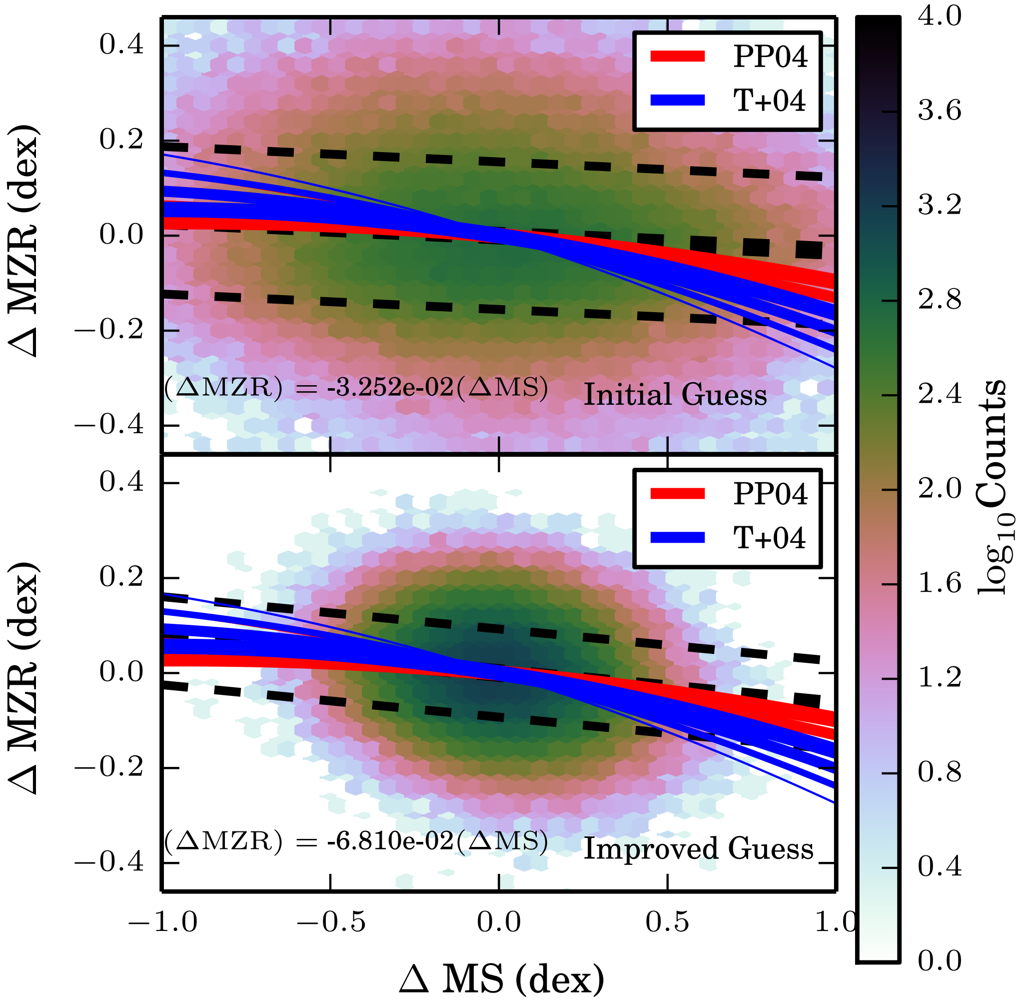

Figure 2.11: Displacements from first-order scaling relations. Here we show the offset of galaxies, in both our initial guess and improved guess models, from fits to their Main Sequences and Mass-Metallicity Relations. The black lines show a linear fit to these data points, with a 1σ. The colored lines show predictions of these quantities from two different fits to the $z = 0$ FMR using different metallicity calibrations, with thicker lines corresponding to larger values of M_*. Here we can see the slight, but significant observed and predicted anti-correlation between star formation rate and metallicity. Note that to fill out this histogram we drew a sample of 100 times the number of galaxies shown in figure 2.9

star formation rate at fixed stellar mass (see figure 2.11), i.e. what we call the 'slope' of the FMR. In particular, the equilibrium metallicity is independent of the accretion rate, and the star formation rate is independent of Z_{IGM} (at least in this simple model), so there would be no slope in the FMR. The situation gets even worse with a scatter in η, since both the equilibrium SFR and Z are inversely related to η, which would tend to create a positive slope in the FMR. Similarly, a scatter in M_* at fixed halo mass tends to induce a positive slope – at a given M_*, galaxies from higher halo mass and lower halo mass will be present, and because both the MS and MZR have positive slopes, the higher (lower) halo mass galaxies will have higher (lower) SFRs and Zs, again leading to a positive slope in the FMR.

Given the difficulty in obtaining a negative FMR by adding scatters in the parameters which enter the equilibrium relations, compared with the natural way the negative slope arises in our statistical equilibrium model, via a time-varying accretion rate, it certainly seems that no alternative model is needed. In fact, by enforcing the requirement that $\partial \log Z / \partial \log \dot{M}_{\rm SF} < 0$, we may be able to obtain limits on the scatter in parameters (M_*, η) which tend to make the slope in the FMR positive.

To accomplish this, we set up a 6-dimensional grid of models. Each point in the grid corresponds to a choice of σ, $\log_{10} \Delta\omega$, scatter in M_*, scatter in $t_{\rm dep}$, scatter in η, and scatter in e^{μ}. For each point in the grid, we simulate a full set of galaxies – 200 per value of $M_h < 10^{12.3} M_\odot$, and compute the scatter in the MS and MZR and the slope in Z vs SFR at fixed M_*. We then compare each of these pieces of information to the observations in a maximally conservative way. We treat the observed scatters in the

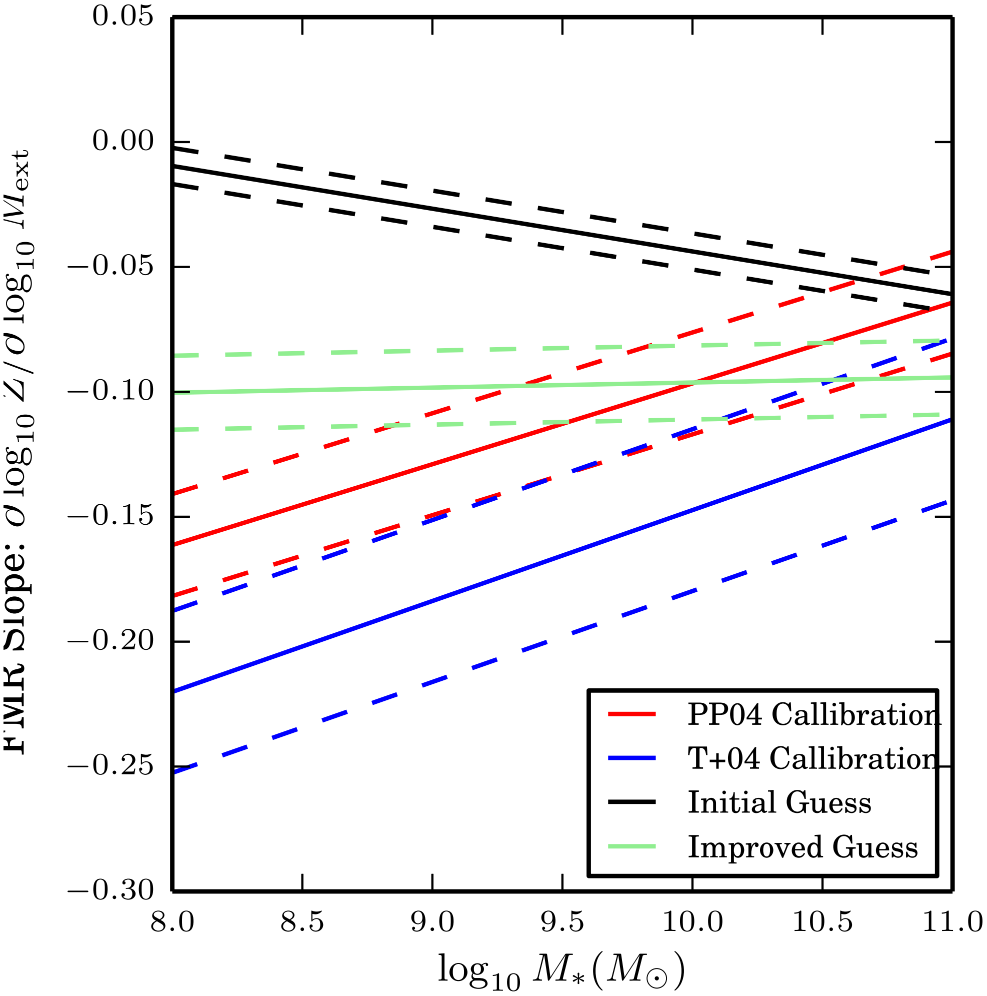

Figure 2.12: Slope in the FMR. For both our improved and initial guesses, as well as the two fits to the $z = 0$ FMR using different metallicity calibrations, we show the slope in the metallicity with respect to the star formation rate, as a function of stellar mass. Since the FMR fits depend explicitly on $\dot{M}_{\rm SF}^2$, we must also choose a star formation rate at which to evaluate this quantity. We choose the main sequence value at that mass (as determined by our fit to the main sequence of our initial guess model), which we plot as the solid lines, and ±0.3 dex, the dashed lines. The two different calibrations predict substantially different values, so to be maximally conservative we have chosen to interpret the observational constraint as $\partial \log Z / \partial \log \dot{M}_{\rm SF} < 0$, a fact on which both observed relations agree.

MS and MZR as upper limits on the intrinsic scatters, and make no assumptions about the purely observational scatter. Although there is an observationally known value of the slope of Z vs SFR at fixed M_*, we make no strong assumptions about the probability distribution function of that parameter – we merely require that it is negative (see figure 2.12). Thus for each point in the grid, we can say how many constraints that model violates: 0, 1, 2, or 3.

In figure 2.13, we project these scores, again in a maximally conservative way. For each point in the 2-d projection, we look up all models in the full 6-d space which have the 2 values under consideration in our projection, and we find the model which violated the fewest constraints. Thus the figure shows the minimum number of constraints violated by *any* model with that combination of values. If any model with those coordinates is allowed by the constraints, the pixel is shown in blue. Each darker shade of red means every model with those coordinates violated at least one more constraint, up to all 3.

This exercise demonstrates that even this conservative interpretation of the observed scatters as upper limits, combined with the weak requirement that Z decrease with increasing SFR at fixed stellar mass, yields non-trivial constraints on the parameters. In particular, $\sigma \lesssim 0.35$ dex, smaller than in our fiducial model. This may point to a smoothing out of the baryonic accretion rate relative to the scatter in dark matter accretion rates implied by the Neistein et al. (2010) formula. Perhaps even more interesting is that there is a minimum σ implied by our observational restrictions, $\sigma \gtrsim 0.1$ dex, which comes from the requirement that the FMR have a negative slope. In par-

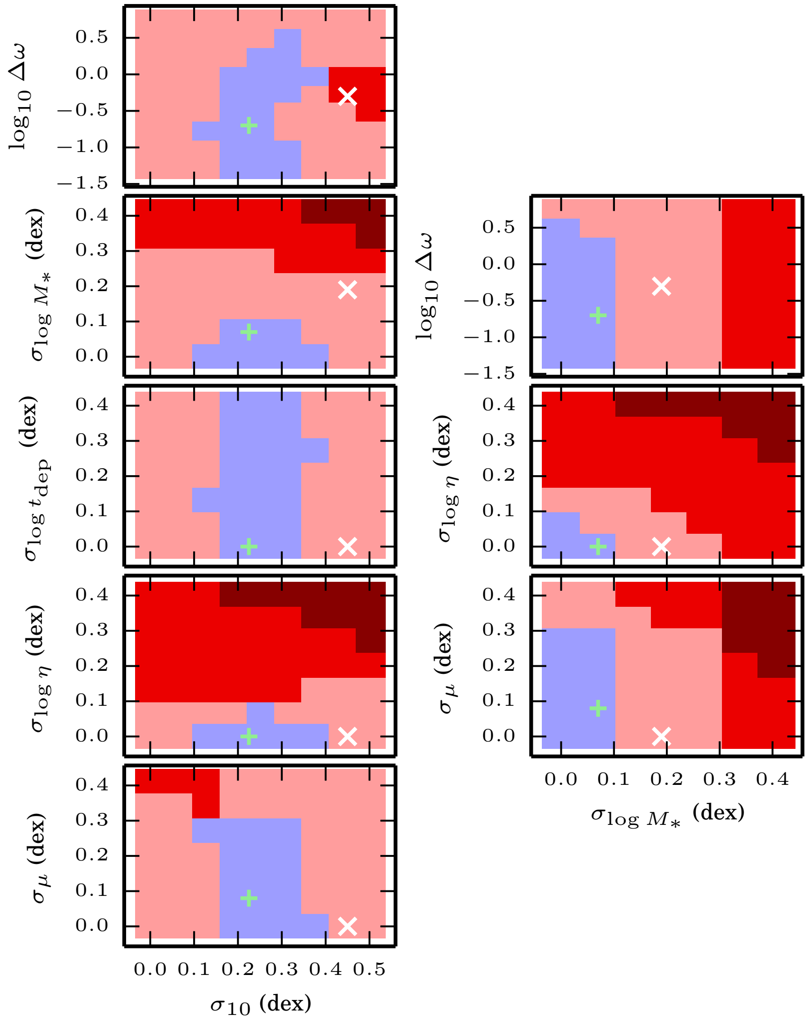

Figure 2.13: The set of parameters conservatively allowed by the observations. In addition to the parameters of the accretion process (σ_{10}, and $\Delta\omega$) we include a variable log-normal scatter in η, M_*, $t_{\rm dep}$, and the median accretion rate e^{μ}. These log-normal scatters have medians equal to the values used in section 2.4.2 to fit the first order relations. Blue pixels indicate that at least one model with that pair of parameters is consistent with the data. Each darker shade of red means the model which violates the fewest constraints for that pair of parameters violates one more constraint, up to all three. The white cross shows our initial guess, while the green '+' shows our "Improved Guess".

ticular, if σ is too small, the subtle feature in the $M_*(M_h)$ relation causes both the star formation rate and metallicity of galaxies with $M_* \sim 10^9 M_\odot$ to be higher than the MS and MZR, and galaxies at other masses to be below those relations, generating a positive correlation between SFR and Z.

The scatter in M_* for a given SGP must be $\lesssim 0.15$ dex. This is a bit at odds with the observational constraints from (Reddick et al., 2013) (0.20 ± 0.03 dex at fixed maximum circular velocity), although we note that our constraint is on scatter in M_* that is *uncorrelated* with everything else in the SGP, whereas in reality it is quite plausible (and in fact predicted by the SGP – see appendix 2.7) that, at a fixed halo mass, M_* is correlated with both SFR and Z.

Another interesting constraint is that the scatter in η must be $\lesssim 0.1$ dex. This is surprisingly small, considering the great deal of theoretical uncertainty as to the actual values and scalings of η in the first place. In our models, this comes from the aforementioned effect that in the equilibrium relations, both Z and $\dot{M}_{\rm SF}$ are inversely related to η, so scatter in η tends to reverse the negative slope in the FMR.

Unsurprisingly, there is virtually no constraint on the scatter in t_{dep}. This is simply because the equilibrium relations for $\dot{M}_{\rm SF}$ and Z are independent of $t_{\rm dep}$ – to constrain this scatter one would need constraints on the scatter in the M_* - M_g/M_* relation, although if such galaxies also had SFR measurements, they would have directly measured depletion times anyway. There is also relatively little constraint on scatter in the median accretion rate, e^μ. Essentially this is because, in the equilibrium relations, only the star formation rate is affected by this scatter, so as long as the scatter in e^μ is smaller than the scatter in the main sequence, there is no problem.

With these constraints in mind, we have altered our initial guess, simply by reducing σ, $\Delta\omega$, and $\sigma_{\log_{10} M_*}$, and slightly increasing σ_μ (see table 2.2). We label this model the "Improved Guess" model. Purely for demonstration, we have also computed scores for the grid of models where we include not only an upper limit on the scatters, but also a much narrower range of acceptable slopes of the FMR, namely $-2 < \partial \log Z / \partial \log \dot{M}_{\rm SF} < -1$. With these stronger restrictions, we get the projections shown in figure 2.14. Our "Improved Guess" model is engineered to adhere to this much stronger constraint, though there are plenty of models which would be ruled out by this strict scoring that are still consistent with the observations. Unsurprisingly the stronger constraint dramatically narrows the range of acceptable models in most of the projections. Particularly striking is that allowed range of $\Delta\omega \sim \Delta z$, the interval in redshift over which galaxies have constant accretion rates in our model, is narrowed substantially from $\Delta\omega \lesssim 3$ to $\Delta\omega \lesssim 0.4$

2.5.2 Domain of applicability

Under what circumstances might a real population of galaxies be in statistical equilibrium? We know that for a constant accretion rate, the star formation rate and the metallicity will equilibrate on the mass loss timescale. A standard equilibrium model therefore requires that $t_{\rm loss}$ be much less than the timescale on which any parameter entering into the equilibrium relations, namely q (i.e., η, ξ, and f_R), Z_{IGM}, t_{dep} and $\dot{M}_{\rm ext}$. The success of these models in understanding the first-order trends in galaxy scaling relations suggests that these requirements, while seemingly numerous, are at

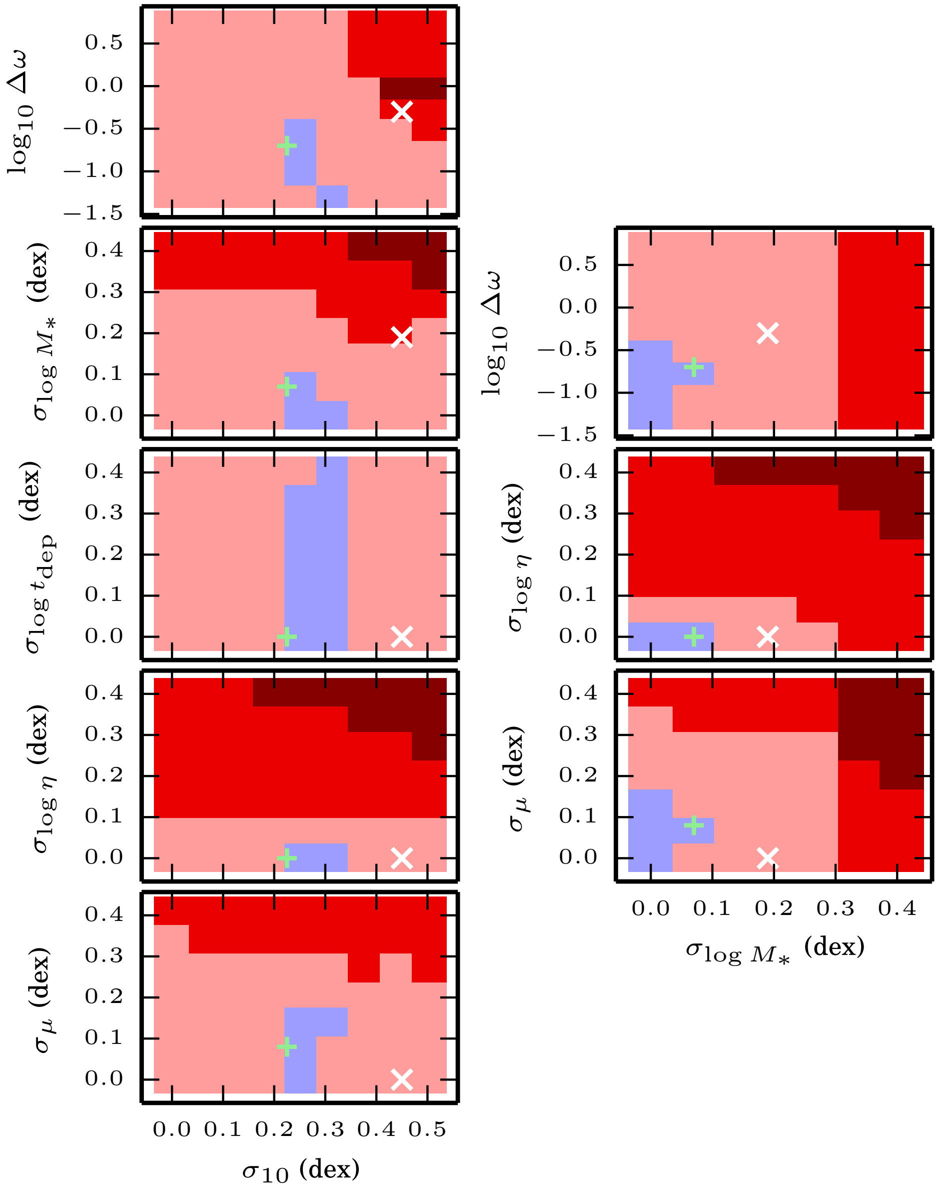

Figure 2.14: A plausible set of constraints on model parameters. Here we make the plausible but uncertain assumption that the FMR slope, $\partial \log Z/\partial \log \dot{M}_{\rm SF}$, is between -2 and -1, in addition to the constraints on the widths of the MS and MZR. As one might expect, narrowing the allowed range of FMR slopes dramatically reduces the allowed regions of parameter space. One should not take these regions to be genuine constraints, but rather to demonstrate the power of the slope of the FMR in constraining these parameters.

least marginally satisfied.

Our statistical equilibrium model relaxes one of these restrictions by splitting $\dot{M}_{\rm ext}$ into a (hopefully) slowly-evolving median e^{μ} and a (potentially) rapidly varying stochastic component $e^{\sigma x(t)}$. Our formulation of this component introduces two timescales, $t_{\rm coherence}$ – the time between new draws from the lognormal distribution, and $\sigma t_{\rm loss}$ – the time for a 1-σ accretion event to be forgotten by the galaxy.

Figures 2.2 and 2.5 show graphically the exponential suppression of old draws of the accretion distribution in their influence on the full distribution of Ψ and $Z^{\dagger}$. In logarithmic space, the separation between the centers of the distribution of each draw is just $\tau_c = t_{\rm coherence}/t_{loss}$, while the width of each distribution is σ. When $\sigma \lesssim \tau_c$, the distributions are well-separated, and we conclude that galaxies may be in statistical equilibrium so long as $t_{\rm loss}$ is appreciably less than the timescale on which any parameters of the SGP change (explicitly μ, σ, $t_{\rm coherence}$, $t_{\rm loss}$, Z_{IGM}, and q). Note that $t_{\rm coherence}$ itself may well be shorter than or comparable to $t_{\rm loss}$.

When $\sigma \gtrsim \tau_c$, the contributions from previous draws begin to matter significantly for the distribution of Ψ. In this case, the number of draws which are important increase from ~ 1 to $\sim \sigma/\tau_c$, so rather than $t_{\rm loss}$ being short, we need $\sigma t_{\rm loss}$ to be short. At least in our initial guess, shown in the previous section, this is a minor effect since $\sigma \sim 1$. Therefore the region of M_h-z space where the statistical equilibrium model is valid should be comparable to the region where an ordinary equilibrium model is capable of reproducing the first-order galactic scaling relations, which in turn is set by the scaling of $t_{\rm loss}$ with halo mass and redshift.

In our fiducial model presented in the previous section, star-forming galaxies at every halo mass were in equilibrium at $z = 0$. As we mentioned, this is not necessarily the case in the real universe - the mass loading factor may well scale weakly with halo mass, in which case low-mass galaxies, with their long depletion times, would be unable to equilibrate to their baryonic accretion rate even in a Hubble time. Either scenario is currently perfectly consistent with observations, since the mass loss timescale is unknown, owing to its dependence on η. In figure 2.15 we show regions of M_h-z space where the equilibrium assumptions are valid – the bluer the color, the better-satisfied the condition that the Hubble time be much longer than t_{eq}, the maximum of the mass-loss time ($t_{\rm loss}$), the coherence time ($t_{\rm coherence}$), and the burn-through time ($\sigma t_{\rm loss}$).

To construct these diagrams, we also need to make an assumption regarding how input parameters vary not only with halo mass, but with redshift. We assume that the mass loading factor and the factor by which we reduce the efficiency, f_ϵ, are independent of redshift, but that the depletion time scales as $t_{dep} \propto (1+z)^{-1}$, a somewhat weaker scaling than if the depletion time scaled with the dynamical time (Davé et al., 2012). This scaling is consistent with recent observational results from CO observations at high redshift (Tacconi et al., 2010, 2013; Saintonge et al., 2013), though of course there are large uncertainties. Moreover, these observations span a very limited range of mass and redshift compared to that shown in these diagrams. We therefore emphasize that these plots represent plausible assumptions, not predictions.

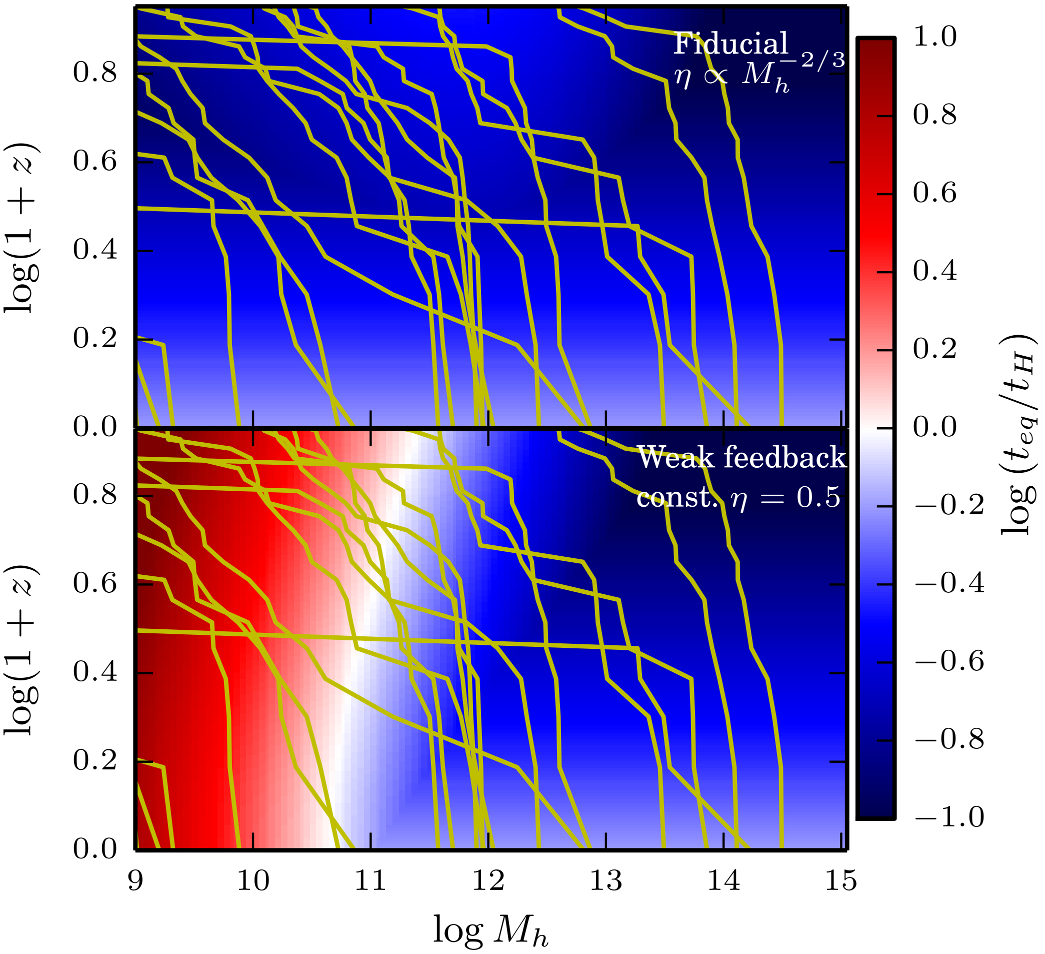

Figure 2.15: The validity of the statistical equilibrium model. Here we show the ratio of t_{eq} to the instantaneous Hubble time, as a function of halo mass and redshift. The bluer the color, the more satisfied the condition that $t_{eq} < t_H$, required for the statistical equilibrium model to be valid. For reference, we overplot the trajectories of 20 haloes as they grow stochastically, as calculated by equations 2.25, 2.26, and 2.27 with $\Delta\omega = 0.5$.

2.5.3 Evolution with redshift

Since it is plausible that a statistical equilibrium model may be used successfully at higher redshift, one may consider extending our analysis beyond the $z \approx 0.1$ data we have considered. There are however both theoretical and observational challenges. First, a great deal of poorly-justified assumptions are required to scale the fiducial model to higher redshift. Observationally, while there are measurements of the MS, MZR, and FMR at higher redshift, the metallicity measurements in particular are fraught with complications arising from the changing set of lines visible from ground-based telescopes and the uncertainty of converting line characteristics to physical metallicities in the substantially different environments of high-redshift galaxies. The high-order quantities we discuss in this paper are therefore both difficult to predict and measure.

We can nonetheless make the basic point under the assumption that the star formation rate and metallicity have reached their equilibrium values, the metallicity will be $Z_{eq} = Z_{IGM} + q$, where q is a combination of f_R, η, and ξ, none of which are expected to change dramatically with redshift. Thus the MZR should not change with redshift if galaxies are in equilibrium. Observational studies tend not to find this result. However given the difficulty in calibrating the zero-point of the MZR even at low redshift, and the possibility that samples of high-redshift galaxies are biased high in SFR and therefore low in Z (Stott et al., 2013), we do not believe equilibrium models have been ruled out at higher redshifts.

Similarly, the slope of the FMR, which one may read off from the quadratic fit to the FMR (equation 2.35),

$$\partial \log Z / \partial \log \dot{M}_{\rm SF} = a_2 + a_4 m + 2a_5 s \tag{2.36}$$

should be independent of the star formation rate. We therefore predict that in general $a_5 \approx 0$, although since m and s are highly correlated, the fact that many best fit models do not yield $a_5 = 0$ is not necessarily an indication that this prediction is wrong.

We can also describe how the fiducial model, which does not quite fit the data at $z = 0$, would scale to higher redshift. The predicted value of σ is not dependent on redshift, at least for the dark matter. Given our results suggesting that baryonic processes likely play an important role in smoothing the accretion, it is unclear how this smoothing process would evolve, so this constant value of σ is highly uncertain. Meanwhile $t_{\rm coherence}$ will evolve strongly with redshift, but the quantity which sets the second-order scatter we consider is $\tau_c = t_{\rm coherence}/t_{\rm loss}$. If we assume η at fixed halo mass does not change, then $t_{\rm loss}$ simply scales as the depletion time, which likely does decrease significantly towards higher redshift. If they decrease at a similar rate, near $1/(1+z)$, then τ_c is unlikely to evolve very much. In this scenario, we would predict that the scatter in the MS and MZR, and the slope in the FMR at fixed stellar mass would be roughly independent of redshift (and as evidenced in figure 2.10, independent of halo mass).

2.5.4 Relationship to other work

Our model bears a resemblance to several recent papers on equilibrium models (Davé et al., 2012; Lilly et al., 2013). Our model reduces to a slightly simpler version of the Davé et al. (2012) model when $\tau_c \to \infty$ and $\sigma \to 0$, i.e. the upper left of the $\sigma - \tau_c$ diagrams we showed in sections 2.2 and 2.3.

One of the criticisms of the Davé et al. (2012) model has been its explicit assumption that $dM_g/dt = 0$. Simple toy models of the growth of galaxies under various star formation laws, (e.g. Feldmann, 2013) point out that for many galaxies at redshift zero, $dM_g/dt < 0$. Indeed in our recent work on the radially-resolved evolution of disk galaxies since $z = 2$ (Chapter 4), we find that much of the galactic disk for many galaxies tends to be moderately out of equilibrium between local sources and sinks, with star formation being somewhat higher than the (local) accretion rate.

The equilibrium model of Lilly et al. (2013) attempts to address this issue by allowing part of the incoming accretion to build up in the gas reservoir. The price they pay is that the star formation rate becomes an input to their model rather than an output. Perhaps even more worrying is that they assume $dZ/dt = 0$ always, even when $dM_g/dt \neq 0$. Note that Feldmann (2013) makes a similar argument. As shown by our sample trajectories in section 2.3, it actually takes Z longer to equilibrate than the star formation rate, since the metallicity can only equilibrate once the star formation rate catches up (or falls off) to the accretion rate. It is therefore odd that they assume the metallicity is in equilibrium while the star formation rate is not. It is only through this oddity that they are able to fit a second order relation, i.e. the FMR, with their model.

We consider our model to be both more self-consistent and more powerful, in that we can generate a scatter in the star formation rate and metallicity, and about the FMR itself.

Another important result, and to our knowledge the only previous theoretical attempt to address the scatter in the main sequence, is Dutton et al. (2010). They use a rather sophisticated semi-analytic model, including cooling from virial shock-heated gas in a dark matter halo and star formation as a function of specific angular momentum (i.e., radius) in the disk, although they do not include any way for the gas to change its angular momentum (as we do in Forbes et al., 2012; Forbes et al., 2014a, i.e. Chapters 3 and 4). They find a significant but small scatter in their model star-forming main sequence arising from variation in halo concentration, which in turn causes differences in the mass accretion histories between different galaxies of the same halo mass. Their model therefore resembles ours in the limit that $\tau_c \to \infty$, but $\sigma \neq 0$. Our model's more flexible treatment of the accretion process and other model parameters (e.g. the mass loading factor) gives us somewhat more insight on the problem of scatter not only in the main sequence but also in the MZR and FMR, although of course our model is far simpler in terms of its treatment of star formation, and we can make no predictions regarding other important and interesting quantities (Dutton, 2012, i.e. galaxy sizes and rotation curves).

2.6 Conclusion

In the past few years, a new view has emerged as a useful way of understanding galaxies. In this picture, galaxies are in a slowly evolving equilibrium between accretion, star formation, and galactic winds regulated by the mass of cold gas in their interstellar media. To the degree that the parameters controlling this balance are well-defined functions of the mass of a galaxy and its redshift, this sort of model may be used to understand the connection between galaxy scaling relations and these physical parameters, which are not known from first principles.

In the spirit of these equilibrium models, we have presented a simple model which relaxes a key assumption in the equilibrium model, namely that the rate at which baryons enter the gas reservoir varies slowly. A population of galaxies in our model has been fed by the same stochastic accretion process for eternity, or at least long enough that the full joint distribution of all galaxy properties has become time-invariant. We therefore refer to our picture as a statistical equilibrium model, since the individual galaxies are not in equilibrium, but the population is.

With this model, we study a number of second-order relationships about the well-known galaxy scaling relations between the stellar mass and the star formation rate (the star-forming main sequence), and the stellar mass and metallicity (the mass metallicity relation). We look at the scatter at fixed stellar mass in both of these quantities, as well as the (anti-) correlation between star formation and metallicity at fixed stellar mass. Our main conclusions are as follows.

1. Including a stochastic scatter in the accretion rate at the level expected from

N-body cosmological simulations naturally produces a scatter in both the star forming main sequence and mass-metallicity relation somewhat larger than the observed scatters. The anti-correlation observed between star formation rates and metallicities at fixed stellar mass is also naturally reproduced.

2. Neglecting the scatter in model parameters, (i.e., the mass loading factor, the depletion time, the scatter in stellar mass at fixed halo mass, etc.) all second-order quantities (the scatter in the main sequence, the scatter in the mass metallicity relation, and the slope in metallicity with respect to the star formation rate at fixed stellar mass) are determined by only two parameters: the scatter in the accretion rate, and the ratio of the timescale on which the accretion varies to the timescale on which the galaxy loses gas mass.

3. Using a maximally conservative interpretation of the available data, we are able to constrain these two parameters as well as a number of "nuisance" parameters, namely the scatter in the mass loading factor at fixed halo mass and the uncorrelated scatter in M_* at fixed halo mass. We find that the log-normal scatter in the baryonic accretion rate is between about 0.1 and 0.4 dex, moderately smaller than what we would have predicted based on N-body simulations and assuming the baryons follow the dark matter. This may point to some process in the halos of galaxies which smooths out variations in the baryonic accretion rate, or a substantial amount of baryon cycling, which has the effect of averaging out the accretion rate over a longer time period. We find that the scatter in the mass loading factor is less than 0.1 dex, remarkably small considering the theoretical uncertainty in

the details of the physics of feedback. Our constraint on the timescale over which the accretion rate varies is much weaker, but could be narrowed considerably by stronger constraints on the Fundamental Metallicity Relation.

We hope that the framework we have presented here motivates new development in both theory and observation. On the theory side, $t_{\mathrm{coherence}}$ (or a more sophisticated quantity describing the timescales on which accretion varies) and σ may be measured with some confidence in both dark matter-only and baryonic cosmological simulations. Semi-analytic models may be altered to include the appropriate level of variability in baryonic accretion rates. Meanwhile, we have shown that observable quantities, e.g., the scatter in the star-forming main sequence, can provide significant constraints on properties of the baryonic accretion process and the galaxy-to-galaxy variability of the mass loading factor. Our inferences are, however, limited by our limited certainty on the intrinsic scatter in the scaling relations we have considered and the true parameters of the FMR. Pinning down these quantities observationally at a variety of masses and redshifts may substantially improve our understanding of the details of baryonic accretion and feedback.

2.7 Details of the Monte Carlo Simulations

Throughout this paper we have presented heatmaps of various quantities as a function of $\tau_c = t_{\mathrm{coherence}}/t_{\mathrm{loss}}$, and $\sigma_{10} = \sigma \log_{10}(e)$. Computing each of these quantities for the model is typically a non-trivial task which requires a monte carlo simulation, in which a large ensemble of galaxies is sampled at random times to sample the underlying true distribution of the quantity in question for galaxies in this model. Here we describe the details of these simulations.

For each pixel in these grids of τ_c vs σ_{10}, we sample an ensemble of 30,000 galaxies. Each galaxy is started at a time $\tau = 0$ with initial values $\Psi = Z^\dagger = 1$, the equilibrium values for those quantities in the limit $\sigma \to 0$ and $\tau_c \to \infty$. The galaxies are then evolved for long enough that, for all practical purposes they forget their initial conditions (formally our model assumes that the galaxy population has been undergoing the same stochastic accretion process for eternity, but this is obviously impractical computationally). To determine 'long enough', we use the analytic results derived in sections 2.2 and 2.3 which show that galaxies forget their initial conditions with an e-folding time of t_{loss}. We also note that for our distribution to represent the true long-term steady-state distribution, as discussed in section 2.5.2, galaxies with large scatters in their accretion rate need to experience enough draws from the accretion rate distribution that even the tail of the probability distribution of past events has no influence on the present distribution.

We therefore define a timescale $\tau_{\mathrm{long}} = 1 + \tau_c + \sigma$, i.e. a time guaranteed to be of order the longest timescale in the problem for any choice of τ_c and σ. We then calculate the number of draws from the accretion distribution necessary to simulate each galaxy out to $15\tau_{\mathrm{long}}$, namely $k = 15\tau_{\mathrm{long}}/\tau_c$. We then draw a pseudo-random number uniformly distributed between 0 and 1, and compute the galaxy's properties (Ψ, $Z^\dagger$) at the time $\tau_{\mathrm{obs}} = (k + r)\tau_c$, where r is the random number. This samples the full distribution of these quantities for the population of galaxies in steady state. In general the computational cost is just proportional to k, since to compute Ψ and $Z^\dagger$, we must first draw k random numbers and compute the sequence $\Psi_1, ..., \Psi_k$ and $Z_1^\dagger, ..., Z_k^\dagger$ before we can calculate those quantities at $\tau_{\rm obs}$. Thus the models with short coherence times, i.e. $\tau_c \ll 1$, are the most expensive.

With the 30,000 samples for each of the 41 by 41 points in the grid of σ and τ_c, we can then compute each of the quantities shown in this paper – the standard deviation of $Z^\dagger$ and Ψ (independently), the correlation between the two quantities, and the linear slope. We also record a quantity which may be regarded as a proxy for stellar mass, defined as

$$\mathcal{M}_* = \int_0^{\tau_{\rm obs}} \Psi(\tau) d\tau \tag{2.37}$$

Naturally the magnitude of this quantity is, on average, proportional to the amount of time we let the simulations run, which is chosen subjectively to be $\gtrsim 15\tau_{\rm long}$. However, one may expect that any statistical properties which remove the mean may be physically relevant. In figures 2.16 and 2.17 we show the correlation between our analogs to star formation and metallicity, and stellar mass. Clearly, over much of parameter space there is a small but appreciable correlation between each quantity and $\mathcal{M}_*$, meaning that the tension between our result that $\sigma_{\log M_*} \lesssim 0.15$ dex and the observational constraint that the scatter in stellar mass at fixed halo mass be 0.19 dex (Reddick et al., 2013), is not a large concern. This is because our constraint is on scatter in M_* that is uncorrelated with any other quantity, whereas in reality, as in the SGP, the stellar mass may well be positively correlated with the quantities in question, in which case at least some of the scatter in M_* will be along the first-order scaling relations, and therefore won't

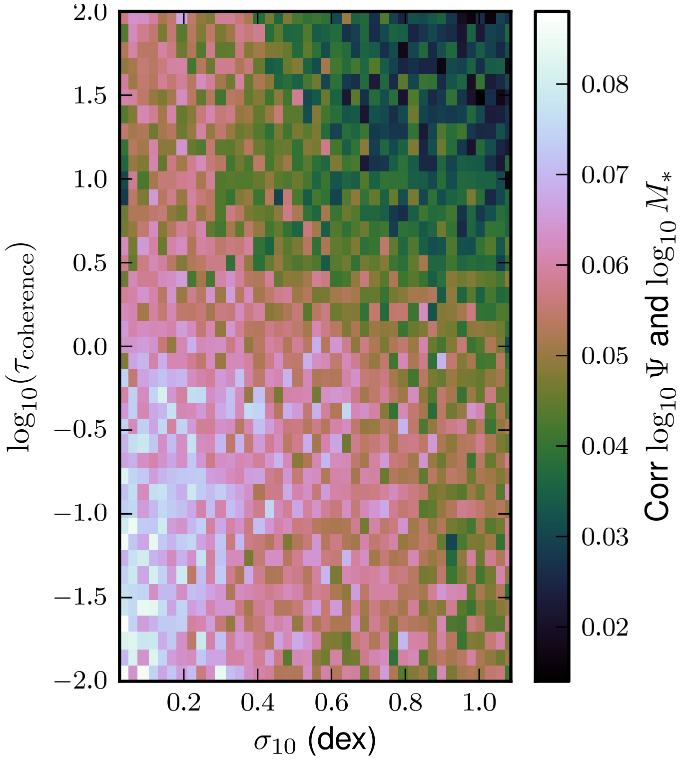

Figure 2.16: The correlation between SGP versions of the star formation rate and stellar mass. The non-zero correlation shows that scatter in stellar mass at fixed halo mass can drive galaxies along the MS, rather than being merely uncorrelated.

contribute to the scatter in the relation at fixed stellar mass.

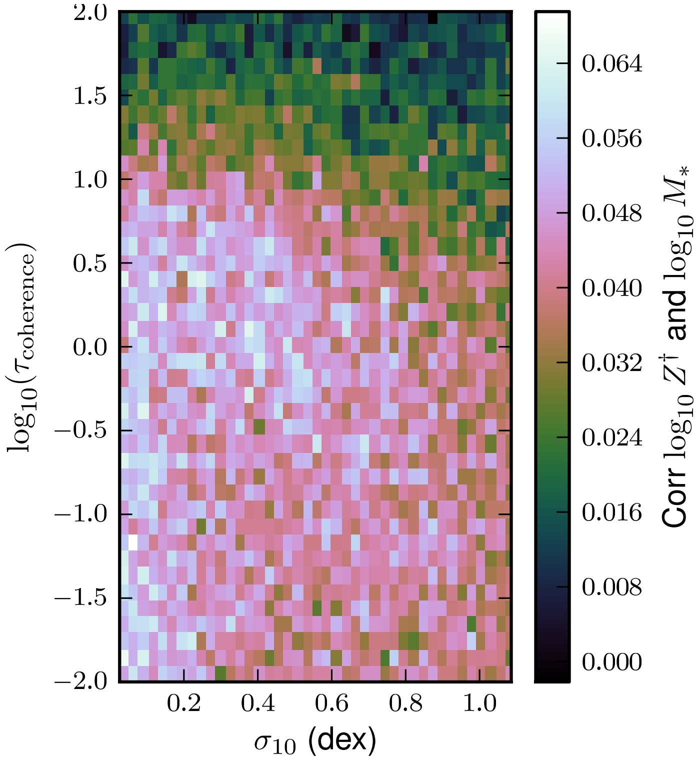

Figure 2.17: The correlation between SGP versions of the metallicity and the stellar mass. Just as with the star formation rate, the correlation is positive everywhere, meaning again that scatter in stellar mass at fixed halo mass scatters galaxies along the MZR.

Chapter 3

The Age Velocity Dispersion Relation

3.1 Introduction

In the past decade, observations of galaxies near $z \sim 2$ have revealed compelling evidence for the importance of gravitational instability in their dynamics and evolution. A whole class of galaxies has been observed whose images are dominated by large luminous clumps of gas (Elmegreen et al., 2004, 2005; Förster Schreiber et al., 2009), while measurements of the velocity dispersion of such massive star-forming galaxies have found values near 50 km/s spread across the entire disk (Cresci et al., 2009; Genzel et al., 2011). This is difficult to reproduce with supernova feedback, which is strongest near the centers of galaxies where the star formation rate peaks, and which is only strong enough to drive velocity dispersions of ~ 10 km/s (Joung et al., 2009).

Other forms of stellar feedback may drive turbulence (Thompson et al., 2005; Elmegreen & Burkert, 2010), but we will concentrate on the case where turbulence is driven by gravitational instability in the disk.

To a first approximation, the gravitational stability of a thin disk to axisymmetric perturbations is described by Toomre's Q parameter $Q = \kappa\sigma/(\pi G \Sigma)$, where κ is the epicyclic frequency, σ is the 1d velocity dispersion, and Σ is the gas surface density. The disk is unstable when $Q \lesssim 1$. The importance of gravitational instability in high redshift galaxies arises from the high cosmological accretion rates they experience, which drive up the value of Σ (Dekel et al., 2009a). This instability gives rise to clumps of the sort observed at high redshift. The inhomogenous and time-varying gravitational field drives turbulence throughout the disk, regardless of the stellar density or supernova rate. The energy source for these random motions must ultimately be the gravitational potential of the galaxy, so gas is transported inwards.

Cosmological simulations with sufficiently high resolution (Bournaud & Elmegreen, 2009; Ceverino et al., 2010) successfully reproduce disks in which gravitational instability forms clumps and causes the inward migration of material through galactic disks. Simulations of isolated galaxies (Bournaud et al., 2011; Dobbs et al., 2011a,b) with initial conditions set such that $Q < 1$ provide a higher resolution view of such galaxies over a few outer orbits of the disk. These studies, while illuminating, are expensive, since they must solve the equations of hydrodynamics in three dimensions over cosmological times. The model we present here solves the hydrodynamical equations in the limit of a thin axisymmetric disk. Since quantities vary only in the radial direction, the problem is computationally much cheaper to solve, allowing us to explore parameter space efficiently, while still solving the full 1D equations of fluid dynamics instead of relying entirely on semi-analytic models (Dekel et al., 2009a; Cacciato et al., 2012)

Past 1D models of gravitational instability in disks have a number of shortcomings. The rate at which mass and angular momentum are transported inwards is often parameterized and fit to the results of hydrodynamical simulations, rather than being derived from first principles. The rotation curves are only allowed to be either Keplerian or flat. Energy is frequently assumed to be instantaneously equilibrated, which neglects the possibility that it might be advected through the disk. The pressure support of the disk is often treated as coming from thermal pressure rather than supersonic turbulence. Few models take into account the stellar component of the disk, which becomes increasingly important as the galaxy evolves towards the present day, and can ultimately provide a variety of observable predictions.

In particular, the age-velocity dispersion-metallicity correlation of stars in the solar neighborhood (Nordström et al., 2004), might well be explained by means of gravitational instability in high redshift disks (Bournaud et al., 2009). The high velocity dispersion in these disks means that the population of stars formed in that epoch will start with a high velocity dispersion (Burkert et al., 1992). The gas disk cools as a result of slowing cosmological accretion rates, so younger stars are formed in a thinner, more metal-rich disk. This mechanism of thick and thin disk formation contrasts with the more common story that various secular processes and minor mergers heat thin disk stars into a thick disk (e.g. Binney & Tremaine, 1987).

Krumholz & Burkert (2010) (hereafter KB10) found an analytic steady-state solution to the full equations of fluid dynamics in the thin disk limit under the assumption that the disk self-regulates to maintain $Q = 1$. To make the problem tractable analytically, however, they required a handful of simplifying assumptions: they use an analytic approximation to Q, which becomes progressively worse at lower redshift as the ratio of gas to stellar velocity dispersion deviates from unity. They also assume that the velocity dispersion of stars and gas are equal, and the gas fraction at all points in the disk remains constant in radius and time. In this paper, we relax these assumptions and include treatments of stellar migration, metallicity, the non-zero thermal temperature of the gas, and evolution of individual stellar populations. These improvements along with an efficient simulation code allow us to realistically evolve disks from high redshift to the present day at minimal computational expense.

In section 2 we derive the equations governing the evolution of the gas over time. Section 3 presents the derivation of the equations governing the stellar dynamics. In section 4, we derive the evolution of metallicity in the gas and stars. In section 5 we discuss how these differential equations are solved numerically, and in section 6 we present the results for fiducial parameters chosen to be similar to the Milky Way. We conclude in section 7. The code we describe, named GIDGET for Gravitational Instability-Dominated Galaxy Evolution Tool, is available at `http://www.ucolick.org/~jforbes/gidget.html`

3.2 Gas Evolution Equations

3.2.1 Basic Equations

We first give a brief overview of the derivation of the evolution equations for the gas column density Σ and velocity dispersion σ. For more details see KB10. The equations of mass, momentum, and energy conservation for a viscous star-forming fluid in a gravitational field are

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= -\nabla \cdot (\rho \mathbf{v}) - (f_R + \mu)\dot{\rho}_*, && (3.1) \\ \rho \frac{D\mathbf{v}}{Dt} &= -\nabla P - \rho \nabla \psi + \nabla \cdot \mathbf{T}, && (3.2) \\ \rho \frac{De}{Dt} &= -P \nabla \cdot \mathbf{v} + \Phi + \Gamma - \Lambda, && (3.3) \end{aligned}$$

where ρ, $\mathbf{v}$, e, and P are the gas density, velocity, specific internal energy, and pressure respectively. The star-formation rate per unit volume at an Eulerian point is $\dot{\rho}_*$, with a mass loading factor μ equal to the ratio of gas ejected in galactic-scale winds to the star formation rate. We will be employing the instantaneous recycling approximation (see section 4.2.4), which approximates all stellar evolution as occurring immediately. Of the mass which forms stars, the gas will only lose the so-called remnant fraction, f_R, to stars, while the remaining $(1 - f_R)$ will be immediately recycled into the ISM. The gravitational potential is ψ, $\mathbf{T}$ is the viscous stress tensor, $\Phi = T^{ik}(\partial v_i / \partial x_k)$ is the rate of viscous dissipation, and Γ and Λ are the rates of radiative energy gain and loss per unit volume.

For a thin disk, we formally have $\rho = \Sigma \delta(z)$ and $v_z = 0$. By expanding the fluid equations in cylindrical coordinates, integrating over z, assuming axisymmetry and $v_r \ll v_\phi$, and dropping time derivatives of the potential and the circular velocity, we can obtain evolution equations for the gas column density and gas velocity dispersion. The evolution of column density is given by

$$\begin{aligned}
\frac{\partial \Sigma}{\partial t} &= \frac{1}{2\pi r}\frac{\partial}{\partial r}\dot{M} - (f_R + \mu)\dot{\Sigma}_*^{SF} \\
&= \frac{1}{2\pi(\beta+1) r v_\phi}\left[\frac{\beta(\beta+1) + r(\partial\beta/\partial r)}{(\beta+1)r}\left(\frac{\partial \mathcal{T}}{\partial r}\right) - \frac{\partial^2 \mathcal{T}}{\partial r^2}\right] \\
&\quad -(f_R + \mu)\dot{\Sigma}_*^{SF} \qquad (3.4)
\end{aligned}$$

where $\beta = \partial \ln v_\phi / \partial \ln r$ is the power law index of the rotation curve, $\mathcal{T} = \int 2\pi r^2 T_{r\phi} dz$ is the viscous torque, $\dot{\Sigma}_*^{SF}$ is the star formation rate per unit area, and

$$\dot{M} = -2\pi r \Sigma v_r = -\frac{1}{v_\phi(1+\beta)}\frac{\partial \mathcal{T}}{\partial r} \qquad (3.5)$$

is the mass flux. The second equality follows from the angular momentum equation, which is in turn derived from the ϕ component of the Navier-Stokes equation (equation 3.2, see KB10).

The derivation of the velocity dispersion evolution equation requires an equation of state, which we take to be $P = \rho\sigma^2$. The velocity dispersion has a thermal and a turbulent component. It is a reasonable approximation to treat both as contributing to the pressure so long as we average over scales much larger than the characteristic size of the turbulent eddies, which will be of order the disk scale height.

Taking the dot product of the velocity with the Navier-Stokes equation and adding it to the internal energy equation yields an equation for the total energy, i.e. internal energy, kinetic energy, and gravitational potential energy. By decomposing the velocity as $v^2 = v_r^2 + v_\phi^2 + 3\sigma_{nt}^2$, the kinetic plus thermal energy may be rewritten

$$\frac{1}{2}v^2 + e = \frac{1}{2}(v_r^2 + v_\phi^2) + \frac{3}{2}\sigma^2 \tag{3.6}$$

where the velocity dispersion is taken to be the quadrature sum of a thermal and non-thermal component, $\sigma^2 = \sigma_t^2 + \sigma_{nt}^2$. Neglecting the v_r^2 term as small compared to both σ^2 and v_ϕ^2 in a thin, rotation-dominated, $Q \sim 1$ disk, employing radial force balance to set $\partial\psi/\partial r = v_\phi^2/r$, assuming a constant potential to set $\partial v_\phi/\partial t = 0$, and integrating over z yields the evolution equation

$$\begin{aligned}\frac{\partial\sigma}{\partial t} \quad = \quad & \frac{\mathcal{G}-\mathcal{L}}{3\sigma\Sigma} + \frac{1}{6\pi r\Sigma}\left[(\beta-1)\frac{v_\phi}{r^2\sigma}\mathcal{T}\right. \\ & + \frac{\beta^2\sigma + \sigma(r\frac{d\beta}{dr}+\beta) - 5(\beta+1)r\frac{\partial\sigma}{\partial r}}{(\beta+1)^2 r v_\phi}\left(\frac{\partial\mathcal{T}}{\partial r}\right) \\ & \left. - \frac{\sigma}{(\beta+1)v_\phi}\left(\frac{\partial^2\mathcal{T}}{\partial r^2}\right)\right] \end{aligned} \tag{3.7}$$

To fully specify the evolution of the gas, we need to set a rotation curve, a prescription for radiative energy gain and loss per unit area, and a procedure for finding the viscous torque. This will allow us to specify v_ϕ, β, $\mathcal{G} = \int \Gamma dz$, $\mathcal{L} = \int \Lambda dz$, and $\mathcal{T}$. The rotation curve is specified at run-time, and $\mathcal{T}$ is set by our treatment of gravitational instability (see section 3.2.2). We set $\mathcal{G} = 0$, which is equivalent to requiring that the energy balance in the gas is completely determined by the effects of the viscous torque and radiative loss. We assume that the loss rate, meanwhile, is proportional to the kinetic energy density per disk scale height crossing time, in agreement with the decay rate of turbulence observed in full 3D MHD simulations of supersonic turbulence (Mac Low et al., 1998; Stone et al., 1998).

$$\mathcal{L} \equiv \frac{d}{dt}\left(\frac{3}{2}\Sigma\sigma^2\right)^{rad} = \frac{d}{dt}\left(\frac{3}{2}\Sigma\sigma_{nt}^2\right)^{rad} = \eta\frac{\Sigma\sigma_{nt}^2}{H/\sigma_{nt}} \tag{3.8}$$

where η is a free parameter of order unity. If the decay time is exactly the crossing time, $\eta = 1.5$, since the kinetic energy surface density is $(3/2)\Sigma\sigma^2$. In dropping the time derivative of σ_t, we have assumed that the thermal velocity dispersion is unaffected by radiative dissipation, i.e. that the gas is isothermal.

The scale height H is approximated by taking the solution to the equations of vertical equilibrium for a single component disk, $H_1 = \sigma^2/\pi G\Sigma$, and adopting it to multiple components:

$$H = \frac{\sigma^2}{\pi G(\Sigma + f\Sigma_*)}, \tag{3.9}$$

where f represents the relative importance of the stellar mass, or the stellar mass within a gas scale height. Taking $f = \sigma/\sigma_*$ interpolates between two extreme cases: when $\sigma/\sigma_* \ll 1$, the scale height should approach the single-component value, i.e. $f = 0$. When $\sigma \sim \sigma_*$, the two-component disk behaves (at least in terms of vertical density) like a single fluid with surface density $\Sigma + \Sigma_*$, i.e. $f = 1$. Note that the stellar scale height, which does not directly affect the dynamics of the disk, is just taken to be the single component solution,

$$H_* = \frac{\sigma_*^2}{\pi G(\Sigma + \Sigma_*)}, \tag{3.10}$$

which is reasonable for the small values of f_g found within the star-forming regions of the disk. In reality the vertical structure of a self-gravitating disk in a dark matter halo is not this simple. However, excluding the effects of dark matter introduces an error of only 13%, even in the dark-matter dominated regions of the outer disk (Narayan & Jog, 2002). Given the uncertainty in η, this approximation is adequate.

Substituting for the scale height and $\sigma_{nt}^2 = \sigma^2 - \sigma_t^2$, we obtain a radiative loss rate of

$$\mathcal{L} = \eta \Sigma \sigma^2 \kappa Q_g^{-1} \left(1 + \frac{\Sigma_* \sigma}{\Sigma \sigma_*}\right) \left(1 - \frac{\sigma_t^2}{\sigma^2}\right)^{3/2} \tag{3.11}$$

In this form, the radiative loss rate is the gas kinetic energy per dynamical time multiplied by a factor to account for the effect of stars on the disk's thickness and a factor to zero out the radiative losses when there is no turbulence. As the gas velocity dispersion falls towards the constant thermal velocity dispersion, non-thermal motions die away, the gas no longer dissipates its energy via shocks, and $\mathcal{L} \to 0$. The gas temperature used to calculate σ_t is a free parameter of the model, but fiducially we assume $T_g = 7000K$, appropriate for the warm neutral medium of the Milky Way. At high redshift when the gas is virtually all molecular, $T \ll 7000K$, but in that regime $\sigma_t/\sigma \ll 1$ anyway, even if we use the higher-than-appropriate gas temperature. The choice of σ_t therefore has virtually no effect on the high-redshift evolution of the disk.

The governing equations for the gas (equations 4.2 and 4.3) are derived under the assumption that $v_z = 0$. We therefore implicitly neglect the gravitational potential energy of the disk associated with its vertical extent, and the associated $P\,dV$ work that the gas performs when it changes its scale height. Qualitatively, the effect of including these terms would be to provide the gas with another place to store energy which it gains when falling down the galaxy's potential well, aside from turbulent motion. Thus with these effects the gas velocity dispersion would be slightly lower, and hence so would the dissipation rate, the gas column density, and the star formation rate. The dissipation rate and star formation rate are each already controlled by a free parameter which is uncertain at the factor of two level, so we are content to neglect these additional repositories of energy.

3.2.2 Gravitational Instability

The stability against gravitational collapse of a self-gravitating disk is given by a Toomre Q-like parameter. Several such fragmentation conditions exist in the literature. We adopt a modified version of the condition determined by Rafikov (2001), wherein the stability of a multi-component disk is considered with the stars treated realistically as a collisionless fluid.

$$Q(q)^{-1} = Q_g^{-1} \frac{2q}{1+q^2} + \sum_i \left[Q_{*,i}^{-1} \frac{2}{q\phi_i} \left(1 - e^{-q^2\phi_i^2} I_0(q^2\phi_i^2) \right) \right] \tag{3.12}$$

where i indexes an arbitrary number of stellar populations, ϕ_i is the ratio of the ith stellar population's velocity dispersion to the gas velocity dispersion, $I_0(x)$ is a modified Bessel function of the first kind, and the Q parameter for each component is defined by

$$Q_j = \frac{\kappa \sigma_j}{\pi G \Sigma_j}. \tag{3.13}$$

The epicyclic frequency is $\kappa = \sqrt{2(\beta+1)}\Omega$, and $q = k\sigma/\kappa$ is the dimensionless wavenumber, where k is the dimensional wavenumber of the perturbation. Values of q, or equivalently k, for which $Q(q) < 1$ are unstable for an infinitely thin disk, and the q which minimizes $Q(q)$ corresponds to the least stable mode. It follows that if $Q_{Raf} = \min(Q(q)) < 1$, the disk is formally unstable, while if $Q_{Raf} > 1$, the disk is stable.

Computing the value of Q requires a minimization with respect to q. Since Q and its partial derivatives must be calculated frequently (see equation 3.16 below), it is computationally expedient to use an approximate formula which does not require such a minimization. KB10 used $Q^{-1} \approx Q_{WS}^{-1} \equiv Q_g^{-1} + Q_*^{-1}$, as proposed by Wang & Silk (1994), but this approximation becomes inaccurate when $\sigma_g/\sigma_* \lesssim 0.5$. Romeo & Wiegert (2011) have proposed a more accurate approximation

$$Q_{RW}^{-1} = \begin{cases} \frac{W}{Q_*T_*} + \frac{1}{Q_gT_g} & \text{if } Q_*T_* \geq Q_gT_g \ , \\ \frac{1}{Q_*T_*} + \frac{W}{Q_gT_g} & \text{if } Q_*T_* \leq Q_gT_g \ ; \end{cases} \tag{3.14}$$

$$W = \frac{2\sigma_*\sigma}{\sigma_*^2 + \sigma^2}. \tag{3.15}$$

This formula includes corrections for the fact that the disk is not razor-thin, T_* and T_g. A disk of finite thickness is more stable against gravitational collapse because its mass is spread out vertically, so larger values of the T_j increase the value of Q for a given set of column densities and velocity dispersions. Romeo & Wiegert (2011) give approximations to these correction factors, $T_j \approx 0.8 + 0.7\sigma_{z,j}/\sigma_{r,j}$. For simplicity we have assumed an isotropic velocity dispersion, so $T_* = T_g = 1.5$. Q_{RW} and its partial derivatives are straightforward to compute and accurate over a wide range of σ_g/σ_*

and Q_g/Q_*. The stability parameter as determined by Q_{Raf} should also be modified to include the effects of disk thickness, so our code can use either $Q \approx Q_{Raf}T$ or $Q \approx Q_{RW}$.

Disks where gravitational instability dominates the dynamics are expected to be self-regulated near $Q = 1$ (Burkert et al., 2010). A disk with $Q \lesssim 1$ develops inhomogeneities in the gravitational field, which exert random forces on gas in the disk, driving turbulence. The ultimate source of this energy is the gravitational potential of the galaxy, so mass must move inwards. If the disk had $Q \lesssim 1$, more mass would gather into inhomogeneities, thereby increasing the driving of turbulence, which stabilizes the disk, driving Q upwards. Meanwhile if $Q \gtrsim 1$, mass transport through the disk slows even if the cosmological accretion rate does not, which tends to add mass and destabilize the disk. We therefore take as a hypothesis that Q is a constant of order unity at all points in the disk at all times. Thus we can set

$$\begin{aligned}\frac{dQ}{dt} &= \frac{\partial Q}{\partial \Sigma}\frac{\partial \Sigma}{\partial t} + \frac{\partial Q}{\partial \sigma}\frac{\partial \sigma}{\partial t} \\ &\quad + \sum_i \left(\frac{\partial Q}{\partial \Sigma_{*,i}}\frac{\partial \Sigma_{*,i}}{\partial t} + \frac{\partial Q}{\partial \sigma_{*,i}}\frac{\partial \sigma_{*,i}}{\partial t} \right) = 0. \qquad (3.16)\end{aligned}$$

The evolution of the gas state variables Σ and σ, derived in the previous section, depends on the viscous torque and its radial derivatives, so we can recast equation (3.16) in the form

$$\frac{dQ}{dt} = f_2 \frac{\partial^2 \mathcal{T}}{\partial r^2} + f_1 \frac{\partial \mathcal{T}}{\partial r} + f_0 \mathcal{T} - F = 0, \qquad (3.17)$$

where the f_i are coefficients which can be read off from the gas evolution equations, and F encompasses all terms which do not depend on the viscous torque, including all stellar processes, discussed in the following section, and the rate of radiative dissipation. In particular,

$$f_2 = -\frac{\sigma}{6\pi r\Sigma v_\phi(\beta+1)}\frac{\partial Q}{\partial\sigma} - \frac{1}{2\pi(\beta+1)rv_\phi}\frac{\partial Q}{\partial\Sigma}, \tag{3.18}$$

$$f_1 = \frac{\beta^2\sigma + \sigma(r\frac{\partial\beta}{\partial r}+\beta) - 5(\beta+1)r\frac{\partial\sigma}{\partial r}}{6\pi(\beta+1)^2r^2v_\phi\Sigma}\frac{\partial Q}{\partial\sigma} + \frac{\beta(\beta+1)+r(\partial\beta/\partial r)}{2\pi(\beta+1)^2r^2v_\phi}\frac{\partial Q}{\partial\Sigma}, \tag{3.19}$$

$$f_0 = \frac{1}{6\pi r\Sigma}(\beta-1)\frac{v_\phi}{r^2\sigma}\frac{\partial Q}{\partial\sigma}, \tag{3.20}$$

$$F = \frac{\eta\pi}{3}G\Sigma\left(1+\frac{\Sigma_*\sigma}{\Sigma\sigma_*}\right)\left(1-\frac{\sigma_t^2}{\sigma^2}\right)^{3/2}\frac{\partial Q}{\partial\sigma} + (f_R+\mu)\dot{\Sigma}_*^{SF}\frac{\partial Q}{\partial\Sigma} - \sum_i\left(\dot{\Sigma}_{*,i}\frac{\partial Q}{\partial\Sigma_{*,i}} + \dot{\sigma}_{*,i}\frac{\partial Q}{\partial\sigma_{*,i}}\right). \tag{3.21}$$

Usually F will be dominated by the first term, the radiative dissipation of energy, which tends to destabilize the disk by “cooling” the gas, making $F > 0$. In this case, one can interpret equation (3.17) as requiring the torques to move gas such that it stabilizes the disk to counter the effects of this cooling.

Equation (3.17) is a second order ODE requiring two boundary conditions. At the outer edge of the disk, we specify the accretion rate of gas onto the disk, $\dot{M}_{ext}$ according to a pre-calculated accretion history, typically a fit to average accretion histories from cosmological simulations (Bouché et al., 2010). The torque is related to $\dot{M}$ through equation (3.5), so by rearranging that equation, evaluating quantities at the outer radius, and requiring a particular $\dot{M}_{ext}$, we obtain the outer boundary condition

$$\left(\frac{\partial \mathcal{T}}{\partial r}\right)_{r=R} = -\dot{M}_{ext} v_\phi(R)(1+\beta(R)). \tag{3.22}$$

Here R is a fixed outer radius of the disk. This condition implicitly assumes that all gas is accreted at the outer edge of the disk, which is not an unreasonable approximation as long as gas accretes mostly through cold streams

At the inner boundary, we require that the disk and bulge exert no torques on each other,

$$(\mathcal{T})_{r=r_0} = 0 \tag{3.23}$$

The inner edge of the computational domain is r_0, chosen for numerical reasons to be non-zero. Note that this boundary condition is somewhat different than the one used in KB10, namely $(\mathcal{T})_{r=r_0} = -\dot{M}_{ext} v_\phi(R)(1+\beta(R))r_0$ for a flat rotation curve. This will approach the physically motivated value of equation (3.23) in the limit that $r_0 \to 0$, and was chosen to satisfy a regularity condition at the inner boundary. However, since our goal here is not to obtain an analytic solution, there is no need to impose such a condition. In practice we have experimented with both choices in our numerical calculations, and we find that the choice of inner boundary condition has negligible effects at radii $r \gg r_0$, which is the great majority of the disk.

3.3 Stellar Evolution Equations

In addition to the gas, we would like to know how stellar populations in the disk evolve with time. The stars will provide most of the observable consequences of the model, in addition to determining, along with the gas, whether the disk is gravitationally unstable. Among the questions we are interested in investigating is the cause of the age-velocity dispersion correlation, namely that older stars have higher velocity dispersions. Therefore it is useful to not only keep track of the stars as a single population with a single column density Σ_* and velocity dispersion σ_* (each a function of radius and time), but also to bin the stars by age, so that $\Sigma_{*,i}$ and $\sigma_{*,i}$ describe the ith age bin.

The overall stellar population, along with each sub-population, will be directly affected by two processes - star formation and stellar migration. The two effects may be added together, recalling that of the gas which forms stars, only a fraction f_R will remain in stars after stellar evolution has taken its course,

$$\dot{\Sigma}_{*,i} = f_R \dot{\Sigma}_{*,i}^{SF} + \dot{\Sigma}_{*,i}^{Mig}. \tag{3.24}$$

Evolution equations for each process will be derived separately below.

3.3.1 Star Formation

The rate of star formation will depend on the properties of the gas from which stars form. In particular, in a sufficiently large region of the disk, the star formation rate will be proportional to the molecular gas mass divided by the free fall time, defined to be $\sqrt{3\pi/(32G\rho)}$. In deriving the gas evolution equations, we assumed that formally the density was given by $\Sigma\delta(z)$, but this is of course an approximation. The disk will have a finite thickness of order the scale height (defined by equation 3.9), so we take the density to be $\rho = \Sigma/H$. Thus we can write the star formation rate density

$$\dot{\Sigma}_*^{SF} = \epsilon_{\text{ff}} f_{H_2} \Sigma \sqrt{32 G\rho/(3\pi)} = \epsilon_{\text{ff}} f_{H_2} \Sigma \kappa \sqrt{32/3} \left(1 + \frac{\Sigma_*}{\Sigma}\frac{\sigma}{\sigma_*}\right)^{1/2} \tag{3.25}$$

For molecular gas, the efficiency of star formation per free-fall time is $\epsilon_{\text{ff}}\sim 0.01$ (Krumholz & McKee, 2005; Krumholz & Tan, 2007; Krumholz et al., 2012), though this may be significantly higher or lower given observational uncertainties. Following progress made by Krumholz et al. (2008); Krumholz et al. (2009), McKee & Krumholz (2010) have analytically approximated the molecular fraction of the gas, f_{H_2} as a function of metallicity and surface density. We adopt this prescription with a slight alteration:

$$f_{H_2} = \begin{cases} 1 - \left(\frac{3}{4}\right)\frac{s}{1+0.25s} & \text{if } s < 388/203 \\ 0.03 & \text{otherwise} \end{cases} \tag{3.26}$$

$$s = \frac{\ln(1 + 0.6\chi + 0.01\chi^2)}{0.6\tau_c} \tag{3.27}$$

$$\chi = 3.1\ \frac{1 + 3.1(Z/Z_\odot)^{0.365}}{4.1} \tag{3.28}$$

$$\tau_c = 320\ c\ (Z/Z_\odot)(\Sigma/1\text{ g cm}^{-2}), \tag{3.29}$$

where Z is the gas metallicity. We take the solar metallicity to be $Z_\odot = 0.02$, and c encapsulates the effects of clumping in the gas when averaging over large regions. Since the model presented in this paper takes averages over large areas of the disk, we take $c \sim 5$, as determined in Krumholz et al. (2009). The modification from McKee

& Krumholz (2010) is that we impose a lower limit on f_{H_2} of 3%, motivated by the observation that even extremely low total gas surface density regions form stars at a rate consistent with a constant H_2 depletion time (Bigiel et al., 2011).

Equation (3.25) is used to update the stellar column density, and it also enters into the gas column density equation (equation 4.2) through the conservation of mass. At any particular time in a simulation, all but one of the $\dot{\Sigma}_{*,i}^{SF} = 0$. Formally we can write this as

$$\dot{\Sigma}_{*,i}^{SF} = \dot{\Sigma}_{*}^{SF}\ \Theta(A(t) - A_{young,i})\ \Theta(A_{old,i} - A(t)) \tag{3.30}$$

where $\Theta(x)$ is a step function, one for $x > 0$ and zero for $x < 0$, $A(t)$ is the age that a star will be at redshift zero if it forms at time t after the beginning of the simulation, and $A_{young,i}$ and $A_{old,i}$ are the boundaries of the ith age bin.

To update the stellar velocity dispersion of a stellar population, we require that the new kinetic energy of the population be equal to the old kinetic energy plus the energy of the newly formed stars,

$$(\Sigma_{*,i}\sigma_{*,i}^2)_{new} = (\Sigma_{*,i}\sigma_{*,i}^2)_{old} + f_R(d\Sigma_{*,i}^{SF})\sigma^2 \tag{3.31}$$

where we have assumed that the newly formed stars have the same velocity dispersion as the gas from which they form. Setting $\Sigma_{*,new} = \Sigma_{*,old} + f_R(d\Sigma_{*}^{SF})$, we can rearrange, solve for $\sigma_{*,new}$, and expand to first order in the small quantity $d\Sigma_*^{SF}/\Sigma_{*,old}$

$$\begin{aligned}\sigma_{*,i,new} &= \sqrt{\frac{(\Sigma_{*,i}\sigma_{*,i}^2)_{old} + f_R(d\Sigma_{*,i}^{SF})\sigma^2}{\Sigma_{*,i,old} + f_R(d\Sigma_{*,i}^{SF})}} \\ &\approx \sigma_{*,i,old} + \frac{f_R(d\Sigma_{*,i}^{SF})}{2\Sigma_{*,i,old}\sigma_{*,i,old}}(\sigma^2 - \sigma_{*,i,old}^2) \qquad (3.32)\end{aligned}$$

Thus in the limit that the time step and therefore the density of new stars produced is small, we may use the definition of a derivative to write

$$\left(\frac{\partial \sigma_{*,i}}{\partial t}\right)^{SF} \approx f_R \, \frac{1}{2\Sigma_{*,i}\sigma_{*,i}}(\sigma^2 - \sigma_{*,i}^2)\dot{\Sigma}_{*,i}^{SF} \quad \text{for } \Sigma_{*,i} > 0 \qquad (3.33)$$

We only need this derivative for its contribution to the torque equation (equation 3.16), in which it will always be multiplied by the term $\partial Q/\partial \sigma_{*,i}$. To actually update the quantity $\sigma_{*,i}$, we use the exact relation of equation (3.31), which holds even if $\Sigma_{*,i} = 0$. Note that when $\Sigma_{*,i} = 0$, this new population of stars will have no effect on the torque equation, since $\partial Q/\partial \sigma_{*,i} = 0$, i.e. non-existent stars do not affect the stability of the disk. Thus equation (3.33) need only be employed when $\Sigma_{*,i} > 0$.

3.3.2 Radial Migration

In addition to star formation, stars are subject to radial migration. In particular, when $Q_* \lesssim 2$, transient spiral arms form which attempt to stabilize the disk (Sellwood & Carlberg, 1984; Carlberg & Sellwood, 1985; Sellwood & Binney, 2002).

N-body simulations (Sellwood & Carlberg, 1984) suggest that this heating is such that

$$\frac{\partial Q_*}{\partial t}^{Mig} = \max\left(\frac{Q_{lim} - Q_*}{T_{mig}(2\pi\Omega^{-1})}, 0\right) \tag{3.34}$$

where T_{mig} is the time scale in local orbital times over which this heating occurs, typically a few orbits, and Q_{lim} is the value of Q_* above which the stars are stable to spiral perturbations. Equation (3.34) assumes that this mechanism acts independently of the torques which act on the gas as a result of the axisymmetric instability described in section 3.2.2. In $z \sim 2$ galaxies with morphologies dominated by clumps containing both gas and stars, one might expect the axisymmetric instability to affect both components equally, as assumed in the models of Cacciato et al. (2012). However, it remains an open question whether these clumps are disrupted on a dynamical timescale by a stellar feedback process, just like giant molecular clouds, their low-redshift analogues (Krumholz & Dekel, 2010; Genel et al., 2012). Even if clumps are long-lived, they contain a relatively small part of the total stellar population (Murray et al., 2010), and thus their impact on stellar migration might be small. Moreover, in most realistic situations, the scale height of stars will be significantly greater than that of the gas, so an instability dominated by the gas will have little traction on the stars. As long as σ_*/σ is appreciably greater than unity, which it is in our fiducial model (section 3.6), we expect this treatment to be reasonable.

The time derivative of Q_* may be re-expressed in terms of the time derivatives of Σ_* and σ_* using the definition of Q_*,

$$\begin{aligned}\frac{\partial Q_*}{\partial t}^{Mig} &= \frac{\kappa}{\pi G}\left(\frac{1}{\Sigma_*}\frac{\partial \sigma_*}{\partial t}^{Mig} - \frac{\sigma_*}{\Sigma_*^2}\frac{\partial \Sigma_*}{\partial t}^{Mig}\right) \\ &= Q_*\left(\frac{1}{\sigma_*}\frac{\partial \sigma_*}{\partial t}^{Mig} - \frac{1}{\Sigma_*}\frac{\partial \Sigma_*}{\partial t}^{Mig}\right) \qquad (3.35)\end{aligned}$$

The partial time derivatives on the right hand side will depend on the mean velocity of stars in the radial direction, $v_{*,r}$, and so the forcing imposed by equation (3.34) will yield an ordinary differential equation for $v_{*,r}$. The value of $v_{*,r}$ is then used to evolve Σ_* and σ_*.

This approach assumes a single bulk velocity of stars in the radial direction at each radius, $v_{*,r}(r)$. It has been well-demonstrated (e.g. Bird et al., 2012) that the Sellwood & Binney (2002) mechanism scatters stars in both directions, i.e. a star born at some galactocentric radius may end up with a guiding center radius multiple kpc away. There are additional scattering mechanisms, such as two-body scattering and the resonant overlap between spirals and the bar (Minchev & Famaey, 2010; Brunetti et al., 2011), which will also redistribute stellar angular momenta. Modeling this redistribution is critical in explaining the detailed properties of Milky Way stellar populations. However, there are no straightforward prescriptions to model all of these effects. We therefore ignore for now the effects of radial mixing and merely require conservation of mass and energy, and that the stars will stabilize themselves if they are subject to spiral instabilities.

The evolution of Σ_* and σ_* as a function of $v_{*,r}$ is determined by the continuity equations for mass and energy of the ith stellar population

$$\frac{\partial \Sigma_{*,i}}{\partial t}^{Mig} + \frac{1}{r}\frac{\partial}{\partial r}(r\Sigma_{*,i}v_{*,r}) = 0 \tag{3.36}$$

$$\begin{aligned}\frac{\partial}{\partial t}\left[\Sigma_{*,i}\left(v_\phi^2 + 3\sigma_{*,i}^2 + 2\psi\right)\right]^{Mig} + & \\ \frac{1}{r}\frac{\partial}{\partial r}\left[r\Sigma_{*,i}v_{*,r}\left(v_\phi^2 + 3\sigma_{*,i}^2 + 2\psi\right)\right] &= 0\end{aligned} \tag{3.37}$$

Expanding the energy equation using the product rule and employing the mass equation to cancel terms leaves

$$\begin{aligned}\Sigma_{*,i}\frac{\partial}{\partial t}\left[\left(v_\phi^2 + 3\sigma_{*,i}^2 + 2\psi\right)\right] + & \\ \Sigma_{*,i}v_{*,r}\frac{\partial}{\partial r}\left[\left(v_\phi^2 + 3\sigma_{*,i}^2 + 2\psi\right)\right] &= 0\end{aligned} \tag{3.38}$$

The time derivatives of v_ϕ and ψ are zero by assumption, so expanding the surviving derivatives, setting $\partial\psi/\partial r = v_\phi^2/r$ and $\partial v_\phi/\partial r = \beta v_\phi/r$, and rearranging yields

$$\frac{\partial \sigma_{*,i}}{\partial t}^{Mig} = -v_{*,r}\left(\frac{(1+\beta)v_\phi^2}{3r\sigma_{*,i}} + \frac{\partial \sigma_{*,i}}{\partial r}\right) \tag{3.39}$$

The corresponding equation for stellar column density follows immediately from the continuity equation:

$$\frac{\partial \Sigma_{*,i}}{\partial t}^{Mig} = -\Sigma_{*,i}\frac{\partial v_{*,r}}{\partial r} - v_{*,r}\frac{\partial \Sigma_{*,i}}{\partial r} - \Sigma_{*,i}v_{*,r}/r \tag{3.40}$$

Substituting the transport equations into equation (3.35) and imposing equation (3.34) yields

$$2\pi r \frac{v_{*,r}}{v_\phi} \left(-\frac{v_\phi^2}{\sigma_*^2} \frac{(1+\beta)}{3r} - \frac{1}{\sigma_*} \frac{\partial \sigma_*}{\partial r} + \frac{1}{\Sigma_*} \frac{\partial \Sigma_*}{\partial r} + 1/r \right) \\ + \frac{2\pi r}{v_\phi} \frac{\partial v_{*,r}}{\partial r} = \frac{\max(Q_{lim} - Q_*, 0)}{T_{Mig} Q_*} \tag{3.41}$$

This is a first order ordinary differential equation (since at any particular time we treat all variables as functions of radius only), requiring a single boundary condition which we take to be $v_{*,r}(r = R) = 0$, which means that no stars are allowed to migrate between the outer edge of the disk and the IGM. This boundary condition guarantees that the bulk velocity of stars in the radial direction will be inwards at all radii, which means this method does not conserve angular momentum; to compensate for a large mass of stars moving inwards, a small mass of stars would need to move outwards. The error we make in conservation of total angular momentum is about 2% in the fiducial case.

3.4 Metallicity Evolution

3.4.1 Advection of Metals in Gas

To describe the evolution of the metal content, we begin by defining Σ_Z, the surface density of metals, so locally the metallicity of the gas is $Z = \Sigma_Z/\Sigma$. The continuity equation for Σ_Z is

$$\frac{\partial}{\partial t}\Sigma_Z = \frac{1}{2\pi r} \frac{\partial}{\partial r} \dot{M}_Z - \dot{\Sigma}_Z^{SF} + S_Z \tag{3.42}$$

where $\dot{\Sigma}_Z^{SF} = \dot{\Sigma}_*^{SF} Z$ is the rate at which metals are incorporated into newly formed stars, and S_Z is a source term for metals injected into the gas by supernovae and AGB stars. Note that, in writing this equation, we neglect transport of metals through the disk by either turbulent diffusion or galactic fountains. The inward flux of metallic mass is

$$\dot{M}_Z = \dot{M} Z = -\frac{Z}{v_\phi (1+\beta)} \frac{\partial \mathcal{T}}{\partial r}, \tag{3.43}$$

which follows from equation 3.5. The left hand side of equation 3.42 can be reexpressed in terms of Z by noting $\partial \Sigma_Z / \partial t = Z \partial \Sigma / \partial t + \Sigma \partial Z / \partial t$. Equation 3.42 then becomes

$$\begin{aligned} Z \frac{\partial \Sigma}{\partial t} + \Sigma \frac{\partial Z}{\partial t} &= \\ \frac{Z}{2\pi r (1+\beta)^2 v_\phi} & \left((1+\beta) \frac{\partial \mathcal{T}}{\partial r} \beta / r - (1+\beta) \frac{\partial \mathcal{T}}{\partial r} \frac{\partial \ln Z}{\partial r} \right. \\ & \left. + \frac{d\beta}{dr} \quad \tfrac{\partial \mathcal{T}}{\partial r} \quad -(1+\beta) \frac{\partial^2 \mathcal{T}}{\partial r^2} \right) - \dot{\Sigma}_Z^{SF} + S_Z \end{aligned} \tag{3.44}$$

Comparing this with the previously derived gas surface density evolution equation, we can cancel most of the terms on the right hand side with $Z \partial \Sigma / \partial t$, leaving only

$$\frac{\partial Z}{\partial t} = -\frac{1}{(\beta+1) r \Sigma v_\phi} \frac{\partial \ln Z}{\partial r} \frac{\partial \mathcal{T}}{\partial r} + \frac{S_Z}{\Sigma}. \tag{3.45}$$

Inflowing gas has some metallicity Z_{IGM}, which we fix at $Z_{IGM} = 0.1 Z_\odot$ for the entire simulation. Simulations (Shen et al., 2012) and observations (Adelberger et al., 2005) suggest that the circum-galactic medium is enriched to this degree as early as $z = 3$.

3.4.2 Metal Production

For simplicity, we adopt the instantaneous recycling approximation, proposed by Tinsley (1980), to specify S_Z, the production rate of metals. First we recognize that metals are produced in supernovae and AGB stars. To a first approximation, we can assume that the lifetimes of stars that dominate metal production are much smaller than the timescales with which we are concerned in this paper, so metals enter the ISM at a rate proportional to the star formation rate. Not all gas which forms stars is returned to the ISM, since low-mass stars do not leave the main sequence in a Hubble time and even high-mass stars form remnants. Defining the remnant fraction f_R as the fraction of gas forming stars which will end up not being returned to the ISM, the surface density of recycled gas appearing in the ISM is $(1 - f_R)\dot{\Sigma}_*^{SF}$. Supernovae and normal stellar evolution will enrich a small fraction of this gas, namely y_M, the yield. The surface density of metal production is therefore

$$S_Z = y_M \zeta (1 - f_R) \dot{\Sigma}_*^{SF}. \tag{3.46}$$

Assuming a Chabrier (2005) initial mass function and a coarse approximation for the ultimate fates of stars as a function of mass, Krumholz & Dekel (2012) compute $f_R = 0.46$. Assuming in addition a production function, the fraction of a star's initial mass converted to a given element, from Maeder (1992), they compute a yield of $y_M = 0.054$ for Solar metallicity stars. The effective yield may be somewhat smaller than this, since galactic winds driven by supernovae tend to eject material which is richer than average in metals. The factor of $\zeta \lesssim 1$ represents the ratio of metallicity in the ISM to metallicity of ejected material. We adopt $\zeta = 1$, corresponding to an assumption that the ejecta are well-mixed with the ISM. This value will in principle depend on the mass of the galaxy considered (Mac Low & Ferrara, 1999). However, owing to the high resolution required, to date no simulation has reliably calculated the degree to which metal-rich gas is preferentially ejected. Changes in the exact value of ζ roughly translate into the normalization of the metallicity distribution in the gas, so our fiducial value of ζ was chosen to give reasonable values for this normalization.

3.4.3 Diffusion of Metals

The metallicity gradients produced when accounting only for metal production by stars and advection by inflowing gas are far steeper than the observed gradient in the Milky Way. Metals are formed in proportion to the star formation rate, which tends to be high towards the center of the simulated galaxies. Meanwhile the inflow of gas throughout the disk concentrates the metals even further. To explain the relatively small observed metallicity gradients, one must allow metals formed at small galactic radii to reach large radii. This may occur either in the plane of the disk (diffusion) or out of the plane (galactic fountains). By assuming a fixed value of $Z_{IGM} = 0.1 Z_\odot$, we have already implicitly assumed some sort of transport of metals from the galaxy into its surrounding medium. However, rather than modeling this transport in any detail, let us consider only the diffusion of metals through the disk.

In general, a diffusion equation will have the form

$$\frac{\partial}{\partial t} M_Z = D \frac{\partial^2}{\partial r^2} M_Z \tag{3.47}$$

where D is the diffusion coefficient and $M_Z = 2\pi r\Delta r\Sigma Z$ is the gas-phase metal mass in a given cell. At an order of magnitude level D may be estimated by taking the typical velocity of gas in the disk, σ, and multiplying by the typical length scale of perturbations, namely the 2d Jeans scale, $\sigma^2/G\Sigma$. For simplicity we simply adopt $D = k_Z v_\phi(R)R$ where k_Z and D will be constant at every time and location in the disk. Numerically we take $k_Z = 10^{-3}$ which is of the correct order of magnitude and yields a metallicity gradient of order 0.1 dex/kpc (see figure 3.2)., which is comparable to observed values in isolated spiral galaxies (e.g. Zaritsky et al., 1994)

3.4.4 Metals Locked in Stars

The metallicity of a given stellar population can be updated when new stars are added to it by again assuming instantaneous missing.The new metallicity is just an average of the old metallicity and the metallicity of the gas, weighted by the surface density of the extant stellar population and the newly formed population respectively.

$$Z_{*,i,new} = \frac{Z_{*,i,old}\Sigma_{*,i} + f_R Z(d\Sigma_{*,i}^{SF})}{\Sigma_{*,i} + f_R(d\Sigma_{*,i}^{SF})} \tag{3.48}$$

Here, as in equation (3.31), $d\Sigma_{*,i}^{SF} = \dot{\Sigma}_{*,i}^{SF} dt$, is the surface density of stars formed in a given time step in a given stellar age bin i.

Besides the formation of new stars, a given stellar population is subject to migration through the disk, as discussed in section 3.3.2 Since the stars migrate through the disk with a mean velocity set by equation (3.41), the metallicity profile of a given

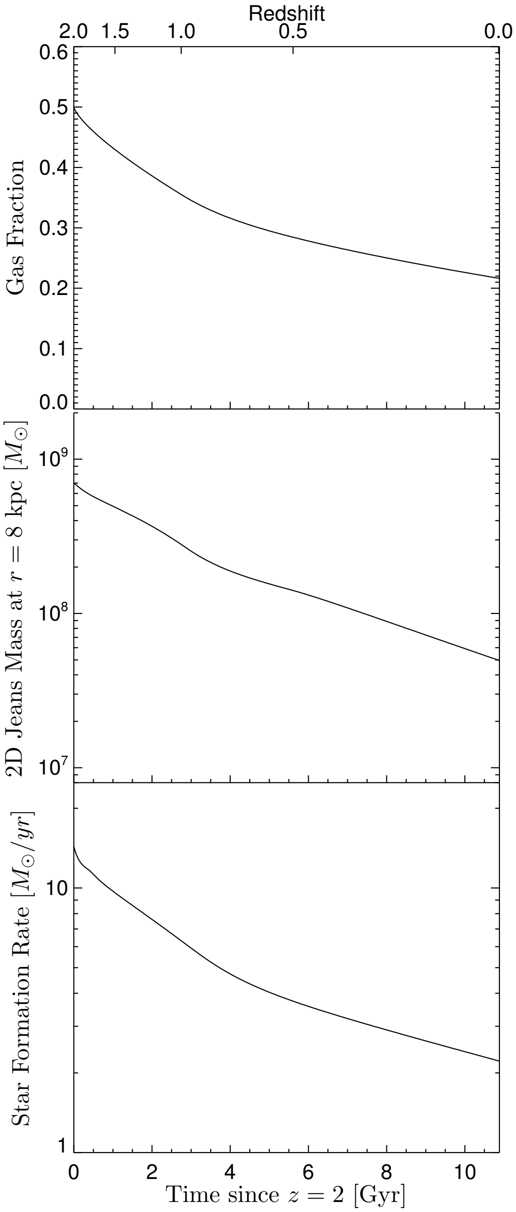

Figure 3.1: Time evolution from the beginning to the end of the fiducial simulation of the radially-integrated gas fraction, 2D Jeans mass at $r = 8$ kpc, and the radially-integrated star formation rate.

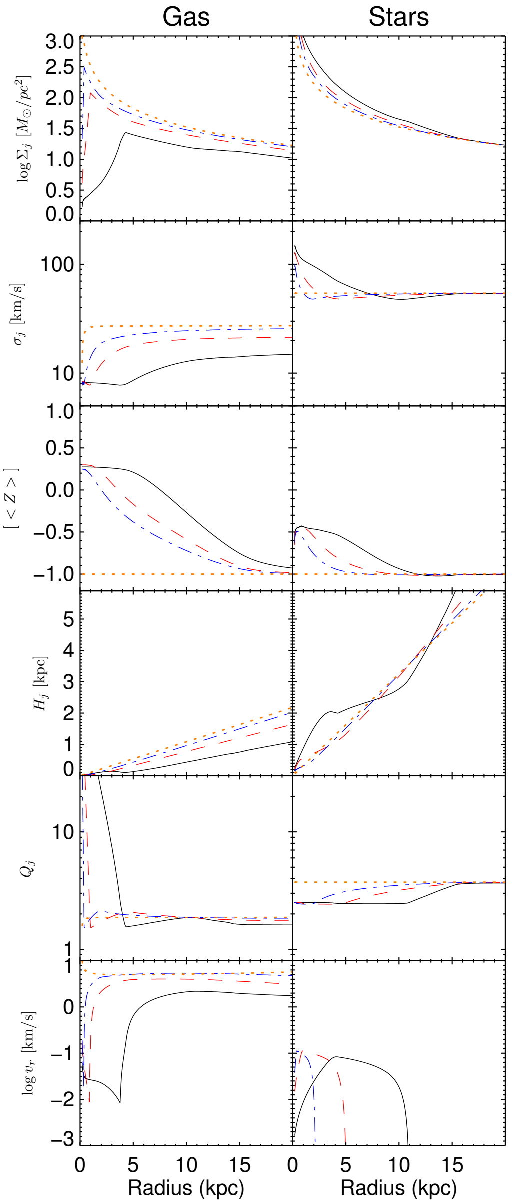

Figure 3.2: A direct comparison of the gas and stellar components as a function of radius at redshifts 2 (orange dotted), 1.5 (blue dot-dash), 1 (red dashed), and 0 (black solid). The gas cools and depletes, while the stars accumulate and heat. The expanding stabilized region of the disk is evident in the dramatic decrease in gas transport velocity, large Q_g, and $\sigma \rightarrow \sigma_t$. The outward movement of the region where stars form and migrate follows the peak in gas column density - Q_* approaches $Q_{lim} = 2.5$, the stellar metallicity gradient steepens, and the stellar scale height flattens.

population of stars evolves under a continuity equation for the metal mass,

$$\frac{\partial}{\partial t}\left(\Sigma_{*,i}Z_{*,i}\right)^{Mig}+\frac{1}{r}\frac{\partial}{\partial r}\left(r\Sigma_{*,i}Z_{*,i}v_{*,r}\right)=0 \tag{3.49}$$

Subtracting the continuity equation for total stellar mass (equation 3.36), we obtain

$$\frac{\partial Z_{*,i}}{\partial t}^{Mig}=-v_{*,r}\frac{\partial Z_{*,i}}{\partial r}, \tag{3.50}$$

for the evolution of stellar metallicity. Equations (3.48) and (3.50) fully describe the evolution of the metallicity of the ith stellar population. Note that these equations neglect radial diffusion of stars, only taking into account the mean velocity $v_{*,r}$. Radial mixing (Sellwood & Binney, 2002; Roškar et al., 2012) is required to explain the spread of metallicities in stars at a fixed age and radius, and undoubtedly leads to a shallower stellar metallicity gradient than what we obtain.

3.5 Numerical Method

3.5.1 Computational Domain

In deriving the gas evolution equations, we assumed the disk to be thin and axisymmetric. Thus the disk is described by variables which change only in radius and time. We therefore define a mesh of radial positions r_i with a fixed number of points, n_x, logarithmically spaced between the outer edge of the disk at a fixed radius R and a fixed inner cutoff r_{min}, usually chosen to be $r_{min} = 0.01R$. Explicitly,

$$r_i = R\left(\frac{r_{min}}{R}\right)^{1-(i-1)/(n_x-1)} \tag{3.51}$$

The highest spatial resolution is therefore given to the region with the shortest dynamical times.

Time, tracked in units of the orbital period at radius R, begins at zero when the simulations are started, typically at $z = 2$, and reach a few tens of orbits at $z = 0$, depending on the assumed radius and circular velocity. The size of the time steps are calculated by first determining all timescales defined by dividing each state variable at each position by its time derivative, picking out the minimum timescale, and multiplying it by a small number TOL, usually taken to be 10^{-4}. Larger values of TOL lead to numerical instabilities near the inner boundary, which is especially susceptible to such issues because the local dynamical timescale in the disk is $\Omega^{-1} \propto r$ for a flat rotation curve.

$$\begin{aligned}\Delta t \quad = \quad & \mathrm{TOL} \times \min_i \left[\frac{\Sigma}{\partial\Sigma/\partial t}(r_i), \frac{\sigma}{\partial\sigma/\partial t}(r_i),\right. \\ & \left.\frac{\Sigma_*}{\partial\Sigma_*/\partial t}(r_i), \frac{\sigma_*}{\partial\sigma_*/\partial t}(r_i), \frac{0.01}{\mathrm{TOL}}\right] \end{aligned} \tag{3.52}$$

A maximum time step of 0.01 outer orbits is imposed to prevent systems extremely close to equilibrium from advancing too quickly.

3.5.2 PDEs

At each time step, the code solves the equations in non-dimensionalized form (see appendix) in the following order. First, we solve equation (3.41) to determine $v_{*,r}$ at all radii. The equation is of the form $\mathcal{H} = h_0 v_{*,r} + \partial v_{*,r}/\partial r$ with

$$\begin{aligned} \mathcal{H} &= \frac{\max(Q_{lim} - Q_*, 0) v_\phi}{2\pi r T_{Mig} Q_*} && (3.53) \\ h_0 &= -\frac{v_\phi^2}{\sigma_*^2}\frac{(1+\beta)}{3r} - \frac{1}{\sigma_*}\frac{\partial \sigma_*}{\partial r} + \frac{1}{\Sigma_*}\frac{\partial \Sigma_*}{\partial r} + \frac{1}{r}, && (3.54) \end{aligned}$$

The boundary condition specifies $v_{*,r}$ at the outer edge of the disk. Thus rewriting the radial derivative as a finite difference and employing a backwards Euler step, we can write an explicit update equation,

$$v_{*,r}(r_{i-1}) \approx \frac{v_{*,r}(r_i) - (r_i - r_{i-1})\mathcal{H}(r_{i-1})}{1 - (r_i - r_{i-1}) h_0(r_{i-1})}, \tag{3.55}$$

which we solve iteratively by starting with the specified value of $v_{*,r}(r_{nx}) = 0$ and moving inwards.

Using the value of $v_{*,r}$ along with the current values of the state variables, we calculate the coefficients of the torque equation (equation 3.17). To solve the resultant linear PDE, we employ a similar finite difference method, which approximates

$$\begin{aligned} \frac{\partial \mathcal{T}_i}{\partial r} &\approx \frac{\mathcal{T}_{i+1} - \mathcal{T}_{i-1}}{r_{i+1} - r_{i-1}} && (3.56) \\ \frac{\partial^2 \mathcal{T}_i}{\partial r^2} &\approx \frac{1}{r_{i+1/2} - r_{i-1/2}} \left(\frac{\mathcal{T}_{i+1} - \mathcal{T}_i}{r_{i+1} - r_i} - \frac{\mathcal{T}_i - \mathcal{T}_{i-1}}{r_i - r_{i-1}} \right) && (3.57) \end{aligned}$$

Since we are using a logarithmically spaced grid, $r_{i+1/2} = \sqrt{r_i r_{i+1}}$. By plugging these approximations into the torque equation, the problem reduces to the inversion of a tridiagonal matrix.

The forcing term in the torque equation, (3.21) generally acts to destabilize the disk, since its largest term comes from radiative cooling of the gas and cooler gas is more prone to gravitational collapse. The torque equation requires that the gravitational torques exactly counteract this effect to maintain $dQ/dt = 0$. However, in the event that the forcing term in the torque equation becomes negative as a result of stellar migration and a reduced rate of cosmological infall leading to $\mathcal{L} \to 0$, we set it to zero so that the gas is not forced to destabilize the disk. This in turn allows positive values of dQ/dt. We do not allow the forcing term to return to the value given by (3.21) until that value is again positive and Q has been allowed to rise and then fall back down to $Q = Q_f$. This allows the simulation to follow disks which stabilize at least temporarily, for example because of a lull in the cosmological accretion rate, and then return to a marginally unstable state. For the smoothed average cosmological accretion history used in our fiducial run, parts of the disk which stabilize remain that way because the accretion rate is monotonically decreasing.

With $\mathcal{T}$, $\partial\mathcal{T}/\partial r$, $\partial^2\mathcal{T}/\partial r^2$, and $v_{*,r}$, we can now evaluate the derivatives of the state variables. Where radial derivatives of the state variables or other quantities appear in the evolution equations or the coefficients of the above differential equations, a minmod slope limiter is used to evaluate them. In particular, if $L =$

$(A(r_i) - A(r_{i-1}))/(r_i - r_{i-1})$ and $R = (A(r_{i+1}) - A(r_i))/(r_{i+1} - r_i)$

$$\frac{\partial A}{\partial r}(r_i) = \begin{cases} L & \text{if } |L| < |R| \text{ and } LR > 0 \\ R & \text{if } |L| > |R| \text{ and } LR > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.58}$$

where A is a stand-in for any quantity. This strongly suppresses noise on the scale of the mesh separation by zeroing out rapid variations in the derivatives.

With the time derivatives calculated at each point, we simply take a forward Euler step to update the state variables, namely Σ, σ, Z, Σ_*, σ_*, and for each age-binned stellar population, $\Sigma_{*,i}$, $\sigma_{*,i}$, and $Z_{*,i}$. Typical runs have time steps limited by the rate of change of the gas state variables near the inner boundary of the disk where the dynamical timescale is shortest. On a single processor, runs take about one day to complete if we numerically evaluate $Q(q)$ and its derivatives using the full Rafikov (2001) formalism. We can shorten this by an order of magnitude by using the approximation to Q suggested by Romeo & Wiegert (2011). This approximation is much more efficient because Q_{RW} and its partial derivatives may be calculated as functions of the state variables alone, without the need to minimize over a wavenumber or compute the partial derivatives $\partial Q/\partial\{\Sigma, \sigma, \Sigma_{*,i}, \sigma_{*,i}\}$ numerically as required by the full Rafikov Q.

3.5.3 Initial Conditions

By assuming a flat rotation curve, fixed gas fraction, equal stellar and gas velocity dispersions, a simple analytic approximation to Q, and ignoring stellar processes (formation and migration), KB10 were able to compute an equilibrium solution to the evolution equations. In particular,

$$\sigma = \frac{1}{\sqrt{2}}\left(\frac{G\dot{M}_{ext,0}}{\eta f_g}\right)^{1/3} \tag{3.59}$$

$$\Sigma = \frac{v_\phi}{\pi G r}\left(\frac{f_g^2 G\dot{M}_{ext,0}}{\eta}\right)^{1/3} \tag{3.60}$$

Here $\dot{M}_{ext,0}$ is the accretion rate of gas onto the outer edge of the disk at the start of the simulation, and f_g is the gas fraction, assumed to be constant in radius. By assumption, $\sigma_* = \sigma$ and $\Sigma_* = \Sigma(1-f_g)/f_g$.

If we relax the assumptions that the velocity dispersions of both components are identical and $Q = 1$, add a factor to correct for finite disk thickness, but retain the approximate form of Q for an infinitely thin disk, $Q^{-1} \approx Q_g^{-1} + Q_*^{-1}$, we obtain a modified version of the equilibrium column density,

$$\Sigma = \frac{T}{Q_f}\frac{v_\phi}{\pi G r}\frac{\phi_0 f_g}{f_g(\phi_0 - 1) + 1}\left(\frac{G\dot{M}_{ext,0}}{\eta f_g}\right)^{1/3} \tag{3.61}$$

where $\phi_0 = \sigma_*/\sigma$ is a free parameter , $T \approx 1.5$ is the thickness correction, and Q_f is the fixed value to which Q will be set everywhere in the disk. To initialize the simulations, we use equations (3.59) and (3.61). We then adjust $\sigma_* = \phi_0\sigma$ keeping ϕ_0 fixed until $Q = Q_f$ exactly at each cell of the grid. Finally, we run the simulation with stellar processes turned off, i.e. $\epsilon_{ff} = Q_{lim} = 0$, and with $\dot{M}_{ext}$ fixed to its initial value, $\dot{M}_{ext,0}$, to allow the gas to adjust to an equilibrium configuration. The greatest effect of this adjustment occurs at the inner edge of the disk, since these relations were derived using a different inner boundary condition and under a more stringent set of assumptions. Once the state variables are changing sufficiently slowly, we have found our initial conditions and therefore return ϵ_{ff}, Q_{lim}, and $\dot{M}_{ext}(t)$ to their user-specified values.

3.6 Fiducial Model

While our code is quite general, here we describe a simple model run using it in order to demonstrate its capabilities. In future work we will explore a much wider part of parameter space, using more realistic cosmological accretion histories.

3.6.1 Setup

The formalism presented here requires a rotation curve, accretion history, and fixed inner and outer radii to be specified before the simulation is run. Since we employ a logarithmic computational grid, there is little cost to extending the outer radius out to 20 (as opposed to 10) kpc. This allows us to follow the transition of the outer disk from somewhat molecular at high redshift to atomic at low redshift. For the inner truncation radius, we take $r_0 = 0.01R = 200pc$. The exact value will affect the quantitative results within a few kpc of the center of the disk, but the exact results of the simulation in this region should be taken with a grain of salt anyway. Here σ_* reaches a similar order of magnitude as the circular velocity, which we take to be independent of radius, $v_\phi(r) = 220$ km/s, so our treatment of this region as a thin disk is not valid. Moreover, the inner boundary value for the torque equation, which we take to be zero - no torque is exerted by the region within the truncation radius on the disk - could easily be some small but non-zero value.

The accretion history employs the fitting formula from Bouché et al. (2010), namely

$$\dot{M}(t) = 7 \ \epsilon_{in} \ f_{b,0.18} \ M_{h,12}^{1.1} \ (1+z)^{2.2} \ M_\odot/yr \tag{3.62}$$

where $M_{h,12}$ is the halo mass in $10^{12} M_\odot$, $f_{b,0.18}$ is the baryon fraction of the accreting matter normalized to 18%, and ϵ_{in} is zero for $M_{h,12} > 1.5$ but varies linearly in time from 0.7 down to 0.35 between redshift 2.2 and 1. Before redshift 2.2, $\epsilon_{in} = 0.7$, and after redshift 1, $\epsilon_{in} = 0.35$. We choose $f_{b,0.18} = 1$, and an initial halo mass which will grow to be about $10^{12} M_\odot$ at redshift zero. The formula governing the growth of the halo mass is given in the same paper,

$$\dot{M}_h = 34.0 \ M_{h,12}^{1.14} \ (1+z)^{2.4} \ M_\odot/yr, \tag{3.63}$$

so an initial halo mass of $M_{h,12} = 0.27$ at $z = 2$ produces a Milky Way-analogue galaxy with $M_{h,12} \approx 1$ at $z = 0$. We note that some of the baryonic accretion may go into expanding the outer radius of the disk, instead of being transported inward, which would reduce the accretion rate below the estimate given in equation (3.62). However, since the baryonic mass of galactic disks outside 20 kpc is generally a negligible fraction of the total, this clearly cannot be a large effect, and the error we make by neglecting it is small compared to the general uncertainty in the cosmological accretion rate.

In addition to these functions, there are several free parameters controlling various physical processes in the disk. The star formation efficiency per freefall time is

$\epsilon_{ff} = 0.01$. The mass loading factor of winds ejected from the galaxy in proportion to the star formation rate is $\mu = 1$, chosen to roughly correspond to observations (Erb, 2008). The fraction of turbulent energy in the gas which will decay in a scale height crossing time is $\eta/1.5 = 1$. The time scale for a $Q_* = Q_{lim} - 1$ population to approach $Q_* = Q_{lim}$ is $T_{mig} = 2$ local orbital periods, and the value of Q_* below which the stars are subject to transient spiral instabilities is $Q_{lim} = 2.5$. For computational convenience, we use $Q \approx Q_{RW}$ to evaluate the disk's stability. We will explore the sensitivity of the results to these parameter choices in future work. Here our goal is merely to demonstrate the method and its results.

The value of Q everywhere in the disk is fixed to Q_f. Theoretically Q is expected to be self-regulated to a value of order unity. Formal stability criteria derived from the perturbed equations of motion for infinitely thin disks find the disks to be unstable when $Q < 1$, so the marginal stability which we assume here would imply $Q = 1$. However, recent work by Elmegreen (2011) has shown that for a realistically thick disk where the gas cools on the order of a dynamical time, a marginally stable value of Q is closer to 2 or 3. This is consistent with the observational evidence compiled by Romeo & Wiegert (2011) for nearby spiral galaxies, namely that when Q_{RW} for these disks is calculated, the values typically fall between 2 and 3 for most galaxies at most radii. Thus we adopt $Q_f = 2$ as a fiducial value.

Finally, to specify the initial conditions fully, one must choose an initial gas fraction and a ratio of stellar to gas velocity dispersion. Since the only way the stellar velocity dispersion can decrease is by mixing it with a lower-velocity dispersion population, it is reasonable to expect this ratio to be greater than unity. The simplified models of gravitationally unstable galaxies evolving from $z \gg 1$ discussed in Cacciato et al. (2012) suggest that by $z \sim 2$, this ratio ϕ_0 is a few, so we adopt $\phi_0 = 2$.

3.6.2 Disk-Average Quantities

Before considering the radial structure of the disk, let us consider the evolution of the galaxy as a whole between $z = 2$ and $z = 0$. Our model does not allow the rotation curve or outer radius of the disk to evolve in time. However, over this redshift range, the circular velocity (assuming a constant spin parameter) will evolve by less than about 10% (e.g. Cacciato et al., 2012). Meanwhile, the position of the outer edge of the disk has a minimal effect on its evolution, so long as the outer edge of the star-forming disk is resolved. At larger radii than this, there is so little star formation that the gas is free to flow inwards at a constant rate and arrive at the edge of the star-forming disk unaltered by its passage through the HI disk.

The primary changes in the disk are the steady decline in the accretion rate, and the steady formation of stars. For the fiducial model, $\dot{M}_{ext}(t)$ drops smoothly from about 13 $M_{\odot}/yr$ at $z = 2$ to 2 $M_{\odot}/yr$ at $z = 0$. This falloff is mirrored in the drop in total gas fraction, two-dimensional Jeans mass, and total star formation rate (figure 3.1). The star formation rate in particular has almost the same numerical value as $\dot{M}_{ext}(t)$, starting off slightly higher and converging to the accretion rate. This is a reflection of the fact that the formed stars can only come from gas that started in the simulation or accreted at a later time, and the initial gas reservoir is depleted in about 1 Gyr.

Stars, once formed, remain in the disk, while the mass of gas in the disk falls with the cosmological accretion rate. This drives a steady decrease in the gas fraction from its initial value, down to 20%. Referring to the equilibrium solution for constant gas fraction (equations 3.59 and 3.60), and noting that f_g has dropped by a factor of a few, while the accretion rate has dropped by a factor of about 6, we might expect σ to decrease by maybe a factor of 2, while Σ might decrease by more than a factor of 3.

The two-dimensional Jeans mass (Kim & Ostriker, 2002) is defined by

$$M_J = \frac{\sigma^4}{G^2\Sigma} \tag{3.64}$$

Physically this represents the characteristic mass of a clump of gas which collapses under gravitational instability to form a cluster of stars. Its steady decrease with time reflects the cooling of the disk, which allows smaller regions to collapse. This is the phenomenon that explains why $z \sim 2$ galaxies contain giant clumps far larger than the biggest GMCs in present-day Milky Way-like galaxies. As a practical matter, this means that the typical size of star clusters steadily decreases, so, coupled with the fact that a clump of gas can form stars with at most tens of percent effiency, clusters with $M > 10^6 M_\odot$ are unable to form in today's quiescent spirals. In the fiducial model, $M_J \sim 2 \cdot 10^7 M_\odot$ at $r = 8$ kpc. The decrease in the upper envelope of cluster mass with time is consistent with the arguments made by Escala & Larson (2008).

3.6.3 Radial Structure of the Disk

We show the radial structure of our fiducial disk in figures 3.2, 3.3 and 3.4. We can understand the qualitative behavior shown in these plots by considering the

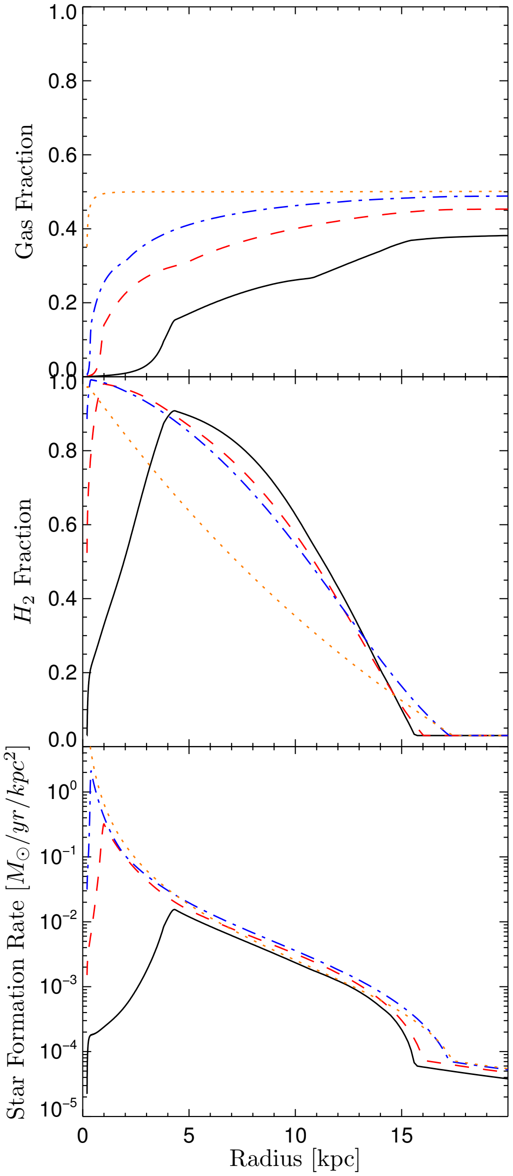

Figure 3.3: Radial profiles of quantities at redshift 2 (dotted), 1.5 (dot-dashed), 1 (dashed), and 0 (solid). The peak of f_{H_2} and hence the star formation rate move outwards as the simulation evolves, as the gas further in has been depleted and cannot be replenished.

processes that drive the evolution. The two most important drivers are that $Q = 1$ almost everywhere at all times, and that stellar migration tends to self-regulate the stars such that $Q_* = Q_{lim}$ - recall that Q_{lim} is a free parameter, below which stars are subject to transient spiral instabilities. If $Q_* > Q_{lim}$, stars will form and drive up Σ_*, decreasing Q_*, while if $Q_* < Q_{lim}$, the stars will migrate inwards increasing σ_* and hence Q_*. These two restrictions set Q_g to a value somewhat less than Q_{lim}, depending on the local ratio σ_g/σ_*. These forces lead the simulations to form three qualitatively distinct regions: a stabilized stellar-dominated region, a star-forming region, and an HI disk.

The radial extent of the star-forming region is more or less set by where the gas is molecular, i.e. $f_{H_2} \approx 1$. This in turn corresponds to where the gas column density is larger than some metallicity-dependent critical value. For our fiducial initial conditions, the disk is molecular out to $r \approx 15$ kpc at $z = 2$. Within this radius, almost the entire disk is vigorously star-forming. As time passes, a stellar-dominated central region begins to appear. This occurs because, towards the center of the disk, the gas has short local dynamical times and hence undergoes rapid star formation. In contrast, the inward mass flux of gas required to maintain $Q \approx Q_f$ is nearly independent of radius. Star formation depletes this gas as it moves inwards, so by the time it reaches the inner region of the disk, not only is there less gas than there would have been neglecting star formation, but it is being consumed faster. In order to maintain a constant Q, given that $Q_* \approx Q_{lim}$, the gas must maintain Q_g close to constant. Star formation decreases the gas column density, so to keep Q_g roughly unchanged, the gas velocity dispersion must fall proportionally. Thus the gas velocity dispersion drops fastest in the center of the disk.

By assuming a fixed gas temperature, we essentially set a floor on the value of σ. When σ hits this floor, which happens first at the inner edge of the computational domain (see figure 3.2), the radiative loss rate $\mathcal{L}$ approaches zero. The gas no longer loses energy through shocks, and therefore ceases to move inwards. In this situation that region of the disk ceases to become gravitationally unstable, and Q is allowed to rise. Without any means of mass transport, the gas simply depletes as it forms stars. As the gas column density drops off, the stars dominate the local stability of the disk. Since they are constrained to $Q_* \approx Q_{lim}$ by our assumptions about stellar migration, the overall value of Q/T of the disk in this region approaches Q_{lim} as well.

The third qualitatively distinct region of the disk may be thought of as the HI disk wherein f_{H_2} is low enough that stars form at a relatively slow rate, and gas flows in adhering even more closely to the equilibrium conditions of equations (3.59) and (3.60), which were derived by neglecting star formation in KB10, than in the star-forming region. In essence, the gas is allowed to flow in with a constant mass flux at each radius, since star formation is not depleting the gas significantly. Depending on the initial conditions of the simulation, the column density of stars may be low enough or the velocity dispersion of the stars high enough that $Q_* > Q_{lim}$ for the duration of the simulation. In this situation the overall stability of the disk is almost exclusively determined by the stability of the gas, therefore the gas properties will correspond more closely to the equilibrium values with the gas fraction set to unity.

Looking at the values for Σ and σ near the solar circle (see figure 3.2), we see that they are too high relative to their observed values of approximately $13 M_{\odot}/pc^2$ and 8 km/s respectively, though not by more than a factor of two. Moreover, the column density of gas near the center of the disk is lower than observed in the Milky Way. Both of these problems stem from the fact that when $\mathcal{L} \to 0$, mass transport due to gravitational instability shuts off, whereas the real Milky Way has a number of mechanisms to transport gas into its central regions even when $\sigma \to \sigma_t$. The gas could be transported by a bar instability from larger radii, or the gas which we assume accretes at the edge of the disk could be accreting directly into the central region of the galaxy. Gas can also be recycled back to the ISM from stars. We assume this occurs instantaneously, so we neglect gas from stars which form farther out in the disk and migrate inwards. Nonetheless, our model qualitatively reproduces the structure of $z = 0$ disk galaxies: a central stellar-dominated bulge, an extended star-forming disk, and an outer HI-dominated disk with very little star formation.

3.6.4 Stellar Populations

As the stars form in the fiducial simulation, one can treat them as adding together into a single population for the purposes of evaluating the torque equation, while at the same time evolving a number of passive populations, binned by age, alongside the single population. Only the active population affects the stability of the disk, while the passive populations simply serve as tracers of the stars formed during a particular epoch. This in turn is a reflection of the state of the gas at that time, with the added effect of gradual stellar heating through radial migration.

Stellar migration occurs locally as the result of star formation, since it is star formation which drives Q_* below Q_{lim}. It is therefore unsurprising that the stellar populations seem to have very similar column density profiles (see figure 3.6) to the star formation rate profile shown in figure 3.3. The primary effect of migration is thus not mass transport inwards, so much as an increase in the velocity dispersion. This can be quite significant - the oldest stars near the center of the disk reach nearly $\sigma_{*,i} = 100$ km/s, which is significantly larger than the gas velocity dispersion at any time in the simulation.

The state of these populations near the solar neighborhood at $z = 0$ is of particular interest, since these populations are well-observed and display well-known correlations. The velocity dispersions of stars in the solar neighborhood vary from about 17 km/s for 1 Gyr-old stars to ~ 10 Gyr-old stars with $\sigma_* \approx 37$ km/s (Nordström et al., 2004; Holmberg et al., 2009). The theoretical explanations for this correlation go back to Spitzer & Schwarzschild (1953) and generally center around the scattering of stars by molecular clouds and spiral structure, which gradually heats the disk. Other explanations have included minor or major mergers (e.g. Dierickx et al., 2010; Bekki & Tsujimoto, 2011; Qu et al., 2011) and popping star clusters (Assmann et al., 2011). All of these explanations are conceptually trying to do the same thing - form a thick disk from a thin disk. However, a gravitationally unstable disk subject to star formation and a decreasing accretion rate will start with a high gas velocity dispersion that will decrease with time. This will also naturally generate an age-velocity dispersion correlation. This is the scenario presented in the simulations of Bournaud et al. (2009), and in the chemodynamical models of Burkert et al. (1992).

The age-velocity dispersion produced in our fiducial model may be explained as the combination of two physical effects. First, the gas velocity dispersion decreases with time as the disk cools. This may be understood from the fact that if Q and Q_* are self-regulated to constant values, then Q_g must remain close to constant, and so if Σ decreases, so must σ. As the gas cools, the stars it forms will be cooler than previous generations of stars, leading to an age-velocity dispersion correlation. The second effect is the heating of stars via transient spirals to maintain $Q_* = Q_{lim}$. Although this is a scattering process which heats stars over time, there is never a thin disk which gradually forms a thick disk.

To better discern the importance of each of these effects, we can compare the stars produced by the fiducial model to runs with certain effects artificially turned off. The high and low constant accretion rate models shown in figure 3.5 have $\dot{M}_{ext}(t) = 12.3 M_\odot/yr$ and $\dot{M}_{ext}(t) = 2.34 M_\odot/yr$ respectively, corresponding to the accretion rates at the beginning and end of the fiducial simulation. For simulations where migration is turned off, we plot the properties of the stars at their epoch of formation, rather than their properties at $z = 0$. Thus the dynamical effects of migration as it affects the stability of the disk remain unchanged as compared with the fiducial simulation. Figure 3.5 shows explicitly that the age-velocity dispersion correlation is strongly affected by the accretion history and the presence of stellar heating. All of the scenarios are able to generate some age-velocity dispersion correlation. Even the case with no stellar heating and a constant accretion rate produces one as the result of a fall in Σ, and hence σ, as a result of star formation.

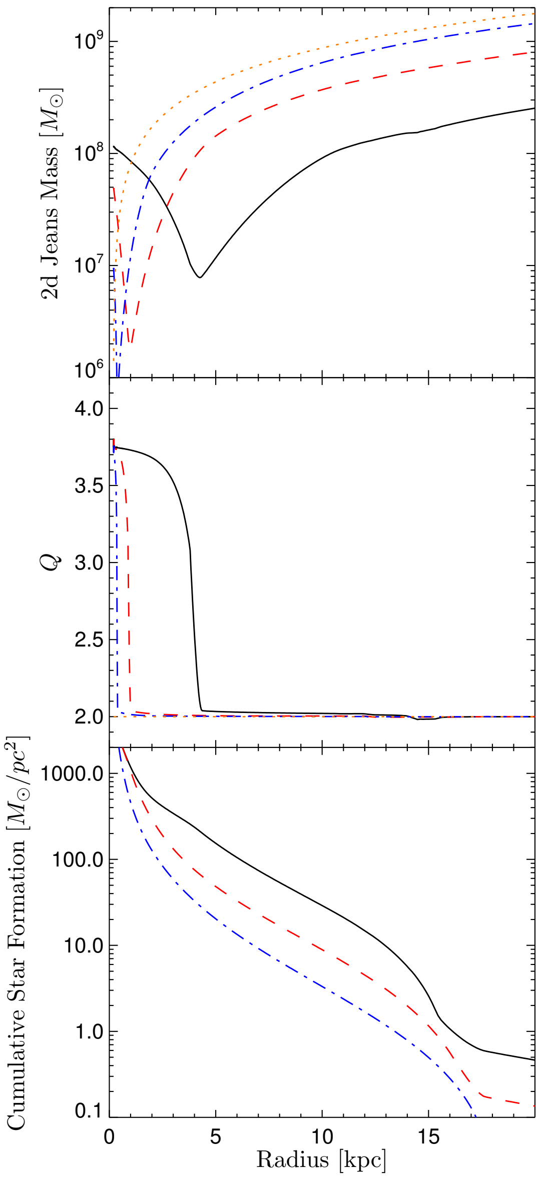

Figure 3.4: Radial profiles of quantities at redshift 2 (dotted), 1.5 (dot-dashed), 1 (dashed), and 0 (solid). Within the star-forming region, the size of the Jeans mass decreases steadily, but increases at the center of the disk owing to the extremely low gas column densities. The two-component Q value transitions from unity in the gas (both H_2 and HI) dominated regions to $Q = Q_{lim}T = 15/4$ in the stellar-dominated component.

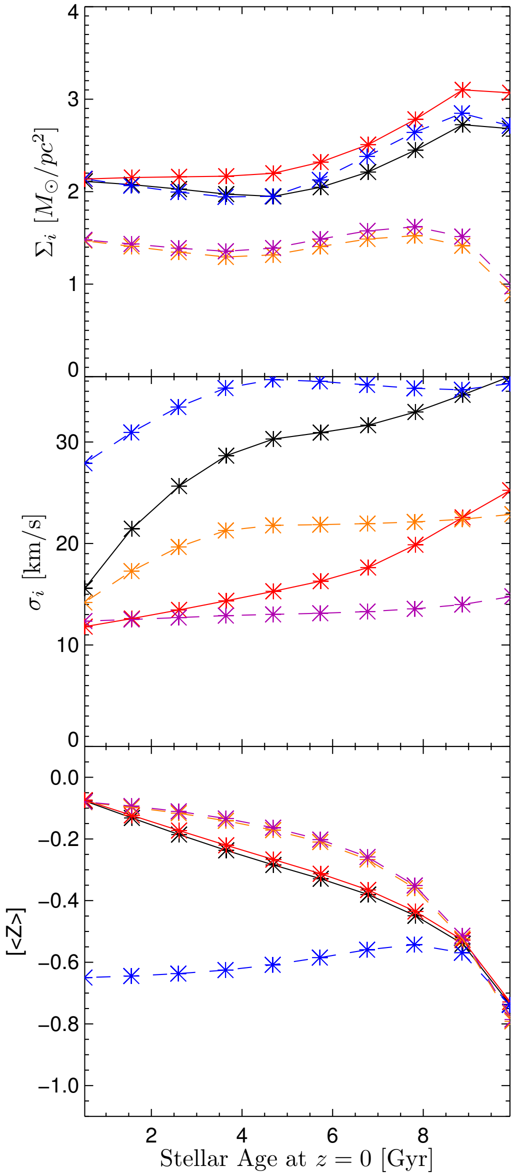

Figure 3.5: Properties of stellar populations as a function of their age at a radius of 8 kpc at redshift zero. Note that the stars comprising the initial condition of the disk are not plotted here. Each line shows the result of a different model: the fiducial model (black), stellar migration off (red), high constant accretion (orange), low constant accretion (blue), stellar migration off and low constant accretion (purple). The models with constant accretion history are dashed. Every simulation produces an age-velocity dispersion correlation via some combination of increasing $\sigma_{*,i}$ of existing stars or decreasing σ, which makes the younger stars dynamically cooler.

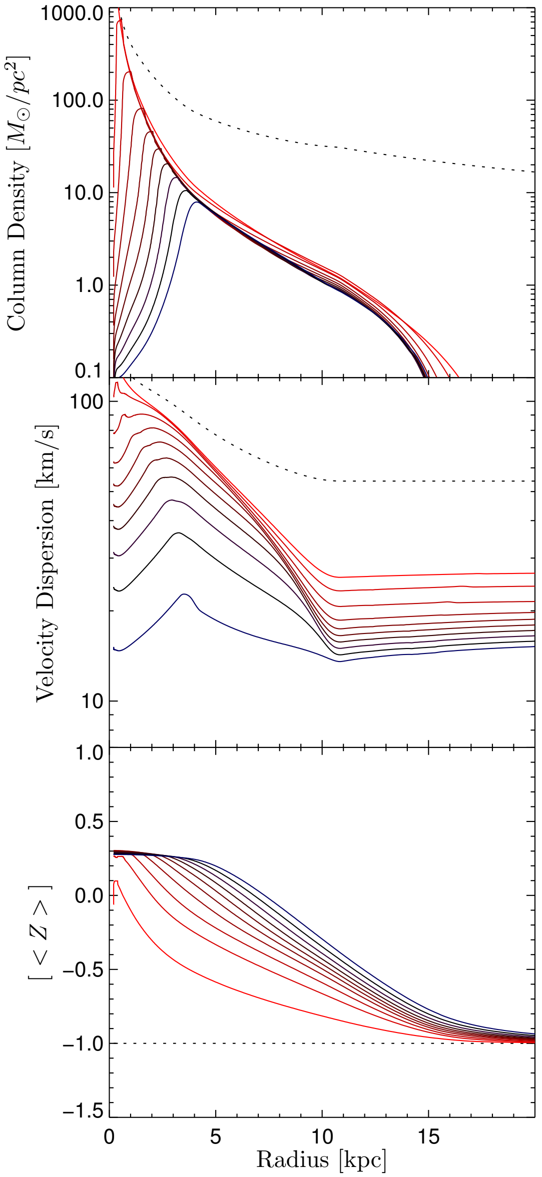

Figure 3.6: All stellar populations produced in the fiducial model at redshift zero, colored by their age with redder stars older. The ages are linearly spaced in time, so each population is about 1 Gyr of star formation. The dotted lines represent the initial population of stars, which has only evolved via stellar migration over the whole course of the simulation. Each newer population is less massive, dynamically colder, and has a steeper metallicity gradient than its older analogues.

3.7 Discussion

Starting from conservation laws and simple assumptions about the gravitational stability of the disk, we have derived evolution equations for the radial profile of a two-component disk. Compared to semi-analytic models, this approach has the advantage that the vast variation in the state variables as a function of radius is resolved rather than averaged over the whole disk. This improvement comes with additional computational costs; however, these are not severe - even using the full Rafikov Q and multiple stellar populations, the code can evolve a disk from $z = 2$ to $z = 0$ on a single processor in a few days, and using the Romeo & Wiegert (2011) approximation to Q reduces the computation time to under an hour.

This paper is primarily meant to introduce our methodology. However, the fiducial model demonstrates a key point which is often overlooked in galaxy evolution and studies of the thick disk, namely that thick disks need not be formed from thin disks. An age-velocity dispersion correlation appears in our simulation, not because of external perturbers, mergers, or gradual heating of a thin disk, but because σ decreases with time and newly formed stars induce transient instabilities in the disk (see also Burkert et al., 1992). Both of these processes are strongly dependent on the cosmological situation in which the disk finds itself, that is, its accretion history. Simulations of isolated thin disks which are gradually heated are therefore unrealistic, in the sense that they are missing the most important drivers of thick disk formation. The smooth increase in stellar velocity dispersion with age produced in our simulations agrees qualitatively with recent observations which demonstrate the lack of a distinctive bimodality between thick and thin disk stars (Bovy et al., 2012a).

This approach has several further applications which we intend to explore in future work. For Milky Way-like galaxies, even modern chemodynamical models with sophisticated treatments of stellar migration and evolution rely on highly parameterized treatments of gas inflow in the disk (Schönrich & Binney, 2009; Spitoni & Matteucci, 2011). If the gas evolves to keep the disk marginally gravitationally unstable, its movement in the disk is not this simple - it depends on the evolution of the full non-linear set of equations we have derived here. By accounting for the diffusion of stars in radius as the result of scattering across corotation resonances (Sellwood & Binney, 2002), our model could be extended to model the Milky Way in detail and compare directly with observations of the metallicity gradient as a function of height above the disk (Cheng et al., 2012), the age-velocity dispersion correlation (Holmberg et al., 2009), the age-metallicity relation or lack thereof (Edvardsson et al., 1993),and the radial and vertical stellar density distributions (Bovy et al., 2012b).

Galaxy bimodality - the separation of galaxies into a blue cloud of star-forming galaxies and a red sequence of ellipticals - is often viewed as an evolutionary sequence. Blue cloud galaxies gradually accrete gas and smaller galaxies, which fuel star formation. Some of these galaxies will undergo major mergers, leaving red and dead elliptical galaxies. These early-type galaxies can subsequently undergo dry mergers, which extend the red sequence to include extremely massive galaxies. Beyond this canonical view, Sánchez Almeida et al. (2011) have noted the existence of a significant population of red spirals.

By taking more realistic accretion histories from cosmological simulations, we expect that a certain fraction of disks in the course of their lifetimes will experience a period of low accretion during which they will exhaust their gas supply and become redder, only to return to the blue cloud with the resumption of higher accretion rates. Given a set of realistic non-smooth but quiescent accretion histories, appropriate for a large fraction of sub-L_* galaxies, we may therefore be able to reproduce aspects of phenomenology from the local universe out to $z = 2$ as semi-analytic models do, but with the added benefit of a physical treatment of the disk dynamics.

3.8 Non-dimensional Equations

For the purposes of implementing the governing equations in a numerical code, it is useful to non-dimensionalize the equations. To do so is straightforward, and basically amounts to rescaling lengths to the radius of the disk, velocities to the circular velocity, and mass fluxes to the initial accretion rate of gas from the IGM. We can make the following substitutions, following KB10: $r = xR$, $t = T[2\pi R/v_\phi(R)]$, $\mathcal{T} = \tau \dot{M}_{ext,0} v_\phi(R) R$, $\sigma_j = s_j v_\phi(R)$, and $\Sigma_j = S_j \dot{M}_{ext,0}/(v_\phi(R)R)$. Here the subscript j may refer to gas or one of possibly many stellar populations. With these substitutions, the gas evolution equations (4.2) and (4.3) become

$$\begin{aligned}
\frac{\partial S}{\partial T} &= \frac{(\beta^2+\beta+x\beta')\tau' - x(\beta+1)\tau''}{(\beta+1)^2 u x^2} - (f_R+\mu)\frac{\partial S_*}{\partial T}^{SF} && (3.65)\\
\frac{\partial s}{\partial T} &= -\frac{s}{3(\beta+1)Sux}\tau'' + \frac{(\beta+\beta^2+x\beta')s - 5s'x(\beta+1)}{3(\beta+1)^2 Sux^2}\tau' \\
&\quad + \frac{u(\beta-1)}{3sSx^3}\tau - \frac{2\pi^2}{3}\eta S K_0 \left(1+\frac{S_*}{S}\frac{s}{s_*}\right)\left(1-\frac{s_t^2}{s^2}\right)^{3/2} && (3.66)
\end{aligned}$$

Primes denote partial derivatives with respect to x, and as with dimensional quantities, S and s with no subscript refer to properties of the gas. The dimensionless initial accretion rate is

$$K_0 = \frac{G\dot{M}_{ext,0}}{v_\phi(R)^3}. \tag{3.67}$$

The dimensionless thermal gas velocity dispersion is s_t.

Employing the same procedure for the evolution equations of each stellar population's column density, we obtain

$$\begin{aligned}
\frac{\partial S_{*,i}}{\partial T} &= f_R \left(\frac{\partial S_{*,i}}{\partial T}\right)^{SF} + \frac{\partial S_{*,i}}{\partial T}^{Mig}, && (3.68)\\
\frac{\partial S_{*,i}}{\partial T}^{SF} &= 8\pi\sqrt{\frac{2}{3}} f_{H_2}\epsilon_{ff} K_0 \frac{S^2}{s}\sqrt{1+\frac{S_*}{S}\frac{s}{s_*}}, && (3.69)\\
\frac{\partial S_{*,i}}{\partial T}^{Mig} &= -2\pi y\left(S_{*,i}\frac{y'}{y} + S'_{*,i} + \frac{S_{*,i}}{x}\right) && (3.70)
\end{aligned}$$

where we have explicitly separated the effects of stellar migration and star formation. The dimensionless radial component of the bulk stellar velocity is $y = v_{r*}/v_\phi(R)$.

Similarly, the velocity dispersion evolution equations are

$$\begin{aligned}
\frac{\partial s_{*,i}}{\partial T} &= \frac{\partial s_{*,i}}{\partial T}^{SF} + \frac{\partial s_{*,i}}{\partial T}^{Mig}, && (3.71)\\
\frac{\partial s_{*,i}}{\partial T}^{SF} &\approx f_R \frac{1}{2S_* s_*}(s^2 - s_*^2)\frac{\partial S_*}{\partial T}^{SF}, && (3.72)\\
\frac{\partial s_{*,i}}{\partial T}^{Mig} &= -2\pi y\left(\frac{(1+\beta)u^2}{3xs_*} + s'_*\right) && (3.73)
\end{aligned}$$

The change in velocity dispersion as a result of star formation is only an approximate relation, since it relies on a first order Taylor series expansion of the exact change in $s_{*,i}$, which in turn requires that $S_{*,i} \gg (\partial S_{*,i}/\partial T)^{Mig} dT$. This condition cannot be satisfied when a completely new population of stars is formed as the simulation crosses into a new age bin, at which time $S_{*,i} = 0$. Therefore we use the exact relation,

$$s_{*,i,new} = \sqrt{\frac{(S_{*,i} s_{*,i}^2)_{old} + f_R(dS_{*,i}^{SF}) s^2}{S_{*,i,old} + f_R(dS_{*,i}^{SF})}} \tag{3.74}$$

where $dS_{*,i}^{SF} = dT(dS/dT)^{SF}$

Finally we have the equations describing the transport of metals in the gas,

$$\frac{\partial Z}{\partial T} = -\frac{2\pi}{(\beta+1)xSu} \frac{\partial \ln Z}{\partial x} \tau' + \frac{y_M(1-f_R)}{S} \frac{\partial S_*}{\partial T}^{SF} \tag{3.75}$$

and in a stellar population,

$$\frac{\partial Z_{*,i}}{\partial T}^{Mig} = -2\pi y S'_{*,i}. \tag{3.76}$$

The stellar metallicity change owing to the formation of new stars can be computed exactly as

$$Z_{*,i,new} = \frac{(S_{*,i} Z_{*,i})_{old} + f_R(dS_{*,i}^{SF}) Z}{S_{*,i,old} + dS_{*,i}^{SF}} \tag{3.77}$$

These equations, given a torque τ and a radial stellar velocity y, fully describe the evolution of the system. To obtain these two quantities, we imposed conditions on the evolution of Q and Q_* (equations 3.16 and 3.34). In dimensionless form these partial differential equations are

$$\begin{aligned} y' + y\left(-\frac{u^2}{s_*^2}\frac{(1+\beta)}{3x} - \frac{s_*'}{s_*} + \frac{S_*'}{S_*} + \frac{1}{x}\right) &= \frac{\max(Q_{lim} - Q_*, 0)u}{2\pi x T_{mig} Q_*} && (3.78) \\ g_2\,\tau'' + g_1\,\tau' + g_0\,\tau &= g_F && (3.79) \end{aligned}$$

where the coefficients of the dimensionless torque equation are

$$\begin{aligned} g_2 &= -\frac{s}{3xSu(\beta+1)}\frac{\partial Q}{\partial s} - \frac{1}{(\beta+1)xu}\frac{\partial Q}{\partial S} && (3.80) \\ g_1 &= \frac{\beta^2 s + s(x\beta' + \beta) - 5(\beta+1)xs'}{3(\beta+1)^2x^2uS}\frac{\partial Q}{\partial s} + \frac{\beta(\beta+1) + x\beta'}{(\beta+1)^2x^2u}\frac{\partial Q}{\partial S} && (3.81) \\ g_0 &= \frac{u(\beta-1)}{3x^3sS}\frac{\partial Q}{\partial s} && (3.82) \\ g_F &= \frac{2\pi^2}{3}\eta S K_0\left(1 + \frac{S_*}{S}\frac{s}{s_*}\right)\left(1 - \frac{s_t^2}{s^2}\right)^{3/2}\frac{\partial Q}{\partial s} + (f_R + \mu)\frac{\partial S}{\partial T}^{SF}\frac{\partial Q}{\partial S} && (3.83) \\ &\quad -\sum_i\left(\frac{\partial S_{*,i}}{\partial T}\frac{\partial Q}{\partial S_{*,i}} + \frac{\partial s_{*,i}}{\partial T}\frac{\partial Q}{\partial s_{*,i}}\right) \end{aligned}$$

Both partial differential equations require an outer boundary condition, which essentially specifies the flux of each type of material at the edge of the disk. The mass flux of the gas is specified by some accretion history $\dot{M}_{ext}(t)$,

$$\tau'(x=1) = -\left(\frac{\dot{M}_{ext}(t)}{\dot{M}_{ext,0}}\right)(1 + \beta(x=1)), \tag{3.84}$$

while the flux of stars is set to zero via $y(x=1) = 0$. The torque equation also requires an inner boundary condition, which we take to be $\tau(x = x_0) = 0$

Chapter 4

Radial Gas Structure and Central Quenching

4.1 Introduction

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

Numerous surveys have shown that the specific star formation rates (sSFR, the star formation rate divided by the stellar mass) of Milky Way (MW) mass galaxies have decreased by roughly a factor of 20 since $z = 2$. With the wide acceptance of ΛCDM cosmology, which entails the hierarchical growth of dark matter haloes, it became common lore that mergers were a major driver of this dramatic change in the nature of galaxies. More recently though, the small scatter in the correlation between the stellar mass M_* and the star formation rate (the star forming main sequence) for galaxies out to $z = 2$ has suggested that most stellar mass growth occurs in galaxies that are not undergoing dramatic merger events, but rather in typical-looking disks (e.g. Noeske et al., 2007; Rodighiero et al., 2011; Kaviraj et al., 2013). Maps of Hα emission in main sequence galaxies confirm that star formation occurs in radially extended disks at $z \sim 1$ (Nelson et al., 2013).

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

High resolution hydrodynamical simulations (Bournaud & Elmegreen, 2009; Ceverino et al., 2010) have strongly suggested that the reason these galaxies are so different from low-redshift disks is rapid gas accretion from the cosmic web through cold dense filaments (Dekel et al., 2009b), which in turn leads to galaxies with low values of the Toomre Q parameter (Toomre, 1964),

$$Q_{\rm Toomre} = \frac{\kappa \sigma_d}{\pi G \Sigma_d} \tag{4.1}$$

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

Alternative scenarios for driving the turbulence and producing clumps have been explored by other authors. Genel et al. (2012) constructed a simple model for the scenario in which the turbulence is driven by the kinetic energy of material as it accretes onto the disk (see also Elmegreen & Burkert, 2010). The details of the origin of the clumpy morphologies has also come under recent theoretical and observational investigation, and the importance of ex-situ clumps from minor mergers is not negligible (Mandelker et al., 2014). Supernovae (Joung et al., 2009), radiation pressure (Krumholz & Thompson, 2012; Krumholz & Thompson, 2013), and the two working in tandem (Agertz et al., 2013), being the primary sources of energy outside of gravitational potential energy, have also been studied as drivers of turbulence and outflows.

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

In this work we build on the physical picture presented in simple toy models (Dekel et al., 2009a; Cacciato et al., 2012) of the gravitational instability and how it evolves over time. Krumholz & Burkert (2010) developed a formalism to show how gravitationally unstable disks behave as a function of radius in steady state and how quickly the disks approach steady state. In the previous Chapter Forbes, Krumholz, & Burkert (2012, hereafter F12), we extended the time-dependent numerical model of Krumholz & Burkert (2010) to include star formation, stellar migration, and metallicity evolution to give a realistic picture for how galaxies evolve over cosmological times with all these processes. In this work, rather than focus on the stellar populations, we explore what sets the gas distribution. Our model includes a number of improvements over the models presented in F12 which we discuss in detail in appendix 4.6, and a new stellar migration formalism (appendix 4.7).

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

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

In section 4.2 we review the equations solved by our 1D code. The results of a wide range of simulations done with the code are presented in section 4.3. We discuss the implications in section 4.4 and summarize in section 4.5.

4.2 The GIDGET code

Our one-dimensional disk galaxy evolution code, GIDGET[1], is described in more detail in F12. The code tracks the surface density, velocity dispersion, and metallicity of one gas component and one or more stellar components, as a function of radius and time. The following subsections will describe the evolution equations for these quantities in some detail; we include a comprehensive list of all parameters used in this

---

[1] The source code repository is freely available from `http://www.ucolick.org/~jforbes/gidget.html`

study, defined below, in table 4.1. The most important physical ingredients are star formation, external accretion onto the disk, and radial transport of gas through the disk.

4.2.1 Gas transport and cooling

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

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

$$\frac{\partial \Sigma}{\partial t} = \frac{1}{2\pi r}\frac{\partial}{\partial r}\dot{M} - (f_R + \mu)\dot{\Sigma}_*^{SF} + \dot{\Sigma}_{cos}. \tag{4.2}$$

The first term represents the flow of mass within the disk, where $\dot{M}$ is defined as the net gas mass per unit time moving towards the center of the disk across cylindrical radius r. Typically $\dot{M} > 0$, representing inward mass flux, but negative values at large radii in the disk are generally necessary to conserve angular momentum. The second term of the continuity equation represents gas forming stars. Only a fraction f_R of that gas will remain in stellar remnants, while the remainder will be recycled to the ISM; we approximate this process as instantaneous as suggested by Tinsley (1980). Mass is also

Table 4.1: An exhaustive list of all parameters used in Chapter 4

Parameter	Fiducial Value	Plausible Range	Description
			Gas Migration (section 4.2.1)
η	1.5	0.5 – 4.5	(3/2) kinetic energy dissipation rate per scale-height crossing time
Q_{GI}	2	1–3	Marginally stable value of Q
T_{gas}	7000 K	3000–10^4	Gas temperature; sets the minimum gas velocity dispersion
α_{MRI}	0.01	0–0.1	Value of $T_{r\phi}/\rho\sigma_{\mathrm{sf}}^2$ without gravitational instability
			Rotation Curve (section 4.2.2)
v_{circ}	220 km s^{-1}	180–250	Circular velocity in flat part of rotation curve
r_b	3 kpc	0–10 kpc	Radius where rotation curve transitions from powerlaw to flat
β_0	0.5	0–1	Powerlaw slope of $v_\phi(r)$ at small radii
n	2	1–5	Sharpness of the transition in the rotation curve
			Star Formation (section 4.2.3)
ϵ_{ff}	0.01	.003–.03	Star formation efficiency per freefall time in the Toomre regime
$f_{\mathrm{H}_2,min}$	0.03	.01–.1	Minimum f_{H_2}.
t_{SC}	2 Gyr	1–3 Gyr	Depletion time of H_2 in the single cloud regime
f_R	0.54	.4+	Mass fraction of a zero-age stellar population not recycled to the ISM
μ	0.5	0–2	Galactic winds' mass loading factor
			Metallicity (section 4.2.4)
y	.054	.05–.07	Mass of metals yielded per mass locked in stellar remnants
ξ	0	0–1	Metallicity enhancement of galactic winds
Z_{IGM}	$0.1Z_\odot = 0.002$	$(.01\text{–}1)Z_\odot$	Metallicity of initial and infalling baryons
k_Z	1	.3–3	Amplitude of metallicity diffusion relative to (Yang et al., 2012)
			Stellar Migration (appendix 4.7)
Q_{lim}	2.5	2–3	Value of Q_* below which spiral instabilities will heat the stars
T_{mig}	4	2–5	Number of local orbital times over which stars are heated by spiral instabilities
			Accretion (section 4.2.5)
$M_{h,0}$	$10^{12}M_\odot$	-	Halo mass at $z = 0$
$\Delta\omega$	0.5	0.1–1	Interval of $\omega \sim z$ over which accretion rate is constant
$r_{\mathrm{acc}}(z=0)$	6.9 kpc	3–20 kpc	Scale length of new infalling gas
β_z	0.38	0–1	Scaling of efficiency with $(1+z)$
β_{M_h}	-0.25	-1–0	Scaling of efficiency with halo mass
ϵ_0	0.31	$\sim$ 0 - .5	Efficiency at $M_h = 10^{12}M_\odot$, $z = 0$
ϵ_{max}	1	0.5–1	Maximum value of efficiency
			Initial Conditions (section 4.2.6)
α_r	1/3	0–1	Scaling of accretion scale length with halo mass
$f_{g,0}$	0.5	0.2–0.7	Initial gas fraction
f_{cool}	1	0.4–1	Fraction of $f_b M_h(z = z_{relax})$ contained in the initial disk
z_{relax}	2.5	2–3	z at which the simulation is initialized
ϕ_0	1	1–5	Initial ratio of stellar to gaseous velocity dispersion
			Computational Domain (see F12)
x_0	.004	$0 < x_0 \ll 1$	Inner edge of domain as a fraction of R
R	40 kpc	10–100 kpc	Outer edge of domain
n_x	200	$\gtrsim 100$	Number of radial cells
tol	10^{-4}	$10^{-5} - -x_0$	Fastest change allowed in state variables, per orbital time at $r = R$
			Cosmology (section 4.2.5)
Ω_m	0.258	-	Average present-day matter energy density as a fraction of the critical density
$1 - \Omega_m - \Omega_\Lambda$	0	-	Deviation from a flat universe
f_b	0.17	-	Universal baryon fraction
H_0	72 km s^{-1} Mpc^{-1}	-	Hubble's constant
σ_8	.796	-	Normalization of the dark matter power spectrum

ejected at each radius in galactic scale winds in proportion to the star formation rate, with mass loading factor μ. Finally, $\dot{\Sigma}_{cos}$ represents the rate of cosmological accretion onto the disk. The winds are assumed to escape the galaxy, though in principle they could be re-accreted later through this final term.

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

$$\frac{\partial \sigma}{\partial t} = \frac{\mathcal{G} - \mathcal{L}}{3\sigma\Sigma} + \frac{\sigma}{6\pi r \Sigma}\frac{\partial}{\partial r}\dot{M} + \frac{5(\partial\sigma/\partial r)}{6\pi r\Sigma}\dot{M} + \frac{(\beta - 1)v_\phi}{6\pi r^3 \Sigma\sigma}\mathcal{T}. \tag{4.3}$$

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

The viscous torque is related to the mass flux via the conservation of angular momentum, as derived from the $\phi-$component of the Navier-Stokes equations:

$$\dot{M} \equiv -2\pi r \Sigma v_r = -\frac{1}{v_\phi(1+\beta)}\frac{\partial \mathcal{T}}{\partial r}. \tag{4.4}$$

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

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

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

$$\begin{aligned} \frac{dQ}{dt} &= \frac{\partial \Sigma}{\partial t}\frac{\partial Q}{\partial \Sigma} + \frac{\partial \sigma}{\partial t}\frac{\partial Q}{\partial \sigma} + \\ &\quad \frac{\partial \Sigma_*}{\partial t}\frac{\partial Q}{\partial \Sigma_*} + \frac{\partial \sigma_{rr}}{\partial t}\frac{\partial Q}{\partial \sigma_{rr}} + \frac{\partial \sigma_{zz}}{\partial t}\frac{\partial Q}{\partial \sigma_{zz}} \\ &= f_{\mathrm{transport}}\left(\Sigma, \sigma, \Sigma_*, \sigma_{rr}, \sigma_{zz}, \mathcal{T}_{GI}, \frac{\partial \mathcal{T}_{GI}}{\partial r}, \frac{\partial^2 \mathcal{T}_{GI}}{\partial r^2}\right) \\ &\quad + f_{\mathrm{source}}\left(\Sigma, \sigma, \Sigma_*, \sigma_{rr}, \sigma_{zz}, Z\right). \end{aligned} \tag{4.5}$$

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

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

The gravitational stability of disks to linear axisymmetric perturbations is roughly determined by the value of Q_{Toomre}. Modern versions of this parameter take into account both gas and stars (e.g. Rafikov, 2001; Romeo & Falstad, 2013), the finite thickness of the disk (Shu, 1968; Romeo, 1992, 1994; Elmegreen, 2011), gas turbulence (Hoffmann & Romeo, 2012) and the fact that gas which can cool to arbitrarily small scales is never formally stable (Elmegreen, 2011). Romeo & Wiegert (2011) have developed an approximate, but analytic, formula for Q taking into account two components of finite thickness, i.e. both gas and stars. In this way the gravitational effects of the stars are included in the instability, subject to what we assume about how the stars self-regulate Q_* (see appendix 4.7). To account for the final complication, we demarcate the stable from the unstable values of Q at $Q_{GI} = 2$, rather than the canonical value of unity, as suggested by Elmegreen (2011). This approximation to Q and its partial derivatives with respect to Σ, σ, Σ_*, and σ_* is extremely cheap to compute, which is advantageous since all of these values must be computed at each (unstable) radius and time to solve $f_{\text{transport}} = -f_{\text{source}}$.

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

$$\mathcal{L} = \eta\Sigma\sigma^2\kappa Q_g^{-1}\left(1 + \frac{\sigma\Sigma_*}{\sigma_{zz}\Sigma}\right)\left(1 - \frac{\sigma_{\text{sf}}^2}{\sigma^2}\right)^{3/2}, \tag{4.6}$$

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

4.2.2 Rotation curve

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

$$v_\phi(r) = v_{circ} \left(1 + (r_b/r)^{|\beta_0 n|}\right)^{-\mathrm{sign}(\beta_0)/n}. \tag{4.7}$$

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

4.2.3 Star formation

Stars form with a constant efficiency per freefall time $\epsilon_{\rm ff}$ from molecular gas, so that $\dot{\Sigma}_*^{SF} \sim \epsilon_{\rm ff} f_{\rm H_2}\Sigma/t_{\rm ff}$, where $t_{\rm ff}$ is the freefall time and $f_{\rm H_2}$ is the molecular fraction (Krumholz & Tan, 2007; Krumholz et al., 2012). Following Krumholz et al. (2012), we posit that there are two regimes: one in which the appropriate time-scale is the freefall time of gas distributed over the full scale height H of the disk, namely $t_{\rm ff} = \sqrt{3\pi/32G\rho} \approx \sqrt{3\pi H/32G\Sigma}$, which we call the 'Toomre regime' and one in which the time-scale is determined by the freefall time of individual molecular clouds, which observations suggest is $t_{\rm ff}/\epsilon_{\rm ff} = \Sigma_{\rm H_2}/\dot{\Sigma}_*^{SF} \equiv t_{SC} \approx 2$ Gyr (Bigiel et al., 2011), the 'single cloud regime'. Then the star formation rate is simply set by which of these two time-scales is shorter [2],

$$\dot{\Sigma}_*^{SF} = \max\left(\epsilon_{\rm ff} f_{\rm H_2}\Sigma\kappa\frac{\sqrt{32/3}}{Q_g\pi}\left(1+\frac{\sigma\Sigma_*}{\sigma_{zz}\Sigma}\right)^{1/2},\ \ f_{\rm H_2}\frac{\Sigma}{t_{SC}}\right) \tag{4.8}$$

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

The molecular fraction $f_{\rm H_2}$ is calculated according to the analytic formula of Krumholz et al. (2009). Their formula predicts $f_{\rm H_2}$ as a function of Σ and Z. Roughly

---

[2] Note that F12 omitted the factor of $1/Q_g\pi$, though the code and the appendix with the dimensionless version were correct

speaking, $f_{\mathrm{H}_2} \to 1$ at high surface densities, and below some transition surface density, there is a sharp cutoff where f_{H_2} rapidly approaches zero. This transition is metallicity dependent, roughly $5M_\odot\ pc^{-2}(Z/Z_\odot)^{-1}$. We include the slight modifications to this formula we used in F12, namely a floor of $f_{\mathrm{H}_2} \geq 0.03$ to account for the fact that star formation is observed even at very low surface densities, (Bigiel et al., 2010; Schruba et al., 2011), likely as a result of the requirement that the FUV flux not fall below a certain floor in order for two-phase equilibrium in the atomic ISM to be possible (Ostriker et al., 2010).

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

Once stars form, they also migrate. In our model, this is treated quite similarly to the gas migration discussed in section 4.2.1, namely the stars experience torques if they are gravitationally unstable to spiral instabilities. Our prescription has improved significantly since F12, so we discuss the new governing equations in appendix 4.7. Overall this typically has a minor effect on the dynamics of the disk, although it can strongly influence the stellar velocity dispersions particularly at small radii.

4.2.4 Metallicity

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

$$\begin{aligned}\frac{\partial M_Z}{\partial t} &= \Delta r \frac{\partial \dot{M} Z}{\partial r} + (y f_R - f_R Z - \mu Z_w)\dot{M}_*^{SF} \\ &\quad + \dot{M}_{acc} Z_{IGM} + \frac{\partial}{\partial r}\kappa_Z \frac{\partial}{\partial r} M_Z. \end{aligned} \tag{4.9}$$

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

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

$$Z < Z_w < \begin{cases} Z + yf_R/(1-f_R) & \text{if } \mu \leq 1-f_R \\ Z + yf_R/\mu & \text{if } \mu > 1-f_R \end{cases} \tag{4.10}$$

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

$$Z_w = Z + \xi \frac{yf_R}{\max(\mu, 1-f_R)}. \tag{4.11}$$

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

The diffusion of metals has received relatively little attention until recently. In F12, we included this diffusion term to prevent the metallicity gradient from steepening excessively, tuning the value of κ_Z to yield a reasonable gradient. Since then, Yang et al. (2012) have measured the value of κ_Z in a 2D shearing box simulation with turbulence driven by thermal instability. They show that to a reasonable approximation $\kappa_Z \propto r_{inj}^2/t_{orb}$, where r_{inj} is the initial wavelength of the metallicity perturbation, and

t_{orb} is the orbital time. Here we make the approximation that $r_{inj} \approx \lambda_J = \sigma^2/G\Sigma$, the 2D Jeans length, since this should be similar to the spacing of the largest giant molecular clouds. We can therefore scale κ_Z in our simulation to their measured value, as

$$\kappa_Z(r,t) = k_Z 1.2\, \frac{\mathrm{kpc}^2}{\mathrm{Gyr}} \left(\frac{\sigma^2/G\Sigma}{3.1\ \mathrm{kpc}}\right)^2 \frac{\kappa}{\sqrt{2}\ (26\ \mathrm{km\ s^{-1}kpc^{-1}})} \tag{4.12}$$

The numerical values are the measured κ_Z and the input parameters r_{inj} and Ω quoted for one of their simulations. We also include a free parameter $k_Z \approx 1$, recognizing that there is some uncertainty in this result. The numerical implementation of the diffusion term is operator-split from the rest of the terms, implicit, and computed in terms of fluxes so that metal mass is explicitly conserved. We also enforce $\kappa_Z < v_{circ}R$, the largest velocity and radius in the problem, which is not guaranteed by equation 4.12 when Σ is very small. This essentially makes sure that the metal injection scale $r_{inj} \lesssim R$, the size of the system.

4.2.5 Accretion

In our model, gas accretes onto the disk at an externally-prescribed rate $\dot{M}_{ext}$ and a profile $\dot{\Sigma}_{cos}$ such that

$$\dot{M}_{ext}(t) = \int_0^\infty \dot{\Sigma}_{cos}(r,t) 2\pi r dr. \tag{4.13}$$

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

$$r_{acc}(z) = r_{acc}(z=0)\left(\frac{M_h(z)}{M_h(z=0)}\right)^{\alpha_r}, \tag{4.14}$$

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

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

$$\dot{M}_h = 39(M_h/10^{12}M_\odot)^{1.1}(1+z)^{2.2}M_\odot/yr, \tag{4.15}$$

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

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

$$S(M_h) = u^2 \left(\frac{c_0 \Gamma}{\Omega_m^{1/3}} \left(\frac{M_h}{1 M_\odot} \right)^{1/3} \right) \frac{\sigma_8^2}{u^2(32\Gamma)}. \tag{4.16}$$

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

$$\begin{aligned} u(x) &= 64.087 \left(1 + 1.074\, x^{0.3} - 1.581\, x^{0.4} \right. \\ &\quad \left. + 0.954\, x^{0.5} - 0.185\, x^{0.6}\right)^{-10}. \end{aligned} \tag{4.17}$$

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

$$\omega(z) = 1.260 \left(1 + z + 0.09(1+z)^{-1} + 0.24 e^{-1.16z}\right). \tag{4.18}$$

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

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

$$\Delta S = \exp\left(x\sigma_k + \mu_k\right) \tag{4.19}$$

where x is a value drawn from a normal distribution with zero mean and unity variance. We use a fixed $\Delta\omega = 0.1$, since this is the timestep used in generating the fitting formulae for σ_k and μ_k in Neistein & Dekel (2008). The fact that we use a fixed $\Delta\omega$ rather than a distribution leads to the distinct steps in Fig. 4.1, where all of the accretion histories change at once.

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

$$\sigma_k = 1.367 + 0.012\log_{10} S + 0.234(\log_{10} S)^2 \tag{4.20}$$

$$\mu_k = -3.682 + 0.76\log_{10} S - 0.36(\log_{10} S)^2 \tag{4.21}$$

Converting each value of S back to M_h one obtains a dark matter accretion history $M_h(\omega_j)$, where the ω_j are the sequence of ω's obtained by incrementing ω by the fixed $\Delta\omega$, namely $\omega_j = \omega_0 + j\Delta\omega$ for $j = 0, 1, 2, ...$ and $\omega_0 = \omega(z = 0)$. We require that the change in M_h over a single step, $M_h(\omega_i) - M_h(\omega_{i+1})$, not exceed $M_h(\omega_{i+1})$ to avoid galaxies 'accreting' a larger mass than their own, i.e. becoming a satellite. Since equations 4.20 and 4.21 were obtained by a fit to the Millenium Run using a cosmology where $(\Omega_m, \sigma_8) = (0.25, 0.9)$, when converting between M_h and S with equation 4.16, we use the parameters from the Millenium Run. Once we have obtained $M_h(\omega_j)$ using this cosmology, we can transform it so that it agrees with the WMAP5 (Komatsu et al., 2009) cosmology $(\Omega_m, \sigma_8) = (0.258, 0.796)$, which is much closer to the current best-fitting values. We use the scaling obtained in Neistein et al. (2010) via a comparison of merger trees from Millenium and an N-body simulation run with WMAP5 cosmology, namely we replace the ω_j with $\tilde{\omega}_j = \omega_0 + 0.86 j \Delta\omega$. The full dark matter mass history of the halo $M_h(t)$ is then obtained by converting the $\tilde{\omega}_j$ to z (with equation 4.18) and subsequently to t, and linearly interpolating the sequence of halo masses $M_h(\tilde{\omega}_j)$ in time. The dark matter accretion history is then just the instantaneous derivative of $M_h(t)$.

The input $\dot{M}_{ext}(t)$ to our simulation is taken to be the average dark matter accretion rate at time t, generated either from the smooth accretion formula 4.15 or the lognormal one 4.19 times $f_b \epsilon_{in}$. For the efficiency, we use a reasonably general parameterization,

$$\epsilon_{in}(M_h, z) = \min\left(\epsilon_0 \left(\frac{M_h}{10^{12}}\right)^{\beta_{M_h}} (1+z)^{\beta_z}, \quad \epsilon_{max}\right) \tag{4.22}$$

Faucher-Giguère et al. (2011) fit the results of a cosmological SPH simulation with no feedback to find $(\epsilon_0, \beta_{M_h}, \beta_z, \epsilon_{max}) = (0.31, -0.25, 0.38, 1)$, though they explicitly only use this fit *above* $z = 2$.

Despite its success at high redshift, the paradigm of cold accretion is fairly uncertain for galaxies which have some hot coronal gas, like the Milky Way, at low redshift. Moreover, Diemer et al. (2013) have pointed out that below $z \sim 1$, nearly all the growth in M_h for haloes with $M_h(z = 0) \sim 10^{12} M_\odot$ corresponds to the fact that

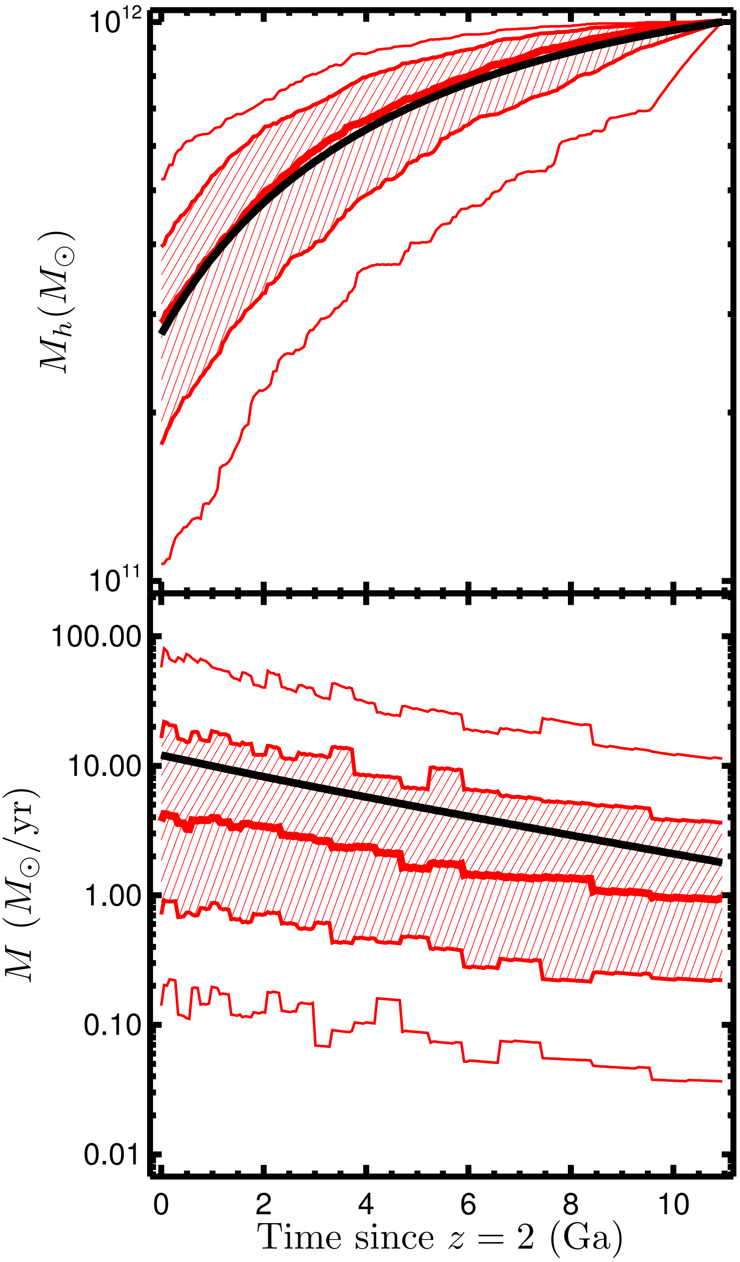

Figure 4.1: The growth of halos. The top panel shows the evolution of the halo mass for the smooth accretion history (black) and the median, central 68 per cent (shaded), and central 95 per cent of 400 stochastic accretion histories (red). The corresponding distribution of the inferred baryon accretion rates, including the efficiency factor (equation 4.22), is shown in the bottom panel. The steps in $\dot{M}$ correspond to our fixed interval $\Delta\omega = 0.086$.

the background density of the universe is decreasing (roughly as $\rho_m \propto (1+z)^3$) while dark matter halos are changing very little. Because the halos are defined in simulations as having a spherical overdensity relative to the background of ~ 200, relatively static halos increase their mass merely because of this drop in the background density. Dekel et al. (2013) have verified that this is not a significant effect at $z > 1$.

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

Given the uncertainties in how gas is accreted at low redshift, our naive approach of setting $\dot{M}_{ext} = \dot{M}_h f_b \epsilon_{in}$ is not unreasonable. In our fiducial model, a MW-mass galaxy accretes roughly $2M_\odot \ yr^{-1}$ at redshift zero, and so yields a star formation rate similar to observations, even if the physical mechanism for this accretion is unclear.

We do retain, in varying the parameters of the accretion efficiency and the accretion profile, a considerable amount of flexibility in the model, which is appropriate given the uncertainties. In Fig. 4.1, we show $M_h(t)$ and the resulting $\dot{M}_{ext}$ for the fiducial smooth model and the stochastic accretion model.

4.2.6 Initial conditions at $z \sim 2$

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

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

$$\Sigma = f_{g,0} f_{cool} f_b \frac{M_h(z = z_{start})}{2\pi r_{IC}^2} \exp\left(-r/r_{IC}\right) \frac{1}{1 - f_{out}} \tag{4.23}$$

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

$$f_{out} = \frac{1}{2\pi r_{IC}^2} \int_R^\infty 2\pi r \; e^{-r/r_{IC}} dr \tag{4.24}$$

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

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

Typically our initial conditions have $Q = Q_{GI}$ in some annulus. At larger radii Σ drops off quickly so $Q \propto \Sigma^{-1}$ increases, while Q also increases at smaller radii through the dependence $\kappa \propto v_\phi/r$. We discuss the (lack of) sensitivity of our results to these choices of initial conditions in appendix 4.8.

4.3 Simulation Results

In this section we discuss some generic features of the galaxies produced by our model. We begin by exploring models with smooth accretion histories and a fiducial choice of parameters, which we summarize in Table 4.1. These are compared with artificial, illustrative models where one important physical ingredient is turned off by hand. We then allow the accretion histories to vary stochastically in a cosmologically realistic way, illustrating the differences between galaxies with identical physical laws but different accretion histories, as one might expect for real galaxies. Finally we compare our models with recent observational results.

4.3.1 Equilibria in smoothly accreting models

There are three terms in the continuity equation (equation 4.2). At a particular radius, gas arrives via $\dot{\Sigma}_{cos}$, departs via $(f_R + \mu)\dot{\Sigma}^{SF}$, and moves to or from other radii via $\dot{\Sigma}_{\mathrm{tr}} \equiv (2\pi r)^{-1}(\partial \dot{M}/\partial r)$. The generic behavior of this equation at a given radius in our fiducial model is that gas will build up, either via direct accretion or as mass arrives from somewhere else in the disk, until an equilibrium is reached such that $\dot{\Sigma} \approx 0$. This equilibrium will then slowly evolve with time as the global gas accretion rate $\dot{M}_{ext}$ falls off.

To aid in understanding how this equilibrium emerges, we have run three

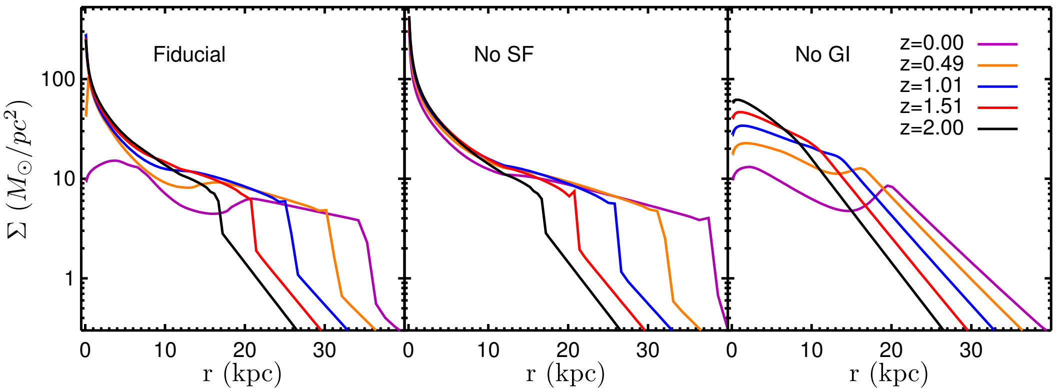

Figure 4.2: The three simplified models. For each of the these simplified models, we show the evolution of the radial gas surface density distribution. We see that the evolution of the fiducial model is a non-trivial combination of the effects of star formation and GI transport.

simple models with identical smooth accretion histories - (i.) the fiducial model - our best guess for physical parameters which will lead to something resembling the Milky Way (see Table 4.1), (ii.) the same model with no star formation, and (iii.) the same model with no gravitational instability, i.e., $\mathcal{T}_{GI} = 0$ everywhere. The features of models i. and ii. are similar at large radii, while the features of i. and iii. bear some resemblance at small radii. This immediately suggests that GI transport is important at large radii and star formation is important at small radii. The gas surface density distributions of each model are shown in Fig. 4.2 as a function of time. The gas is supplied via an exponential distribution, $\dot{\Sigma}_{cos} \propto e^{-r/r_{acc}}$. Without gravitational instability (model iii), star formation carves out the inner parts of the distribution, leaving a hole in the gas at galactic centers, while without star formation (ii), gas is redistributed into a powerlaw distribution, following roughly $\Sigma \propto 1/r$.

A useful way to understand what sets the surface density is to examine the relative effects of each term in the continuity equation. In particular, at each time and

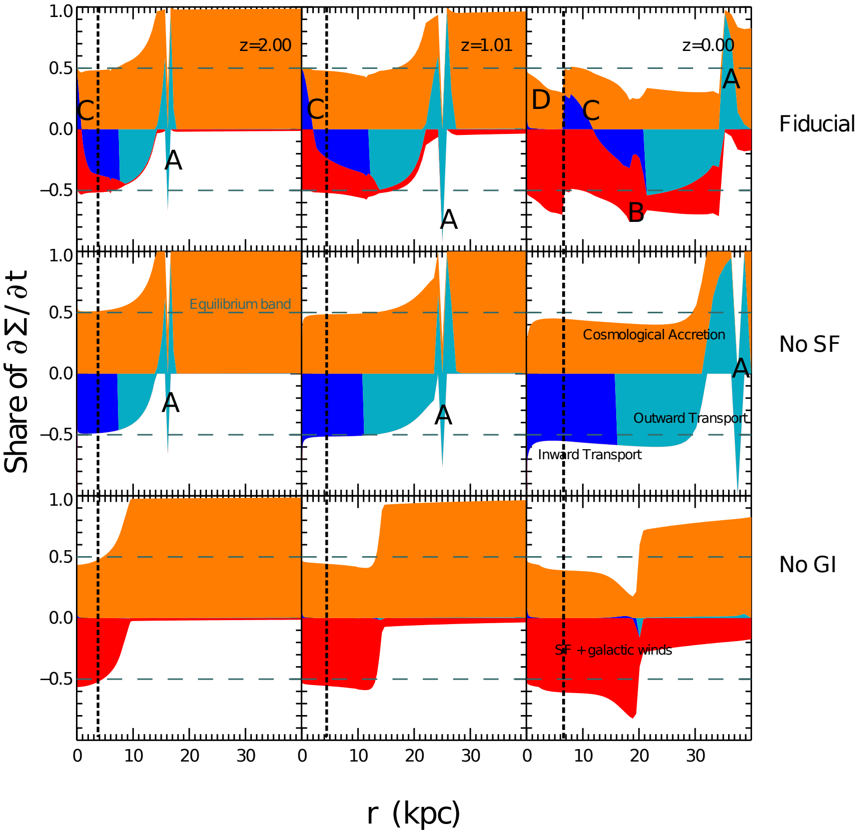

Figure 4.3: The balance of terms in the continuity equation. The terms contributing to $\partial\Sigma/\partial t$ are split into those which increase Σ at a particular radius and time and those which decrease Σ. The former appear above zero and the latter below. Each term is represented by a different color - orange for $\dot{\Sigma}_{cos}$ (with the exponential scale length marked as a vertical dashed line), red for $(\mu + f_R)\dot{\Sigma}_*^{SF}$, and blue for $\dot{\Sigma}_{\rm trans}$. Light (dark) blue indicates gas being transported outwards (inwards). At each radius the height of the colored band is normalized to unity, and its position shows how close the disk is to equilibrium (equal positive and negative contributions) at that radius – a radius is in equilibrium if the colored band falls exactly between the dashed lines labelled "Equilibrium band". The columns show different redshifts ($z = 2, 1, 0$) and the rows show different models: fiducial, no star formation, and no GI transport. The features labelled "A" at e.g. $r = 15$ kpc and $z = 2$ come from gas from the unstable region heating when it piles up in a single stable cell. Real galaxies will not have such a sharp transition since there will be some breaking of axisymmetry and some overshoot from the unstable region, neither of which we model here. The feature labelled "B" at $z = 0$ around 15-20 kpc in the fiducial model is caused by an enhancement in the surface density distribution - this section of the galaxy has had the direction of GI transport reverse from outward to inward. Features labelled by "C" show where gravitational instability changes from removing gas from a given radius to adding it to that radius (see text for more details), and "D" shows GI quenching, wherein star formation has exhausted the supply of inflowing gas.

radius, we can divide each term by $A \equiv |\dot{\Sigma}_{\rm tr}| + \dot{\Sigma}_{cos} + (\mu + f_R)\dot{\Sigma}^{SF}$. In Fig. 4.3 we compute these contributions, including the sign of their effect on the overall value of $\partial\Sigma/\partial t$, so at each radius the fraction of the colored region occupied by (red, orange, blue) represents the fraction of A from (star formation, cosmological accretion, transport). The different shades of blue show which way the mass is flowing in the disk, i.e. the sign of $\dot{M}$ - dark blue indicates gas flowing towards the center of the disk, and light blue outward motion.

When the colored band in Fig. 4.3 stretches from -0.5 to 0.5, that region of the disk has reached an equilibrium configuration. In each case shown here, the equilibration proceeds from inside outwards. This is a combination of two effects- the especially efficient star formation in the center of the disk, and the fairly centrally-concentrated distribution of accreting gas. The equilibrium does not last forever- at $z = 0$, there can be significant deviations as the disk processes past accretion and the instantaneous accretion rate falls owing to the expansion of the universe on time-scales potentially shorter than the gas depletion time at these large radii.

Equilibrium between SF and accretion: the No GI Model

We first focus on the model with no GI. In this model, at a given radius, gas builds up until the local star formation rate $\propto \Sigma$ can balance the incoming accretion. This happens first in the center of the disk. Not only is the cosmological accretion rate per unit area larger there, but the star formation time-scale is shortest (Fig. 4.4). In this model there is in fact a huge range of depletion times, from roughly 100 Myr at $z = 2$ at small radii to 60 Gyr in the outer disk. There are two effects driving this diversity. For depletion times between 100 Myr and 2 Gyr, the disk is in the Toomre regime of star formation (see equation 4.8), for which the depletion time scales as κ^{-1}. This region is typically small, $\lesssim 3$ kpc, outside of which the time-scale would become longer than 2 Gyr if it continued to follow the κ^{-1} scaling. At this point the disk transitions to the single-cloud regime of star formation. At the transition, the disk still tends to be dominated by molecular gas. In the mostly-molecular but still single-cloud regime, the depletion time is roughly 2 Gyr, the single-cloud molecular depletion time - this can be seen as a flattening in the t_{dep} distribution with radius. There is then a transition from molecular to atomic gas, which accounts for the difference between parts of the disk with a 2 Gyr depletion time and a 60 Gyr depletion time - this maximum depletion time is set by $f_{\mathrm{H}_2,min}$, which is quite uncertain.

A generic feature of the No GI model is that at the edge of the star-forming region, star formation occurs at a slightly faster rate than new gas is accreted at that radius (Fig. 4.3, bottom right panel). All of the models, particularly at lower redshift exhibit a slight tendency to fall just below the 'equilibrium band' after they have initially equilibrated at a given radius, since the accretion rate is externally imposed and falling monotonically. The feature at $z = 0$ in the No GI model goes beyond this, however, and may be explained by a small feedback loop in the star formation law introduced by the dependence on metallicity. The demarcation between the star-forming part of the disk and the outskirts is set by the molecular to atomic transition. Typically the star formation rate at a given radius is able to the incoming material only if the molecular fraction there is above the minimum allowed value - otherwise star formation would be too slow. When enough gas has accumulated to satisfy $f_{\mathrm{H}_2} > f_{\mathrm{H}_2,min} = 0.03$, the star formation rate rises steeply with column density and new metals are produced, which in turn catalyze star formation by reducing the amount of gas needed to maintain a molecular, star-forming phase. Thus the extra gas, which is now no longer necessary for the star formation rate to balance the accretion rate, can be consumed, though this generally takes a significant amount of time, $t_{dep} \gtrsim 2$ Gyr.

Equilibrium between GI transport and accretion: the No SF Model

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

$$\Sigma \approx \Sigma_{GI} \equiv \frac{3}{2} \frac{\sqrt{2(\beta+1)} v_\phi \sigma}{\pi G r Q_{GI}} \tag{4.25}$$

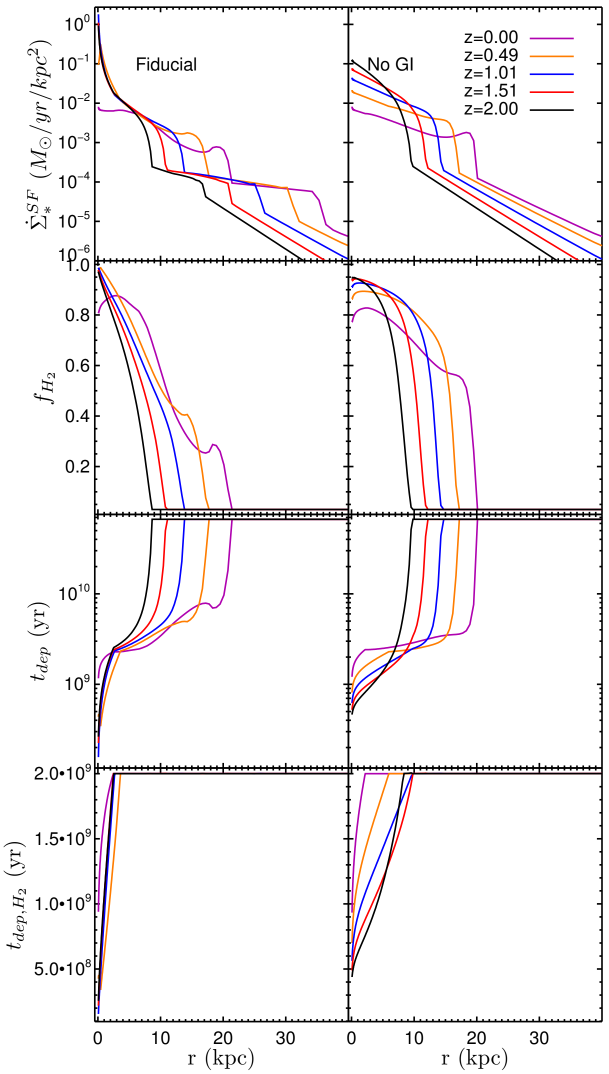

Figure 4.4: Star formation in the fiducial and No GI models. The star formation rate (top) is proportional to the surface density of gas modulated by other factors reflected in the depletion time, $t_{dep} = \Sigma/\dot{\Sigma}_*^{SF}$ (second row). The first is the molecular fraction (third row), itself determined by Σ and Z, and the second is the regime of star formation (bottom row), either single-cloud ($t_{dep,\mathrm{H}_2} = 2$ Gyr), or Toomre ($t_{dep,\mathrm{H}_2} \propto 1/\sqrt{G\rho} < 2$ Gyr). For each quantity, the left panel shows the fiducial model, while the right shows the model with GI transport turned off. Star formation in the fiducial model is much more concentrated and reaches much higher surface densities $\dot{\Sigma}_*^{SF}$ through the action of GI transport. The absence of GI causes so much gas to build up at larger radii that at high redshift the Toomre regime of star formation extends to nearly 10 kpc, instead of just the inner few kpc.

At a given radius, β, v_ϕ, r, and Q_{GI} are all fixed, so equation 4.25 may be considered a direct mapping between Σ and σ. If σ does not vary by much and the velocity dispersions of the gas and stars are similar, then Σ will simply follow a $1/r$ powerlaw over a wide range of radii.

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

$$\Sigma \gtrsim \Sigma_{crit} \equiv \frac{3}{2}\frac{\sqrt{2(\beta+1)}v_\phi\sigma_{\mathrm{sf}}}{\pi G r Q_{GI}}. \tag{4.26}$$

The maximum is determined by the gas supply – for a given $\dot{M}_{ext}$ to be transported to the center of the disk in a quasi-steady state, it must dissipate the gravitational potential energy between where it arrives and the center of the galaxy, and it must experience enough torque to lose its angular momentum. In a steady state, local heating by torques balances local cooling by turbulent dissipation (see section 4.3.1). Note that 'heating' and 'cooling' refer to changing the turbulent velocity dispersion of the gas, not its kinetic temperature. The rate at which the gas cools (and hence experiences torques) $\mathcal{L}$, depends on the velocity dispersion. The maximum velocity dispersion is therefore set by assuming that 100 per cent of the gas arriving from an external source flows towards the center in steady state. Since some gas never reaches the center because of star formation, and other gas moves outwards rather than inwards, this is an upper limit. As shown in section 4.4.1, at $z \sim 2$ for galaxies accreting at $\sim 10 M_\odot\ yr^{-1}$ the velocity

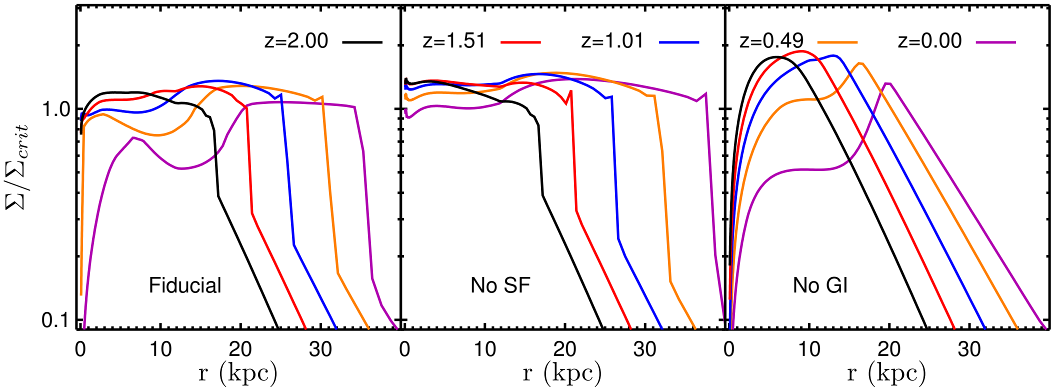

Figure 4.5: The ratio of Σ to the minimum surface density necessary for gravitational instability, Σ_{crit}. Gas spreads out to keep the surface density above this critical value but below the maximum value, Σ_{GI} evaluated at σ_{max}. This ratio falls below ~ 1 where the disk has not yet destabilized (left two panels, large radii) or has stabilized due to GI quenching (left panel, small radii, low redshift). Note however that in the fiducial model, especially at low redshift, Σ can fall slightly below Σ_{crit} even in gravitationally unstable regions because in deriving Σ_{GI} and Σ_{crit} we assumed that $\sigma \approx \sigma_{rr} \approx \sigma_{zz}$, which is no longer true at low redshift. This is a factor of two level effect – the gravitationally stable regions always have Σ well below Σ_{crit}. Interestingly, even the 'No GI' run does not reach values far larger than $\Sigma/\Sigma_{crit} = 1$ (although it reaches significantly higher values than the other two models), since the star formation rate increases as Q_g decreases – essentially the gas is compressed under its own weight and forms stars faster.

dispersion is restricted to 8 km s$^{-1} < \sigma \lesssim 20$ km s^{-1}. This value is low compared to the measured velocity dispersions in the SINS galaxies. As we will see in section 4.4.1, some small fraction of MW-progenitors do have much higher accretion rates in our stochastic accretion model. Moreover, the SINS galaxies are likely somewhat more massive than the MW progenitors we consider here.

As more gas arrives at a region of the disk in a marginally unstable state, the surface density is fixed in the profile given by Σ_{GI}. Since there is a maximum velocity dispersion for a fixed accretion rate, gas is not allowed to accumulate, lest $\sigma \propto \Sigma_{GI}$ exceed this maximum, so the only thing the gas can do is move elsewhere. The gas will then be transported away from where it arrives until it reaches part of the disk which is stable, where it will pile up until that region too becomes unstable. This 'wave' of gravitational instability can be seen propagating outwards in Fig. 4.3 in both the fiducial model and the model without star formation, until essentially the entire disk is unstable. The equilibrium between GI transport and accretion appears originally at $r \leq 7$ kpc both with and without star formation. This location is picked out by the maximum in $\dot{\Sigma}_{acc}/\Sigma_{GI} \propto r\exp(-r/r_{acc})$, i.e. where gas piles up fastest relative to the amount necessary to be gravitationally unstable, which occurs at r_{acc} for a flat rotation curve.

The fiducial model

Having examined the simplified models where we disabled GI transport or star formation, we now turn to our fiducial model which includes both. Recalling the surface density distributions shown in Fig. 4.2, it seems that the fiducial model behaves largely like a superposition of the model without star formation and the model without gravitational instability.

In the previous section we point out that an equilibrium between GI transport and infalling accretion arises when $\Sigma \approx \Sigma_{GI} > \Sigma_{crit}$ (see Fig. 4.5) and more gas is added. The new gas will be whisked away until it piles up somewhere in the disk that is not yet unstable. If we also include star formation, then rather than being pushed out into a stable region, the gas can be consumed by star formation. Comparing the model without star formation to the fiducial model at $z = 2$ and $z = 1$ in Fig. 4.3, we can see this effect in action. Gas arrives around r_{acc}, and on its way inwards it is consumed by star formation. The balance is then between cosmological accretion and both star

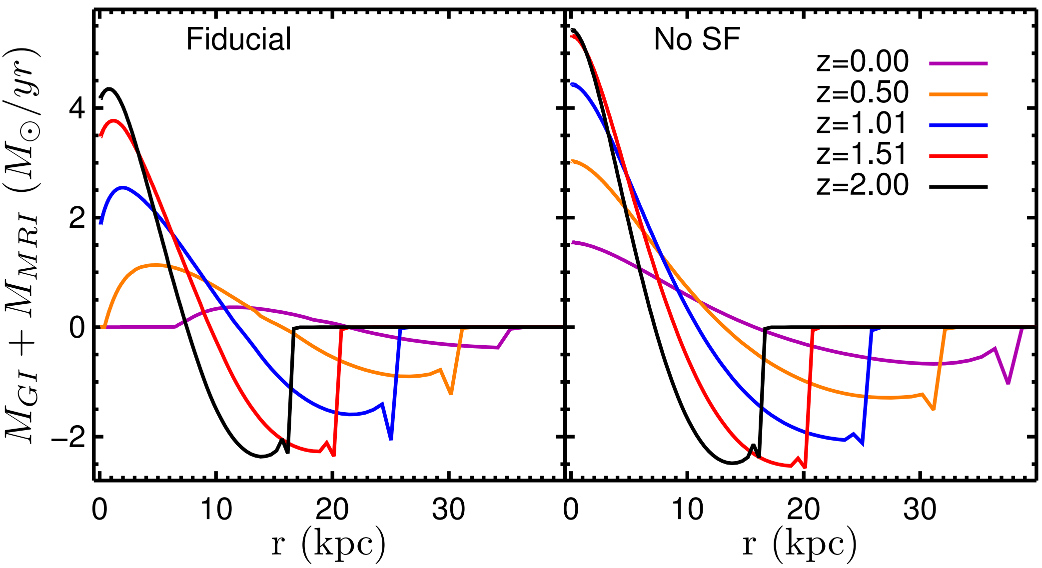

Figure 4.6: The inward mass flux through a given radius. Negative values are outward flux. We compare the fiducial model (left) to the same model with star formation turned off (right). As usual, black, red, blue, orange and purple are $z = 2.0, 1.5, 1.0, 0.5, 0.0$. The outward mass flow is modestly affected by star formation at late times, whereas the inward flow is completely consumed by low redshift. The disk stabilizes in the center, i.e. $\dot{M} \approx 0$, simply because all of the available mass has been consumed by stars.

formation and GI transport, rather than just GI transport alone. In other words, if the disk can get rid of some gas via star formation, it no longer has to transport it away as fast to maintain $\Sigma \approx \Sigma_{GI}$. Eventually all of the infalling gas at a given radius can be consumed by star formation, and GI transport briefly has no net effect. Just interior to this point though, the cosmological accretion rate is low enough and the star formation rate is fast enough that accretion alone can no longer supply the star formation at that radius, and the stars start forming not from material falling directly onto that radius, but from gas arriving from other parts of the disk via GI transport. These are the points in Fig. 4.3 (labelled by "C") where $\dot{\Sigma}_{\rm trans}$ goes from negative to positive. Visually it is clear that the star formation (red) is being supplied by inflowing material (dark blue).

In this situation, where the star formation is depleting the inflowing gas, the surface density is affected but not necessarily drastically. In steady state, the surface density and velocity dispersion (related via $\Sigma \approx \Sigma_{GI}$) are primarily set by the amount of energy that needs to be dissipated by turbulence, which is set by the amount of torque which must be exerted on the gas to maintain the steady state of matter flowing through the disk at rate $\dot{M}$ (see Fig. 4.6), which is set by the profile and rate of external accretion. If star formation is removing some of this gas supply, less energy needs to be dissipated and both Σ and σ will decrease. Eventually, if the star formation rate is fast enough, the inflowing gas (plus the much smaller supply of directly accreting gas) will be entirely depleted and GI will be shut down within that radius. The MRI or some other torque may operate within that radius, and there is certainly still gas within that radius. For $\alpha_{MRI} \lesssim 0.1$, the supply of gas from transport is essentially negligible compared to the supply from continued cosmological accretion. Once the gas supply is shut off in this manner, the gas will burn through the previously $\sim 1/r$ surface density until it reaches equilibrium with the infalling material. At this point newly accreted material is immediately consumed by star formation, and it would take a large burst of accretion to re-activate the GI. In the fiducial model, this shutoff occurs between $z = 1$ and $z = 0$. For quantitative estimates of when this is important, see section 4.4.2.

The fiducial model also shows a peculiar peak in the star formation rate around $r = 17$ kpc at $z = 0$ (visible in Fig. 4.3 and 4.4). This corresponds to a peak in the surface density where gas has built up in a ring, which in turn is caused by the fact that the stagnation point in the GI transport flow (i.e. where $\dot{M} = 0$) passes through this region. At first gas arrives at this radius from a smaller radius, but at late times it arrives from a larger radius. The location of this stagnation point is set by the boundaries of the GI region, which move outward with time (as a result of GI quenching and the steady viscous spread of the disk), and the particular choice of accretion profile. We therefore expect this feature to exist in many galaxies, but its location and prominence is quite parameter-dependent in our model.

Energy equilibrium

Thus far we have been concerned mostly with the surface density distribution. It is clear that GI transport plays a significant role in setting this surface density. For regions of the disk which are gravitationally unstable, we have asserted that $\Sigma \approx \Sigma_{GI} \propto \sigma/r$. In section 4.4.1 we will show that there is a maximum velocity dispersion set by the mass accretion rate; there is also a minimum velocity dispersion, σ_{sf} set by the

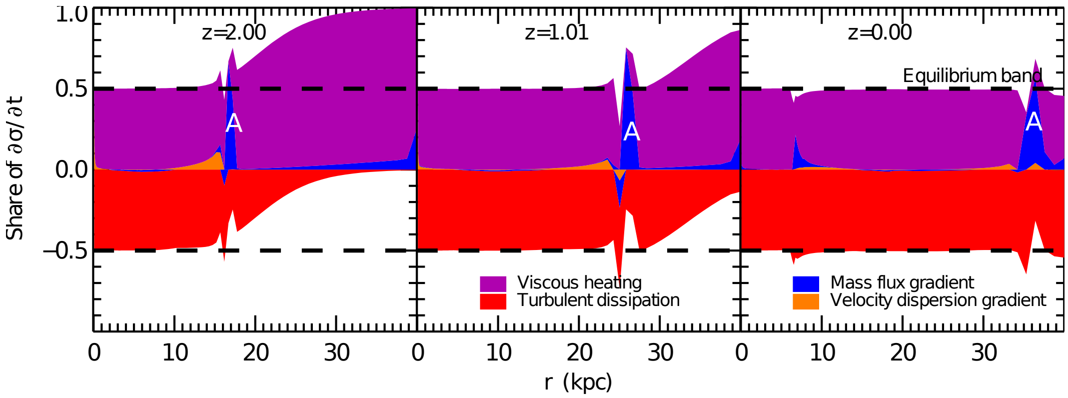

Figure 4.7: The balance of terms in the energy equation for the fiducial model. For each term in $\partial\sigma/\partial t$ (equation 4.3), the contribution it makes relative to all the other terms is shown as a function of radius and time. The red region shows cooling $\propto \mathcal{L}$, purple is heating $\propto \mathcal{T}$, orange and blue are terms associated with advection, $\propto \partial\sigma/\partial r$ and $\partial\dot{M}/\partial r$ respectively. Within the regions which are gravitationally unstable, heating and cooling balance almost perfectly. The advection terms are only relevant right where the gravitationally unstable region borders a stable region, where the velocity dispersion and especially the mass flux change dramatically, leading to the spikes at the locations labelled “A” in the figure.

temperature of the gas. This is an adequate first-order understanding of what sets the surface density in the gravitationally-unstable regions, but we have yet to explore what sets σ and hence Σ between the minimum and maximum values.

Just as with the surface densities, we can show which terms dominate the evolution of σ as a function of radius and time (Fig. 4.7) for the fiducial model. The equilibrium here is even more striking than for the surface densities. Nearly everywhere in the disk, the advection terms (blue and orange) are negligible, and the disk equilibrates between local heating via GI and MRI torques, $\propto \mathcal{T}$ and cooling $\propto \mathcal{L}$ from turbulent dissipation. The exception is at the wave of gas moving outwards to maintain $\Sigma \approx \Sigma_{GI}$ in the inner disk. Here advection becomes important because gas is being transferred from an unstable cell to a stable one with much lower surface density. This stable cell does not pass any mass to the next cell since both have $\mathcal{T} \approx 0$, so $\partial \dot{M}/\partial r$ can be quite large. In reality, the radius separating the gravitationally unstable region from a stable region would be much less well-defined, both because real galaxies are not axisymmetric, and because there may be some ‘overshoot’. Our model overlooks these effects, so our transition is quite sharp – a single cell in our simulation. This is the cause of the spikiness, not only in Fig. 4.7, but also Fig. 4.3, 4.6, and 4.9.

Another exception to the otherwise-good approximation that local heating balances local cooling is at $z = 2$ at large radii, where the disk is not gravitationally unstable and the only torque comes from the MRI. This region takes a long time to equilibrate because the dynamical time is quite long, and the MRI is weak, so building up enough turbulent velocity dispersion to be countered by turbulent dissipation takes a few Gyr. Note that this is not the case in the central region at $z = 0$ where the disk is again gravitationally stable, but this time the dynamical time is short. Note also that our model implicitly assumes that gas near $\sigma \approx \sigma_{\rm sf}$ is in equilibrium between radiative cooling and heating, so the terms we don’t show here, e.g. cooling due to metal lines or heating due to the grain photoelectric effect, may dominate in the regions stable to GI.

Based on Fig. 4.7, it is safe to approximate the energy balance as entirely local, i.e. to neglect the advection terms, in regions of the disk where GI transport is important. Though our simulations keep all of the relevant terms, we will make this approximation in section 4.4.1 to understand exactly what sets σ and Σ in gravitationally unstable regions.

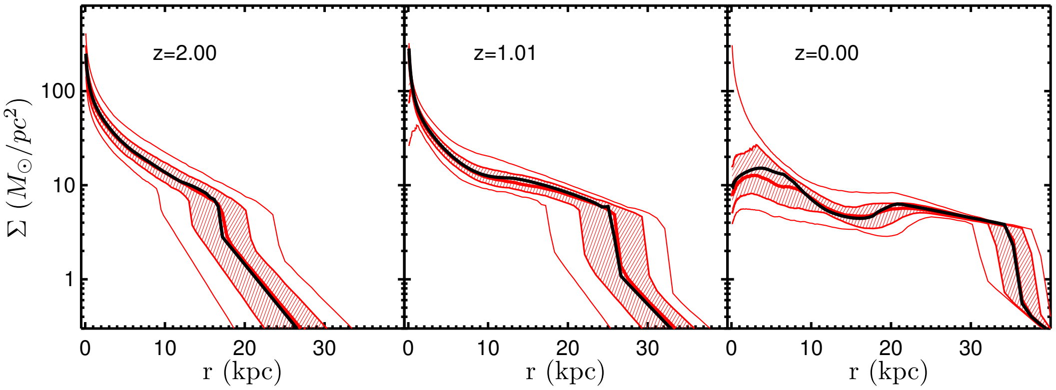

Figure 4.8: The radial surface density profile of gas for different redshifts. The black line shows the fiducial model, and the red lines show the median and central 68 per cent (shaded) and 95 per cent of the models with stochastic accretion histories. The variation between surface density profiles at a given radius and time depends mostly on whether the galaxy is gravitationally unstable there. The variation in external accretion rate is largely responsible for the differences between galaxies in regions of the disk which are gravitationally unstable. Additional dependences on parameters of the model and physical properties of the galaxy are shown in appendix 4.8.

4.3.2 Stochastic accretion

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

Fig. 4.8 shows the distribution of surface densities for the same 400 galaxies whose accretion histories were shown in Fig. 4.1, plus the fiducial smooth model for reference. These galaxies all have the same radial scale, namely $r_{acc} = 6.9$ kpc. At high redshift the galaxies have similar profiles - $1/r$ profiles at small radii and exponential profiles at large radii. The variation is mostly due to the different gas masses of each galaxy, largely the result of the variation in initial halo mass. Regions of galaxies that are gravitationally unstable have similar $\Sigma \approx \Sigma_{GI}$, since Σ_{GI} varies only weakly with accretion rate (see section 4.4.1). As a consequence, the radii over which the galaxy is gravitationally unstable is just a matter of how far the gas needs to be pushed away from where it arrives to maintain $\Sigma \approx \Sigma_{GI}$.

By low redshift, the galaxies have become remarkably similar at large radii but with more than an order of magnitude variation near the center. At large radii, the disk tends to be gravitationally unstable, but in contrast to the high redshift case, these galaxies all have the same halo mass and so are quite similar in terms of the available gas budget. Meanwhile at small radii, some galaxies, namely those with a recent burst of accretion, are still gravitationally unstable and so exhibit the same $1/r$ profiles seen at high redshift, while others have stabilized and are in an equilibrium between infalling gas and star formation. Thus GI transport greatly magnifies the different accretion rates, causing a wide range of column densities near the center of the galaxy, but at the same time gravitational instability enforces remarkable similarity at large radii.

Whether the galaxies are in equilibrium is shown explicitly in Fig. 4.9. As with the fiducial model, the ensemble of disks tends to equilibrate from the inside out. The most remarkable difference is the significant fraction of galaxies which are out of equilibrium, not because they are building up gas, but because they are burning through

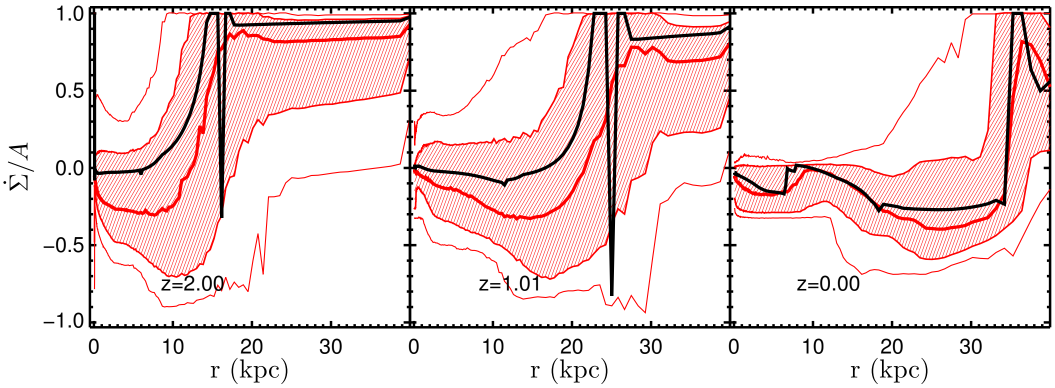

Figure 4.9: Inside-out equilibration. Here we show, for the smooth accretion model (black) and the median, central 68 per cent (shaded) and central 95 per cent of the stochastically accreting ensemble of galaxies (red), the radial distribution of $\dot{\Sigma}$ divided by A, where A is the sum of the absolute value of each term contributing to $\dot{\Sigma}$. Values of $\dot{\Sigma}/A$ near 1 or -1 indicate that the surface density is changing entirely due to a single term in the equation, while values near zero mean terms of opposing sign are canceling and the surface density is close to equilibrium. Equilibration occurs from inside out, though significant deviations from equilibrium are possible - in fact the typical galaxy is in a low-accretion-rate state and burning through the gas from a past accretion event. Galaxies are also out of equilibrium at large radii where the gas is mostly atomic and hence star formation is slow.

excess gas. These are galaxies which had a burst of accretion followed by a lull. Most galaxies in our stochastic sample are in this state because of the lognormal distribution of accretion rates, which vary on time-scales that are typically short compared to the depletion time. At any given time, a galaxy is therefore likely to be accreting gas slowly but still working through gas that was accreted in a recent burst.

4.3.3 Comparison with observations

Using high resolution and high sensitivity data to infer the HI and H_2 distributions in nearby spiral galaxies, Bigiel & Blitz (2012) found that these galaxies have neutral gas surface density profiles well-approximated by a simple exponential,

$$\Sigma_{UP} = 2.1\Sigma_{tr}e^{-1.65r/r_{25}}. \tag{4.27}$$

Here Σ_{tr} and r_{25} are empirical quantities derived from the data, respectively the surface density at which a particular galaxy has $\Sigma_{HI} = \Sigma_{\mathrm{H}_2}$ and the radius of the 25 magnitude per square arcsecond B-band isophote. To compare to our simulations, we need to determine these quantities in our own simulated data. We can find Σ_{tr} in our simulations by searching for the location where $f_{\mathrm{H}_2} = 0.5$. In our model this is determined by the Krumholz et al. (2009) formula, in which this transition surface density is set by the metallicity. The value we should use for r_{25} is somewhat more ambiguous. B-band luminosities are, roughly speaking, set by the star formation rate averaged over at least gigayear time-scales, and the exact luminosity derived for a particular star formation history is somewhat model-dependent. To avoid this issue, we note that if the universal profile is correct, it can be written just as well

$$\Sigma_{UP} = 2.1\Sigma_{tr} \exp\left(-0.74r/r_{tr}\right) \tag{4.28}$$

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

For each of our galaxies, we can easily compute r_{tr} and Σ_{tr} (Fig. 4.10), each as a function of time, to construct the corresponding Σ_{UP} (Fig. 4.11). The agreement is reasonable, within a factor of two of the empirical relation at $z = 0$ for most of the simulated galaxies. At large radii, the effects of photoionization may be important- namely the observations are sensitive only to neutral gas, whereas for the low surface densities $\sim 1M_\odot\ pc^{-2}$, UV radiation may ionize a significant portion of the gas. As in the observed galaxies, the largest scatter occurs within the central region. We argue that this is a consequence of variations in the accretion histories which allow some galaxies to continue to transport gas to their centers via GI torques, while others have stabilized.

The agreement between Σ and Σ_{UP} is not a trivial consequence of the exponential cosmological accretion profile we use. In particular, the universal profile predicts that the gas surface density profile should have a scale length equal to $r_{25}/1.65$. Reading off from figure 4.10, we see that $r_{25}/1.65 \sim 12 - 18$ kpc, whereas in our fiducial model, the scale length of the exponential accretion only reaches 6.9 kpc at $z = 0$, and is smaller at higher redshift. In other words, the scale length of the accretion is

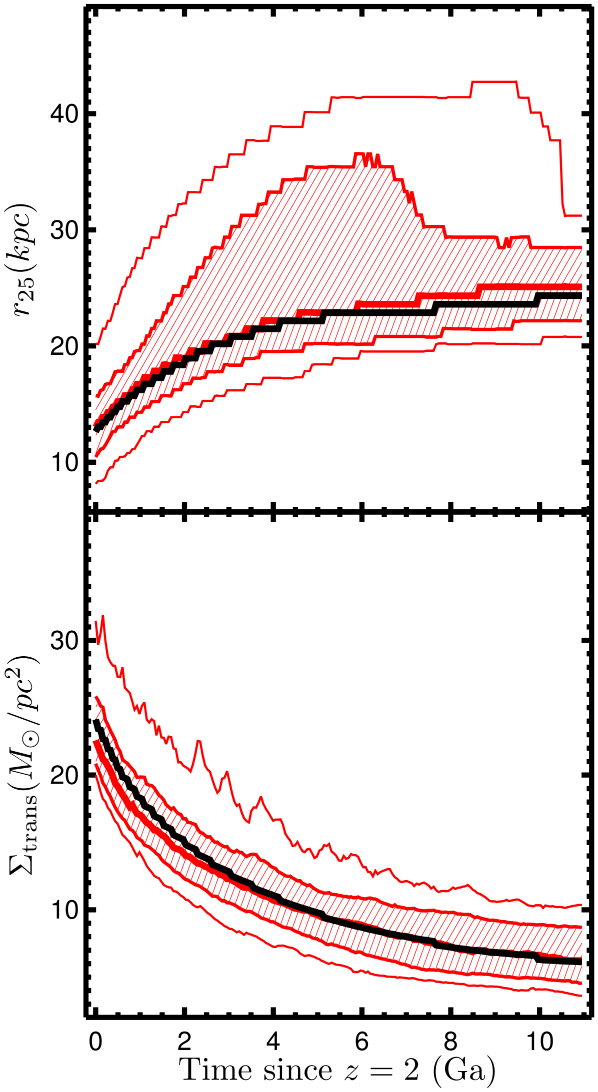

Figure 4.10: The parameters that define the universal profile. The median, central 68 per cent (shaded) and 95 per cent of the stochastic ensemble of galaxies are shown in red, with the smooth accretion model (black) for comparison. As the metallicity of the galaxies increases, the column density Σ_{tr} at which $f_{\rm H_2} = 0.5$ falls. As metals build up in the outer disk from local star formation, and advection and diffusion from star formation nearer to the center, the radius at which the molecular-atomic transition occurs, r_{tr}, and hence, $r_{25} = r_{tr}/0.45$ steadily increase.

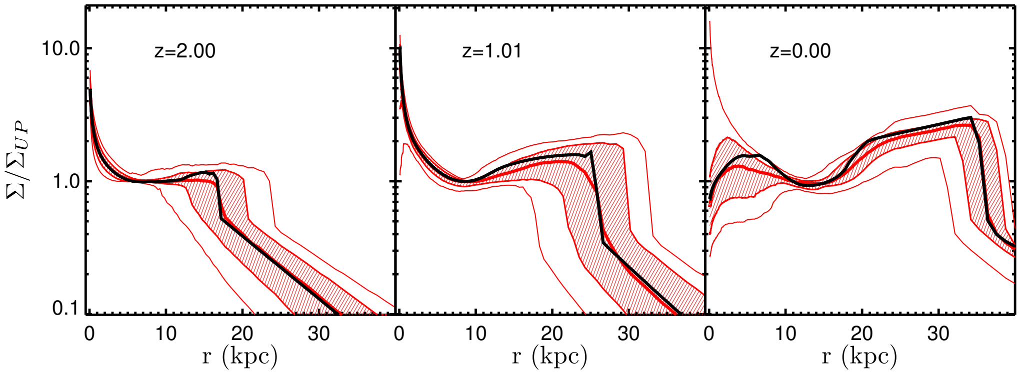

Figure 4.11: The ratio of the surface density to the universal profile inferred by measuring r_{tr} and Σ_{tr} for each simulation for each time. Black shows the smooth accretion model, while red shows the median, 68 per cent (shaded) and 95 per cent of the distribution for a sample of 400 stochastically accreting galaxies.

always substantially smaller than the universal profile scale length in our simulations. Therefore star formation and gravitational instability must be responsible for altering the profile such that we find a reasonable agreement with the observations.

4.4 Discussion

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

We discuss, roughly in chronological order, or more to the point, in order of decreasing external accretion rate the implications of this slowly evolving equilibrium. At high redshift, the galaxy experiences the maximum surface density it can obtain via an equilibrium between cosmological accretion and GI transport. (section 4.4.1). GI transport is eventually shut off via star formation (section 4.4.2), after which each annulus near the center of the disk reaches an equilibrium between local gas supply and local star formation (section 4.4.4)

4.4.1 Maximum velocity dispersion

Conservation of angular momentum requires that $\partial\mathcal{T}/\partial r = -\dot{M}v_\phi(1+\beta)$ (equation (4.4)). At a particular time, we see that the torque at a given radius can be calculated by integrating

$$\mathcal{T}(r) = \mathcal{T}(r = r_0) - \int_{r_0}^{r} \dot{M}(r')v_\phi(r')(1+\beta(r'))dr'. \tag{4.29}$$

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

$$|\mathcal{T}| < \mathcal{T}_{max} \equiv \dot{M}_{ext}v_{circ}r. \tag{4.30}$$

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

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

$$\frac{1}{3}\eta\Sigma\sigma^2\kappa\left(1-\sigma_{\mathrm{sf}}^2/\sigma^2\right)^{3/2}=\frac{(\beta-1)v_\phi}{6\pi r^3}\mathcal{T} \tag{4.31}$$

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

$$\mathcal{T}=6r\eta(\beta+1)v_\phi\sigma_{\mathrm{sf}}^3(\sigma^2/\sigma_{\mathrm{sf}}^2-1)^{3/2}/((\beta-1)GQ_{GI}) \tag{4.32}$$

Again specializing to a flat rotation curve and defining the dimensionless number $\mathcal{N} \equiv Q_{GI}G\dot{M}_{ext}/6\eta\sigma_{\mathrm{sf}}^3 = 1.8\dot{M}_{ext}/(1M_\odot\, yr^{-1})$ and imposing the requirement that $-\mathcal{T} \lesssim \mathcal{T}_{max}$, we arrive at the condition

$$\sigma \lesssim \sigma_{max} \equiv \sigma_{\mathrm{sf}}(\mathcal{N}^{2/3}+1)^{1/2} \tag{4.33}$$

Thus we see that the velocity dispersions of galactic disks are a direct consequence of cosmological accretion and energy equilibrium. We compare this prediction with the maximum measured values of σ in our simulations in Fig. 4.12. From the decay of

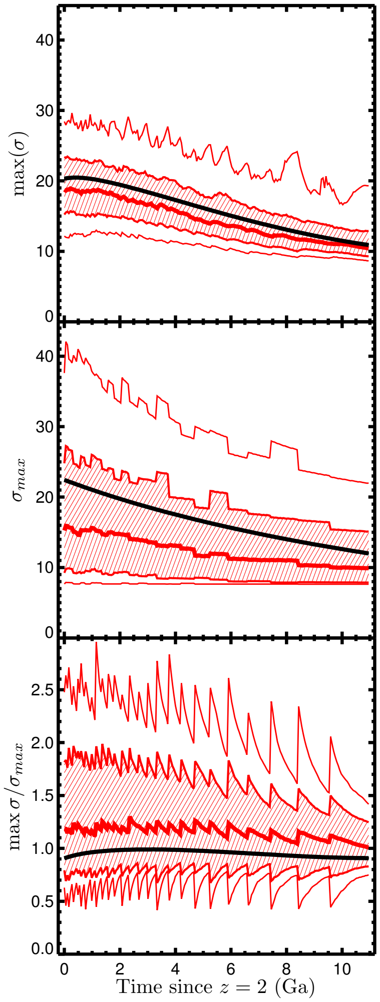

Figure 4.12: Simulated vs predicted velocity dispersion. Here we show the maximum value of σ measured in our simulations (top), the value of σ_{max} predicted by equation 4.33 (middle panel) and their ratio, max(σ)/σ_{max} (bottom). At every change in the accretion rates, the predicted σ_{max} jumps and it takes some time for each galaxy to adjust to its new accretion rate. As usual the black line shows the fiducial model and the red lines show the median, central 68 per cent and 95 per cent of the distribution for the stochastically accreting models.

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

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

The exact way that σ varies between $\sigma_{\rm sf}$ and σ_{max} (Fig. 4.13) depends on the particular accretion profile feeding the galaxy (which roughly determines the shape of the σ(r) profile), the total amount of gas accreted previously (which sets the outer boundary of the GI region), and star formation (which sets the inner boundary). Qualitatively, the velocity dispersion is highest near the center of the galaxy, since most of the accreted mass arrives near the center of the galaxy and flows inwards. At low redshift this is no longer true because the center of the galaxy becomes gravitationally stable, so the velocity dispersion is forced towards its value from star formation feedback $\sigma_{\rm sf}$. The outer edge of the unstable region moves outwards as well, since GI transport will always move some gas outwards to conserve angular momentum. This gas is barely touched by star formation given the low molecular fraction at large radii, so over cosmological time that gas will continue to build up and the edge of the gravitationally unstable region will march outwards. Stabilization at small radii and destabilization at large radii lead the whole unstable region to move outwards in time. The lower velocity dispersions in the unstable region, the result of the decreasing cosmological accretion rate, leads to lower characteristic clump masses as estimated by the 2D Jeans mass, $M_J = \sigma^4/G^2\Sigma$, shown in Fig. 4.14.

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

$$\Sigma \lesssim \frac{(3/2)v_{circ}\sigma_{\rm sf}}{\pi G r Q_{GI}}(\mathcal{N}^{2/3}+1)^{1/2} = \Sigma_{crit}\frac{\sigma_{max}}{\sigma_{\rm sf}}. \tag{4.34}$$

Since Σ_{crit} for a given model is a fixed function of radius, we immediately see that at a given radius Σ in a gravitationally unstable region will also only vary by a factor of a few. However Σ, unlike σ, may fall below the value corresponding to $\sigma = \sigma_{\rm sf}$. This typically happens because some process has shut off GI transport (section 4.4.2), at

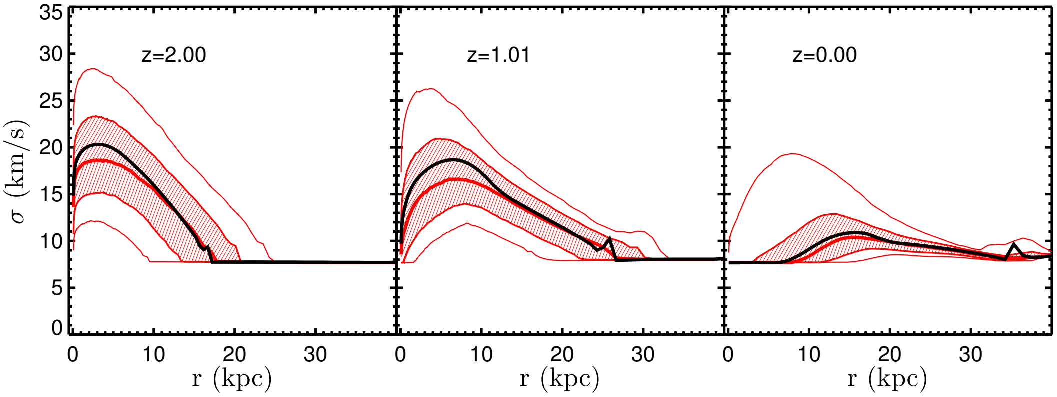

Figure 4.13: The velocity dispersion distribution for our models as a function of radius and time. The median, central 68 per cent (shaded) and 95 per cent for the stochastic accretion models are shown in red, with the smooth accretion model in black for reference. The high velocity dispersions at high redshift characterize galaxies undergoing ‘violent disk instability’, which manifests itself most dramatically with giant clumps. At low redshift, the turbulent velocity dispersions driven by GI are much lower and occur further out in the disk, largely as a result of the falling accretion rate. Galaxies undergo this transition, from violent, dynamical evolution to a ‘secular’ evolution, smoothly.

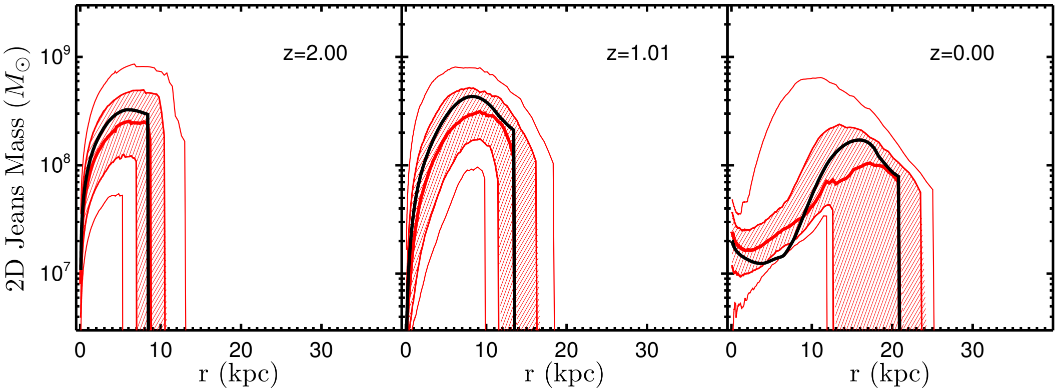

Figure 4.14: The characteristic size of clumps in the star-forming disk. Here we show the distribution of the 2D Jeans mass in regions where the molecular fraction is larger than $f_{\mathrm{H}_2,min}$. The typical mass of gravitationally bound clumps decreases with time, and the peak moves outward in radius. The median, central 68 per cent (shaded) and 95 per cent of the values at each radius and time for the ensemble of stochastically accreting galaxies is shown in red, along with the fiducial model in black.

which point the disk will equilibrate to a new, lower value of Σ (section 4.4.4). We also note that, at least for galactic disks, this maximum column density is likely to be much more restrictive than the one proposed by Scannapieco (2013), which is based upon the requirement that the rate of turbulent energy dissipation must be removable by radiative cooling.

4.4.2 GI quenching

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

$$\frac{\dot{M}_{\rm supply}}{\dot{M}_{SF}} \approx \frac{\dot{M}(r_B)}{\int_{r_A}^{r_B} 2\pi r (f_R + \mu) \dot{\Sigma}_*^{SF} dr} \tag{4.35}$$

When this ratio is $\ll 1$, the region in question would easily deplete the gas supply and shut down GI transport, while when it is $\gg 1$, star formation makes no difference and gas flows through the region unharmed. To evaluate this ratio, we use the star formation rate for the Toomre regime, on the grounds that once star formation is slow enough to be in the single-cloud regime, it is unlikely to be hugely important anyway and this ratio will just be $\gg 1$. On similar grounds, we can also assume $f_{\rm H_2} \approx 1$, $\Sigma \approx \Sigma_{GI}$ and $Q_g \approx (2/3) Q_{GI}$. In that case, our ratio becomes

$$\frac{\dot{M}_{\rm supply}}{\dot{M}_{SF}} \approx \frac{\dot{M}(r_B) G Q_{GI}^2 \pi \left(1 + 2Q_{GI}/3Q_{lim}\right)^{-1/2}}{\int_{r_A}^{r_B} 36\sqrt{2/3}(f_R + \mu)\epsilon_{\rm ff}(\beta + 1) v_\phi^2 \sigma r^{-1} dr}, \tag{4.36}$$

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

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

$$r_A = r_B \exp\left(-0.24 v_{220}^{-2} \epsilon_{0.01}^{-1} \frac{1.54}{f_R + \mu} \frac{\dot{M}_1}{\left(1.5\dot{M}_1^{2/3} + 1\right)^{1/2}}\right) \tag{4.37}$$

where we have neglected the additional scalings with $\sqrt{1 + Q_g/Q_*}$ and we have introduced a few scaled parameters, $v_{220} = v_{circ}/(220 \text{ km s}^{-1})$, $\epsilon_{0.01} = \epsilon_{\rm ff}/0.01$, and $\dot{M}_1 = \dot{M}(r_B)/(1 M_\odot\ yr^{-1})$. We caution that this formula is for illustrative purposes only, since v_ϕ and σ are unlikely to be constant. For these values, it turns out that the exponent is fairly close to zero and so relatively insensitive to the exact values. The exponential evaluates to 0.86, 0.64, and 0.43 for $\dot{M}_1 = 1,\ 4,\ 10$.

This is actually somewhat surprising, since in our fiducial model Fig. 4.6 shows that the mass flux at a few kpc is near $3 M_\odot\ yr^{-1}$, yet the gas reaches the inner edge of the computational domain at $r = 80$ pc easily and GI transport is not shut off until much later. This illustrates the dramatic effect of the rotation curve. The essence of the effect is visible even in equation 4.37, namely by the time we reach radii well within the turnover in the rotation curve at $r_b = 3$ kpc, v_ϕ is appreciably smaller than 220 km s^{-1}, meaning r_A/r_B should be much smaller. The two powers of v_ϕ come from (i) the dynamical time's proportionality to the star formation time - stars form more slowly if the freefall time $\propto r/v_\phi$ is longer, and (ii) the requirement that $Q = Q_{GI}$, which implies $\Sigma \approx \Sigma_{GI} \propto v_\phi$ - lower velocities and hence smaller shear means less gas is required to destabilize the disk. Thus lowering v_ϕ decreases both the surface density and the star formation rate for a fixed surface density.

Our simulations use a fixed rotation curve which increases as a powerlaw with index $\beta_0 = 0.5$ near the center, but galaxies with prominent bulges have what we would term negative values of β_0, i.e. their rotation curves fall with radius near their centers (see e.g. Dutton & van den Bosch (2009)). As gas approaches the center, it would see higher rotation velocities, which, just the opposite of above, would increase the gas surface density required to maintain GI transport and speed up star formation for fixed gas surface density, hence increasing the GI quenching radius r_A. We suggest that this may be a specific physical mechanism for morphological quenching (Martig et al., 2009). In our estimation, the formation of a bulge acts to quench the innermost regions of the galaxy by shutting off GI transport through the increase in v_ϕ, but other factors contribute, namely the available supply of gas $\dot{M}$ and the radius at which stars begin to form efficiently in a galaxy, r_B. We also note that in our model this quenching is not caused by an increase in Q - the increase in Q and the decrease in SFR are both caused by the shutdown of GI transport.

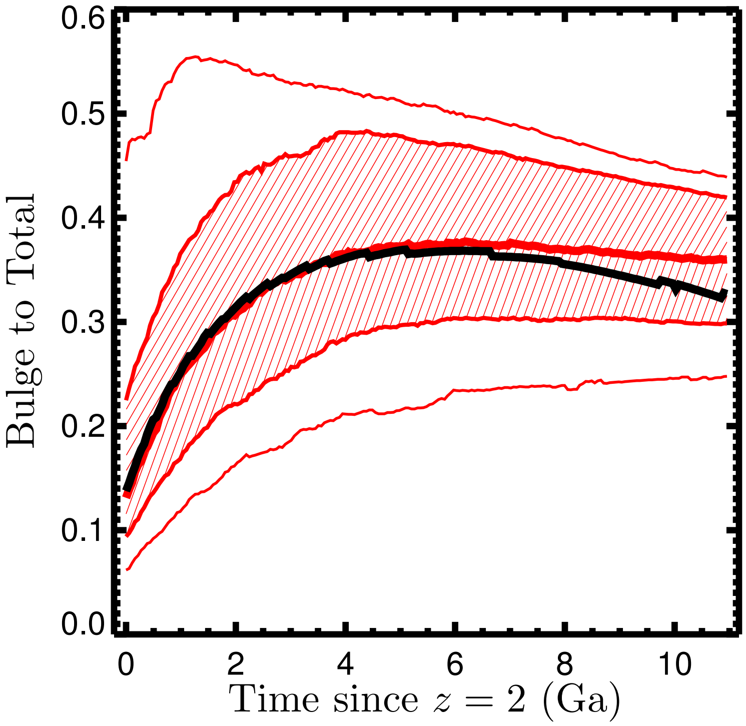

Figure 4.15: Estimated bulge to total ratio of the stellar profile. Both the fiducial model (black) and the stochastic ensemble (red) follow similar trends, growing their bulges through GI transport at high redshift, then forming stars preferentially in the disk since $z = 1$.

4.4.3 The growth of bulges

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

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

$$\begin{aligned} M_B(t) &= \int_0^t \left(\dot{M}_*(r=r_0) + \frac{f_R}{\mu + f_R} \left(\dot{M}(r=r_0) \right. \right. \\ & \left. \left. + \int_0^{r_0} 2\pi r \dot{\Sigma}_{cos} dr \right) \right) dt \\ & + \int_{r_0}^{r_g} 2\pi r (\Sigma_* - \Sigma_{*,exp}) dr. \end{aligned} \tag{4.38}$$

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

$$\log \Sigma_{*,exp} = \log \Sigma_*(r_g) + r \frac{m \log(\Sigma_*(r_g)/\Sigma_*(r_g - \Delta r))}{\Delta r} \tag{4.39}$$

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

The growth of bulges measured by the bulge to total (BT) ratio, with the bulge mass estimated by equation 4.38, is shown in Fig. 4.15. Although gravitational instability funnels gas to the centers of these galaxies, our simulations have star formation efficient enough and a mass loading factor large enough, that the BT ratios tend to lie near 1/3, a fairly reasonable value for MW-mass galaxies. The trend with redshift seems to be a steep rise between $z = 2$ and $z = 1$, followed by a very gradual decrease from $z = 1$ to $z = 0$. This may be attributable to the efficient action of GI at high redshift and its subsequent quenching at lower redshift. Moreover, it is clear that galaxies for which GI transport is important at $z \sim 2$ need not end up as bulge-dominated galaxies at $z = 0$. These specific numbers are sensitive to both the angular momentum distribution of infalling gas, and to the parameters which influence star formation, and hence GI quenching, near the center of the galaxy. The galaxies in our sample all have the same accretion scale length at $z = 0$, but if we include a 0.4 dex scatter in this parameter, comparable to the scatter in spin parameters observed for dark matter halos in N-body simulations (Bullock et al., 2001), the central 95 per cent of $z = 0$ BT ratios for those galaxies stretches from 0.05 to 0.72.

4.4.4 Equilibrium between accretion and SF

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

$$\Sigma = \dot{\Sigma}_{cos}(f_R + \mu)^{-1}\epsilon_{\mathrm{ff}}^{-1} f_{\mathrm{H}_2}^{-1} t_{SC} \tag{4.40}$$

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

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

$$\begin{aligned}
\Sigma_{eq} &= \left(\frac{3\pi \dot{\Sigma}_{cos}^2 Q_{lim} \sigma_{\rm sf}}{32 \epsilon_{\rm ff}^2 f_{\rm H_2}^2 \kappa G (f_R + \mu)^2} \right)^{1/3} \\
&\approx 16 \frac{M_\odot}{\rm pc^2} \frac{\dot{\Sigma}_{cos,.01}^{2/3}}{(f_R+\mu)^{2/3} f_{\rm H_2}^{2/3}} r_1^{1/3} v_{220}^{-1/3} \sigma_{th,8}^{1/3} \epsilon_{.01}^{-2/3}. \qquad (4.41)
\end{aligned}$$

We have used typical values of $\dot{\Sigma}_{cos,.01} = \dot{\Sigma}_{cos}/(0.01\ M_\odot {\rm kpc}^{-2}{\rm yr}^{-1})$, $r_1 = r/(1\ {\rm kpc})$, and $\sigma_{th,8} = \sigma_{\rm sf}/(8\ {\rm km/s})$. Note that other sources of gas may be added to $\dot{\Sigma}_{cos}$, although if they depend on the star formation rate (e.g., for a galactic fountain) the form of the solution will be a bit different. The assumed accretion rate corresponds closely to the redshift zero value for the smooth accretion history model, and the numerical value of Σ_{eq}, despite the approximations made, agrees quite well with the simulation. We see that as long as $\dot{\Sigma}_{cos}$ is sufficiently flat, as is the case for an exponential on radial scales much less than the scale length, the value of Σ_{eq} will have a moderate increase with radius. This relation will break down if radial transport of gas is operating, and if $f_{\rm H_2}$ is appreciably smaller than unity there will be an implicit dependence on Σ_{eq} on the right hand side, since $f_{\rm H_2}$ is a function of Σ (and Z).

We saw in section 4.3.1 that in our smooth accretion model, the inward mass flux from GI transport is exhausted beginning around $z = 0.5$, after which the central gas surface density is rapidly depleted by star formation. We refer to this process as 'GI quenching'. When GI transport is active, it essentially collects cosmological infall from all radii and sends most of that gas inwards and some outwards. This can concentrate most of the star formation in the center of the disk, i.e. gas does not form stars at the location it arrives, but in the center of the galaxy. When GI transport is shut off, the center of the galaxy loses this vast supply of gas virtually instantaneously. The surface density falls from $\Sigma \approx \Sigma_{GI} \propto 1/r$ to $\Sigma \approx \Sigma_{eq} \propto r^{1/3}\dot{\Sigma}_{cos}^{2/3}$ in a few depletion times, which may be significantly faster than 1 Gyr (Fig. 4.4).

We have found that even for large values of an α viscosity (see appendix 4.8), and even for a rotation curve quite favorable for transporting gas to the central regions of galaxies, the supply of gas to the central regions of galaxies at $z = 0$ via transport through the disk is negligible for a large fraction of the galaxy population. Moreover, gas within this region is unable to move any significant distance radially via these mechanisms. Therefore the equilibrium which develops there is a balance between the *local* star formation in some annulus and the *local* gas supply. In our model this comes from cosmological infall, but it could in principle also come from supernova-induced accretion (Hobbs et al., 2012; Marinacci et al., 2010) or gas recycling from old stellar populations (Leitner & Kravtsov, 2011). Therefore we suggest that measuring the star formation rate and profile in the centers of local galaxies with low star formation rates should directly determine the rate and profile with which those particular regions (regardless of the rest of the galaxy) are being supplied with cold gas.

4.5 Summary

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

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

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

2. At a given time, a galaxy will tend to have the following **progression of regions**, from outside inwards. First there is an out-of-equilibrium, low column density region, where gas is building up from cosmological accretion but is not yet gravitationally unstable. Next, the galaxy is in equilibrium between infalling material and GI transport. Further in, star formation takes up an increasing share of

the responsibility for balancing incoming accretion - at this point all three terms in the continuity equation are important. Eventually star formation is so efficient that it outstrips the direct supply of gas and can only be balanced by GI transport from larger radii. Finally, if star formation can use up the entire supply of GI-transported gas, there is a quenched region at the center of the galaxy where star formation balances only the direct accretion onto that radius.

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

4. In equilibrium, new accretion must be balanced by the available sinks: star formation (plus galactic winds) and transport through the disk. Even at radii where star formation is inefficient, GI transport alone is sufficient to balance accretion. GI transport operates through torques which redistribute angular momentum, allowing gas to be removed from where it accretes. To balance the accretion rate, the gas must lose angular momentum in proportion to the accretion rate, so in steady state the accretion rate specifies the torque. The heating caused by these torques is balanced by turbulent dissipation. The turbulent dissipation rate is

proportional to the kinetic energy in the gas, so this balance picks out a velocity dispersion. In summary, the mass flux sets the torque and hence a dissipation rate, which in turn sets the velocity dispersion, so **the cosmological accretion rate sets the velocity dispersion.**

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

6. The process of GI quenching becomes more effective at higher rotational velocities, which increase the star formation rate. Massive bulges increase the rotational velocity near the center of a galaxy, so we propose that **morphological quenching** may occur through the following physical channel: GI transport moves gas to the

center of a galaxy forming a bulge, the central concentration of matter increases the rotational velocity, GI transport is quenched by the increased star formation rate (and the decreasing cosmological accretion rate), and so the star formation in the central region drops dramatically as its gas supply is removed. The value of Q and Q_{gas} will rise as the gas surface density drops to its new, much lower, equilibrium value. This is distinct from the mechanism proposed by Martig et al. (2009), wherein they claim that the formation of a spheroid removes the stellar disk and causes the gas disk to stabilize and hence star formation to cease. In our model, the self-gravity of the stars has very little effect on the gas because the stars are assumed to be separately self-regulated to a fixed Q_*. Both models predict a rise in Q and a drop in star formation rate; in our model, both of these are effects of the shutoff in GI transport (which may be hurried by an increased circular velocity from the formation of a bulge), whereas in Martig et al. (2009), Q increases through the removal of the stars' contribution to the self-gravity to the disk, which then causes the star formation rate to drop.

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

8. Our simulations show that at $z = 0$, some galaxies will be gravitationally unstable

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

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

10. Although we have emphasized the equilibration of galaxies, we also observe situations where some region of the galaxy is **out of equilibrium** (meaning that the surface density is changing at a rate $\gtrsim 5\%$ of the instantaneous accretion rate), even in the gravitationally unstable region, and even if the accretion history is perfectly smooth. This occurs primarily in the outer regions of the galaxy where the depletion time and even the dynamical time can be long enough for the cosmological accretion rate to change significantly - in other words the sinks for

gas are in equilibrium with a past accretion rate. Equilibrium also breaks down near the moving boundaries between gravitationally stable and unstable regions - for instance when a new region of the galaxy has just lost its gas supply via GI quenching, it takes a depletion time to burn through the (now stationary) gas and reach a new equilibrium with direct accretion. As the surface density of gas, and hence the molecular fraction, decline with time, even the inner parts of the disk may experience depletion times much longer than 2 Gyr, and they too may drop out of equilibrium.

11. The turbulent velocity dispersion of gas in the galaxy falls over time, and the region of the disk subject to GI-driven transport and turbulence moves outwards. This may be interpreted as a **smooth transition from violent to secular instability**. The high velocity dispersions and shorter dynamical times of gas at small radii and high redshift leads to giant clumps (since the Jeans mass $\propto \sigma^4/\Sigma$) evolving rapidly, while at low redshift the gravitationally unstable region has a much longer dynamical time and is characterized by lower clump masses. As predicted in the simpler models of Cacciato et al. (2012), the violent disk instability which operates at $z = 2$ no longer operates today in most MW-mass galaxies, but we show that the transition is gradual and the outskirts of the disk remain unstable even at $z = 0$.

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

4.6 Changes since F12

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

4.6.1 Finite volume / explicit mass conservation

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

$$\dot{M}_{i+1/2} = \frac{-1}{v_\phi(r_{i+1/2})(1+\beta(r_{i+1/2}))}\frac{\mathcal{T}_{i+1}-\mathcal{T}_i}{r_{i+1}-r_i}, \tag{4.42}$$

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

$$\left(\frac{\partial\Sigma}{\partial t}\right)_{\text{transport}} = \frac{\dot{M}_{i+1/2}-\dot{M}_{i-1/2}}{2\pi r_i(r_{i+1/2}-r_{i-1/2})} \tag{4.43}$$

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

The reason this is an improvement is that, written in terms of $\mathcal{T}$ and not $\dot{M}$, $\partial\Sigma/\partial t_{\rm transport} \propto \partial^2\mathcal{T}/\partial r^2$. This derivative was computed using a minmod slope limiter, so for example if material attempted to enter or exit a cell from both directions (i.e., the value of $\dot{M}_{i+1/2}$ and $\dot{M}_{i-1/2}$ had opposite signs), $\partial\Sigma/\partial t_{\rm transport} = 0$, and so the entering mass would be lost or the exiting mass would remain in the cell. For monotonic solutions of $\mathcal{T}$ this is a small effect, and so only became apparent when mass was added inside the computational domain instead of at its outer edge (see 4.6.3), which meant some regions would have mass flowing outwards.

4.6.2 Treatment of stable regions where $Q > Q_{GI}$

In our previous work, when $Q > Q_{GI}$, we solved $f_{\rm transport} = 0$. In contrast, in this work we simply set $\mathcal{T}_{GI} = 0$ in those regions. The difference between the two is somewhat subtle. The two treatments would be equivalent if the boundary conditions around the stable region were $\mathcal{T}_{GI,\rm boundary} = 0$, but this will generally not be the case in our disks because the neighboring unstable regions will have nonzero torques. Thus our previous approach would lead to small but non-zero mass fluxes in stable regions. Our new approach is more consistent with the physical picture we're presenting, namely that radial motion is caused by gravitational instability-induced turbulence.

4.6.3 Accretion onto the disk instead of at the outer boundary

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

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

4.7 New Stellar Migration Equations

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

$$\frac{\partial \rho_*}{\partial t} + \nabla \cdot (\rho_* \langle \mathbf{v}^* \rangle) = 0. \tag{4.44}$$

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

$$\frac{\partial \rho_* \langle v_\phi^* \rangle}{\partial t} + \frac{\partial \rho_* \langle v_r^* v_\phi^* \rangle}{\partial r} + \frac{\partial \rho_* \langle v_\phi^* v_z^* \rangle}{\partial z} + \frac{2\rho_* \langle v_r^* v_\phi^* \rangle}{r} = 0 \tag{4.45}$$

As with the gas, we have assumed axisymmetry.

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

$$\frac{\partial}{\partial t} \int_{-\infty}^{\infty} \rho_* dz = -\frac{1}{r}\frac{\partial}{\partial r}\left(r \int_{-\infty}^{\infty} \rho_* \langle v_r^* \rangle dz \right) - \int_{-\infty}^{\infty} \frac{\partial}{\partial z} \rho_* \langle v_z^* \rangle dz \tag{4.46}$$

We will assume that $\langle v_i^* \rangle$ does not vary much over the scale height of the disk, that the disk does not change orientation (so $\langle v_z^* \rangle = 0$), and that $\rho_* \to 0$ for large and small values of z, so integrals over z of the z-derivative of a quantity weighted by ρ_* will vanish. Defining $\Sigma_* \equiv \int_{-\infty}^{\infty} \rho_* dz$, we have

$$\frac{\partial \Sigma_*}{\partial t} = -\frac{1}{r}\frac{\partial}{\partial r}\left(r \Sigma_* \langle v_r^* \rangle \right) = \frac{1}{2\pi r}\frac{\partial}{\partial r}\dot{M}_* \tag{4.47}$$

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

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

$$\frac{\partial}{\partial t}\Sigma_* \langle v_\phi^* \rangle + \frac{1}{r^2}\frac{\partial}{\partial r} r^2 \int \rho_* \langle v_r^* v_\phi^* \rangle dz = 0 \tag{4.48}$$

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

$$\begin{aligned} \langle v_\phi^* \rangle \frac{\partial \Sigma_*}{\partial t} + \Sigma_* \frac{\partial \langle v_\phi^* \rangle}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial r} r^2 \Sigma_* \langle v_r^* \rangle \langle v_\phi^* \rangle = \\ -\frac{1}{r^2}\frac{\partial}{\partial r} r^2 \int \rho_* \sigma_{r\phi}^2 dz. \end{aligned} \tag{4.49}$$

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

$$2\pi r \Sigma_* \frac{\partial j_*}{\partial t} + 2\pi r \Sigma_* \langle v_r^* \rangle \frac{\partial j_*}{\partial r} = \frac{\partial}{\partial r}\mathcal{T}_*, \tag{4.50}$$

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

$$-\dot{M}_* \frac{\partial j_*}{\partial r} = -\dot{M}_* v_\phi (1+\beta) = \frac{\partial}{\partial r}\mathcal{T}_*, \tag{4.51}$$

so it is clear that the time derivative of Σ_* is proportional to the second derivative of the torque. At this point we have also assumed that $\langle v_\phi^* \rangle = v_\phi$, the circular velocity of the gas, so that here, as in e.g. equation (4.3), $\beta = \partial \ln v_\phi / \partial \ln r$.

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

$$\frac{\partial f}{\partial t} + v_i^* \frac{\partial f}{\partial x_i} - \frac{\partial \psi}{\partial x_i} \frac{\partial f}{\partial v_i^*} = 0 \tag{4.52}$$

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

$$\begin{aligned} \int v_j^* v_j^* f d^3\mathbf{v}^* &+ \frac{\partial}{\partial x_i} \int v_i^* v_j^* v_j^* f d^3\mathbf{v}^* \\ &- \frac{\partial \psi}{\partial x_i} \int v_j^* v_j^* \frac{\partial f}{\partial v_i^*} d^3\mathbf{v}^* = 0 \end{aligned} \tag{4.53}$$

The final term may be integrated by parts,

$$\begin{aligned} \frac{\partial \psi}{\partial x_i} \int v_j^* v_j^* \frac{\partial f}{\partial v_i^*} d^3\mathbf{v}^* &= -\frac{\partial \psi}{\partial x_i} \int \frac{\partial v_j^* v_j^*}{\partial v_i^*} f d^3\mathbf{v}^* \\ &= -2 \frac{\partial \psi}{\partial x_i} \int v_i^* f d^3\mathbf{v}^*, \end{aligned} \tag{4.54}$$

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

$$\begin{aligned} \langle v_i^* v_j^* v_j^* \rangle &= \langle v_i^* \rangle \langle v_j^* \rangle \langle v_j^* \rangle + \langle v_i^* \rangle \sum_j \sigma_{jj}^2 \\ &\quad + 2 \langle v_j^* \rangle \sigma_{ij}^2 + \rho_*^{-1} \int f \delta v_i \delta v_j \delta v_j d^3\mathbf{v}^* \end{aligned} \tag{4.55}$$

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

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

$$\begin{aligned} 0 &= \frac{\partial}{\partial t} \rho_* \left(\langle \mathbf{v}^* \rangle^2 + \sum_i \sigma_{ii}^2 \right) + \nabla \cdot \rho \langle \mathbf{v}^* \rangle \left(\langle \mathbf{v}^* \rangle^2 + \sum_i \sigma_{ii}^2 \right) \\ & \quad + \nabla \cdot \left(2\rho \langle \mathbf{v}^* \rangle \cdot \underline{\underline{\sigma^2}} \right) + 2\rho_* \nabla \psi \cdot \langle \mathbf{v}^* \rangle \end{aligned} \tag{4.56}$$

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

$$\begin{aligned} \nabla \cdot (\rho_* \langle \mathbf{v}^* \rangle \psi) &= \psi \nabla \cdot (\rho_* \langle \mathbf{v}^* \rangle) + \rho_* \langle \mathbf{v}^* \rangle \cdot \nabla \psi \\ &= -\frac{\partial (\rho_* \psi)}{\partial t} + \rho_* \frac{\partial \psi}{\partial t} + \rho_* \langle \mathbf{v}^* \rangle \cdot \nabla \psi \end{aligned} \tag{4.57}$$

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

$$\frac{\partial}{\partial t} \rho_* \mathcal{A} + \nabla \cdot \rho \langle \mathbf{v}^* \rangle \mathcal{A} + \nabla \cdot \left(2\rho \langle \mathbf{v}^* \rangle \cdot \underline{\underline{\sigma^2}} \right) - 2\rho_* \frac{\partial \psi}{\partial t} = 0 \tag{4.58}$$

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

$$\rho_* \frac{\partial}{\partial t}\left(\mathcal{A} - 2\psi\right) + \rho_* \langle \mathbf{v}^* \rangle \cdot \nabla \mathcal{A} + \nabla \cdot \left(2\rho_* \langle \mathbf{v}^* \rangle \cdot \underline{\underline{\sigma^2}}\right) = 0 \tag{4.59}$$

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

$$\begin{aligned} 0 &= \rho_* \frac{\partial}{\partial t}\left(\sigma_{rr}^2 + 2\sigma_{zz}^2\right) + \rho_* \langle v_r^* \rangle \frac{\partial}{\partial r}\left(v_\phi^2 + \sigma_{rr}^2 + 2\sigma_{zz}^2 + 2\psi\right) \\ &\quad + \frac{2}{r}\frac{\partial}{\partial r} r \rho_* \left(\langle v_r^* \rangle \sigma_{rr}^2 + v_\phi \sigma_{r\phi}^2\right) \end{aligned} \tag{4.60}$$

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

$$\begin{aligned} \frac{\partial \sigma_{rr}}{\partial t} &= \frac{1}{2\pi r \Sigma_* (\sigma_{rr} + \sigma_{zz})} \left(\frac{v_\phi (\beta - 1)}{r^2} \mathcal{T}_* + \sigma_{rr}^2 \frac{\partial \dot{M}_*}{\partial r} \right. \\ &\quad \left. + \dot{M}_* \left(3\sigma_{rr} \frac{\partial \sigma_{rr}}{\partial r} + 2\sigma_{zz} \frac{\partial \sigma_{zz}}{\partial r} \right) \right) \end{aligned} \tag{4.61}$$

This is very similar to equation (4.3), since the procedures used to derive the two are quite similar. The primary distinction is that here we have split the velocity dispersion into a radial and non-radial component whereas for the gas they are assumed to be identical (a reasonable approximation since the gas is collisional). Besides that the only difference is in the numerical values of the coefficients.

With equation 4.61, its counterpart for $\partial\sigma_{zz}/\partial t$, and the continuity equation (equation 4.47) we can follow a similar procedure as detailed in section 4.2.1 to solve for $\mathcal{T}_*$. In particular, we can again split the terms which appear in dQ_*/dt into those which contain $\mathcal{T}_*$ and its radial derivatives, and those which do not.

$$\begin{aligned}\frac{dQ_*}{dt} &= \frac{\partial\Sigma_*}{\partial t}\frac{\partial Q_*}{\partial\Sigma_*} + \frac{\partial\sigma_{rr}}{\partial t}\frac{\partial Q_*}{\partial\sigma_{rr}} + \frac{\partial\sigma_{zz}}{\partial t}\frac{\partial Q_*}{\partial\sigma_{zz}} \\ &= f^*_{\text{transport}}\left(\Sigma_*, \sigma_{rr}, \sigma_{zz}, \mathcal{T}_*, \frac{\partial\mathcal{T}_*}{\partial r}, \frac{\partial^2\mathcal{T}_*}{\partial r^2}\right) \\ &\quad + f^*_{\text{source}}(\Sigma, \sigma, Z, \Sigma_*, \sigma_{rr}, \sigma_{zz}) \qquad (4.62)\end{aligned}$$

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

4.8 Sensitivity to parameters

Thus far we have used only the fiducial parameters, but each one is at least somewhat uncertain (e.g. ϵ_{ff} may vary by a factor of 3 in either direction), or may change for physical reasons (e.g., lower-mass haloes will likely have smaller r_{acc} and

v_{circ}, and larger μ). To explore the effects of each parameter, we have varied them one at a time from their fiducial values for the smooth accretion history. Fig. 4.16 shows the $z = 0$ surface density distribution when each parameter is varied. Essentially, all of the models are gravitationally unstable over a wide range of radii and parameter choices. We also show the metallicity distribution for all of these models in Fig. 4.17. The metallicities are in general hugely sensitive to the parameters, so much so that any attempt to draw a physical conclusion by fitting a metallicity gradient should be treated with extreme caution, since the same metallicity gradient can be produced by varying any number of parameters. We discuss each of the parameters in more detail in the following sections.

4.8.1 Initial conditions - α_r, $f_{g,0}$, f_{cool}, ϕ_0

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

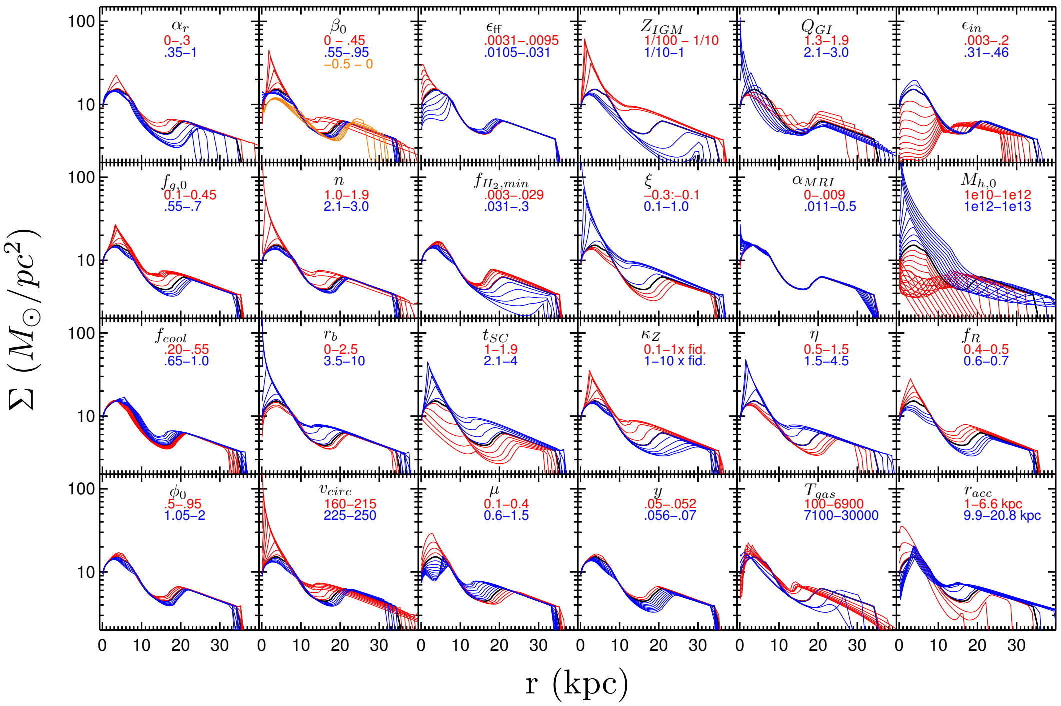

Figure 4.16: Surface density at $z = 0$. Each pane shows models where the given parameter is varied within the quoted range - red models have lower values of the parameter, blue higher. The models are arranged in 6 columns according to what the parameters are controlling- from left to right: initial conditions, rotation curve, star formation, metallicity, gas transport, and gas supply.

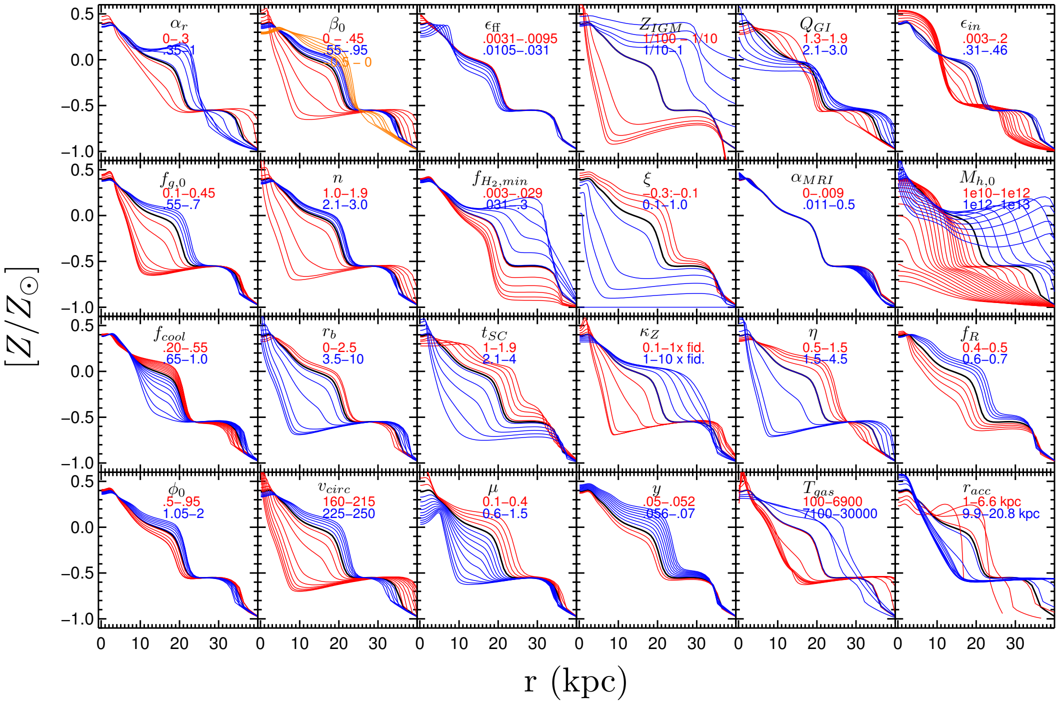

Figure 4.17: Metallicity for the same models as in Fig. 4.16

4.8.2 Rotation curve - β_0, n, r_b, v_{circ}

The shape of the rotation curve is controlled by these four parameters - inner powerlaw slope, the sharpness of, and location of the turnover from flat, and the overall normalization. The most dramatic effect of changing these parameters is in the inner region of the disk, where different values can change the surface density by an order of magnitude. This is because the rotation curve influences both the surface density in gravitationally unstable regions and the star formation rate in the central region, where $\dot{\Sigma}_*^{SF} \propto \kappa \propto v_\phi$, and hence has a strong influence on where exactly the disk is able to form stars fast enough to shut off GI transport to the innermost region. Negative values of β_0 have an effect at large radii too. In general, however, the qualitative behavior of our models is largely insensitive to these parameters except near galactic centers.

4.8.3 Star formation - $\epsilon_{\rm ff}$, $f_{{\rm H}_2,min}$, t_{SC}, μ

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

4.8.4 Metallicity - Z_{IGM}, ξ, κ_Z, y

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

4.8.5 Transport - Q_{GI}, α_{MRI}, η, T_{gas}

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

4.8.6 Gas supply - ϵ_{in}, $M_{h,0}$, f_R, r_{acc}

Here we come to the most important parameters in setting the surface density - the quantity and distribution of the gas supply. These parameters respectively are the efficiency at $z = 0$ (assuming a fixed efficiency at $z = 2$), the halo mass at redshift zero (where we change only M_h and hence the accretion history, but no other parameters), the remnant fraction, and finally the accretion scale length. These models obey the trends one might expect. Less gas means the region over which the gas is gravitationally unstable is smaller. Parts of the disk beyond this radius retain the exponential character of the accretion profile, and parts of the disk interior have their surface densities set by the balance between local accretion and local star formation.

Part II

Detailed Simulations

Chapter 5

Suppresion of Star Formation via Photoelectric Heating

5.1 Main Text

Photoelectric heating has long been recognized as the primary source of heating for the neutral interstellar medium(Draine, 1978). Simulations of spiral galaxies (Bekki, 2015a) found some indication that photoelectric heating could suppress star formation. However, simulations that include photoelectric heating have typically found that it has little effect on the rate of star formation in either spiral galaxies(Tasker, 2011; Tasker et al., 2015) or dwarfs(Hu et al., 2015) suggesting that supernovae and not photoelectric heating are responsible for setting the star formation law in galaxies(Hopkins et al., 2011, 2013; Hayward & Hopkins, 2015). This result is in tension with recent work(Krumholz et al., 2009; Krumholz & Dekel, 2012; Christensen et al., 2012; Krumholz, 2013; Makiya et al., 2014) indicating that a star formation law that depends on galaxy metallicity, as expected for photoelectric heating but not for supernovae, reproduces the present-day galaxy population better than a metallicity-independent one. Here we report a series of simulations of dwarf galaxies, where the effects of both photoelectric heating and supernovae are expected to be strongest. We simultaneously include space- and time-dependent photoelectric heating, and we resolve the Sedov phase of every supernova blast wave, allowing us to make a direct measurement of the relative importance of momentum injection by supernovae and dust heating by FUV photons in suppressing star formation. We find that supernovae are unable to account for the long observed(Hunter et al., 2012) gas depletion times in dwarf galaxies. Instead, ordinary photoelectric heating is the dominant means by which dwarf galaxies regulate their star formation rate at any given time, suppressing the star formation rate by more than an order of magnitude relative to simulations with only supernovae.

To investigate whether the depletion times in dwarf galaxies, which are longer than for Milky Way-like galaxies by more than an order of magnitude(Bigiel et al., 2011; Hunter et al., 2012), are set by the momentum injection from supernovae or by photoelectric heating, we perform a series of high-resolution hydrodynamic simulations using the Enzo adaptive mesh refinement code(Bryan et al., 2014). We include a new prescription for supernova and pre-supernova stellar feedback, and a new method for self-consistent spatially-dependent photoelectric heating (see Methods). We use two sets of initial conditions. Both correspond to isolated dwarf galaxies with an initially laminar exponential disk, a stationary hot halo, and collisionless particles representing

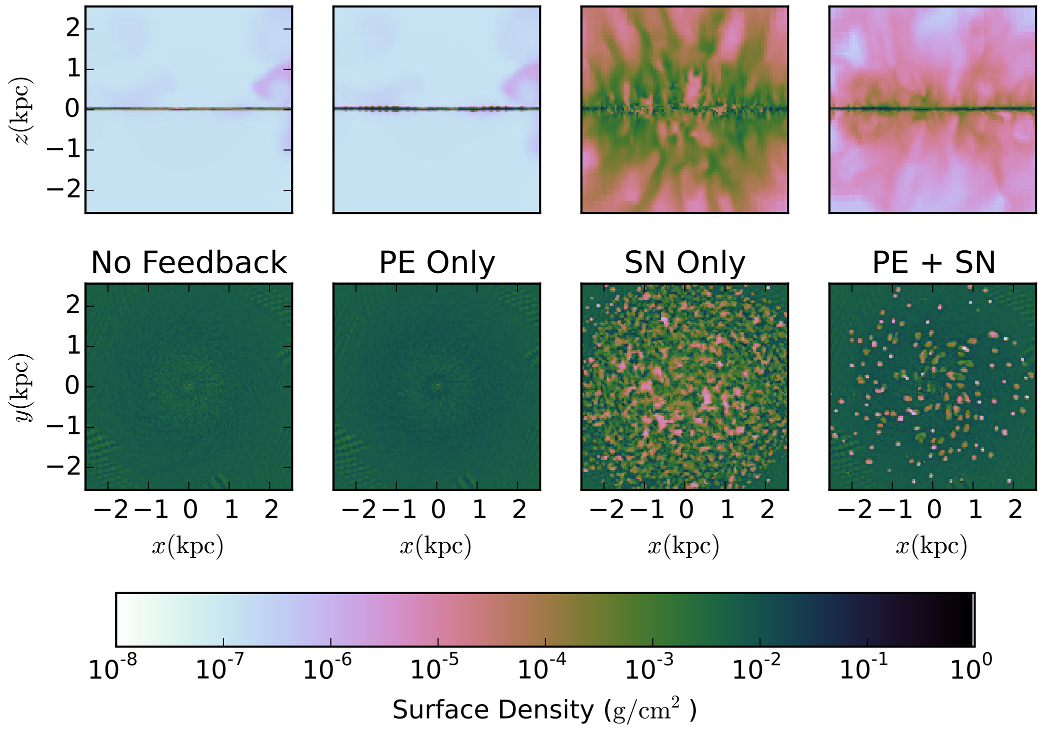

Figure 5.1: The morphology of the gas. For each of the 5 kpc scale length simulations after 90 Myr of evolution, the density is integrated 200 pc in the y dimension (top panels) and in the z dimension (bottom panels). The morphology of the disk is essentially determined by the presence of supernovae, despite the fact that the PE Only and PE+SN runs have nearly identical star formation rates (Figure 2). The SN Only simulation has an order of magnitude higher SFR than the PE+SN simulation, which is why the outflow and the disruption of the cold disk is more dramatic.

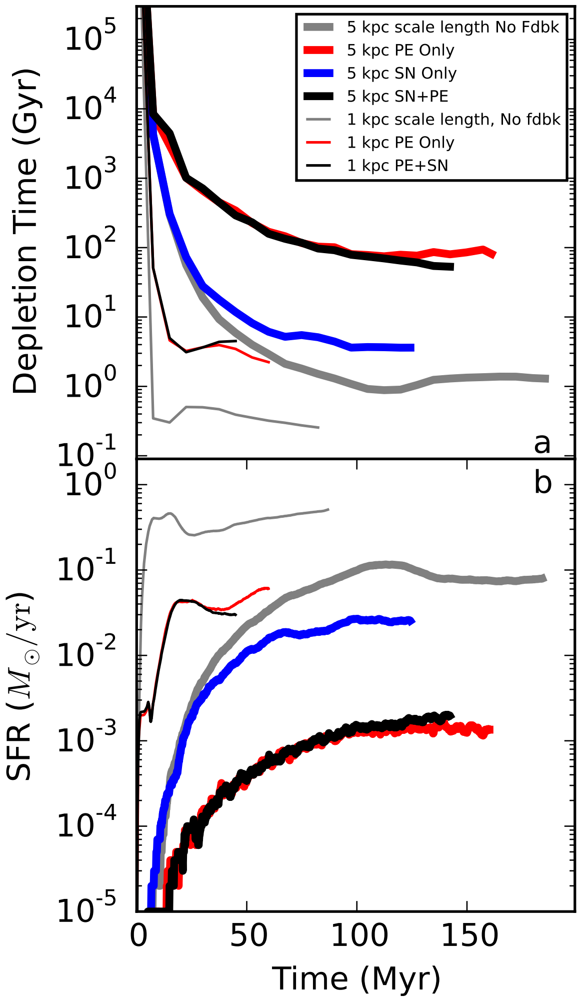

Figure 5.2: The star formation rates of the simulations. The long depletion times frequently observed in dwarfs (see Extended Data Figure 4) are reproduced when both photoelectric heating and supernova feedback are included. Remarkably, turning off supernovae has almost no effect, meaning that photoelectric heating alone is responsible for the long depletion times in the simulations. More compact galaxies have higher star formation rates and shorter depletion times, but again disabling supernovae has a much smaller effect than disabling photoelectric heating.

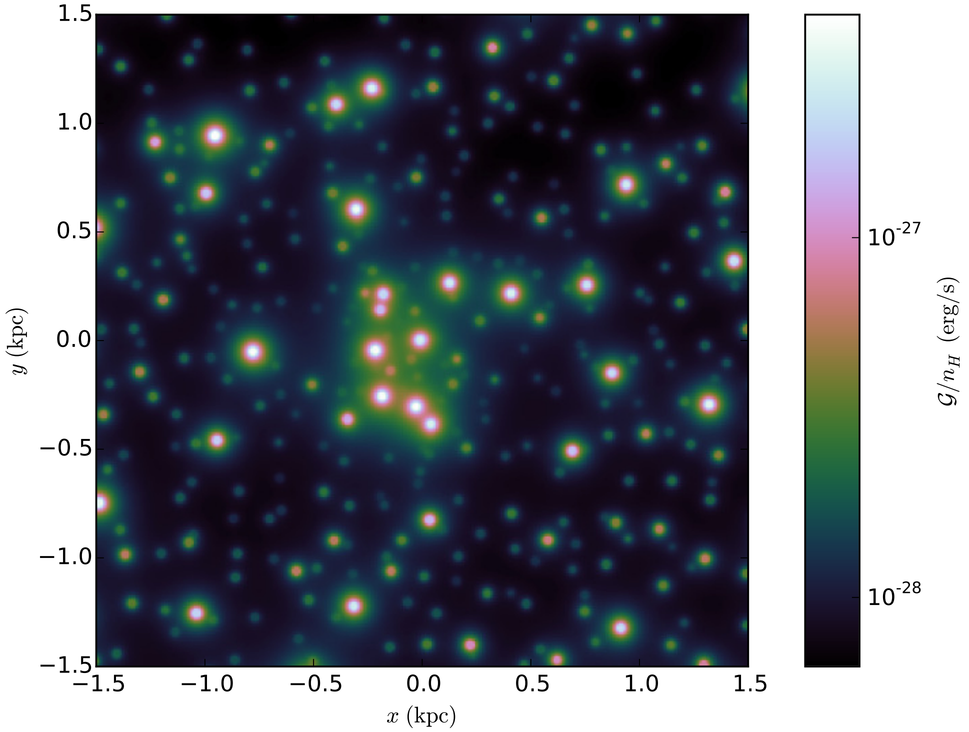

Figure 5.3: The photoelectric heating rate. Every young star particle in the simulation produces FUV radiation. The flux from all of these stars is summed in each cell (see Methods), yielding the volumetric heating rate shown here. Given the low surface density of young stars, this distribution is highly inhomogeneous. The low metallicity and small size of the galaxy means that the mean free path of the FUV photons is large, so the flux from each star and hence the heating rate falls off as r^{-2} from each source.

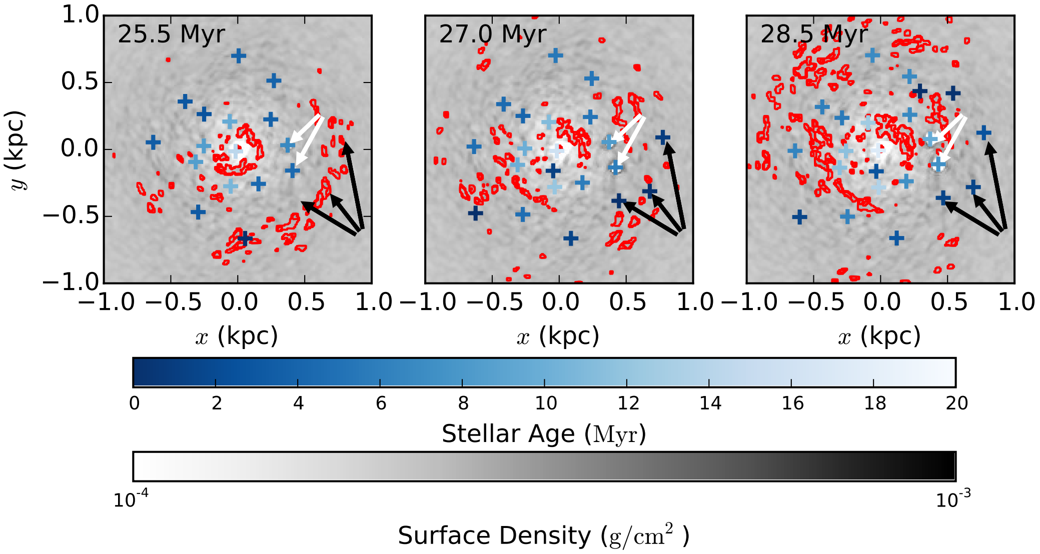

Figure 5.4: The effect of photoelectric heating. The surface density of gas is shown in grayscale, and the red contours show regions where the density is at least 80% of the Jeans density, i.e. the threshold density for forming stars at the gas's temperature. Star particles formed within the 20 Myr preceding each snapshot are shown as crosses, with darker colors indicating younger stars. The stars indicated with arrows, formed between the first and second snapshots, by the third snapshot have heated nearby gas preventing star formation. Supernovae remnants (white arrows) have clear morphological signatures, but do not affect star formation substantially.

stars and dark matter. The galaxies have a dark matter halo mass of $10^{10}M_\odot$, a stellar mass of $10^7M_\odot$, and a cold gas mass of $10^8M_\odot$. Galaxies in this mass range are comfortably above the limit where star formation can be quenched by the cosmological UV background(Okamoto et al., 2008), but small enough that the effects of both heating by FUV photons and supernova feedback(Dekel & Silk, 1986) are plausibly extreme. The initial conditions differ in the exponential scale length chosen for the gas; one set uses 5 kpc, designed to mimic recently-discovered nearly-starless galaxies(Janowiecki et al., 2015), and toward the high end of the range observed for field dwarf galaxies(Hunter et al., 2012). The other set uses 1 kpc, toward the low end of the observed field dwarf range. For a galaxy with an HI mass of $10^8M_\odot$, assuming an exponential HI profile, the observed relation (Broeils & Rhee, 1997) between HI mass and HI size suggests an HI scale length of about 1.9 kpc.

To understand how supernovae and photoelectric heating each contribute to the evolution of these galaxies, we perform a straightforward numerical experiment. We run a fiducial simulation of the 5 kpc scale length initial conditions including both supernovae and photoelectric heating, and simulations where each of these effects is turned off in turn. We refer to these as the "SN+PE", "PE Only," "SN only," and the "No feedback" simulations. We also run the "SN+PE", "PE Only", and "No feedback" cases for the 1 kpc initial conditions. The simulations with supernovae also include pre-supernova stellar feedback from winds and HII regions. For the 5 kpc case, we re-run with the four different feedback models at three different maximum spatial resolutions – 10 pc, 5 pc, and 2.5 pc. These resolutions are high enough, and the typical densities in which supernovae explode in these simulations are low enough, that our simulations do not suffer from the overcooling problem(Katz, 1992), whereby poorly-resolved simulations overestimate the rate at which SN-heated gas cools (see Figure 5.5). We focus first on the 10 pc resolution simulations for the 5 kpc scale length initial conditions, since we have run these for the longest time. In the Methods section we compare to the higher resolution runs to evaluate the level of convergence of our results. We compare the 5 kpc and 1 kpc initial conditions below.

We find that all of the simulations follow a similar initial transient behavior. The gas disk cools from its center outwards, causing the disk to collapse vertically. Stars form first in the center, then further and further out. The central region of the galaxy after 90 Myr is shown in Figure 5.1 for each of the four feedback models. In terms of large-scale morphology, the supernovae have the most dramatic effect, driving large outflows with mass loading factors of order 100. Photoelectric heating slightly alters the global structure of the gas, but the two simulations without supernovae look quite similar.

Figure 5.2 shows the star formation rates and depletion times as a function of time for each of the four feedback models. After each simulated galaxy experiences its initial transient as the gas collapses from its initial state, their instantaneous star formation rates and depletion times are strikingly different depending on whether photoelectric heating was included. However, if two simulations are different only in their inclusion of supernova feedback, they end up with similar depletion times. This immediately shows that photoelectric heating, not feedback from supernovae, is primarily responsible for the long depletion times observed in dwarf galaxies. Supernovae, or even a lack of feedback, can result in low star formation rates in the long run by rapidly ejecting gas or locking gas into stellar remnants. This can be seen in the central kpc of the simulations without photoelectric heating – the star formation rate falls in this region, but the depletion time is relatively unaffected. Only the simulations including photoelectric heating produce depletion times in reasonable agreement with the large values frequently observed in dwarfs, as illustrated in Figure 5.8.

Do these results depend on the gas scale length? In Figure 5.8, we show the results for our simulations with a 1 kpc gas scale length. We find that the more compact galaxies have higher star formation rates and shorter depletion times, explaining the wide range of depletion times visible in the observational data. However, even for the 1 kpc simulations we find that, when supernovae are disabled but photoelectric heating is left active, the star formation rate is virtually unaffected. In contrast, disabling photoelectric feedback again causes the depletion time to drop by an order of magnitude, to values inconsistent with the observed sample. This indicates that photoelectric heating and not supernovae regulate star formation over a wide range of gas surface density in dwarf galaxies, and not just in the potentially-extreme(Janowiecki et al., 2015; Cannon et al., 2015) nearly starless dwarfs.

The means by which photoelectric heating suppresses star formation in our simulations is simple and intuitive. When a new star cluster is formed in the simulation, its most massive stars will emit photons with energies between 8 and 13.6 eV. These photons dominate the heating rate owing to the grain photoelectric effect, since they have energies high enough to liberate electrons from dust grains, but low enough not to be absorbed by intervening neutral hydrogen (Figure 5.3). The heating rate in the vicinity of a newly formed star increases the equilibrium temperature of the gas at fixed pressure or density. This in turn increases the Jeans mass of the gas and makes star formation more difficult (Figure 5.4).

We conclude that the physics responsible for setting the instantaneous star formation law in dwarf galaxies, i.e. the depletion time, is nearly independent of the physics determining the properties of the outflows. Ultimately young stars are responsible for both photoelectric heating and supernovae, but it is the moderate, local, instantaneous, volumetric heating of the former which controls the depletion time. Supernovae, despite their ability to eject mass from the galaxy at a rate more than 100 times the star formation rate, are unable to shut down star formation locally and instantaneously. In the long run, even though dwarf galaxies have 10-100 Gyr depletion times, large mass loading factors mean that the timescale on which gas is lost from the ISM is shorter – between 1 Gyr and 10 Gyr, implying that these galaxies may be in equilibrium between gas inflow and star formation plus outflows(Davé et al., 2012; Lilly et al., 2013; Forbes et al., 2014b). In the long run, therefore, the star formation rate in these galaxies will be set by the value of the mass loading factor, whereas the mass of gas in the ISM will be set by the depletion time. The former, in turn, is set by supernovae, and the latter by photoelectric heating.

5.2 Methods

Our simulations follow the evolution of an isolated dwarf galaxy for ~ 100 Myr using the AMR code Enzo(Bryan et al., 2014). We use a piecewise-parabolic mesh hydrodynamics solver, with an HLLC Riemann solver to follow the motion of the gas. The gravitational potential is computed on the same mesh used to solve the hydrodynamics. Dark matter and stars are included as collisionless particles acting independently of the hydrodynamics, except when new particles are formed or existing particles inject mass, metals, and energy back into the gas according to our feedback model.

We use the same sort of initialization as in previous work(Goldbaum et al., 2015), though of course with parameters appropriate for a dwarf. In particular, we create a set of stellar and dark matter particles with the MakeGalaxy code(Springel & White, 1999). The density of a stellar disk and bulge are pre-specified, and their velocities are set so that the galaxy begins in approximate Jeans equilibrium. The dark matter has an initial Hernquist profile. We use a halo mass of $10^{10} M_\odot$, a concentration of 10, a spin parameter of 0.04 a stellar mass of $10^7 M_\odot$, and a stellar scale length of 300 pc.

We initialize the gas disk according to

$$\rho_d(r,z) = \rho_0 \exp\left(-\frac{r}{r_d}\right) \exp\left(-\frac{|z|}{H}\right) \tag{5.1}$$

This equation applies until the pressure $\rho_d T_d$ falls below the halo pressure $\rho_h T_h$, at which point the density and temperature are set to ρ_h and T_h. While ρ_d is spatially-dependent, ρ_h, T_d, and T_h are all taken to be constant. In these simulations we set $T_h = 10^6 K$, $T_d = 1000 K$, $\rho_h = 2.34 \times 10^{-30}\mathrm{g/cm^3}$, and $\rho_0 = 1.41 \times 10^{-24}\mathrm{g/cm^3}$. The scale length and scale height are $r_d = 5\mathrm{kpc}$ and $H = 30\mathrm{pc}$. These initial conditions are chosen to minimize the time the disk spends in its initial collapse phase, and minimize the influence of the galactic halo on the dynamics. In particular, at this temperature the halo does not monolithically cool onto the galaxy over the timescale on which the simulation is run.

These initial conditions are evolved under the influence of hydrodynamics, gravity, and cooling, with additional subgrid prescriptions for the creation of new star particles and feedback from young stars, specifically stellar winds, type II SNe, and a rough treatment of heating from photoionization. We also include a new prescription for self-consistently calculating the volumetric heating rate from the grain photoelectric effect.

5.2.1 Feedback

Our feedback prescription is designed to avoid the ad-hoc modifications to the physics that are typically necessary to produce feedback strong enough to affect the galaxy's properties. Our goal is to show that none of these modifications are necessary to have effective feedback at sufficiently high resolution. The basic physical picture is that once the adiabatic radius of the SNe can be resolved, they will expand in a somewhat realistic way, rather than immediately losing all their energy to the cell, as in the famous overcooling problem (Katz, 1992).

When each star particle is formed in the simulation with initial mass M_p, we draw a number from a Poisson distribution to determine how many type II SNe will be produced by the particle. The rate parameter of this Poisson distribution is taken to be $\lambda = 1104 M_p/(10^6 M_\odot)$, where the prefactor is the number of SNe produced over the lifetime of a $10^6 M_\odot$ stellar population formed in a burst of star formation according to Starburst99 (Leitherer et al., 1999) using a Chabrier IMF(Chabrier, 2003).

For each SN produced, we next draw a number from the delay time distribution of type II SNe, again determined from the output of Starburst99. This is essentially just the convolution of the lifetime of massive stars with the IMF. At the hydrodynamical time step during which the SN explodes, we add 10^{51} ergs to the internal energy of the cell where the SN resides. The mass of ejecta and its metallicity are also determined from the SB99 output, and are fit to piecewise-polynomial functions for use in Enzo. To ensure that the supernova goes off on the highest level of refinement in the simulation, we make the star particles "must-refine" if the particle has any SNe remaining in its future. This means that the cells containing such particles are marked for refinement to the highest resolution level, guaranteeing that they and at least the surrounding two cells in each direction will be on the highest resolution level.

In addition to the energy from the SNe added to the cell at the appropriate time step, we also add energy to the cell prior to the supernova itself. From the delay time for each SNe, we infer the mass of the star that will be exploding, from which we can estimate the total ionizing luminosity for the particle by adding up the contribution from every massive star that has yet to explode in that particle. Given an ambient density, i.e. the density in the surrounding gas cell, we may then estimate the volume of gas which can be ionized by the calculated luminosity, as in the standard Strömgren analysis. If this volume is larger than the cell, and the cell has a temperature below 10^4 K, internal energy is added to the cell to set the temperature to 10^4 K. The cell is not modified by this photoionization prescription if its temperature is above 10^4 K.

For cells with Strömgren volumes smaller than the cell volume, we compute the internal energy of the Strömgren sphere and compare it to the internal energy across the entire volume of the cell. If the whole cell has less internal energy than the Strömgren sphere alone, we add internal energy to the cell to make the two equal, and otherwise we do nothing.

This treatment of photoionization feedback is extremely conservative. In the limit of small cell sizes or low densities, we may vastly underestimate the size of the HII region, since only one cell will be at 10^4 K. In high-density or large cells, the total energy being injected into the gas is also conservative, in that the cell's energy is only changed if the entire internal energy of the cell is smaller than the internal energy associated with the Strömgren sphere on its own.

In addition to photoionization, we include feedback from stellar winds. Once again we rely on Starburst99 to compute the wind luminosity, with several small modifications. The specific energy of the wind is taken from Starburst99 only when the particle will in the future experience a SN from the death of a star with a delay time less than 10 Myr. If no such SNe will explode in the future, the cell still loses mass to winds, but their specific energy is set to $(10 \text{ km/s})^2$. This bifurcation in wind temperatures takes place in the real universe as a result of a sharp change in opacity in the atmospheres of stars as a function of mass.

One additional change we make to the specific energy of the wind occurs right at the beginning of the star particle's life. During this time the most massive stars in the population drive extremely hot winds, exceeding 10^9 K. When the outflowing winds dominate the material in a single cell as sometimes happens in our simulations, this high temperature can slow down the simulation substantially and cause other numerical problems associated with a high density contrast. To ameliorate this issue we cap the wind specific energy at $10^{8.5}$ K. We keep the total wind energy injected constant by slightly increasing the mass lost during this time. This changes the total mass returned by a tiny amount since this phase is so short.

Figure 5.5 shows that every supernova in the 10 pc resolution simulation with the 5 kpc initial scale length explodes in a cell with a density below the critical density at which a supernova remnant would cool before expanding to be the size of a single cell in the simulation. We calculate this critical density by adopting the following value for the radius at which a supernova remnant exits the Sedov phase and enters the pressure-driven snowplow (PDS) phase(Thornton et al., 1998),

$$R_{\mathrm{PDS}} = 14 \frac{E_{51}^{2/7}}{n_H^{3/7} (Z/Z_{\odot})^{1/7}} \mathrm{pc} \tag{5.2}$$

We take the energy of the supernova in units of 10^{51}ergs to be $E_{51} = 1$, and set the mass fraction of elements heavier than helium to a value appropriate for the sun, $Z = Z_{\odot} = 0.02$. In our initial conditions, the disk component has $Z/Z_{\odot} = 0.1$, and the halo has $Z/Z_\odot = 0.01$, but for the purposes of calculating the critical density we use a higher value because supernovae produce enough metals that locally Z may be substantially higher than its initial value. By setting $R_{\rm PDS}$ to the size of a single cell in the simulation, we can solve for the value of n_H, the number density of hydrogen atoms in units of $\rm cm^{-3}$, at which the PDS phase would be marginally resolved.

By resolving this crucial piece of physics in these galaxies, we find that our results are relatively insensitive to the resolution at which we run the simulations. We compare the depletion time for each physical scenario, run at 10 pc, 5 pc, and 2.5 pc resolution, in Figure 5.6, and we find that the results of the simulations tend to become independent of resolution after roughly 100 Myr of evolution. This is less clear for the runs which include SNe, for which the 2.5 pc simulations have not advanced as far as their lower-resolution counterparts, but even here there is reasonable agreement between the 10 pc and 5 pc runs.

5.2.2 Photoelectric Heating

FUV photons from young stars liberate electrons from dust grains in the interstellar medium. This is the primary means by which the neutral atomic gas in the interstellar medium is heated in the Milky Way. To include it in our simulation, we assume the following proportionality.

$$\mathcal{G} \propto F_{FUV}\ Z\ n_H\ [\rm erg/s/cm^3] \tag{5.3}$$

The heating rate from FUV photons is proportional to their flux, and the density of metals. At low densities and high temperatures, there is an additional dependence on the electron density and gas temperature, but these effects are negligible in the cold, dense gas where FUV heating is important for suppression of star formation, so we omit that effect. We also do not include cosmic ray heating, since this is roughly an order of magnitude less important than FUV heating under the optically thin conditions that prevail in the low density, dust poor galaxies we are simulating.

We calculate the FUV flux in the simulation by taking the luminosity for each star particle to be

$$\log_{10} L(t_7) = \log_{10}(M_p/10^6 M_\odot) + \begin{cases} \sum_{j=0}^{2} p_j t_7^j \text{ if } t_7 \leq 3 \\ \sum_{j=0}^{6} q_j t_7^j \text{ if } 3 \leq t_7 \leq 10 \\ 0 \text{ otherwise} \end{cases} \tag{5.4}$$

where t_7 is the age of the star particle in units of 10 million years, and M_p is the mass of the star particle. L is in units of erg/s. The coefficients of this polynomial equation are given in Extended Data Table 1. This is the result of integrating the output spectrum of a Starburst99 single-burst model for a cluster mass of $10^6 M_\odot$ (hence the pre factor in the previous equation) over the range 8 to 13.6 eV at finely spaced time intervals out to 100 Myr. Note that this function is somewhat sensitive to the IMF – recent indications of a bottom-light IMF(Geha et al., 2013) in dwarfs would somewhat increase it by a factor of ~ 2.

The FUV flux is then simply $F_{FUV} = L(t)/4\pi r^2$, where r is the three-dimensional distance from the center of the gas cell in question to the star in question. Note that this neglects any effects from self-shielding, which should be negligible in the galaxies we have simulated, given their low column densities and low metallicities. The total FUV flux at a given cell is the sum of this quantity over all stars in the simulation. If a given star's contribution to the FUV flux varies by less than 10% across a given grid (the computational element one step above cells in Enzo), then we approximate that star's contribution as constant across the grid, to avoid doing the full order N by M computation (where N is the number of cells and M is the number of particles) in regions far from FUV-emitting particles. To compute $\mathcal{G}/n_H$, we scale $\mathcal{G}/n_H$, F_{FUV} and Z to the known values of these quantities in the solar neighborhood,

$$\frac{\mathcal{G}}{n_H} = 8.5 \times 10^{-26} \frac{\mathrm{erg}}{\mathrm{s}} \frac{F_{FUV}}{0.0015859021\ \mathrm{erg/s/cm^2}} \frac{Z}{Z_\odot} \tag{5.5}$$

The numerical constant in the denominator is simply the Habing(Habing, 1968) estimate of the intensity of the interstellar radiation field in the solar neighborhood, multiplied by c to convert to a flux. The pre factor of 8.5×10^{-26} is the photoelectric heating rate normalized to the Habing value for the solar neighborhood from a radially-dependent model of the Milky Way's ISM(Wolfire et al., 2003). This quantity is computed for each cell in the simulation, and fed to the Grackle library[1] , which computes the rate of change of the internal energy density(Bryan et al., 2014; Kim et al., 2014).An example of $\mathcal{G}/n_H$ in the PE Only simulation is shown in Figure 5.7.

This method is a substantial improvement over many current implementations

---

[1] https://grackle.readthedocs.org/en/latest/

of the interstellar radiation field (ISRF). Many simulations do not include this source of diffuse heating at all. Some include it as constant throughout the simulation volume(Tasker, 2011; Tasker et al., 2015), sometimes with a correction for self-shielding(Hu et al., 2015), but without regard to the individual sources or time-dependence of the FUV photons. Other simulations explicitly approximate the radiative transfer of these photons(Oñorbe et al., 2015), while others go even further and explicitly model the dust particles via which this radiation interacts with the gas(Bekki, 2015a,c,b). Although the latter two methods are better approximations to the ISRF than what we have implemented here, our simulations have a higher resolution. As a result, we resolve the Sedov-Taylor phase of the SNe (Figure 5.5), and hence can definitively show whether supernova feedback or the ISRF is dominant in the regulation of star formation in dwarf galaxies.

5.2.3 Star Formation

The star formation prescription we use is similar to many commonly-used schemes. At each time step the probability of forming a star in each cell is taken to be

$$p_* = \begin{cases} 0 & \text{if excluded} \\ \min\left(1, \epsilon_{\rm ff} \frac{dt}{t_{\rm ff}} \frac{M_{\rm cell}}{M_p}\right) & \text{otherwise} \end{cases} \tag{5.6}$$

where the simulation time step (at the refinement level in question) is dt, $\epsilon_{\rm ff}$ is the efficiency of star formation per free fall time, and the free fall time $t_{\rm ff} = \sqrt{3\pi/32G\rho}$. Cells are excluded if their mass is less than the Jeans mass in that cell or if the cell is not on the maximum refinement level. In other words, cells must have a density exceeding

$$\rho > \frac{\gamma \pi k_B T}{N_J^2 G \mu m_{\rm H} (\Delta x)^2} \tag{5.7}$$

in order to form stars. Here $\gamma = 5/3$ is the ratio of specific heats, μ is the mean molecular weight in units of the Hydrogen mass, Δx is the size of the cells on the maximum refinement level, and $N_J = 4$ is the number of cells by which we require that the Jeans length be resolved throughout the simulation to avoid artificial fragmentation.(Truelove et al., 1998) This criterion corresponds to a diagonal line in the density-temperature phase diagram above which stars are not allowed to form (Figure 5.7). The value of p_* is such that the average star formation rate in the non-excluded cells follows a volumetric Schmidt Law $\dot{\rho_*} = \epsilon_{\rm ff} \rho / t_{\rm ff}$. This model has a few parameters that we must set, although we argue that we do not have a huge amount of freedom to change them.

The efficiency per free fall time is constrained by observations to be within a factor of a few of 0.01 (Krumholz & Tan, 2007; Krumholz et al., 2012), so we simply adopt this central value. The mass of individual star particles M_p should in principle be low enough to be irrelevant – this would have the advantage of sampling the star formation rate density very well and forming stars if and only if the cell were Jeans unstable. This introduces two numerical difficulties. The first is that the Jeans mass for the coldest gas in our simulations can be comparable to the mass of a single star. It would be problematic for our feedback recipe to form such low-mass stars, in that a 120 solar mass star or a supernova which ejects 10 solar masses of material could never exist within a 10 solar mass particle. Moreover the number of star particles we would have to follow increases as M_p^{-1}. Each of these issues would be avoided by choosing a large particle mass. However, using a large mass increases the chance that the cell in which the star formed would have insufficient mass to supply all of the gas needed to form the particle. When this happens, in order to maintain global mass conservation we gather the mass from neighboring cells. Each cell supplying gas contributes the same fraction of its mass to the new particle. We have found that a particle mass of $M_p = 50 M_\odot$ strikes a reasonable balance between keeping star formation local to 1 cell and keeping the supernova ejecta mass less than a single star particle's mass.

5.2.4 Comparison to observations

In Figure 5.8, we compare the star formation rate and depletion time of the simulations to those of dwarf galaxies from three different samples (Hunter et al., 2012; Cannon et al., 2015; Janowiecki et al., 2015). We exclude four blue compact dwarf galaxies from this sample, since these are known to be undergoing starbursts. We do not include estimates for the errors on each point. The statistical standard errors are of order 10%, but this under-predicts the true error, which is dominated by systematic uncertainty in distance (HI mass and star formation rate) and the assumptions made in converting UV or Hα luminosities to star formation rates(Chomiuk & Povich, 2011). For the two smaller galaxies in Janowiecki et al. (2015), no strict upper limits are given on the star formation rate, but since they are not detected in the same Hα image in which an SFR of $4 \times 10^{-3} M_\odot/\text{yr}$ was measured for the largest galaxy in the sample, we take this as a very conservative upper limit on the SFR for these two galaxies (Janowiecki, priv. comm.).

j	p_j	q_j
0	41.5709926107	40.7875024388
1	2.40501751872	-0.227682606645
2	-9.19544984847	0.0078916423535
3	10.5203892767	0
4	-5.72637964222	0
5	1.50479521662	0
6	-0.153355377095	0

Table 5.1: Extended Data Table 1. Parameters for the piecewise polynomial fit to the FUV luminosity of a simple stellar population as a function of its age (equation 5.4).

5.2.5 Data Availability

The numerical experiments presented in this work were run with a fork of the enzo code available from https://bitbucket.org/jforbes/enzo-dev-jforbes, in particular change set daed04d1e5e6. This altered version of enzo also requires an altered version of the grackle cooling library, available from https://bitbucket.org/jforbes/grackle, particularly change set 12d3856. A subset of the raw data files are also available online from www.johncforbes.com/dwarfs.html.

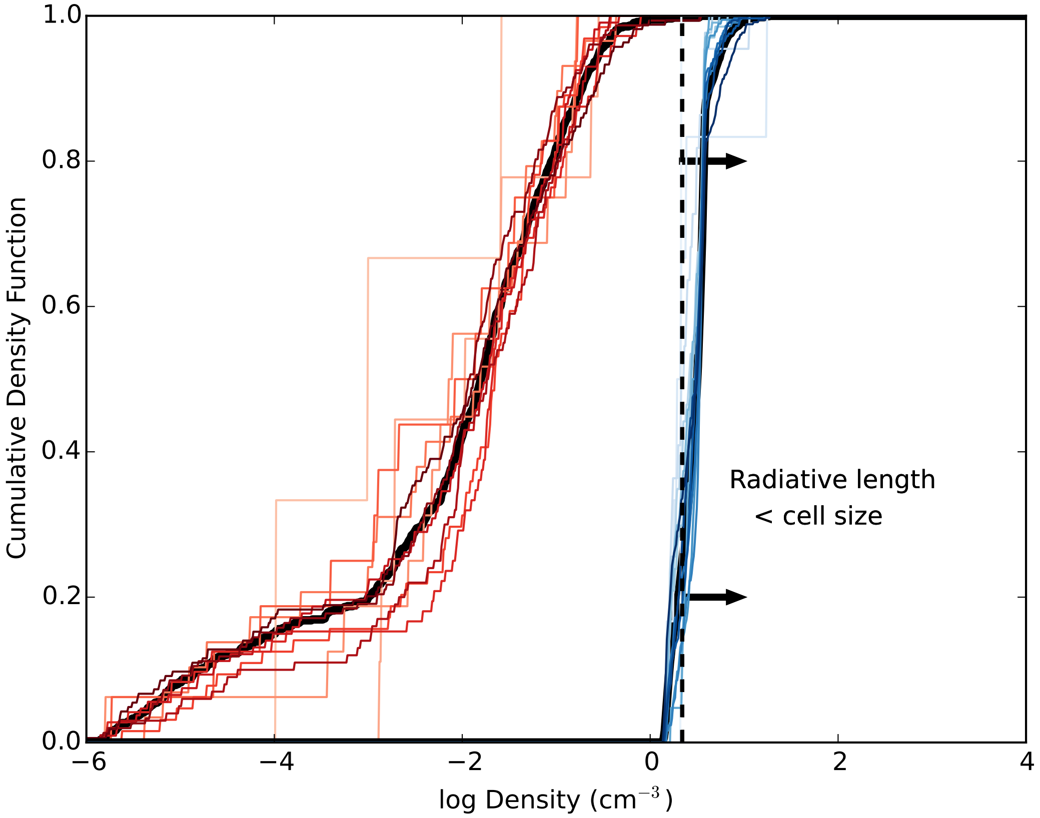

Figure 5.5: Supernovae are well-resolved. The cumulative distribution of the density of cells in which supernovae explode is shown in red, and in which stars form is shown in blue, with the thin lines showing these distributions in different 10 Myr intervals. The vertical dashed line indicates the density at which a supernova remnant would radiate all of its energy before it expanded to the size of a single cell (10 pc) in the simulation, assuming solar metallicity. Nearly every supernova in the simulation occurs to the left of this line, indicating that the simulation does not suffer from the overcooling problem.

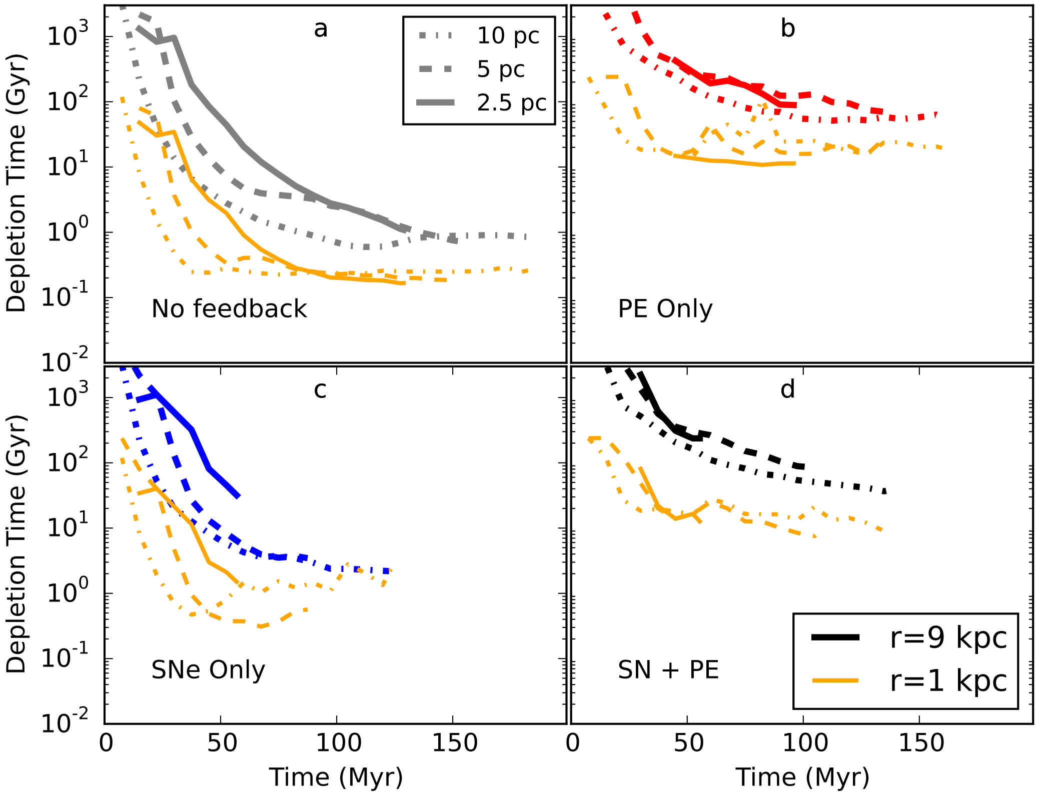

Figure 5.6: A resolution study. The depletion time of all 12 simulations with 5 kpc gas scale length is plotted over time. The panels show the four different feedback models, while the lines of a given color in each plot show the result for different resolutions. The orange lines show the depletion time when the measurement is carried out in a cylinder with 1 kpc radius, whereas the other lines use a 9 kpc radius. Regardless of the aperture, the simulations quickly converge; differences between simulations with factor-of-two differences in resolution are small compared to the differences resulting from changing the physics.

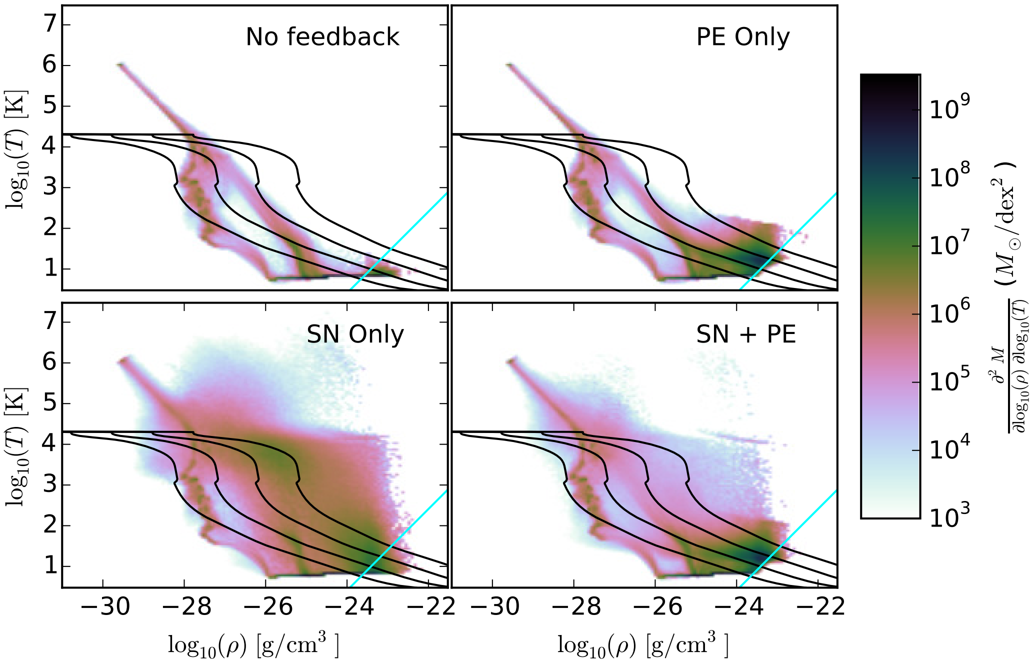

Figure 5.7: Phase diagrams after 90 Myr of evolution. The panels show runs with different feedback models and 5 kpc gas scale length, all at 10 pc resolution. The light blue diagonal lines show the threshold for star formation, namely where the gas becomes Jeans unstable on the highest refinement level. The black lines trace where the net cooling rate is zero, assuming different values for the volumetric heating rate, from 10^{-26} erg/s (highest line), to 10^{-29} erg/s. Photoelectric heating raises the typical temperature of gas near the star formation threshold such that moderate star formation can stabilize nearby gas against collapse.

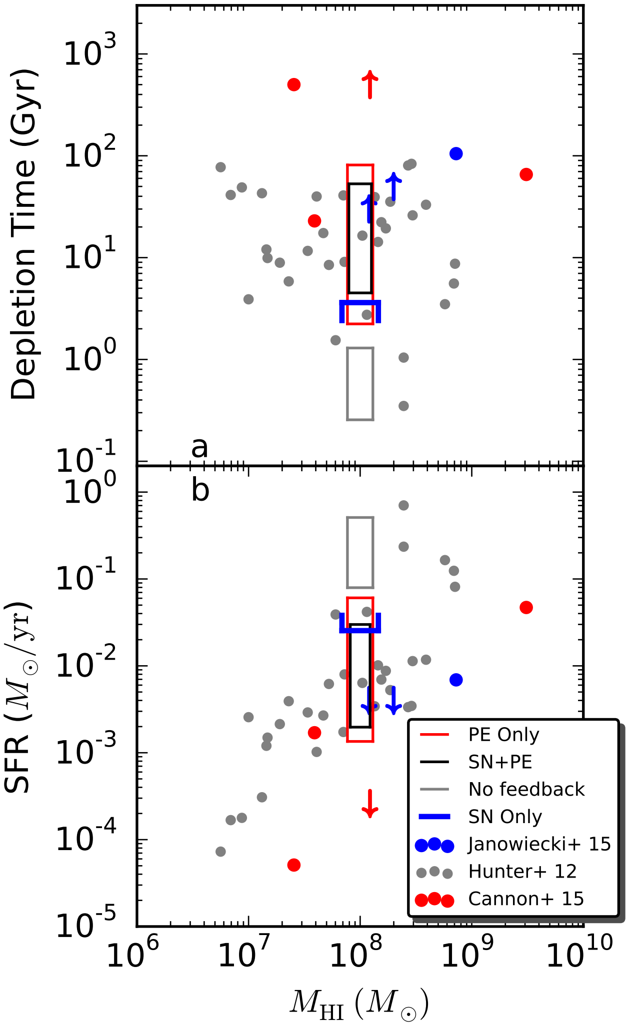

Figure 5.8: Comparison to observations. Star formation properties for a heterogeneous set (Hunter et al., 2012; Janowiecki et al., 2015; Cannon et al., 2015) of galaxies are shown as a function of gas mass. Boxes representing the range of values covered by our simulations are overplotted. The vertical range of the box is determined by the final snapshot for each simulation; the high-(low-) SFR extremum represents the 1 (5) kpc disks. No "SN Only" simulation was run for the 1 kpc case, so the blue box is not closed. Only simulations that include photoelectric heating agree with the depletion times observed for bulk of galaxies in the mass range we simulated.

Chapter 6

Galactic Winds

6.1 Introduction

Galactic winds are simultaneously among the most important and least-understood ingredient in the evolution of galaxies. Not only are they ubiquitously observed (Erb et al., 2006; Rubin et al., 2014), but they are likely responsible for setting the average star formation rates (SFRs), metallicities, and gas fractions of star-forming galaxies according to a simple but powerful class of models variously known as "bathtub," or "equilibrium" models (Davé et al., 2012; Lilly et al., 2013; Forbes et al., 2014b).

Although we can infer their importance observationally, we remain profoundly ignorant when it comes to very basic properties of galactic winds. Estimates of the mass loading factor (MLF), the rate at which mass is ejected in the wind to the SFR of the galaxy, can vary by orders of magnitude at a given galaxy mass (Zahid et al., 2014; Muratov et al., 2015). The degree to which winds are preferentially metal-enriched compared to the bulk of the interstellar medium is completely unknown. Even the physical mechanism responsible for launching the winds is unclear, with supernovae (e.g. Dekel & Silk, 1986), radiation pressure acting on dust grains (Murray et al., 2005), and cosmic rays interacting with small-scale features in the magnetic field (Socrates et al., 2008) each being proposed as the potentially dominant driver of winds.

Much of the uncertainty about the strength and driving mechanisms of galactic winds stems from the difficulty of both simulating and observing them. Although simulating the effects of supernova (SN) blast waves is straightforward in principle, i.e. it does not require including physics much beyond ordinary hydrodynamics, their effects have been impossible to model in hydrodynamic simulations. The Sedov-Taylor phase, the most important phase of their evolution, occurs on very small scales relative to the galaxy as a whole (Thornton et al., 1998). As a result, adding the energy of the SNe to the gas without sufficiently high resolution spreads the energy over too large a mass of gas, causing the simulated SN remnant to skip the Sedov-Taylor phase entirely. In simulations subject to this "overcooling problem," SNe have virtually no effect on the galaxy (Katz, 1992; Tasker et al., 2015). Simulators have turned to ad-hoc solutions, for example temporarily disabling the ability of the gas to radiate energy in the vicinity of a SN (Stinson et al., 2006; Governato et al., 2010), or saving up energy from SNe that should have exploded and then releasing it all at once (Dalla Vecchia & Schaye, 2012). These methods have made simulated galaxies look more realistic by making the SNe more effective at driving outflows. However, by using these unphysical workarounds the predictive power of the simulations is compromised. It is therefore unclear to this day whether SNe are truly responsible for driving galactic winds.

Here we present high resolution simulations of a single isolated dwarf galaxy in a non-cosmological context. First results from these simulations were presented in the previous chapter. Our simulations simultaneously have a high enough resolution and a low enough ambient gas volume density in the interstellar medium that individual supernova explosions are well-resolved without any ad-hoc modification to the physics of the simulation. Our simulations therefore represent a high-fidelity physical model of a dwarf galaxy taking into account thermal feedback from supernovae, pre-supernova mass loss from stellar winds, local photoionization heating, and as discussed at length in the previous chapter, photoelectric heating of dust grains by FUV photons. Other physical processes that may be responsible for launching winds, i.e. cosmic ray and radiation pressure, are not included. The details of the physics governing these processes is sufficiently uncertain that any treatment of these processes will necessarily be prohibitively expensive or likely incorrect. We consider our simulations to be a useful benchmark for what would happen assuming the most straightforward, least uncertain scenario.

6.2 Simulations

Simulations are carried out with the Enzo adaptive mesh refinement code (Bryan et al., 2014). A cartesian box with a side length of 1.3 Mpc is initialized with a uniform density of $\rho_h = 10^{-6}$g/cm^3, a temperature of $T_h = 10^6$ K and a root grid size of 32^3. Near the center of the box, we initialize a dark matter halo and an 'old' stellar population, each represented by collisionless particles which only interact gravitationally with the remainder of the simulation. The dark matter halo has a mass of $10^{10} M_\odot$, while the stars have a total mass of $10^7 M_\odot$, in reasonable agreement with the inferred stellar mass-halo mass relations derived by abundance matching (Moster et al., 2010; Behroozi et al., 2013a). Finally, a gaseous disk with temperature $T_d = 100$ K which falls off exponentially in density both vertically and radially as $\rho_d(r, z) \propto \exp(-|z|/H) \exp(-r/r_d)$ is added. The disk extends until its pressure would fall below that of the hot halo, i.e. until $\rho_h T_h = \rho_d T_d$, at which point the disk component is truncated and only the halo is left. The total mass of this cold disk is taken to be $M_g = 10^8 M_\odot$, in agreement with the observed relation between gas and stellar mass (Papastergis et al., 2012). The disk begins the simulation with a metallicity of $Z_d = 0.1 Z_\odot$, while the halo has $Z_h = 0.01 Z_\odot$.

The initial coarse grid is refined when any of the following criteria are met for each individual cell

1. The Jeans length is not resolved by 4 cells
2. The cell contains a star which will at a future point in the simulation experience a supernova
3. The cell exceeds a particle mass of $3 \times 10^5 \times 2^{-0.444916 L} M_\odot$, where L is the level of refinement of the cell in question.
4. The cell is within 2 cells of a cell satisfying any of the above conditions.

We include 12 levels of refinement, so that the finest resolution available in the simulation is 10 pc.

Stars are allowed to form only on this final level of refinement, and only if the

Jeans length of the cell is not resolved by 4 cells. In other words, stars form in cells that would have refined further but were unable to do so owing to already being at the maximum refinement level. In cells that meet this condition, star particles are created stochastically with a probability such that the average star formation rate is $\epsilon_{\rm ff}\rho/t_{\rm ff}$, with the star formation efficiency per freefall time $\epsilon_{\rm ff} = 0.01$ (Krumholz & Tan, 2007; Krumholz et al., 2012).

Stars formed during the simulation time add energy to the cells in which they are contained. The wind luminosity for a Starburst99 (Leitherer et al., 1999) is added to the cell as thermal energy if the star particle will experience a supernova in the future (indicating that the particle contains a massive star). Which particles contain stars that will experience supernovae are chosen stochastically such that the supernova rate is equal to $10616/10^6$ supernovae per solar mass of stars formed, as expected for a Kroupa IMF (Kroupa, 2001) and a low initial metallicity according to Starburst99. For each supernova experienced by the particle, a delay time (between the particle's formation and the supernova) is drawn from the supernova delay time distribution. Additionally, particles containing massive stars set the temperature of their cell to 10^4K if their Stromgren volume exceeds the cell's volume.

6.3 Results

Given the artificial initial conditions where no attempt was made to set up hydrostatic balance, the gas disk quickly collapses in the center, and begins to form stars. The star formation rate rises steadily until plateauing to about $10^{-3}M_\odot$/yr after

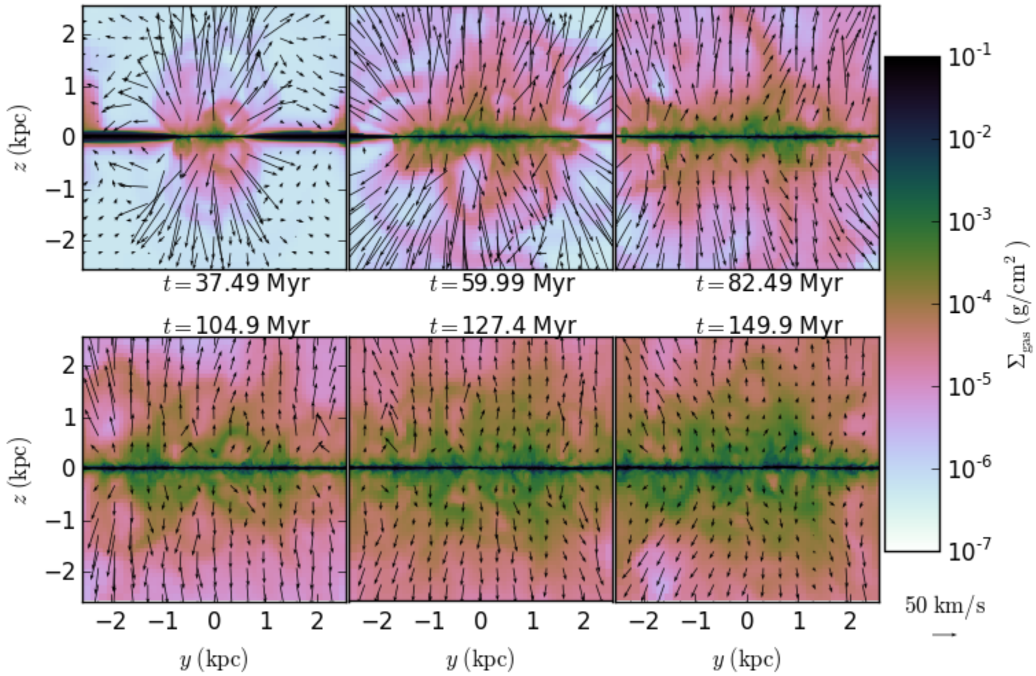

Figure 6.1: A shallow edge-on projection of the galaxy over time. Material outflowing from the galaxy is clearly visible at all times. It begins as high-speed and low-density, but as the simulation develops, the region above and below the disk increases in density and decreases in typical velocity. The net effect is a moderate increase in the mass outflow rate, but much of the mass is close to the threshold of being unbound.

around 100 Myr of evolution. During this time an outflow develops from the disk driven by the thermal energy input of supernovae. The morphology of this outflow is shown in Figure 6.1 as a temporal sequence of snapshots. The quantity displayed is the column density of gas, with the galaxy viewed edge-on, and centered on the center of the galaxy. The column density is the integral of the volume density extending ±500 pc along the line of sight, i.e. the depth of the column is shallow so that the structure of the wind is more clearly visible. Arrows in each panel show the mass-averaged velocity in the y-z plane. The column density and velocity scales are the same in each panel.

To understand what portion of the outflowing gas has the potential to escape

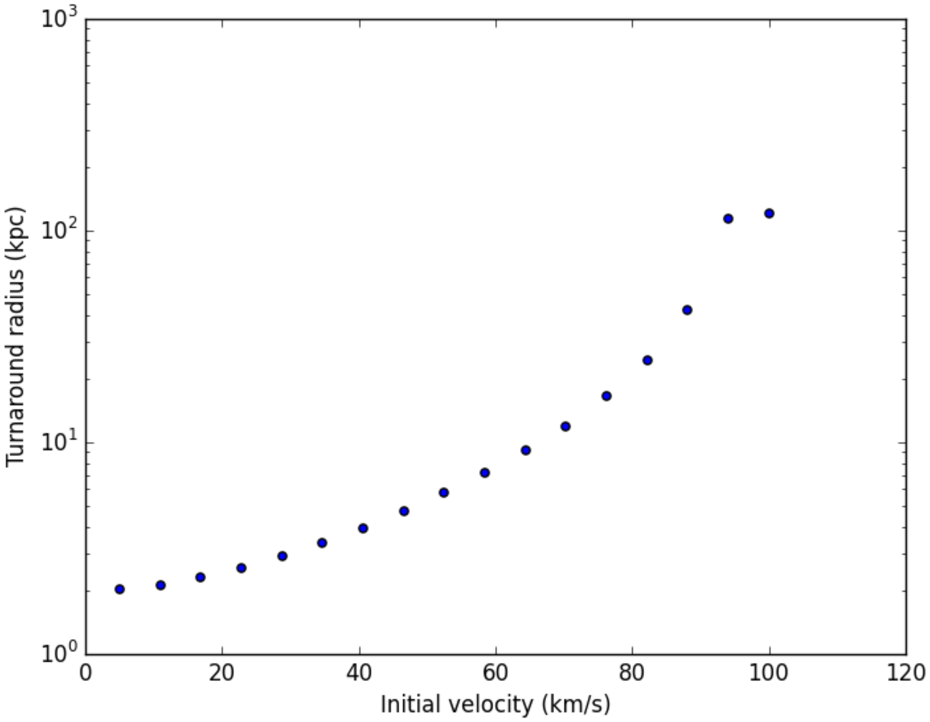

Figure 6.2: Turnaround radius. For a particle on a radial orbit starting at a radius of 2 kpc from the center of the dark matter halo, we integrate the particle's trajectory and find the radius where the particle's velocity drops to zero. The final point at 100 km/s is above the escape velocity, so the particle never actually turns around. To even temporarily escape the halo, assuming no further acceleration of the particle except via gravity, the gas needs to be going around 80 km/s.

the galaxy, we examine the radius a collisionless particle can reach before turning around and falling back to the galaxy in Figure 6.2. Particles are numerically integrated on radial orbits using an interpolated version of $GM(<r)/r^2$ measured from the simulation (note that this profile is essentially constant over the course of the simulations, and nearly identical among all simulations presented here). Assuming that no further forces apply to the gas besides gravity, most (though not all) of the outflow in our simulations is formally bound to the halo.

The gas mass escaping from the galaxy may be measured by counting up the flux of outflowing mass in cells that intersect some surface in the simulation. For simplicity, we choose horizontal surfaces (parallel to the plane of the disk), above and below the disk at various heights, with a radius encompassing the star-forming region of the disk. The mass flux through these planes is summed up on a cell-by-cell basis,

$$\dot{M}_{\mathrm{out}} = \sum_i \rho_i v_{\mathrm{out},i} A_i \tag{6.1}$$

where ρ_i is the gas density of the ith cell, $v_{\mathrm{out},i}$ is the z-component of the velocity with a sign such that positive indicates gas moving away from the disk, and A_i is the cross-sectional area of the cell in question. Cells are only included in this sum if their $v_{\mathrm{out},i}$ exceeds some threshold.

The outgoing mass flux obtained in this manner is shown in Figure 6.3, normalized to the star formation rate. This normalization can be done in several ways. Naively, at a given time the mass loading factor is just the current outflow rate divided by the current star formation rate. However, the star formation rate does change over time, and there is a finite travel time between where the wind is launched and where the outflow rate is measured, so we also include a mass loading factor where the star formation rate in the denominator is not the current star formation rate, but rather the star formation rate when the wind was launched. This time is estimated by dividing the height at which the outgoing mass flux has been computed by the average velocity of the outflowing gas which exceeds the velocity threshold.

The mass loading factors we compute here are intermediate between various

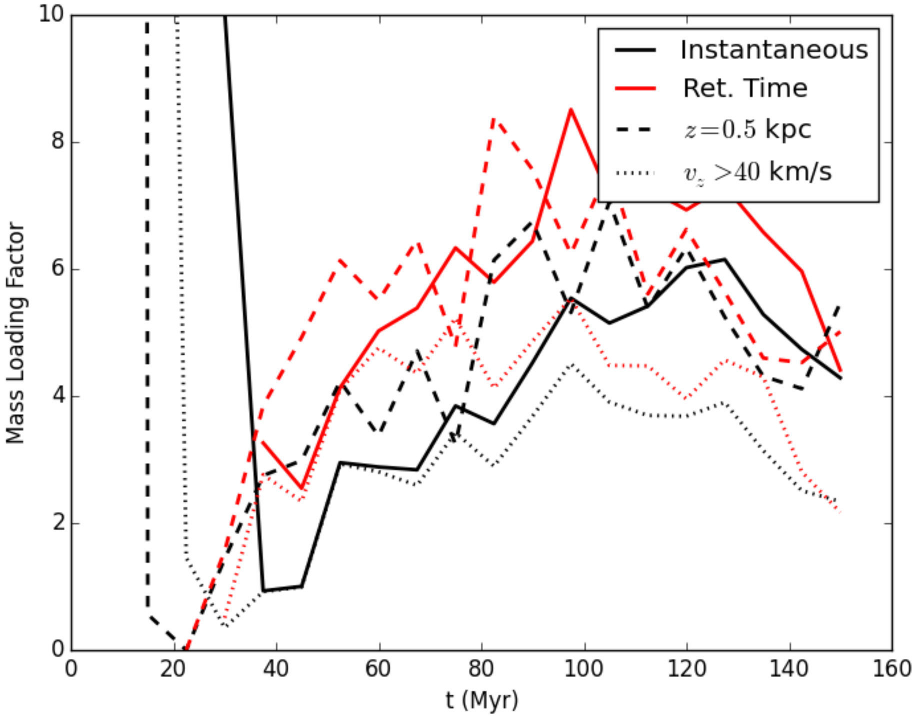

Figure 6.3: The Mass Loading Factor. Shown are the ratio of the mass outflow rate to the star formation rate. The black line uses a naive estimate, namely the instantaneous outflow rate divided by the current star formation rate. The red line shows an estimate based on the star formation rate at a previous time, namely the current time minus the time it would take the outflowing gas to travel to the height where the outflow rate is being measured.

extreme expectations (Schroetter et al., 2015; Muratov et al., 2015), and are much smaller than the values necessary to explain galaxy scaling relations with the equilibrium model discussed in Chapter 2.

In addition to the mass outflow rate, we would also like to know the metal content of the wind. This is particularly important because when modellers attempt to fit observed galaxy data, one of the main tools to discriminate between different feedback scenarios is the mass metallicity relation. However, the strength of outflows is degenerate with whether or not the winds are preferentially metal-enriched, in which case outflows do not need to be as strong to explain a given deficit in metallicity.

To quantify this, we use a variable defined in Chapter 2, and discussed in somewhat more detail in Chapter 4.

$$\xi = (\dot{M}_{Z,\mathrm{out}} - Z_0\dot{M}_{\mathrm{out}})/(y f_R \dot{M}_{\mathrm{SFR}}) \tag{6.2}$$

where Z_0 is the metallicity of the gas in the disk, $\dot{M}_{Z,\mathrm{out}}$ is the flux of metal mass through the same surface as the outflowing gas mass $\dot{M}_{\mathrm{out}}$ is measured, y is the yield, and f_R is the remnant fraction. This parameter is equal to zero if the wind is the same metallicity as the disk, and 1 if all of the metals produced in supernovae are ejected in the wind. Figure 6.4 shows the evolution of this quantity over time in the simulation, both with a naive estimate and using the star formation rate at a retarded time. The time during which $\xi > 1$ is due either to a mismatch in the relevant star formation rate (i.e. the fact that the SFR is not constant, but rather increasing), or due to an error in the values used for y and f_R. Despite this uncertainty, it seems probable that ξ is of order unity, not zero. This suggests that imperfect metal mixing, i.e. preferential ejection of metals, may play a major role in the evolution of low-mass galaxies.

In addition to the raw outflow rates and metal enhancement of the wind, we can look at the wind's detailed properties by examing the probability distribution function of various wind properties. These PDFs are weighted by mass, metal mass, and ejecta mass. Ejecta refers specifically to material which is added to a cell in which a supernova explodes. This supernova ejecta is tagged with a passive scalar tracer field, which is

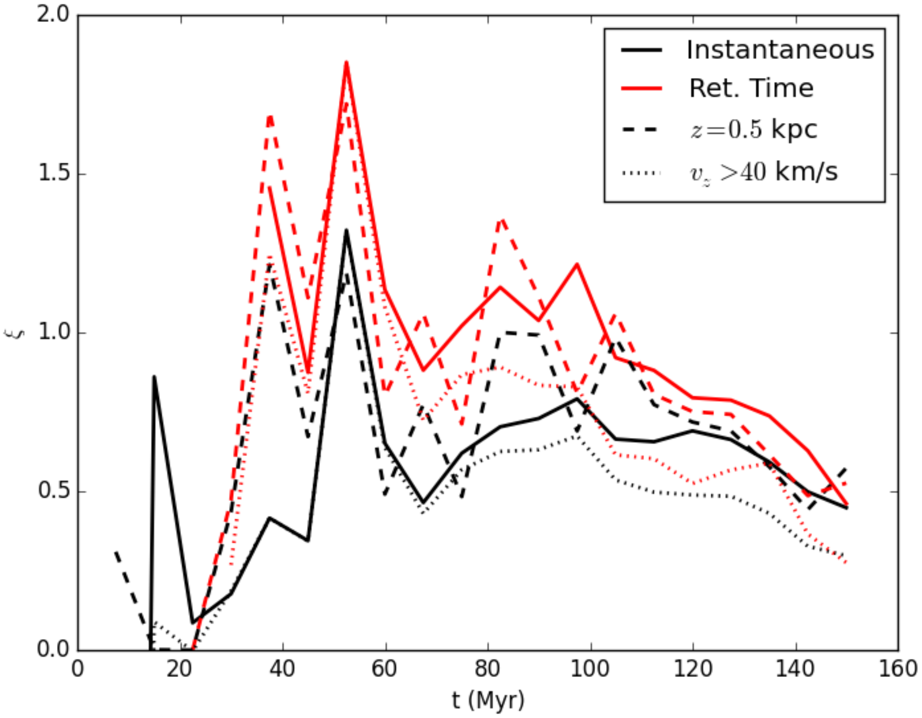

Figure 6.4: The Metal Mixing Factor. The metal enhancement of the wind. The dimensionless quantity ξ shows the metal content of outflowing material. When $\xi = 0$, the wind metallicity is identical to the gas phase metallicity of the disk, whereas when $\xi = 1$, 100% of the metals produced in supernovae leave the galaxy in the wind.

advected along with the gas, just like metals and other chemical species. Although we do not use any lagrangian tracer particles, this method allows us to track the ejecta as it leaves the galaxy.

Figures 6.5, 6.6, 6.7, and 6.8 show PDFs for the outflowing velocity, transverse velocity, density, and temperature of material in a volume 1 kpc above and below the disk, each with a thickness of 0.25 kpc. I.e. these PDFs represent wind material, or at least gas interacting with the wind, in bulk. These plots show that the wind is composed of a diverse array of gas, with velocities largely near the rotational velocity of the galaxy, but with a tail extending to much higher velocities. The density of the ejecta-rich gas ranges from much lower to much higher than the bulk of the gas above and below the disk, which agrees with the morphology shown in Figure 6.1 of structures resembling bubbles, shells, and filaments.

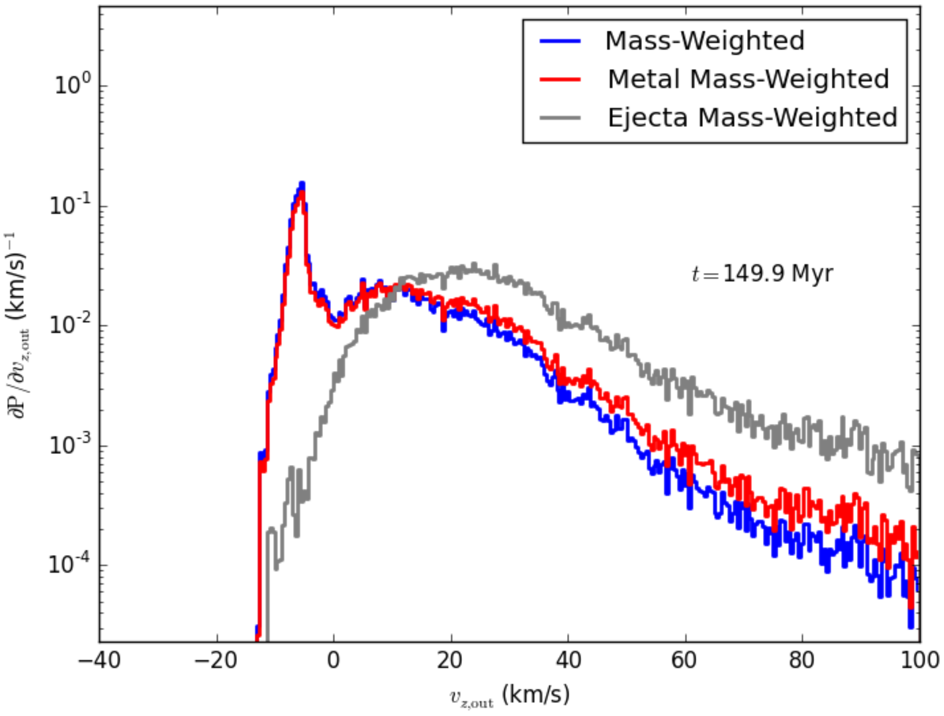

Figure 6.5: The PDF of Outflowing Velocities. In a volume of thickness 0.25 kpc, height 1 kpc above and below the disk, and radius 10 kpc, we show the probability distribution function of the z velocity of the gas, weighted by mass, metal mass, and ejecta mass, where a positive z velocity indicates material flowing away from the disk. This PDF is computed at a single time, $t = 150$ Myr into the simulation. Material that is part of the wind clearly shows an exponential tail to large velocities, with a peak around the rotational velocity of the galaxy, 40 km/s.

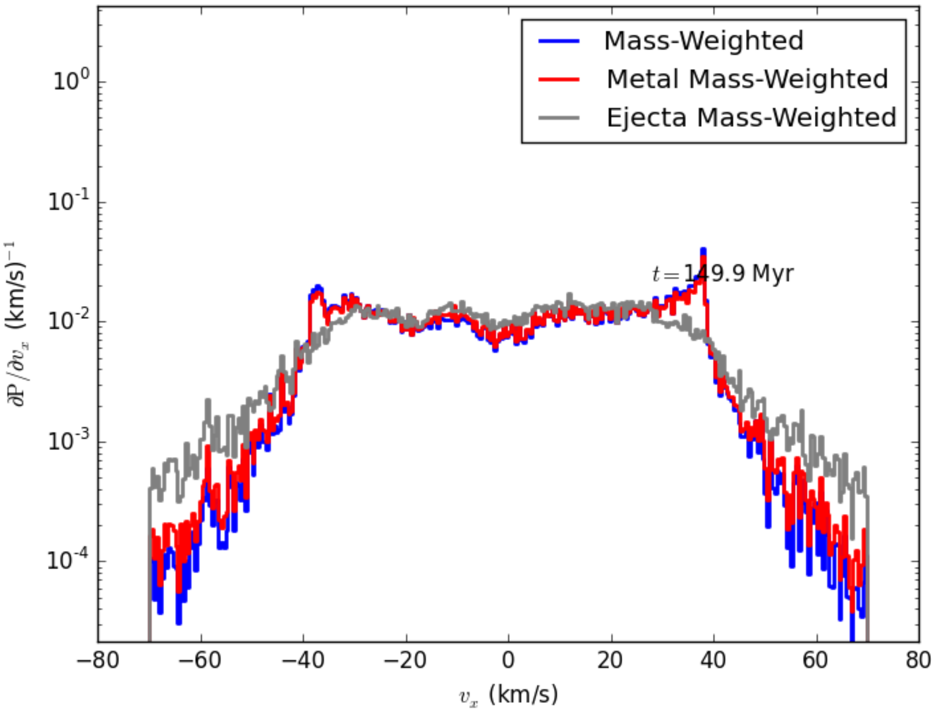

Figure 6.6: The PDF of Transverse Velocities. In the same volume used in Figure 6.5, we show the PDF of the x-component of the velocity. For the bulk of the mass and metals, the velocity shows a double-peaked distribution characteristic of a rotating disk. Interestingly, the initial conditions contain no rotation for gas at this height. Rather, the rotation of the cold disk has transferred angular momentum to the halo gas. However, the ejecta material itself is more consistent with a spherical expansion, showing little evidence of rotation.

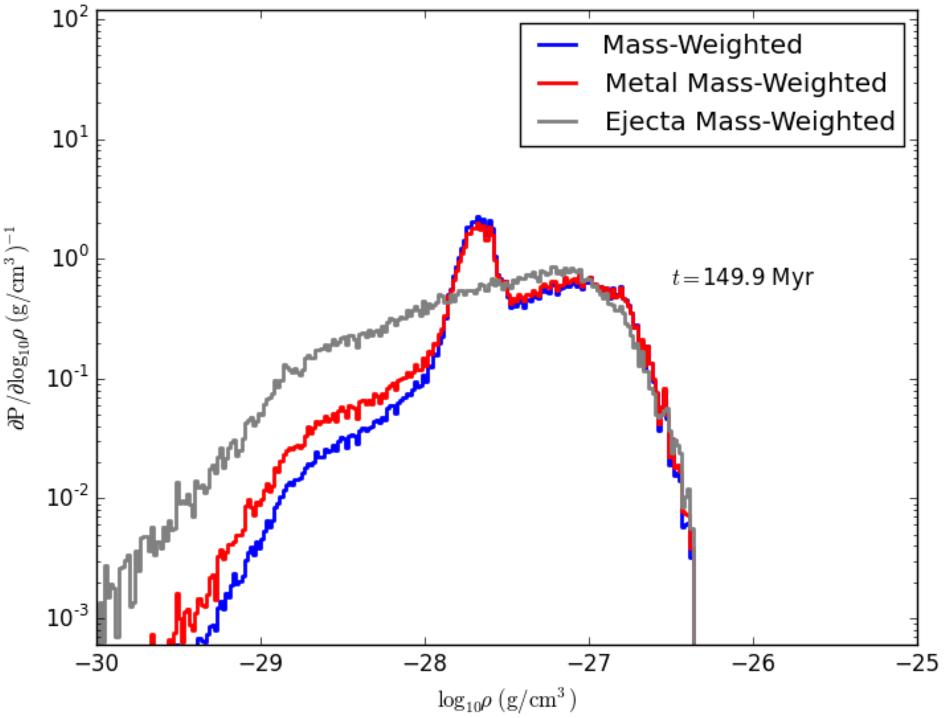

Figure 6.7: The PDF of Gas Density. In the same volume used in Figures 6.5 and 6.6, we show the PDF of the gas density.

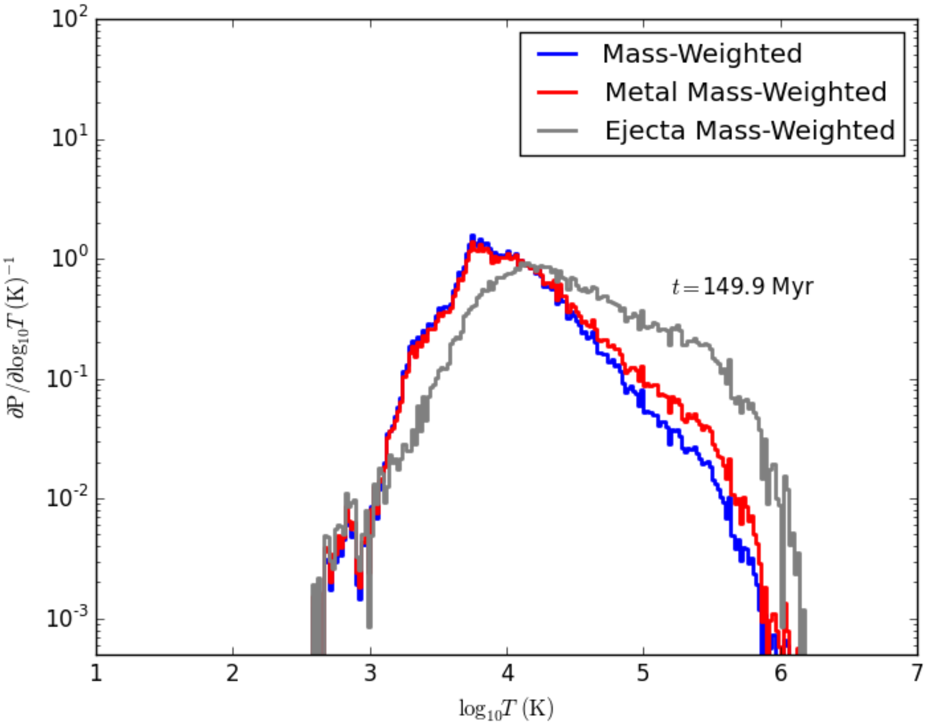

Figure 6.8: The PDF of gas temperatures. In the same volume used in Figures 6.5 and 6.6, we show the PDF of the gas temperature.

6.4 A comparison to other simulations

In addition to the fiducial simulation discussed above, we have examined the winds produced in two alternative simulations. In the first, as described in the previous chapter, we turn off photoelectric heating. The primary effect of disabling photoelectric heating is that the bulk of the ISM becomes extremely cold, and the star formation rate increases dramatically. In Figures 6.9, 6.10, and 6.11 we show the basic properties of the wind in this artificial circumstance. We find that not only is the outflow rate higher, but it is proportionally higher and proportionally more metal-enriched than in the fiducial case. Even though the star formation rate is an order of magnitude higher, the mass loading factor and the metal mixing factor are only moderately larger than in the fiducial case, suggesting that the absolute magnitude of the star formation rate has some influence on the properties of the wind, but the outflow rate is nearly linear with the star formation rate if the galaxy is otherwise the same.

Another simulation we ran and discussed in the previous Chapter was identical to the fiducial simulation, except that the scale length of the gas disk was much smaller, 1 kpc instead of 5 kpc. The star formation rate in this simulation is also much higher than in the fiducial case because the gas surface density is 25 times larger. Figures 6.12 and 6.13 show the basic results for this simulation. With the caveat that this simulation has not been run as long as the other two discussed here, the results are shockingly different. There is some outflowing mass, roughly comparable in flux to the fiducial simulation, but the higher star formation rate means that the mass loading factor is much smaller, less than unity. This dramatically changes the nature of the galaxy's

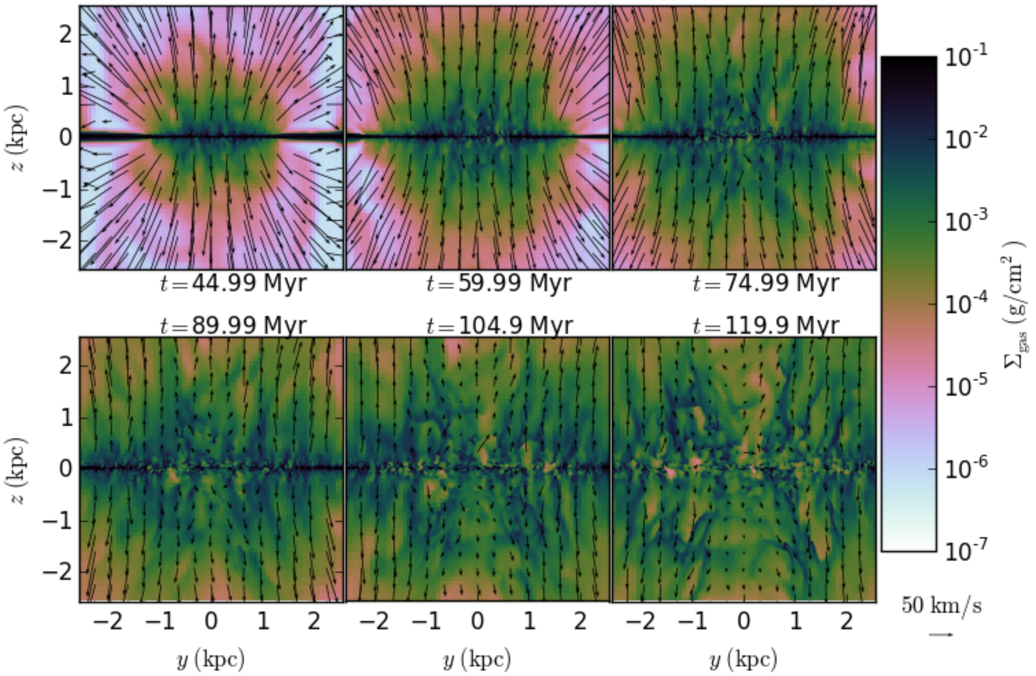

Figure 6.9: The morphology of the wind in the simulation with photoelectric heating disabled.

cosmological evolution, which may ultimately lead to a very high variance in e.g. the stellar mass - halo mass relation (Garrison-Kimmel et al., 2016) or the mass-metallicity relation (Guo et al., 2016) in the dwarf regime.

Finally, we use the 1D disk evolution code described in Chapters 2 and 3 to explore the cosmological evolution of a set of galaxies chosen to represent three basic scenarios that are all consistent with observed data. One simulation has a high mass loading factor of 30, perfect metal mixing ($\xi = 0$), and efficient cooling of gas from the cosmic web (i.e. the baryonic accretion rate follows the dark matter accretion rate, reduced only by the cosmolgical baryon fraction). This is the most typical way simulators create dwarf galaxies that obey observational constraints at redshift zero. The next model uses parameters more closely aligned with the results of the hydrodynamic

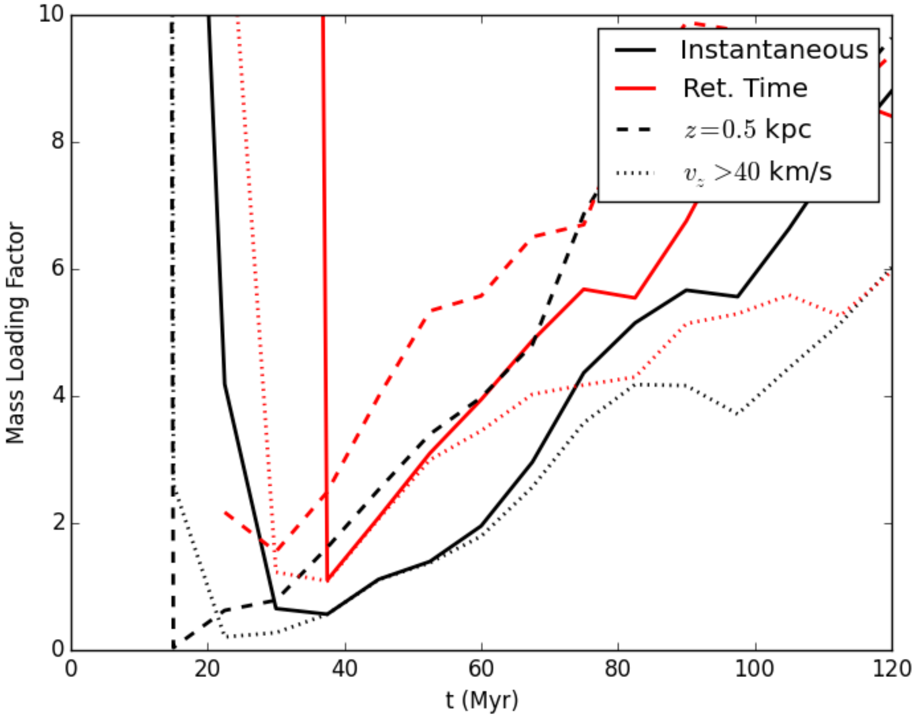

Figure 6.10: The Mass Loading Factor in the absence of photoelctric heating. The increased star formation rate in this simulation where photoelectric heating is artificially disabled results in a moderately higher mass loading factor than in the fiducial case.

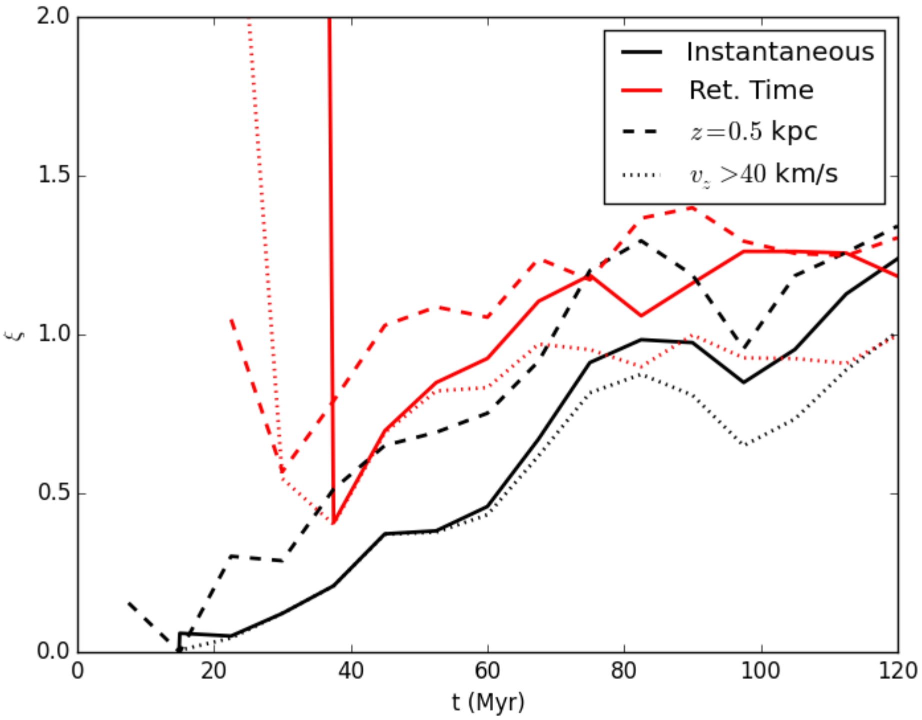

Figure 6.11: The Metal Mixing Factor in the absence of photoelectric heating. The increased outflow rate in this simulation is accompanied by a wind which is, relatively speaking, even more metal-enhanced than in the fiducial case. Essentially all metals produced by supernovae are leaving the galaxy in the wind.

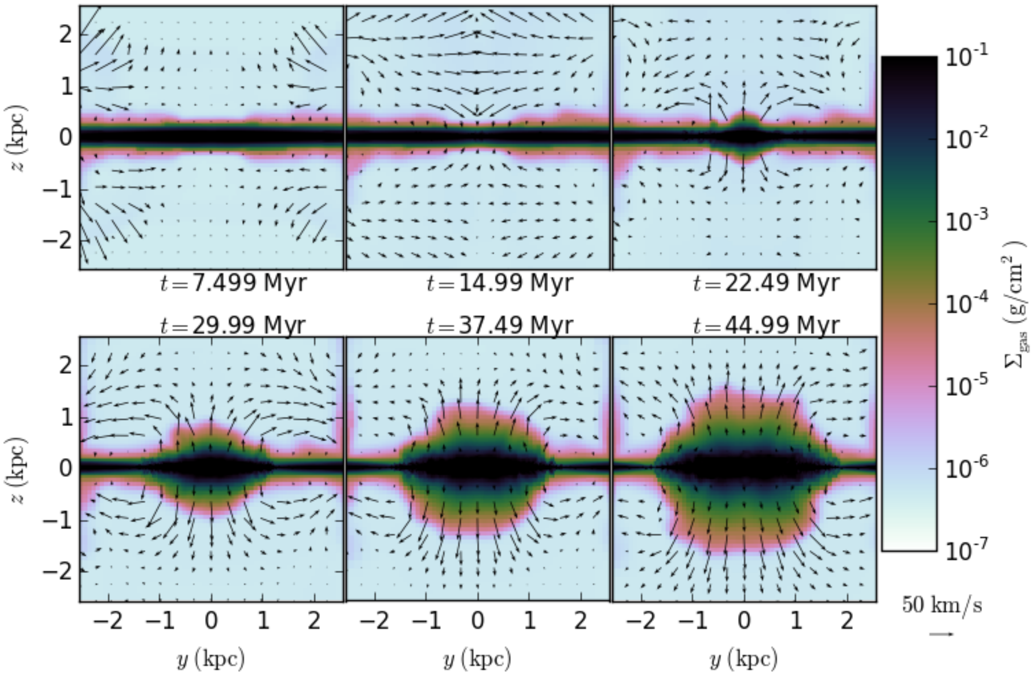

Figure 6.12: The morphology of the wind in the simulation with a 1 kpc scale length. A wind is clearly developing, but it is unclear whether it will end up carrying substantial mass compared to the star formation rate.

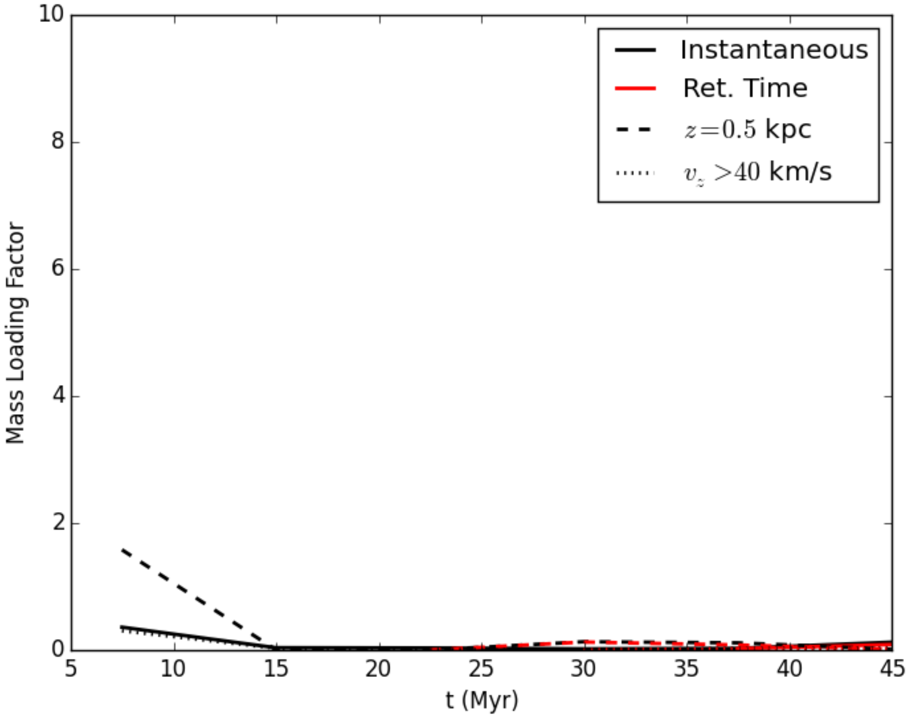

Figure 6.13: The Mass Loading Factor in the simulation with a 1 kpc scale length. The increased star formation rate in this simulation results in a negligible mass loading factor.

simulation. In particular, we set $\eta = 6$ and $\xi = .5$. Finally, we use a model in line with the preventative feedback picture proposed by Lu et al. (2015), in which low-mass galaxies are preferentially unable to accrete material from the IGM because that gas was heated at an earlier epoch. For this model, we use $\eta = 1$ and $\xi = 0$, and we only allow 10% of the baryons expected to accrete onto a halo of this size to enter the galaxy.

Figure 6.14 shows the evolution of the gas mass of these three representative models over cosmological time. Despite all ending up in the same area at $z = 0$, these different sets of physical parameters imply vastly different evolutionary histories, with high mass loading factors allowing the galaxy to be in equilibrium, including a burst of rapid star formation at high redshift. For low mass loading factors, the star formation history steadily rises with time as mass and metals build up in the galaxy. In this case the current state of the galaxy strongly depends on the galaxy's history.

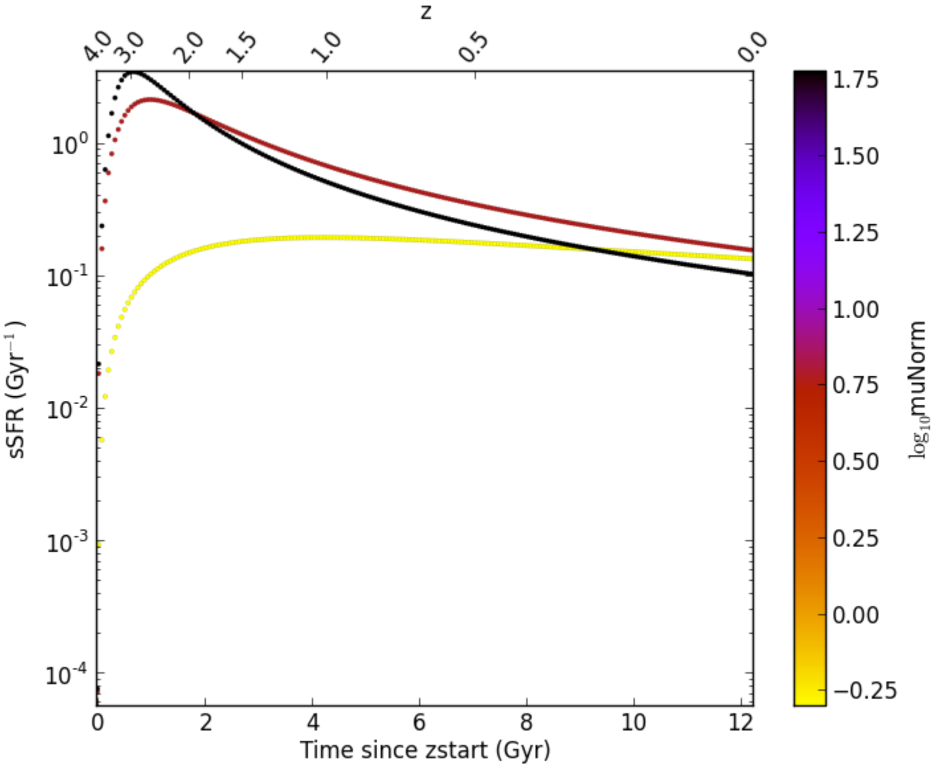

Figure 6.14: Cosmological evolution of the specific star formation rate of several representative models, with varying mass loading factors.

6.5 Conclusion

We have run a high resolution hydrodynamic simulation of a dwarf galaxy with well-resolved supernova feedback. We find that a galactic wind develops in this galaxy with a mass-loading factor of order 5 and a highly metal-enriched wind, in strong contrast to the typical model of a dwarf galaxy employed to explain the mass-metallicity relation, namely one in which winds are much stronger, and are perfectly mixed with the ISM of the galaxy.

We find that these results are likely to depend strongly on the scale length of the galaxy, with more compact galaxies having much weaker feedback given their star formation rates. This implies that galaxies at the same halo mass may end up being vastly different, leading to large scatter in relations that are tight for more massive galaxies, signifying the selective breakdown of the equilibrium model of galaxy evolution.

Chapter 7

Conclusions

7.1 The scatter in galaxy scaling relations

Star-forming galaxies have fundamental properties that lie along several well-known powerlaw-like relations. In the present-day universe, knowing only a galaxy's stellar mass, one can predict with reasonable confidence the galaxy's star formation rate (e.g. Whitaker et al., 2012), metallicity (Kewley & Ellison, 2008, e.g.), and gas fraction (Papastergis et al., 2012). To understand why galaxies follow this locus of points, simulators have employed hydrodynamical simulations (e.g. Davé et al., 2011) or semi-analytic models (e.g. Lu et al., 2014), wherein sub-grid parameters are tuned to match these and other relations. These models can be quite computationally intensive and non-intuitive. However, in the past 5 years it has been realized (e.g. Davé et al., 2012; Lilly et al., 2013) that these relations can be understood as a consequence of rough equilibrium in a galaxy's mass budget between new accretion, and gas being locked in long-lived stellar remnants or ejected from the galaxy. So long as the timescale to reach equilibrium is reasonably short compared to the age of the universe, galaxies will have gas masses, star formation rates, and metallicities determined by simple algebraic formulae.

Equilibrium models provide a powerful and simple way to understand the origin of galaxy scaling relations. In Chapter 2, I set about trying to understand the scatter in galaxy scaling relations in this context. To do so, I wrote down the mass and metal continuity equations, which are in general differential equations that should be integrated over the entire lifetime of a galaxy. I solved these differential equations analytically for the case of accretion rates that remained constant for some timescale $t_{\mathrm{coherence}}$ before switching to a new random value drawn from a log-normal distribution. With this simple *ansatz* for the behavior of the external accretion rate, I mapped out the scatter one would expect in the star formation rate and metallicity, and the anti-correlation between star formation and metallicity, all at fixed stellar mass. Each of these quantities is directly observable, so I was able to place surprisingly strong constraints on the parameters of the accretion process and other possible sources of scatter.

With this procedure, I found that my initial guess for the intrinsic scatter in the accretion rate (based on Neistein & Dekel, 2008), was too large to be accommodated by the observations. This suggests that the gas accretion rate may be smoothed out by some process in the halos of galaxies. I also found that the scatter in the mass loading factor at fixed mass needs to be surprisingly small, less than about 0.1 dex. This is truly remarkable, because independent theoretical and observational estimates for the mass loading factor at any particular mass are hugely uncertain, sometimes by several orders of magnitude. We do not even know the physical mechanism responsible for launching galactic winds, and yet in galaxies of a fixed mass, somehow the spread in mass loading factors must be tiny.

7.2 The radial profiles of galaxies over cosmic time

The *Hubble Space Telescope* discovered that broadband images of high-redshift galaxies are frequently dominated by a handful of large clumps (Elmegreen et al., 2007). This discovery, unanticipated by theory, led to the development of a new paradigm in the evolution of galaxies, in which massive gas-rich disks are fueled by cold cosmological streams of accreting gas (Kereš et al., 2005; Dekel et al., 2009b), leading to gravitationally unstable disks (Dekel et al., 2009a; Ceverino et al., 2010). In this picture, the clumps are largely caused by the instability of the disk to the collapse of structures on the Toomre scale.

I set out to understand the long-term consequences of this picture of galaxy evolution by writing a fast code that could evolve the radial profiles of gas and stars in galaxies subject to this instability. Based on the formalism presented in (Krumholz & Burkert, 2010), I wrote a brand new code that solved their evolution equations orders of magnitude more quickly, and I added new physics, transforming the calculations from an idealized toy model to a realistic model of galaxy evolution.

Chapter 3 discusses the first set of results using this code. We found that the correlation of stellar age and velocity dispersion in the solar neighborhood (Nordström et al., 2004; Holmberg et al., 2007) could be understood as a natural consequence of the gravitational instability-regulated picture of galaxy evolution, since at high redshift the gas has a high velocity dispersion, and hence so will the stars formed in that era. We also found that our galaxies developed central depressions in their gas surface density at low redshift, as a result of gas being consumed by star formation and outflows on its journey to the center of the galaxy. We explored this result further in Chapter 4, showing under what circumstances these central holes could be produced, and predicting their spatial extent. We also showed that this model could easily recover the "universal profile" of gas in local spiral galaxies discovered by Bigiel & Blitz (2012).

7.3 High-resolution simulations of dwarf galaxies with self-consistent photoelectric heating

Dwarf galaxies can have incredibly long depletion times (Hunter et al., 2012), sometimes exceeding the age of the universe by an order of magnitude. In some models, supernovae are responsible for slowing down star formation in these galaxies, but supernovae are notoriously difficult to model correctly in hydrodynamic simulations owing to the very small scales on which they evolve initially. To definitively understand whether supernovae are responsible for setting the long depletion times in galaxies, we have been running high resolution 3D simulations of isolated dwarf galaxies which have a resolved Sedov-Taylor phase. We have shown in Chapter 5 that supernovae have virtually no effect on the depletion time in these galaxies. Instead, the long depletion times arise from the spatially- and time-dependent interstellar radiation field acting to heat the gas via the grain photoelectric effect (e.g. Bakes & Tielens, 1994).

Gas which is locally Jeans unstable (red) eventually collapses and forms a new star cluster. Immediately, stars begin to emit radiation in the FUV, between 8 and 13.6 eV. This radiation is non-ionizing, and so has a large mean free path, particularly in a dwarf galaxy with low metallicity. Almost immediately, gas near the location of the newly formed star is heated, and ceases to be Jeans unstable. Supernova remnants are completely irrelevant for preventing the collapse of nearby gas – the star formation rates between identical simulations with and without supernovae, but including photoelectric heating, are nearly identical.

These simulations also launch strong winds, with mass loading factors exceeding unity. Moreover the metal content of the winds substantially exceeds the metallicity of the galactic disk, meaning that metals are being preferentially ejected in the winds.

Bibliography

Adelberger, K. L., Shapley, A. E., Steidel, C. C., et al. 2005, ApJ, 629, 636

Agertz, O., Kravtsov, A. V., Leitner, S. N., & Gnedin, N. Y. 2013, ApJ, 770, 25

Assmann, P., Fellhauer, M., Kroupa, P., Brüns, R. C., & Smith, R. 2011, MNRAS, 415, 1280

Bakes, E. L. O., & Tielens, A. G. G. M. 1994, ApJ, 427, 822

Balbus, S. A., & Hawley, J. F. 1998, Reviews of Modern Physics, 70, 1

Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013a, ApJ, 770, 57

—. 2013b, ApJLetters, 762, L31

Bekki, K. 2015a, MNRAS, 449, 1625

—. 2015b, MNRAS, 454, L41

—. 2015c, ApJ, 812, L14

Bekki, K., & Tsujimoto, T. 2011, ApJ, 738, 4

Bigiel, F., & Blitz, L. 2012, ApJ, 756, 183

Bigiel, F., Leroy, A., Walter, F., et al. 2010, The Astronomical Journal, 140, 1194

Bigiel, F., Leroy, A. K., Walter, F., et al. 2011, ApJLetters, 730, L13

Binney, J. 1977, ApJ, 215, 483

Binney, J., & Tremaine, S. 1987,

Bird, J. C., Kazantzidis, S., & Weinberg, D. H. 2012, MNRAS, 420, 913

Birnboim, Y., & Dekel, A. 2003, MNRAS, 345, 349

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Bothwell, M. S., Maiolino, R., Kennicutt, R., et al. 2013, MNRAS, 433, 1425

Bouché, N., Dekel, A., Genzel, R., et al. 2010, ApJ, 718, 1001

Bournaud, F., Dekel, A., Teyssier, R., et al. 2011, ApJ, 741, L33

Bournaud, F., & Elmegreen, B. G. 2009, ApJ, 694, L158

Bournaud, F., Elmegreen, B. G., & Martig, M. 2009, ApJ, 707, L1

Bournaud, F., Elmegreen, B. G., Teyssier, R., Block, D. L., & Puerari, I. 2010, MNRAS, 409, 1088

Bovy, J., Rix, H.-W., & Hogg, D. W. 2012a, ApJ, 751, 131

Bovy, J., Rix, H.-W., Liu, C., et al. 2012b, ApJ, 753, 148

Broeils, A. H., & Rhee, M.-H. 1997, A&A, 324, 877

Brunetti, M., Chiappini, C., & Pfenniger, D. 2011, A&A, 534, A75

Bryan, G. L., Norman, M. L., O'Shea, B. W., et al. 2014, ApJSupplement Series, 211, 19

Bullock, J. S., Kolatt, T. S., Sigad, Y., et al. 2001, MNRAS, 321, 559

Burkert, A., Truran, J. W., & Hensler, G. 1992, ApJ, 391, 651

Burkert, A., Genzel, R., Bouché, N., et al. 2010, ApJ, 725, 2324

Cacciato, M., Dekel, A., & Genel, S. 2012, MNRAS, 421, 818

Caldu-Primo, A., Schruba, A., Walter, F., et al. 2013, ArXiv e-prints, 1309, 6324

Cannon, J. M., Martinkus, C. P., Leisman, L., et al. 2015, AJ, 149, 72

Carlberg, R. G., & Sellwood, J. A. 1985, ApJ, 292, 79

Ceverino, D., Dekel, A., & Bournaud, F. 2010, MNRAS, 404, 2151

Chabrier, G. 2003, PASP, 115, 763

Chabrier, G. 2005, in Astrophysics and Space Science Library, Vol. 327, The Initial Mass Function 50 Years Later, ed. E. Corbelli, F. Palla, & H. Zinnecker, 41

Cheng, J. Y., Rockosi, C. M., Morrison, H. L., et al. 2012, ApJ, 746, 149

Cheung, E., Faber, S. M., Koo, D. C., et al. 2012, ApJ, 760, 131

Chomiuk, L., & Povich, M. S. 2011, AJ, 142, 197

Christensen, C., Quinn, T., Governato, F., et al. 2012, MNRAS, 425, 3058

Cox, T. J., Jonsson, P., Primack, J. R., & Somerville, R. S. 2006, MNRAS, 373, 1013

Cresci, G., Hicks, E. K. S., Genzel, R., et al. 2009, ApJ, 697, 115

Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156

Dalla Vecchia, C., & Schaye, J. 2012, MNRAS, 426, 140

Danovich, M., Dekel, A., Hahn, O., & Teyssier, R. 2012, MNRAS, 422, 1732

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

Davé, R., Hernquist, L., Weinberg, D. H., & Katz, N. 1997, ApJ, 477, 21

Davé, R., Oppenheimer, B. D., & Finlator, K. 2011, MNRAS, 415, 11

Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2

Dekel, A., Sari, R., & Ceverino, D. 2009a, ApJ, 703, 785

Dekel, A., & Silk, J. 1986, ApJ, 303, 39

Dekel, A., Zolotov, A., Tweed, D., et al. 2013, MNRAS, 435, 999

Dekel, A., Birnboim, Y., Engel, G., et al. 2009b, Nature, 457, 451

Diemer, B., More, S., & Kravtsov, A. V. 2013, ApJ, 766, 25

Dierickx, M., Klement, R., Rix, H.-W., & Liu, C. 2010, ApJ, 725, L186

Dobbs, C. L., Burkert, A., & Pringle, J. E. 2011a, MNRAS, 417, 1318

—. 2011b, MNRAS, 413, 2935

Draine, B. T. 1978, ApJS, 36, 595

Dutton, A. A. 2012, MNRAS, 424, 3123

Dutton, A. A., & Bosch, F. C. V. D. 2009, MNRAS, 396, 141

Dutton, A. A., Bosch, F. C. v. d., & Dekel, A. 2010, MNRAS, 405, 1690

Dutton, A. A., & van den Bosch, F. C. 2009, MNRAS, 396, 141

Dutton, A. A., van den Bosch, F. C., Dekel, A., & Courteau, S. 2007, ApJ, 654, 27

Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, A&A, 275, 101

Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, Astronomy and Astrophysics, 468, 33

Elmegreen, B. G. 2011, ApJ, 737, 10

Elmegreen, B. G., & Burkert, A. 2010, ApJ, 712, 294

Elmegreen, D. M., Elmegreen, B. G., & Hirst, A. C. 2004, ApJ, 604, L21

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

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

Erb, D. K. 2008, ApJ, 674, 151

Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006, ApJ, 644, 813

Escala, A., & Larson, R. B. 2008, ApJ, 685, L31

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

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

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

Faucher-Giguère, C., Kereš, D., & Ma, C. 2011, MNRAS, 417, 2982

Feldmann, R. 2013, MNRAS, 433, 1910

Fenton, L. 1960, Communications Systems, IRE Transactions on, 8, 57

Forbes, J., Krumholz, M., & Burkert, A. 2012, ApJ, 754, 48

Forbes, J. C., Krumholz, M. R., Burkert, A., & Dekel, A. 2014a, MNRAS, 438, 1552

—. 2014b, MNRAS, 443, 168

Förster Schreiber, N. M., Genzel, R., Bouché, N., et al. 2009, ApJ, 706, 1364

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

Fu, J., Kauffmann, G., Huang, M., et al. 2013, arXiv:1303.5586

Gabor, J. M., & Bournaud, F. 2013, ArXiv e-prints, 1310, 1923

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

Garcia-Appadoo, D. A., West, A. A., Dalcanton, J. J., Cortese, L., & Disney, M. J. 2009, MNRAS, 394, 340

Garrison-Kimmel, S., Bullock, J. S., Boylan-Kolchin, M., & Bardwell, E. 2016, ArXiv e-prints, arXiv:1603.04855

Geha, M., Brown, T. M., Tumlinson, J., et al. 2013, ApJ, 771, 29

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

Genzel, R., Newman, S., Jones, T., et al. 2011, ApJ, 733, 101

Genzel, R., Schreiber, N. M. F., Lang, P., et al. 2013, The SINS/zC-SINF survey of z˜2 galaxy kinematics: evidence for gravitational quenching,

Goldbaum, N. J., Krumholz, M. R., & Forbes, J. C. 2015, ApJSubmitted

Governato, F., Brook, C., Mayer, L., et al. 2010, Nature, 463, 203

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

Guo, K., Zheng, X. Z., & Fu, H. 2013, ApJ, 778, 23

Guo, Y., Koo, D. C., Lu, Y., et al. 2016, ArXiv e-prints, arXiv:1603.04863

Habing, H. J. 1968, Bull. Astron. Inst. Netherlands, 19, 421

Hayward, C. C., & Hopkins, P. F. 2015, ArXiv e-prints, arXiv:1510.05650

Hobbs, A., Read, J., Power, C., & Cole, D. 2012, ArXiv e-prints, 1207, 3814

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

Holmberg, J., Nordström, B., & Andersen, J. 2007, A&A, 475, 519

—. 2009, A&A, 501, 941

Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008, ApJS, 175, 356

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

Hopkins, P. F., Narayanan, D., & Murray, N. 2013, MNRAS, 432, 2647

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

Hu, C.-Y., Naab, T., Walch, S., Glover, S. C. O., & Clark, P. C. 2015, ArXiv e-prints, arXiv:1510.05644

Hunter, D. A., Ficut-Vicas, D., Ashley, T., et al. 2012, AJ, 144, 134

Janowiecki, S., Leisman, L., Józsa, G., et al. 2015, ApJ, 801, 96

Joung, M. R., Mac Low, M.-M., & Bryan, G. L. 2009, ApJ, 704, 137

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

Kassin, S. A., Weiner, B. J., Faber, S. M., et al. 2012, ApJ, 758, 106

Katz, N. 1992, ApJ, 391, 502

Kaviraj, S., Cohen, S., Windhorst, R. A., et al. 2013, MNRAS, 429, L40

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

Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183

Kim, J.-h., Abel, T., Agertz, O., et al. 2014, ApJS, 210, 14

Kim, W.-T., & Ostriker, E. C. 2002, ApJ, 570, 132

Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJSupplement Series, 180, 330

Kroupa, P. 2001, MNRAS, 322, 231

Krumholz, M., & Burkert, A. 2010, ApJ, 724, 895

Krumholz, M. R. 2013, MNRAS, 436, 2747

Krumholz, M. R., & Dekel, A. 2010, MNRAS, 406, 112

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

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

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

Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2008, ApJ, 689, 865

Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 693, 216

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

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

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

Kuhlen, M., Krumholz, M. R., Madau, P., Smith, B. D., & Wise, J. 2012, ApJ, 749, 36

Kuhlen, M., Madau, P., & Krumholz, M. R. 2013, ApJ, 776, 34

Lara-López, M. A., López-Sánchez, A. R., & Hopkins, A. M. 2013, ApJ, 764, 178

Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, ApJS, 123, 3

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

Leroy, A. K., Walter, F., Brinks, E., et al. 2008, The Astronomical Journal, 136, 2782

Leroy, A. K., Walter, F., Bigiel, F., et al. 2009, AJ, 137, 4670

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

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

Lotz, J. M., Davis, M., Faber, S. M., et al. 2008, ApJ, 672, 177

Lu, Y., Mo, H. J., & Wechsler, R. H. 2014, arXiv:1402.2036 [astro-ph]

Lu, Y., Mo, H. J., & Wechsler, R. H. 2015, MNRAS, 446, 1907

Mac Low, M.-M., & Ferrara, A. 1999, ApJ, 513, 142

Mac Low, M.-M., Klessen, R. S., Burkert, A., & Smith, M. D. 1998, Physical Review Letters, 80, 2754

Maeder, A. 1992, A&A, 264, 105

Makiya, R., Totani, T., Kobayashi, M. A. R., Nagashima, M., & Takeuchi, T. T. 2014, MNRAS, 441, 63

Mandelker, N., Dekel, A., Ceverino, D., et al. 2014, MNRAS, 443, 3675

Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115

Mannucci, F., Salvaterra, R., & Campisi, M. A. 2011, MNRAS, 414, 1263

Marinacci, F., Binney, J., Fraternali, F., et al. 2010, MNRAS, 404, 1464

Martig, M., & Bournaud, F. 2010, ApJLetters, 714, L275

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

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

McKee, C. F., & Krumholz, M. R. 2010, ApJ, 709, 308

Minchev, I., & Famaey, B. 2010, ApJ, 722, 112

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

Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, ApJ, 710, 903

Muratov, A. L., Keres, D., Faucher-Giguere, C.-A., et al. 2015, ArXiv e-prints, arXiv:1501.03155

Murray, N., Quataert, E., & Thompson, T. A. 2005, ApJ, 618, 569

—. 2010, ApJ, 709, 191

Narayan, C. A., & Jog, C. J. 2002, A&A, 394, 89

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

Neistein, E., Macciò, A. V., & Dekel, A. 2010, MNRAS, 403, 984

Nelson, E. J., van Dokkum, P. G., Momcheva, I., et al. 2013, ApJ, 763, L16

Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJLetters, 660, L43

Nordström, B., Mayor, M., Andersen, J., et al. 2004, A&A, 418, 989

Oñorbe, J., Boylan-Kolchin, M., Bullock, J. S., et al. 2015, MNRAS, 454, 2092

Ocvirk, P., Pichon, C., & Teyssier, R. 2008, MNRAS, 390, 1326

Okamoto, T., Gao, L., & Theuns, T. 2008, MNRAS, 390, 920

Oppenheimer, B. D., Davé, R., Kereš, D., et al. 2010, MNRAS, 406, 2325

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

Papastergis, E., Cattaneo, A., Huang, S., Giovanelli, R., & Haynes, M. P. 2012, ApJ, 759, 138

Peeples, M. S., & Shankar, F. 2011, MNRAS, 417, 2962

Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425

Qu, Y., Di Matteo, P., Lehnert, M. D., & van Driel, W. 2011, A&A, 530, A10

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

Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, ApJ, 771, 30

Rees, M. J., & Ostriker, J. P. 1977, MNRAS, 179, 541

Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, ApJLetters, 739, L40

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

—. 1994, A&A, 286

Romeo, A. B., & Falstad, N. 2013, arXiv:1302.4291

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

Roškar, R., Debattista, V. P., Quinn, T. R., & Wadsley, J. 2012, MNRAS, 426, 2089

Rubin, K. H. R., Prochaska, J. X., Koo, D. C., et al. 2014, ApJ, 794, 156

Rubin, V. C., Ford, W. K. J., & . Thonnard, N. 1980, ApJ, 238, 471

Saintonge, A., Kauffmann, G., Wang, J., et al. 2011, MNRAS, 415, 61

Saintonge, A., Lutz, D., Genzel, R., et al. 2013, ApJ, 778, 2

Sánchez Almeida, J., Aguerri, J. A. L., Muñoz-Tuñón, C., & Huertas-Company, M. 2011, ApJ, 735, 125

Scannapieco, E. 2013, arXiv:1302.3626, 2013, ApJ, 763, L31

Schiminovich, D., Catinella, B., Kauffmann, G., et al. 2010, MNRAS, 408, 919

Schönrich, R., & Binney, J. 2009, MNRAS, 396, 203

Schroetter, I., Bouché, N., Péroux, C., et al. 2015, ApJ, 804, 83

Schruba, A., Leroy, A. K., Walter, F., et al. 2011, The Astronomical Journal, 142, 37

Sellwood, J. A., & Binney, J. J. 2002, MNRAS, 336, 785

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

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

Shen, S., Madau, P., Aguirre, A., et al. 2012, ApJ, 760, 50

Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119

Shu, F. H. 1968, The Astronomical Journal Supplement, 73, 201

Silk, J. 1977, ApJ, 211, 638

Socrates, A., Davis, S. W., & Ramirez-Ruiz, E. 2008, ApJ, 687, 202

Somerville, R. S., & Primack, J. R. 1999, MNRAS, 310, 1087

Spitoni, E., & Matteucci, F. 2011, A&A, 531, A72

Spitzer, Jr., L., & Schwarzschild, M. 1953, ApJ, 118, 106

Springel, V., & White, S. D. M. 1999, MNRAS, 307, 162

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Stinson, G., Seth, A., Katz, N., et al. 2006, arXiv:astro-ph/0602350, Mon.Not.Roy.Astron.Soc.373:1074-1090,2006

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

Stott, J. P., Sobral, D., Bower, R., et al. 2013, MNRAS, 436, 1130

Tacconi, L. J., Genzel, R., Neri, R., et al. 2010, Nature, 463, 781

Tacconi, L. J., Neri, R., Genzel, R., et al. 2013, ApJ, 768, 74

Tasker, E. J. 2011, ApJ, 730, 11

Tasker, E. J., Wadsley, J., & Pudritz, R. 2015, ApJ, 801, 33

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

Thornton, K., Gaudlitz, M., Janka, H.-T., & Steinmetz, M. 1998, ApJ, 500, 95

Tinsley, B. M. 1980, Fundamentals of Cosmic Physics, 5, 287

Toomre, A. 1964, ApJ, 139, 1217

Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898

Truelove, J. K., Klein, R. I., McKee, C. F., et al. 1998, ApJ, 495, 821

van den Bosch, F. C. 2001, MNRAS, 327, 1334

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

van Dokkum, P. G., Leja, J., Nelson, E. J., et al. 2013, arXiv:1304.2391

Vogelsberger, M., Genel, S., Sijacki, D., et al. 2013, MNRAS, 436, 3031

Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, The Astronomical Journal, 136, 2563

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

Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, ApJLetters, 754, L29

White, S. D. M., & Frenk, C. S. 1991, ApJ, 379, 52

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341

Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, ApJ, 587, 278

Wuyts, S., Förster Schreiber, N. M., van der Wel, A., et al. 2011, ApJ, 742, 96

Yang, X., Mo, H. J., van den Bosch, F. C., Zhang, Y., & Han, J. 2012, ApJ, 752, 41

Zahid, H. J., Torrey, P., Vogelsberger, M., et al. 2014, Ap&SS, 349, 873

Zaritsky, D., Kennicutt, Jr., R. C., & Huchra, J. P. 1994, ApJ, 420, 87

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110