Formation of the First Galaxies

by

Xi Meng

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Astronomy and Astrophysics)
in the University of Michigan
2022

Doctoral Committee:

Professor Oleg Y. Gnedin, Chair
Professor Joel N. Bregman
Professor Dragan Huterer
Associate Professor Keren Sharon

Xi Meng

xim@umich.edu

ORCID iD: 0000-0002-8276-4164

© Xi Meng 2022

To Orion.

ACKNOWLEDGMENTS

I would like to thank my advisor, Prof. Oleg Gnedin, for his guidance over the past five (actually seven) years. I met him by chance in an undergraduate visiting program between Peking University (my undergraduate university) and the University of Michigan in my junior year in 2015. I had always wanted to work on cosmological simulations and that was my first chance of getting started. That was my first time to come to the US and I could hardly communicate with people in English, but he speaks slowly and repeats his ideas in different words. To some extent he is also my English teacher. After that, I did my undergraduate thesis with Oleg remotely. Over the years, he impressed me with his patience, encouragement and knowledge. He is one of the most considerate people I have met. He would sense my tension before giving a talk or taking an exam, and my disappointment if I do not get some fellowship. He always consoles me and encourages me, and when there was a problem, he would communicate with me genuinely. He really wants me to success following my own will. He has helped me in many aspects, from solving physics problems, presenting results, making plots, to writing papers and proposals, making oral presentation and communicating with people. He is also very hard-working. I learnt from him not only what he says, buy also how he behaves and how he conveys his ideas. He has spent a lot of time discussing with me, helping me revise my writing, improving my plots, teaching me how to give presentations. The time and effort he spent on helping me, the patience and kindness he showed, and the way of thinking he taught me are invaluable to me and I will remember for a lifetime.

I would also like to thank the other members of my thesis committee: Joel Bregman, Dragan Huterer, and Keren Sharon, for their advice and guidance, not only on finishing my thesis, but also on becoming a successful scholar.

I would like to thank my collaborators and people in our group, including Hui Li, Gillen Brown, Maria Han Veiga, Nick Gnedin, and Yingtian Chen. I met Hui in 2015 in the undergraduate visiting program. He helped me a lot with coding, python, simulations, using super computers, and so on. He also provided me with a lot of experience, encouragement and suggestions in the career. Gillen has been running the simulations and nicely updating me with information of our simulations. He writes code in a very organized way and make detailed documentations that are easy to read. He is also a nice person who would console and encourage me for things happening in life.

I would also thank people in the stellar halo group, including but not limited to Eric Bell, Monica Valluri, Ian Roederer, Adam Smercina and Richard D'Souza. They broadened my horizon by introducing their work and interesting papers and news on arXiv, and provided me with very helpful advice when I presented my work.

I would like to thank the other professors in the department. They are all friendly. Some of them have taught me. Some of them have been the professor for the course I GSI'ed for. I would like to thank Nuria Calvet specifically, for her help on preparing a presentation. She sat down at 411 and watched me give my talk, and gave me very useful advice on the pace, volume, and structure of the talk. I would also like to thank the other grad students in the department, for their company in coursework, research, and effort in making the department a more inclusive place.

I would like to thank my Chinese friends in astronomy: Zhijie Qu, Chaoran Wang, Yingyi Song, Jiangtao Li, Huanqing Chen, Xiaohan Wu, Haowen Zhang, Zefeng Li, and a lot more. We have discussed astronomy questions, hung out together, having fun together. They also did their best to help me in my job search. They also helped

me in many aspects, including driving me to supermarkets and parks. Zhijie has been a very close friend of mine. He has influenced me in many aspects of life. He offered me great help in astronomy, both in coursework and research. We discussed questions about coding, statistics, writing, algorithms, and so on. He is a good teacher in many ways, and he is also a good cook.

I would also like to thank my other friends, in and out of Ann Arbor. They showed me a world outside of astronomy and helped me maintain my sanity during the years with their company.

Finally, I would like to thank my family for their support and understanding, although I am away from them in these years.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

ABSTRACT

Galaxies at high redshift have different structure and star formation properties from low-redshift galaxies. While most nearby L* galaxies present disk-like morphologies, deep observations with the *Hubble Space Telescope* (HST) reveal irregular and clumpy shapes of high-redshift galaxies at rest-frame UV and optical wavelengths. It is also observed that high-redshift galaxies are more turbulent and have thicker disks than their low-redshift counterparts. Investigating the evolution of galaxy structure across cosmic time can reveal the interplay between star formation and the galaxy as a whole.

In this thesis I explored the formation and evolution of high-redshift galaxies using cosmological simulations. I have investigated the structure of simulated high-redshift Milky Way-sized galaxies. Specifically, their irregular and prolate shapes, turbulent orbits, thick disks, and clustered and bursty star formation make their properties distinct from those of the local galaxies. Despite all these differences, I found that we could still apply Toomre stability analysis to these high-redshift galaxies.

Simulations can prove nothing if they are far from the reality. Therefore, I made close comparisons of the simulated galaxies to observations. The simulations correctly reproduced the non- spatial coincidence of young stars and molecular gas, as well as the huge star-forming clumps observed in high-redshift galaxies. I compared the observed star-forming clumps with the intrinsic 3D clumps of young stars and found that the stellar masses and sizes of the observed clumps are significantly overestimated due to a combination of spatial resolution and projection. I studied the growth of the stellar disk thickness and found that in our simulations stars always form in relatively

thin disks that quickly thicken in a few hundred Myr. These results validated the star formation recipe in the simulations and helped interpret high-redshift observations that are limited by resolution and sensitivity.

Another way to study galaxy evolution is through massive star clusters, which formed in the high-redshift universe and could become globular clusters that are still observable today. I traced the tidal field of the massive star clusters and found that they experience the strongest tidal force in the first few hundred Myr after formation, and then plateaus at a lower value. I also found that the survival rate of clusters from different progenitors is different due to the tidal field they experience at late times. These results could be used in globular cluster formation models and help reconstruct the assembly history of our galaxy using globular clusters.

CHAPTER I

Introduction

The formation of our Milky Way galaxy is full of mysteries and unknowns. How did the originally almost homogeneous cosmic gas form the current complex structure of disk, spiral arms, halo, and streams containing trillions of stars, large amount of gas, planets and life? We cannot go back in time for our own Galaxy, but luckily, the universe itself is a time machine that allows us to see objects in the past as we look into greater distances. Thus, we can look at high-redshift galaxies that are similar to our Milky Way in the past to infer the properties of the young Milky Way – in fact, the word 'galaxy', which comes from the Greek word '$\gamma\alpha\lambda\alpha\xi\acute{\iota}\alpha\varsigma$', literally means 'milky', suggesting their similarity. The study of high-redshift galaxies, however, is also hard and in lack of data. Figure 1.1 shows composite RGB images of high-redshift galaxies observed with HST (Guo et al., 2015). From these somewhat vague images we can see some spiral structures, but more turbulent and distorted from their low-redshift counterparts. We know that high-redshift galaxies have very different properties from local galaxies: higher gas fraction, more irregular shape, clustered star formation, more turbulent interstellar medium (ISM), higher star formation rate (SFR). How do these factors play together? How does enhanced star formation, and thus enhanced stellar feedback, affect galaxy structure? How does the galaxy evolve from the irregular shape into the thin and regular disk we see today? Moreover, what

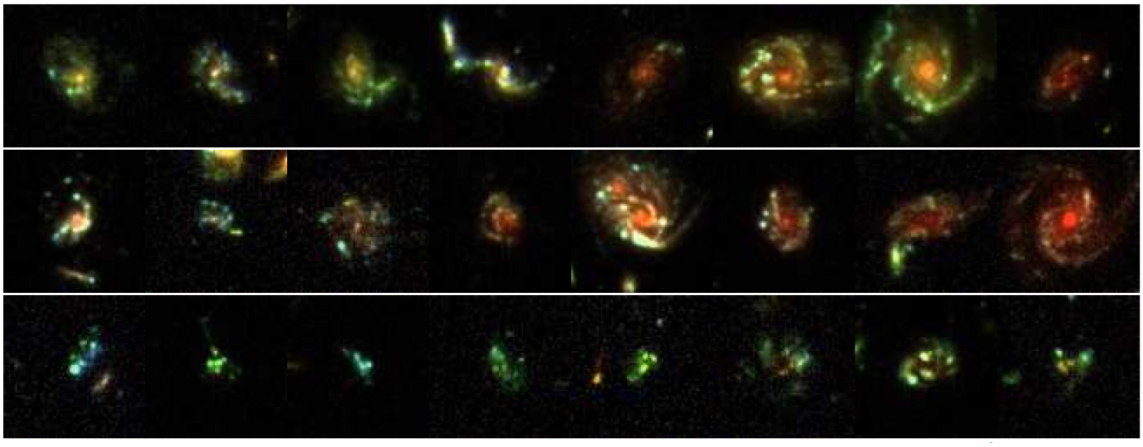

Figure 1.1 Examples of observed high-redshift galaxies. Composite RGB images made by the F435W, F606W, and F850LP images of the galaxies from HST observation. From the images we can see different structures from low-redshift galaxies. Figure taken from Guo et al. (2015).

can high-redshift observations, which are largely limited by resolution and sensitivity since these galaxies are so far away from us, tell us about galaxy properties at that time? What can the current day remnant of what is formed at the peak of cosmic star formation, e.g. globular clusters, reveal in the picture of galaxy formation at high redshift? In this dissertation, I describe my work using cosmological simulations to contribute to answering these questions.

In this introduction, I will first introduce the big picture of galaxy formation in a cosmological context. Then I will describe observational results of the star formation and galaxy structure properties of high-redshift galaxies. I will also include current status of simulation studies on galaxy formation, since this is the approach I use for my dissertation.

1.1 Galaxy formation in a cosmological context

1.1.1 Hierarchical structure formation

Galaxies are believed to form in dark matter halos, while the formation of large scale structure of dark matter is the result of primordial fluctuations in cosmic density

after the Big Bang. The primordial fluctuations are caused by quantum fluctuations in the scalar field, and are believed to follow the power spectrum $P(k) \propto k$ in most inflation theories (e.g. Harrison, 1970; Zeldovich, 1972; Hawking, 1982; Starobinsky, 1982). These small perturbations then grow due to gravity and start to form dark matter halos in the expanding universe. After the decoupling of photons and electrons at the epoch of recombination, baryons fell into the potential wells of dark matter, stars began to form in the earliest galaxies. These galaxies then, as dark matter halos gather and merge due to gravity, assemble and merge to form galaxy clusters and large scale structure (White & Rees, 1978).

The distribution of galaxies in the local (therefore, current day) universe show way more small galaxies than large galaxies. The luminosity function of nearby galaxies can be well fitted by the Schechter function (Schechter, 1976). Figure 1.2 shows the optical and near-UV luminosity functions, stellar mass function, and HI mass function of nearby galaxies (Blanton & Moustakas, 2009). Large elliptical galaxies reside mostly in the high mass (large luminosity) end, while galaxies with a distinct disk structure and dwarf elliptical and irregular galaxies are more abundant in the low-mass end than the high-mass end. In terms of stellar mass to halo mass ratio, neither the low-mass nor the high-mass end galaxies are effective in turning baryon mass into stars. Figure 1.3 shows the mean stellar mass and stellar mass to halo mass ratio as a function of halo mass. We can see that around Milky Way mass galaxies are the most efficient in turning baryons into stars. Low-mass galaxies have too strong stellar feedback compared to their shallow potential wells, while high-mass galaxies experience strong active galactic nuclei (AGN) feedback (e.g. Naab & Ostriker, 2017). The peak of stellar mass to halo mass ratio falls at Milky Way-mass galaxies, and it is these galaxies that this dissertation focuses on. In the next subsection I will describe the theoretical picture of the formation process of Milky Way-like galaxies.

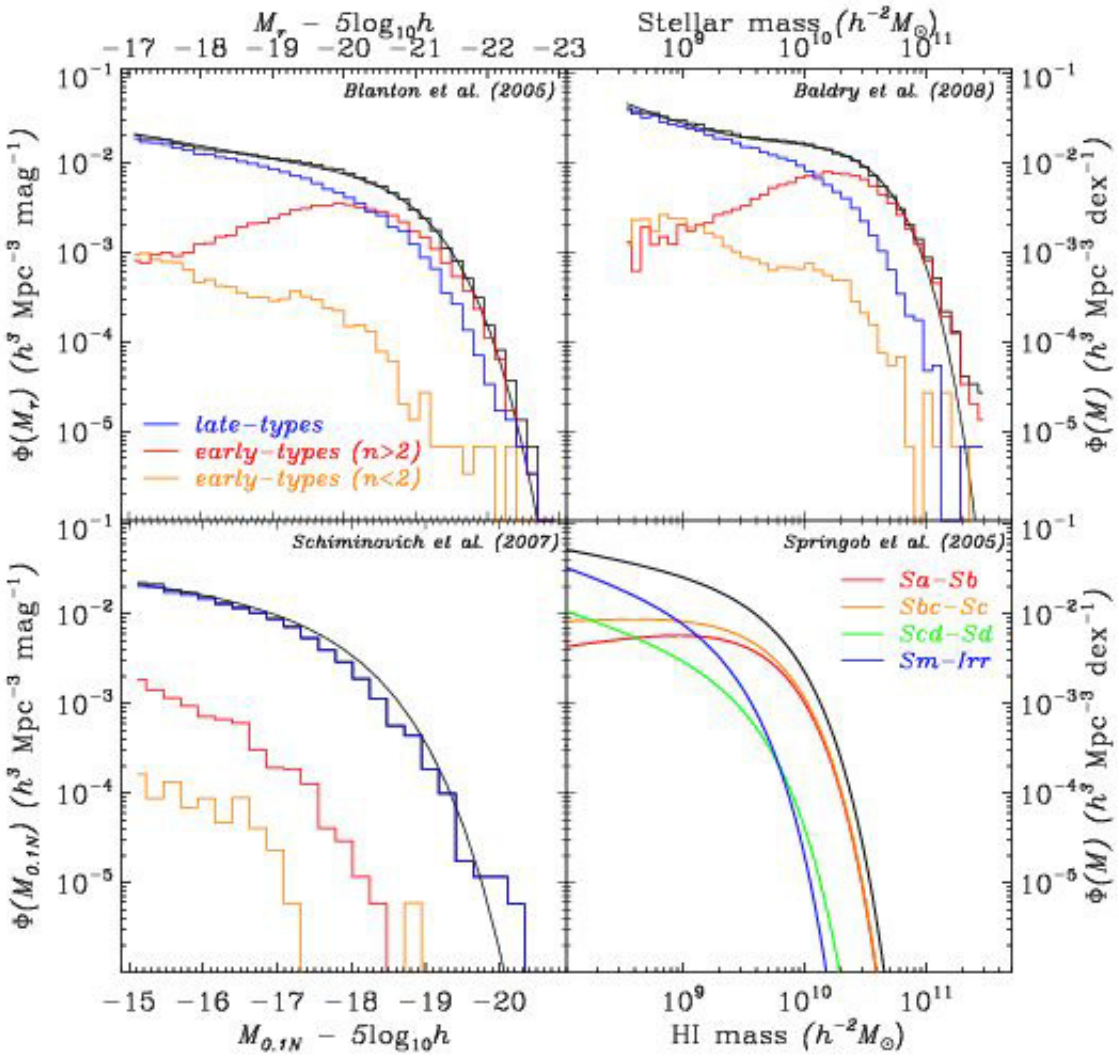

Figure 1.2 Optical and near-UV luminosity functions, stellar mass function, and HI mass function of nearby galaxies. The overlapped smooth curve is the double Schechter function. Figure taken from Blanton & Moustakas (2009).

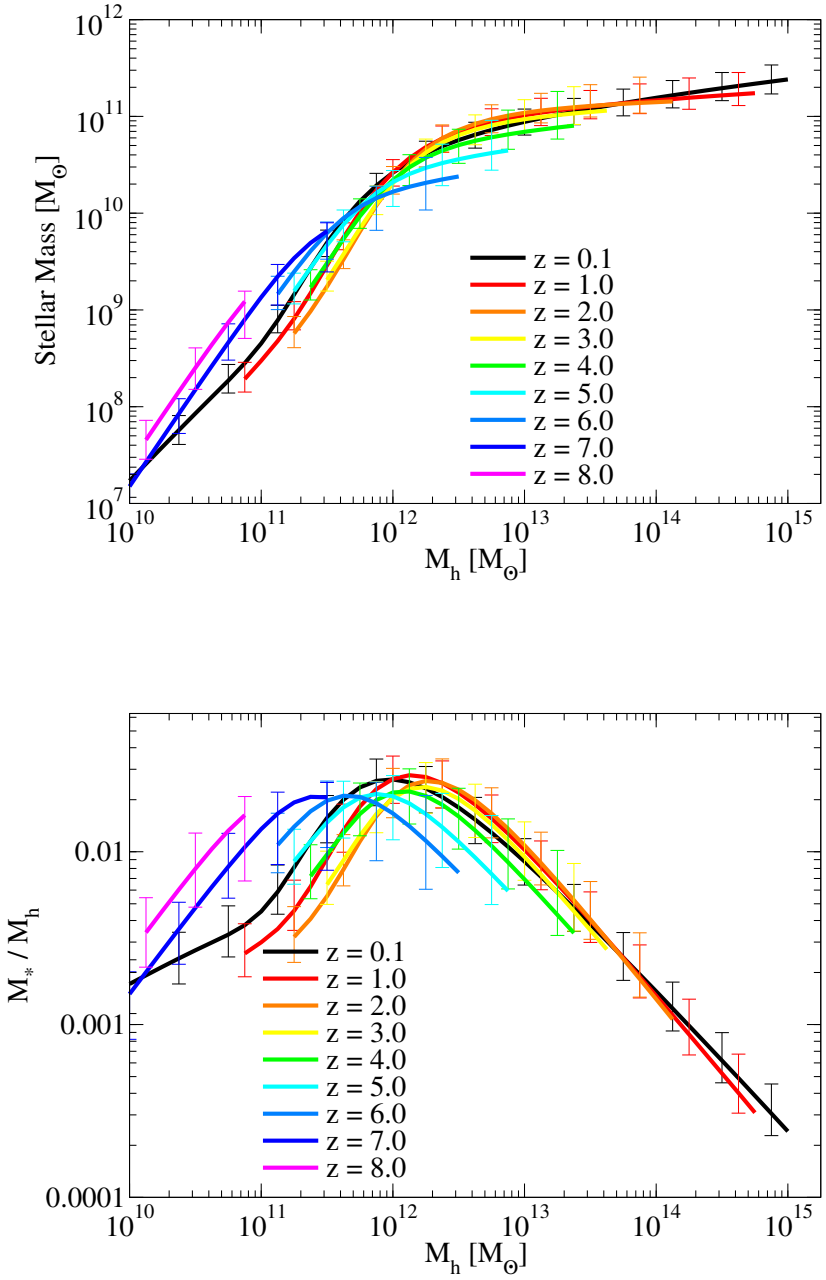

Figure 1.3 Stellar mass (upper panel) and stellar mass to halo mass ratio (lower panel) as a function of halo mass. Lines are mean values for central galaxies. Figure taken from Behroozi et al. (2013).

1.1.2 Gas, star and galaxy formation

Galaxy formation starts when gas falls into dark matter halos. Gas with angular momentum falls into the dark matter halo along filaments, giving away their potential energy through radiative cooling. When gas reaches the central region of the halo, it usually forms a dense, rotating disk due to angular momentum conservation. As gas continues to cool in the disk, gravitational instability comes into play and forms dense clumps in the disk. When these clumps are dense enough, they form stars and star clusters in the galaxy. These stars, especially high-mass stars, produce feedback in the form of radiation, stellar winds, and supernovae, giving energy and momentum to their surrounding ISM, pushing gas out of the galaxy, and injecting the heavy elements they produce into the ISM. The outflowing gas, if it is not energetic enough to escape the halo potential or become part of the hot halo, will eventually cool and fall back into the galaxy disk, participating in another run of star formation. Galaxies thus grow by accreting gas, turning the gas into stars, and at the same time, suffer from feedback from stars.

When galaxies are small, their stellar feedback is usually too strong compared to their gravitational potential. Therefore, they show episodic star formation and sometimes 'breathing mode' – changing their sizes dramatically with star formation and feedback cycles turning on and off. At that time, star formation is usually bursty: when star formation turns on, stellar feedback will soon comes in, blowing all the gas out and stopping star formation, until the gas cools and fall back into the galaxy and starts a new run of star formation. As galaxies grow larger, they gradually use up their gas. The decreased cold gas fraction in the galaxy leads to smaller clumps, more quiescent star formation, and therefore less violent stellar feedback and more stable disk. The cold gas fraction in Milky Way-like galaxies keeps decreasing over time. The peak of star formation in Milky Way-like galaxies (and also cosmic star formation) happens at redshift $z \sim 1 - 2$, often nicknamed 'cosmic noon', which is

the epoch I am interested in. In the next section I will introduce the properties of high-redshift galaxies.

1.2 Galaxy structure evolution

Since gas is collisional and has angular momentum, it is natural for gas to form a disk on a few dynamical timescales. Therefore, it is not so trivial to understand why high-redshift galaxies have different structures and star formation properties from low redshift galaxies. While most nearby L* galaxies present disk-like (oblate) morphologies, deep observations with the Hubble Space Telescope (HST) reveal irregular, clumpy, and sometime prolate shapes of high-redshift galaxies at rest-frame UV and optical wavelengths. Figure 1.4 shows an illustration of oblate and prolate shapes. Just like the cosmic star formation rate, the fraction of galaxies that are clumpy tends to decrease from $z \sim 1-2$ (e.g. Shibuya et al., 2016; Madau & Dickinson, 2014). High-redshift galaxies also appear thicker than low-redshift galaxies, i.e. they do not have as thin disks as usually observed in the local universe. Moreover, the clumpy fraction decreases with stellar mass in galaxies at $z \simeq 0.8-2$ (Tadaki et al., 2014). Given that more massive galaxies usually have less bursty star formation, this suggests a correlation between the morphology of galaxies with star formation activity. The reason for this correlation and how exactly galaxies transform from turbulent thick disks or irregular shapes at high redshift to the present day thin axisymmetric disks, is still under active research. Here I describe recent results on the structure of high-redshift galaxies.

1.2.1 Disky or not disky?

Observing high-redshift galaxies is extremely difficult due to their large distances, small sizes, dim brightness and high redshifts. Only with the most powerful telescopes in the world can we observe and coarsely resolve them. The irregular morphologies

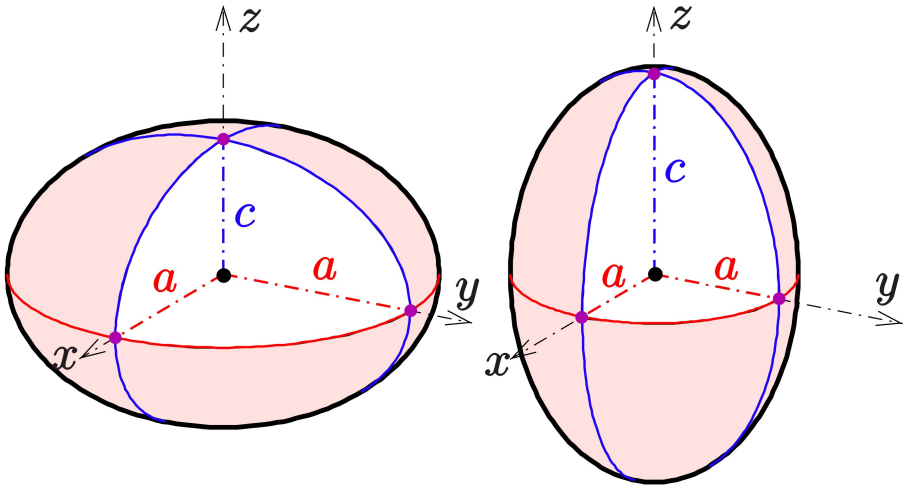

Figure 1.4 An illustration of oblate (left) and prolate (right) shapes. Oblate shapes have two long axes and one short axis, while prolate shapes have two short axes and one long axis.

of star forming galaxies (SFGs) were first noted by HST observations in the 1990s and 2000s in the rest-frame UV and optical (e.g. Griffiths et al., 1994; Windhorst et al., 1995). Later, surveys such as CANDLES (Conselice, 2014; Huertas-Company et al., 2015) observed large sample of high-redshift galaxies, providing quantitative results of galaxy morphology. Studies found that these star-forming high-redshift galaxies are more prolate than the oblate, different from SFGs at low redshift. Going toward higher mass and lower redshift, the majority of galaxies gradually transform from prolate to oblate shapes. Going toward longer wavelengths, galaxies become smoother.

Studies have also been performed to quantify the disk thickness of SFGs. Elmegreen & Elmegreen (2006) found the ratio of scale height to radial scale length to be $\sim$1/3 for chain and spiral galaxies in the HST Ultra Deep Field (UDF). The scale heights of galaxies are also found to increase with galaxy mass and decrease with redshift (Elmegreen et al., 2017). Clues for galaxy disks being thicker at high redshift can also be found in the MW and other nearby galaxies, with the stellar disk showing two distinct geometric components, the thin disk and the thick disk (e.g. Reddy et al., 2006; Jurić et al., 2008; Yoachim & Dalcanton, 2008b,a). The difference in chemical abundances for the thin disk and the thick disk – thin disk more metal rich with lower [α/Fe] and thick disk more metal poor with higher [α/Fe] – indicates that the thick disk is older than the thin disk (Bensby et al., 2005).

Not only are high-redshift galaxies thicker in morphology, but they are also 'thicker' in kinematics. High redshift SFGs are found to have larger velocity dispersion, i.e. more turbulent motions, than their low-redshift counterparts (e.g. Förster Schreiber et al., 2006; Shapiro et al., 2008; Wisnioski et al., 2015). The disk 'thickness' in terms of velocity dispersion to rotation, σ/V, is larger at high redshift and decreases with cosmic time and galaxy mass.

1.2.2 Clumps in high-redshift galaxies

As is stated above, unlike the disk-like morphologies of SFGs at low redshift, people have found high-redshift SFGs to show irregular and clumpy shapes at rest-frame UV and optical wavelengths since as early as the 1990s and referred these clumpy galaxies as 'chain', 'tadpole' and 'clump cluster' galaxies (Elmegreen et al., 2004a, 2005). These clumpy galaxies usually contain several clumps with large stellar masses $\sim 10^7 - 10^9\,M_\odot$ and up to kpc scale sizes (e.g. Förster Schreiber et al., 2011; Guo et al., 2012; Soto et al., 2017). Clumps are also observed in Hα (e.g. Livermore et al., 2012) and CO maps (e.g. Swinbank et al., 2010a). Clumps found at different wavebands do not necessarily overlap. Since these are wavebands associated with star formation, clumps can be seen as clumped star formation. But this is again different from the majority of low-redshift galaxies, in which star formation looks much smoother, at least does not clump on such large scale. The physical mechanism behind this is unclear. How are they associated with the high-redshift environment? Is there any correlation with the global SFR? Could it be related to the frequent galaxy mergers happening at high redshift?

Whether the observed clumps properties are affected by resolution is also under question. In observations of lensed galaxies, people identified smaller clumps with boosted magnification (e.g. Adamo et al., 2013; Johnson et al., 2017; Girard et al., 2018). These findings raised questions to the nature of the clumps.

The fate of these high-redshift clumps is also unclear. Since in low-redshift star-forming galaxies we do not observe stellar clumps of this huge size, if they ever existed, they must have gone somewhere. Whether these clumps survive for a long time and potentially migrate to the galaxy center (e.g. Noguchi, 1999; Ceverino et al., 2010; Bournaud et al., 2014) or dissolve in a relatively short time and potentially contribute to the thick disk (e.g. Murray et al., 2010; Oklopčić et al., 2017), is also under active debate.

1.3 Star cluster formation and evolution

Globular clusters (GCs) are some of the oldest objects in the universe, well known for being old, red and massive. Therefore, they maintain some information about the environment of their host galaxies at the time they formed, and can serve as useful tools to study the early history of their host galaxies. The mass of GC systems scales almost linearly with the host galaxy halo mass, indicating a possible connection between the GC system and the galaxy assembly history (e.g. Spitler & Forbes, 2009; Forbes et al., 2018). Although most GCs are old, their distributions in metallicity, density and mass still extend to overlap with those of young star clusters, suggesting that they have similar formation mechanisms, which differ in formation epochs in cosmological history (e.g. Krumholz et al., 2019). GCs therefore serve as a bridge between the present day universe and the high-redshift universe, the later is much more difficult to probe directly than the local volume.

Although the exact formation scenarios of GCs are still under debate (e.g. Peebles & Dicke, 1968; Katz & Ricotti, 2014; Shapiro et al., 2010), it is widely acknowledged that at least a significant number of GCs are the products of normal star formation in the early universe, when SFGs have high cold gas fraction and gas pressure in the ISM (e.g. Elmegreen, 2010; Kruijssen, 2015). By 'normal' I mean the formation mechanisms of GCs and present day young star clusters share the same physical processes. After the proto-GCs are formed, they experience various forms of dynamical evolution and lose their mass. These include internal processes such as stellar evolution and two-body relaxation, as well as external processes, namely tidal disruption. Due to these mass loss processes, and also that less massive clusters are initially less bound than more massive ones, low mass clusters usually get disrupted and become part of the field stars, while high mass clusters have a chance to survive. Since density is higher closer to the galaxy center, clusters far from the galaxy center are safer from being tidally disrupted. These processes reshape the initial cluster mass function to

the final observed GC mass function. By looking into the formation and evolution of GC systems, we can get some information on galaxy evolution and assembly.

1.4 Numerical simulations

The exact solution for the motion of a gravitating system of three point masses cannot be obtained analytically, let alone a galaxy that contains large amounts of dark matter, trillions of stars, and gas of different phases, not to mention the many physical processes that are more complicated than gravity but equally crucial to galaxy formation. Naturally, computational methods become the possible solution. The ideas of all simulations are simple, and similar: put in initial conditions, and let the system evolve under the laws of physics. From the first galaxy-galaxy encounter simulation performed with light bulbs (Holmberg, 1941) two decades before the first N-body simulation (von Hoerner, 1960), to the state-of-the-art large cosmological simulations performed on super computers (e.g. Springel et al., 2018), galaxy simulations have involved more and more ingredients and gone to larger and larger volumes. They not only include dark matter and cosmology, which is responsible for the most part of structure formation, but also include baryon physics, such as gas heating and cooling, star formation and feedback, supermassive black holes and feedback, magnetic fields, cosmic rays, radiation, etc. Figure 1.5 shows visualizations of some recent structure and galaxy formation simulations taken from Vogelsberger et al. (2020).

There are also simulations that do not track the whole history of galaxy formation, but put in the galaxy at some condition in the beginning and see how it evolves under different physical models. These simulations are usually cheaper than cosmological simulations, and are useful for testing analytical and semi-analytical models, as well as decoupling the effects of multiple physical processes.

Although adding these physical processes makes the simulations more realistic, it also brings in more complications: many of the processes are entangled, so if any

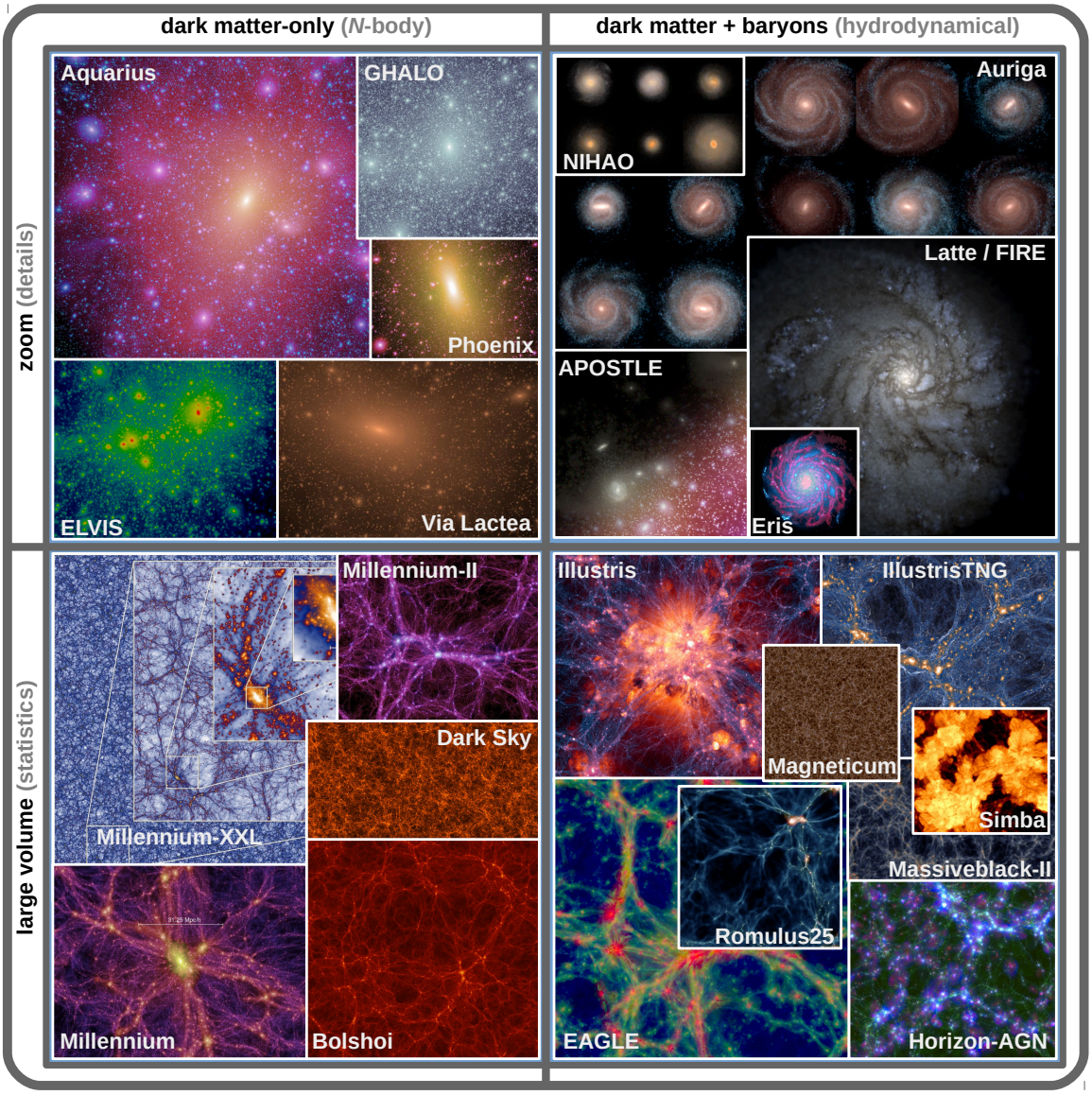

Figure 1.5 Visualization of different kinds of structure formation and galaxy simulations. There are large volume simulations, as well as zoom simulations that focus on specific galaxies. Current simulations in general can produce large-scale structure, dark matter halo structure, and galaxy properties consistent with observations. Figure is taken from Vogelsberger et al. (2020).

of them goes wrong, the results could be completely different. It also makes it non-trivial to figure out what exactly, in all of these processes involved, caused the result. What should also be kept in mind are the caveats and limitations of the simulations, which could lead to unrealistic performance. Therefore, it is important to interpret simulation results with care, and to calibrate simulation results with observations – only in this way can we make sense of our results.

Here I introduce some of the basic physical processes and frameworks used in current galaxy simulations.

1.4.1 Simulation framework

It is unrealistic for modern computers to resolve every single object in a galaxy. Numerically solving equations for physical processes always involves some discretization. Therefore, simulations usually use much more massive particles than in the real world to represent masses (particle-based method), or cut space into girds with finite sizes (grid-based method), The spatial and time resolution are also limited. Current best simulations typically can resolve the scale of giant molecular clouds (GMCs). To best utilize the limited computation resources, simulations usually seek to distribute computing power according to physical importance. There are several ways to do this. For grid-based simulations Adaptive Mesh Refinement (AMR) is most often used. This method adaptively refine regions with larger density, giving them better spatial and time resolution. For particle-based simulations, Smoothed Particle Hydrodynamics (SPH) is the common way to avoid too large force in close encounters, and also distributes resolution according to density. There is also the moving mesh method, which combines the advantages of grid-based method in fluid calculation and particle-based method in adaptivity. Famous realizations of this method include the Arepo and Gizmo code (Weinberger et al., 2020; Hopkins, 2015).

If the simulation is not the single disk galaxy that already has a galaxy in the

beginning, it usually needs to consider the cosmology framework. Most current simulations of galaxy formation use the ΛCDM model, which assumes collisionless cold dark matter, with cosmological parameters consistent with measurement on the cosmic microwave background (CMB) combined with other observations (flat universe, ∼95% dark matter and dark energy, and ∼5% baryon). Some galaxy simulations also explored the possibility of using warm dark matter (Bozek et al., 2019).

The primordial fluctuation field left after inflation seeds the growth of cosmic structures, and therefore is used to generate the initial conditions for cosmological simulations. The post-recombination density field is the convolution of the primordial fluctuation field predicted by inflation with a transfer function $T(k)$. The power spectrum of the density fluctuations is then $P(k) = Ak^n|T(k)|^2$, where $n \approx 1$. The linear density fluctuation field is usually generated at redshift $z \sim 100$ in cosmological simulations, and dark matter particle positions and velocities are then assigned accordingly, alongside with baryon positions, velocities, and temperatures. Simulations of a large enough volume of the universe usually take periodic boundary conditions. For some zoom simulations that focus on a specific volume of the universe that is not representative of the average cosmic density (e.g. a galaxy), the initial conditions are set such that the simulation produces the desired galaxy(-ies).

1.4.2 Dark matter

Dark matter contains most of the mass in the universe and controls the bulk of structure formation. For cold dark matter, we only need to solve the collisionless Boltzmann equation:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \mathbf{v}\frac{\partial f}{\partial \mathbf{r}} - \frac{\partial \Phi}{\partial \mathbf{r}}\frac{\partial f}{\partial \mathbf{v}} = 0,$$

where $f(\mathbf{r}, \mathbf{v}, t)$ is the phase-space density, and the integral or differential form of Poisson's equation:

$$\bigtriangledown^2\Phi = 4\pi G \int f \mathrm{d}\mathbf{v}.$$

In simulations, dark matter is usually represented by dark matter particles with gravitational softening. In principle, calculating the force on each particle by all the other particles can give accurate results of each particle's motion, but this method is not computationally efficient. People use numerical methods that take some approximations to simplify the calculation, e.g. the particle-mesh method (Hockney & Eastwood, 1981).

In general, dark-matter-only simulations can predict the large-scale structure of the universe relatively well, but in terms of the detailed internal halo structure and substructure, it has some deviations from observations. People have figured out that the $\sim$5% of baryon mass plays a big role in solving the problems of dark-matter-only simulations. However, adding baryon physics is non-trivial and much more complicated than calculating dark matter alone. Here I describe how simulations deal with baryon physics.

1.4.3 Baryonic physics

While dark matter accounts for most of the mass in the universe, baryons are responsible for light in the universe. Only with baryons added can we compare simulation results with most observations. It is also baryon physics that makes the universe so interesting and complicated. Physical processes that are most important to galaxy formation include hydrodynamics, gas heating and cooling, chemical evolution, multi-phase interstellar medium (ISM), star formation, stellar feedback, super-massive black hole growth, AGN feedback, magnetic fields, radiation fields, cosmic rays, etc. Depending on the demands and ability of simulations, part or all of these mechanisms are considered. Simulations also vary the parameters of these

physical models to see how simulation results are different.

Hydrodynamics is the most basic baryon physics to consider. Gas in simulations is usually considered to be inviscid ideal gas. Hydrodynamical equations are solved for gas in the simulation, whether in Eulerian or Lagrangian formulation. If magnetic field or radiation field is included and calculated in runtime, the computation would be more complicated and expensive, but the equations for this calculation are certain based on the well-established theoretical framework.

Unlike hydrodynamics, other physical processes such as star formation and feedback, or gas cooling, rely on physical scales that are usually smaller than typical galaxy simulation resolution, or involves multiple processes that are best treated with a more or less empirical model. Gas cooling processes include collisional excitation and ionization, inverse Compton, recombination and free-free emission, and is dependent on gas metallicity. In simulations, gas cooling is usually calculated with cooling functions that are either tabulated or extracted from chemical networks. Some simulations also consider photo-ionization by radiation fields. Combined with gas cooling and heating is the multi-phase ISM. The ionization state of gas can be calculated in a relatively realistic way. In some simulations, molecular H_2 formation and destruction is also considered using phenomenological models.

Star formation and feedback involve spatial scales that are too small for most simulations to model from first principles, thus they are treated with sub-resolution models. When gas cools to a cold and dense enough state, star formation happens. Simulations usually impose some density and temperature threshold for star formation to trigger. There are also other complementary criteria, e.g. 'collapse' condition (gas flowing into the star forming region), gravitationally-bound star forming region, etc. In simulations, star formation process is usually represented with some mass of gas

being turned into stars, with the star formation rate

$$\dot{M}_* = \frac{\epsilon_{\rm ff}}{t_{\rm ff}} M_g$$

where M_g is the gas mass of the cloud, $t_{\rm ff}$ is the free-fall time, and $\epsilon_{\rm ff}$ is the star formation efficiency per free-fall time. Star formation needs to be calibrated with empirical star formation laws, e.g. the Kennicutt-Schmidt law (Kennicutt, 1998b):

$$\Sigma_{\rm SFR} = (2.5 \pm 0.7) \times 10^{-4} \left(\frac{\Sigma_{\rm gas}}{1\,{\rm M_\odot pc^{-2}}} \right)^{1.4\pm0.15} {\rm M_\odot yr^{-1} kpc^{-2}},$$

where $\Sigma_{\rm SFR}$ and $\Sigma_{\rm gas}$ are the surface densities of SFR and gas, respectively.

As is with all the other ingredients in galaxies, individual stars also cannot be resolved in galaxy formation simulations. Therefore, stellar particles in simulations usually represent star clusters in which stars are born roughly at the same time. Some simulations employ specific star cluster formation models, for example, the continuous cluster formation model in our simulations. Some initial mass function (IMF) is usually assumed for the stellar particles, and the later stellar evolution is also calculated according to this assumed IMF. The IMF is defined as the fraction of stars per mass interval:

$$\psi(m) = \frac{{\rm d}N}{{\rm d}m},$$

where N is the number of stars, and m is the initial stellar mass. The first power-law parameterization of the IMF was proposed by Salpeter (1955):

$$\psi(m) \propto m^{-\alpha};\ \alpha = 2.35.$$

Currently more complicated and accurate forms are taken for the IMF, for example

the Chabrier IMF (Chabrier, 2001):

$$\psi(m) = Am^{-\alpha} \exp\left[-\left(\frac{m_0}{m}\right)^{\beta}\right],$$

where $\alpha = 3.3$, $m_0 = 716.4\,M_{\odot}$, and $\beta = 0.25$. There is no guarantee that the IMF is unchanged across the universe and for all time, but there is increasing evidence that the IMF is unchanged, and most simulations, if not all, use a universal IMF.

After stellar particles are created, stellar evolution models are performed on them. Stellar particles lose mass and produce heavy elements. In real world, the most massive stars explode as supernovae, so do stars in galaxy formation simulations. After the typical lifetime of the high-mass stars, stellar particles go supernovae through the injection of energy (both thermal and kinematic) and momentum on their ambient ISM. Star formation and feedback together regulate the galaxy growth. Due to limited resolution, some sub-resolution models are applied to make stellar feedback more efficient and match the cosmic star formation history. Stellar feedback also distributes the heavy elements stars produced to the ISM: Type Ia supernovae are responsible for the Iron peak elements, while other types of supernovae are important for α elements. There are also other feedback channels, such as stellar winds, photoionization and radiation pressure.

These baryon physics models, especially sub-resolution models, are, and should be calibrated with smaller scale simulations. Albeit calibrated, they still have a large range of freedom in terms of parameter choice, etc, and affected by the coarser resolution when they are implemented in galaxy simulations. Making changes to any of them could lead to completely different results. Therefore, close enough comparison with observations in every aspect is necessary and important.

1.4.4 Current galaxy formation simulation results

Current galaxy formation simulations, e.g. the IllustrisTNG simulations (Springel et al., 2018; Pillepich et al., 2018; Nelson et al., 2018; Naiman et al., 2018), the EAGLE simulations (Schaye et al., 2015; Crain et al., 2015), the FIRE simulations (Hopkins et al., 2018), etc., produce fairly consistent results with observations in terms of statistical and global properties of galaxies. They produce realistic halo mass functions, galaxy stellar mass functions, stellar mass-halo mass relation, galaxy hierarchy, correlation functions, across halo mass and cosmic time. They also produce star formation history, galaxy color bimodality, and sometimes chemical evolution consistent with those from observations for galaxies of all masses. Simulations are now trying to also match the smaller scale properties of galaxies, such as ISM and circumgalactic medium (CGM) properties, gas inflow and outflow, star formation on hundred parsec scale, star formation and gas association, star cluster properties, galaxy morphology evolution, etc. Not every simulation can reproduce realistic results for the above mentioned properties, and they do not necessarily reproduce all of these properties at the same time. Comparison to observations with care and in depth should be taken, and not only at $z = 0$, but also across cosmic time.

Some simulations also employ specific models (e.g. star cluster formation and evolution models (Kruijssen et al., 2019a)) in their simulations and rerun the simulation to get results in realistic cosmological galaxy environment. These results can potentially be used to infer the history of galaxies with current observations.

1.5 Formation of our Milky Way

1.5.1 Our Milky Way

Our Milky Way consists of gas and stars embedded in a dark matter halo. Figure 1.6 shows a schematic view of the current Milky Way. Most of the stars and cold

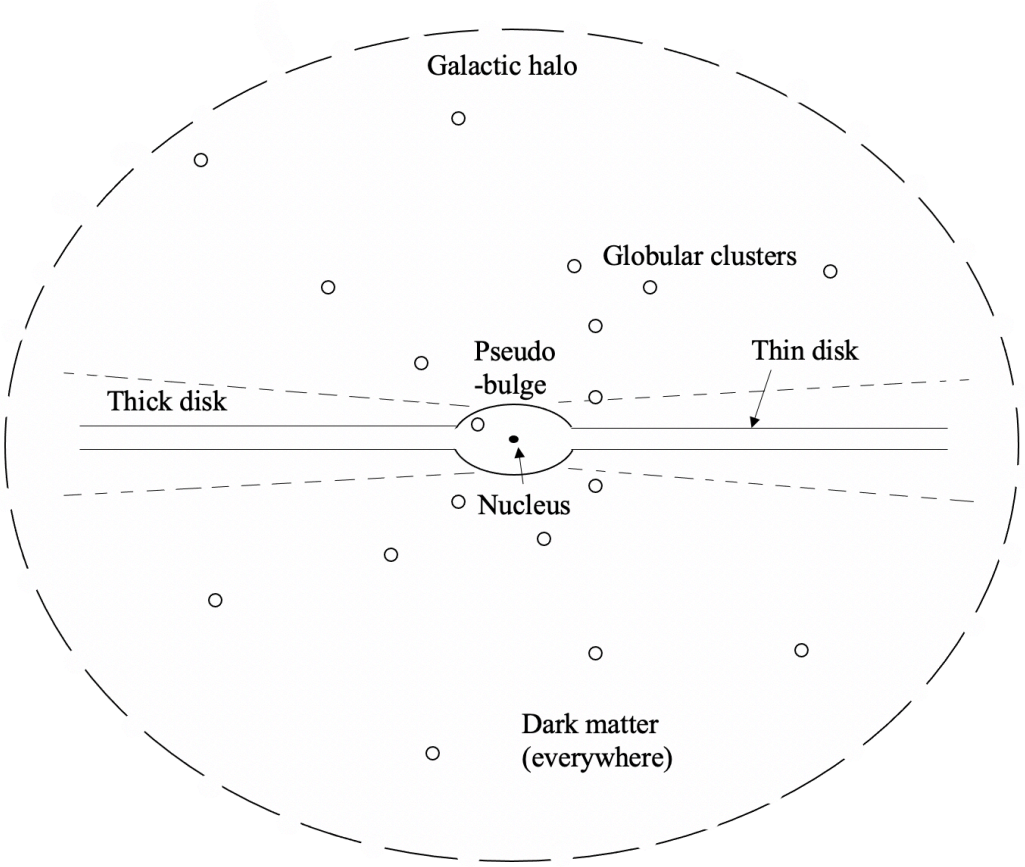

Figure 1.6 A schematic view of our Milky Way, showing the thin disk and the thick disk, the nucleus and the pseudo-bulge, and globular clusters.

gas are in the galactic disk that extends to at least 15 kpc in radius, with a mass of $\sim 5 \times 10^{10}\, M_\odot$. 95% of the disk stars, and all of the young massive stars are in the thin disk, while the thick disk contains older stars. The thin disk has a radial scale length of ~ 3.5 kpc and a vertical scale height of ~ 0.3 kpc, while the vertical scale height of the thick disk is ~ 1 kpc. Our Milky Way also has a nucleus and a pseudo-bulge with a half-light radius of ~ 1 kpc at the center, and a black hole with mass $M_{\rm BH} \approx 4 \times 10^6\, M_\odot$. Globular clusters are distributed in the halo.

The Milky Way grows to what it is like today by accreting cold gas and turning it into stars, as well as assembling smaller galaxies. During its growth, its size, stellar mass and metallicity increase, while the gas fraction decreases. Figure 1.7 shows a schematic figure of the evolutionary trend of these quantities with time.

1.5.2 Toomre analysis

Just like dark matter halos form out of gravitational instability in dark matter, stars and star clusters form out of gravitational instability in the galactic disk. There is also velocity dispersion and shear motions that can to some extent resist gravitational instability and make the disk stable. In a thin, axisymmetric, single component rotating disk, stability criterion to linear perturbations can be described by the Toomre parameter (Safronov, 1960; Toomre, 1964):

$$Q = \frac{\sigma\kappa}{\pi G \Sigma},$$

where κ is the epicycle frequency describing the shear motions, σ is the velocity dispersion, and Σ is the mass surface density. A disk is unstable if $Q \lesssim 1$ and stable if $Q \gtrsim 1$.

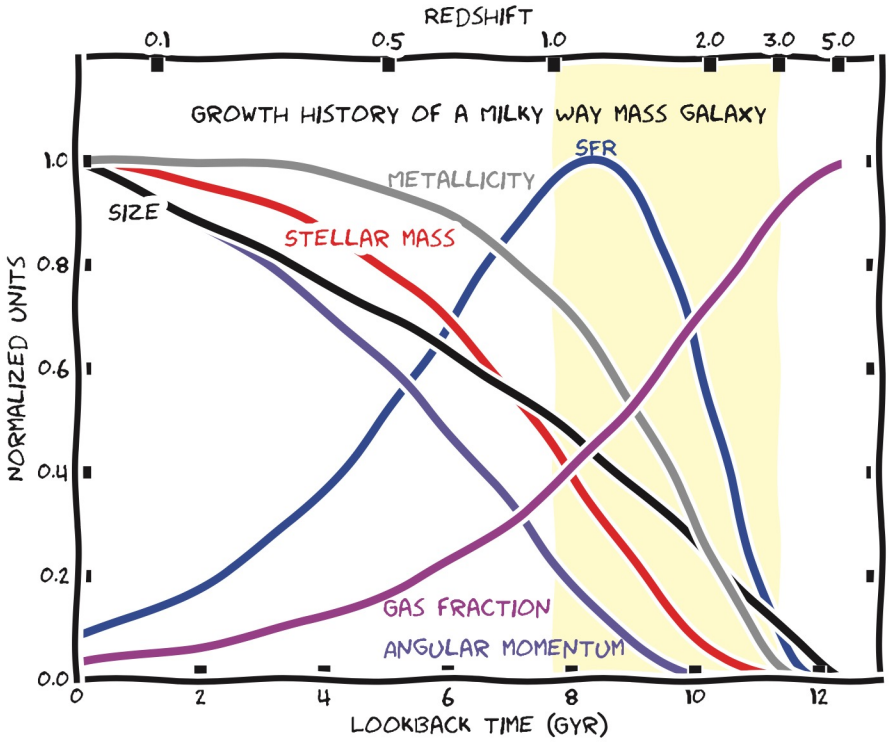

Figure 1.7 Overall evolution of size, stellar mass, metallicity, gas fraction, and SFR of a Milky Way-like galaxy. Figure taken from Förster Schreiber & Wuyts (2020).

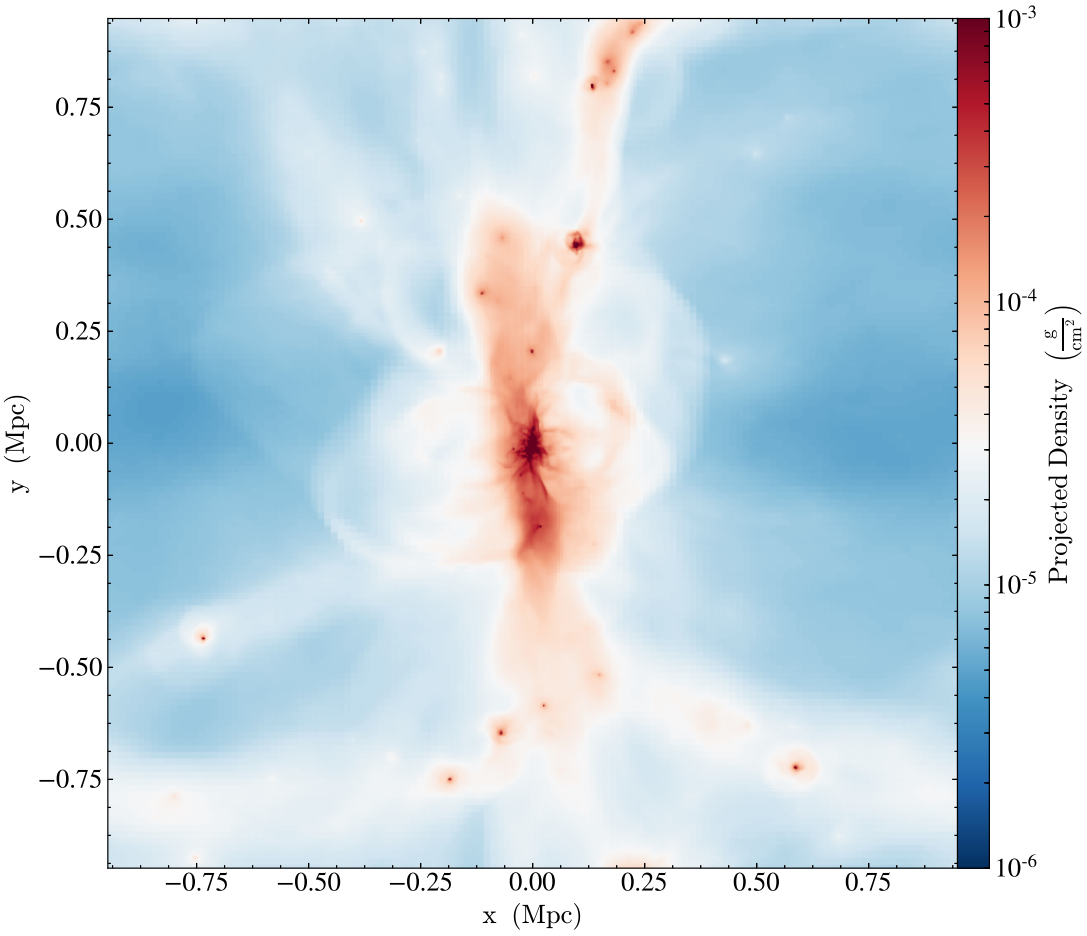

Figure 1.8 A galaxy in our simulations at $z = 1.5$.

1.6 Dissertation Overview

In this dissertation, I describe my work on galaxy formation and evolution, specifically galaxy structure at high redshift, using simulations. Figure 1.8 shows a galaxy in our simulations at $z = 1.5$. Chapter II (Meng et al., 2019) measures high-redshift galaxy shapes and spatial properties of star formation in our simulations and discusses the possibility of applying the theoretical Toomre analysis to the complicated high-redshift environment. In Chapter III (Meng & Gnedin, 2020) and IV (Meng & Gnedin, 2021) I describe my work on comparing our simulated galaxies to observations in terms of galaxy structure and star formation in a more accurate and 'observational'

way, and provided possible insights to simulations. In Chapter V (Meng & Gnedin, 2022) I analyse the tidal influence on star clusters and give implications on the relation of star cluster survival and their origins.

CHAPTER II

Structure and stability of high-redshift galaxies in cosmological simulations

This chapter was published as: ***Meng, X.****, Gnedin, O. Y. & Li H. 2019, MNRAS, 486, 1574*

Abstract

We investigate the structure of galaxies formed in a suite of high-resolution cosmological simulations. Consistent with observations of high-redshift galaxies, our simulated galaxies show irregular, prolate shapes, which are dominated by turbulent motions instead of rotation. Yet molecular gas and young stars are restricted to a relatively thin plane. We examine the accuracy of applying the Toomre linear stability analysis to predict the location and amount of gas available for star formation. We find that the Toomre criterion still works for these irregular galaxies, after correcting for multiple gas and stellar components: the Q parameter in H_2 rich regions is in the range $0.5 - 1$, remarkably close to unity. Due to the violent stellar feedback from supernovae and strong turbulent motions, young stars and molecular gas are not always spatially associated. Neither the Q map nor the H_2 surface density map coincide with recent star formation exactly. We argue that the Toomre criterion is a

better indicator of future star formation than a single H_2 surface density threshold because of the smaller dynamic range of Q. The depletion time of molecular gas is below 1 Gyr on kpc scale, but with large scatter. Centering the aperture on density peaks of gas/young stars systematically biases the depletion time to larger/smaller values and increases the scatter.

2.1 Introduction

Recent analytical models of star formation (SF) in galaxies (e.g., Dekel et al., 2009; Vollmer & Leroy, 2011; Kruijssen & Longmore, 2014; Krumholz et al., 2018) have used the Toomre criterion $Q \lesssim 1$ to identify regions of dense gas in disc galaxies that should collapse and form stars. The original Toomre analysis (Toomre, 1964) considers stability of an isothermal gas layer to linear axisymmetric perturbations in regular, thin, axisymmetric discs. These conditions are not strictly satisfied even in many galaxies at low redshift, and are certainly violated in clumpy, turbulent disc galaxies at high redshift. Without invoking the Toomre criterion, however, it is difficult to construct predictive models that could help interpret observations of galaxies at $z \gtrsim 2$ that are expected from powerful oncoming and future facilities. In this paper we test the validity of the Toomre analysis for turbulent irregular galaxies using new state-of-the-art simulations of galaxy formation.

Deep HST-UDF observations indicate that normal spiral galaxies appear only at redshifts $z < 1.5$ (Elmegreen & Elmegreen, 2014). At higher redshifts galaxies tend to have thick stellar discs fragmented into dense stellar clumps (e.g., Genzel et al., 2010; Elmegreen et al., 2017). These galaxies have kinematics dominated by turbulent motions and show comparable amounts of molecular gas and stars (Tacconi et al., 2013). In some cases enhanced spatial resolution afforded by gravitational lensing reveals that the locations molecular gas and UV emission from young stars are spatially decoupled on sub-kpc scales (Dessauges-Zavadsky et al., 2017a). Measurements

of the Toomre Q parameter find values exceeding unity for ionized Hα gas but giant clumps appear to be unstable with $Q < 1$ (Genzel et al., 2014). Similarly, in a study of low-redshift analogues of turbulent disk galaxies Fisher et al. (2017a) find that large Hα clumps exist only in regions with $Q < 1$.

Numerical simulations have also studied stability of galactic discs. Li et al. (2005, 2006) calculated the Toomre Q for a combination of collisional gas and collisionless stars in isolated disc simulations. Ceverino et al. (2010) and Inoue et al. (2016) used cosmological simulations of clumpy high-redshift galaxies and found values of $Q > 1$ in large parts of the discs. More recent isolated galaxy simulations show that stellar clumps fragment further on smaller sub-kpc scales, below the characteristic Toomre mass (Tamburello et al., 2015). These simulations of high-redshift galaxies show significantly irregular structure of the discs due to fast gas accretion and stellar feedback. The kinematics of the interstellar medium includes strong turbulent motions on all scales.

Given this complexity, is it correct to apply the Toomre analysis to select regions undergoing gravitational collapse and star formation? How well can we predict the amount of star-forming gas and star formation rate (SFR) in high-redshift galaxies?

Our goal is to estimate the accuracy of applying the Toomre criterion and Toomre fragmentation mass. Using a recent suite of cosmological simulations of Milky Way-type galaxies, we verify the validity of the Toomre analysis. Unlike regular disc galaxies observed at low redshift, our simulated high redshift galaxies are turbulent and show irregular shapes. To explore the structure of our simulated galaxies, we study the radial profiles of these galaxies and use moments of inertia to obtain their axis ratios in order to check whether they can be considered discs. We argue that the Q parameter is a better predictor of the star formation sites than simply the molecular gas density, because it covers a smaller dynamic range. We also study the depletion time of cold gas as a function of spatial scale and, in agreement with previous non-

cosmological studies, show that the two commonly used approaches (gas-centred and star-centred) give very different estimates on $\sim$ 100 pc scales but converge on kpc scale.

We describe the simulations used for this analysis and the spatial and kinematic structure of simulated galaxies in Section 5.2. We apply the Toomre analysis and calculate maps of the Q parameter in Section 2.3. In Section 2.4 we study the depletion time of cold molecular gas on different spatial scales. We discuss the implications of our results for modeling galactic star formation in Section 2.5 and present our conclusions in Section 2.6.

Table 2.1: Global properties of the main galaxy

Run	z	$M_{\rm h}\,(M_\odot)$	$M_*\,(M_\odot)$	$M_{\rm HI+H_2}\,(M_\odot)$	$M_{\rm H_2}\,(M_\odot)$	$R_{\rm vir}$ (kpc)	$R_{\rm h,*}$ (kpc)	$R_{\rm h,HI+H_2}$ (kpc)	$R_{\rm h,H_2}$ (kpc)
SFE200	1.78	2.37×10^{11}	4.06×10^{9}	2.65×10^{9}	1.77×10^{8}	67.9	2.6	3.2	2.3
SFE100	1.50	4.36×10^{11}	8.19×10^{9}	7.23×10^{9}	8.54×10^{8}	91.5	2.7	4.7	2.2
SFE50	1.50	4.35×10^{11}	7.13×10^{9}	8.35×10^{9}	7.47×10^{8}	91.5	4.0	8.6	2.8
SFE10	1.78	2.50×10^{11}	6.46×10^{9}	4.98×10^{9}	1.15×10^{9}	69.0	2.4	3.6	2.3
SFEturb	2.85	1.21×10^{11}	2.44×10^{9}	2.78×10^{9}	2.59×10^{8}	39.2	2.2	4.0	2.1
SFE50-SNR3	1.98	2.40×10^{11}	1.37×10^{10}	3.29×10^{9}	1.88×10^{9}	63.5	0.7	1.3	0.9

2.2 Structure of high redshift galaxies

2.2.1 Simulation Suite

We use a suite of cosmological simulations described in Li et al. (2017a, 2018a). These simulations were run with the Adaptive Refinement Tree (ART) code (Kravtsov et al., 1997; Kravtsov, 1999, 2003; Rudd et al., 2008) in a periodic box of 4 comoving Mpc. All runs start with the same initial conditions but use different sub-grid model parameters of star formation and stellar feedback. The initial conditions were selected to produce a main halo with total mass $M_{200} \approx 10^{12}\,M_{\odot}$ at $z = 0$, similar to that of the Milky Way. The ART code uses adaptive mesh refinement to increase spatial resolution in the dense galactic regions. The lowest resolution level is set by the root grid, which in these runs was 128^3 cells. This sets the dark matter particle mass $m_{\rm DM} = 1.05 \times 10^6\,M_{\odot}$. The refinement strategy is quasi-Lagrangian, which keeps cell mass within a narrow range. The finest refinement level is chosen so that the physical size of gas cells at that level is between 3 and 6 pc. This required increasing the number of additional refinement levels gradually with time, from 9 levels at $z > 9$ to 10, 11, and 12 refinement levels at $z \approx 9$, 4, 1.5, respectively.

The simulations include three-dimensional radiative transfer using the Optically Thin Variable Eddington Tensor approximation (Gnedin & Abel, 2001) of ionizing and ultraviolet radiation from stars (Gnedin, 2014) and the extragalactic UV background (Haardt & Madau, 2001), non-equilibrium chemical network that deals with ionization states of hydrogen and helium, and phenomenological molecular hydrogen formation and destruction (Gnedin & Kravtsov, 2011). The simulations also incorporate a subgrid-scale (SGS) model for unresolved gas turbulence (Schmidt et al., 2014; Semenov et al., 2016a). The star formation is implemented with a new method that follows the formation of individual star clusters. In this continuous cluster formation (CCF) algorithm (Li et al., 2017a, 2018a) each star particle represents a star cluster

that forms at a local density peak and grows mass via accretion within a spherical region of fixed physical size, until the feedback of young stars terminates the growth of the star cluster.

In addition to the early radiative and stellar wind feedback (similar to that described in Agertz et al. 2013), the simulations include a supernova (SN) remnant feedback model (Martizzi et al., 2015; Semenov et al., 2016a). As the SN remnant model was calibrated by simulations of isolated SN explosion, rather than multiple SNe that appear in star clusters, its momentum feedback is underestimated. In addition, some momentum is lost due to advection errors as the SN shell moves across the simulation grid. Cosmic rays accelerated by the SN remnant could also boost the momentum deposition (Diesing & Caprioli, 2018). To compensate for these effects, the momentum feedback of the SN remnant model is boosted by a factor $f_{\rm boost}$. The default value $f_{\rm boost} = 5$ is chosen to match the star formation history expected from the abundance matching method, but the simulation suite contains also runs with different $f_{\rm boost}$ (see Section 2.2.8 and Figure 5 of Li et al. 2018a).

We adopt a ΛCDM cosmology with $\Omega_m = 0.304, \Omega_b = 0.048, h = 0.681, \sigma_8 = 0.829$ (Planck Collaboration et al., 2016).

In this paper we focus on several runs with different local star formation efficiency $\epsilon_{\rm ff}$ and SN momentum boost factor $f_{\rm boost}$. Table 2.1 contains basic information of these simulations. The last three columns in Table 2.1 are the radii containing half the mass of stars, neutral gas and molecular gas, respectively. The number after "SFE" in the names corresponds to the percentage of local $\epsilon_{\rm ff}$. In SFEturb run $\epsilon_{\rm ff}$ is variable and turbulence-dependent (as implemented by Semenov et al., 2016a). SFE50-SNR3 run is a weaker feedback run, with the lower SN boost factor $f_{\rm boost} = 3$. All other runs have $f_{\rm boost} = 5$. We focus on the main galaxy in the last available output of each run and list their masses and sizes in Table 2.1. For run SFE200 we analyze the snapshot at $z = 1.78$ (same as for SFE10 run), because at the last output

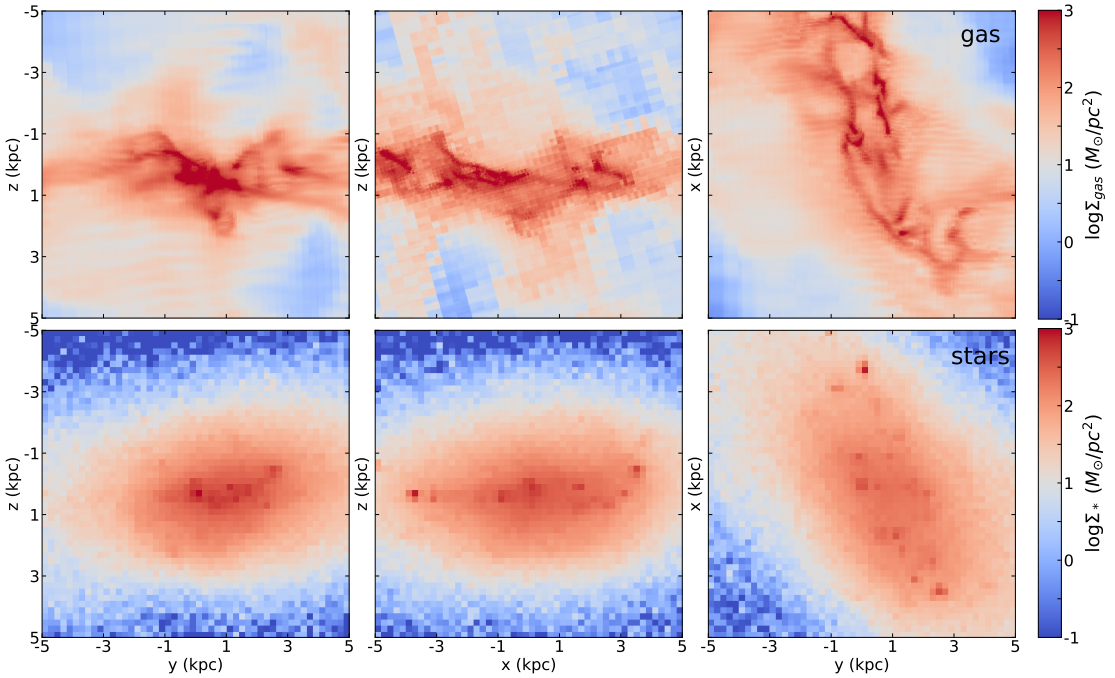

Figure 2.1 Density projection of all gas (upper panels) and stars (lower panels) along three principle axes given by the tensor of inertia of neutral gas, for run SFE50 at $z = 1.5$. Z-coordinate corresponds to the minor axis. Thus the right panels are "face-on" while the left and middle panels are "edge-on" views. The projection depth is ± 5 kpc.

($z = 1.44$) the main galaxy is experiencing a major merger and its morphology is strongly perturbed.

We note that because of strong stellar feedback, at these outputs there are few cells at the highest refinement level in a 10 kpc cube centered on the main galaxy. Therefore, the spatial resolution for our study is limited to about 100 pc.

2.2.2 Surface density profiles

To study the surface density profile of the simulated galaxies, we need to determine the centre of a galaxy and orientation of the projection plane. We define the galaxy centre to be at the location of maximum stellar density, which is found iteratively using smaller and smaller smoothing kernels in Brown et al. (2018). To determine the

galaxy orientation, we examined several alternative definitions based on the angular momentum of neutral gas and the principle axes of the tensor of inertia of gas and stars.

We use the following definition of the inertia tensor (also called "shape tensor" in Zemp et al. 2011)

$$\mathbf{I} \equiv \sum_{k,i,j} M_k \, r_{k,i} \, r_{k,j} \; \mathbf{e}_i \otimes \mathbf{e}_j$$

where M_k is the mass of k-th stellar particle or gas cell, $r_{k,i}$ are its coordinates in the galactocentric reference frame ($i = 1, 2, 3$), and $\mathbf{e}_i$ are the three unit vectors of the coordinate axes. The tensor can be diagonalized by a rotation matrix, to calculate the principle moments of inertia $I_1 \geq I_2 \geq I_3$. From these we calculate the axis ratios $b/a = (I_2/I_1)^{1/2}$ and $c/a = (I_3/I_1)^{1/2}$. The orientation of the galaxy plane is given by the eigenvector corresponding to the smallest eigenvalue.

We find that the angular momentum of neutral ($\mathrm{HI} + \mathrm{H_2}$) gas enclosed within a sphere of a given radius can suddenly change direction between $0.1\,R_{\rm vir}$ and $0.3\,R_{\rm vir}$ in some runs. For example, in SFE50 run it changes by 82°, and in SFEturb run by 52°, while in the other runs is remains stable within 10°. At large radii the gas typically falls along cosmic filaments, whereas closer to the centre mergers of galactic clumps can significantly perturb the galaxy plane orientation. Since most of the stars and neutral gas in the simulated galaxies are located within $0.1\,R_{\rm vir}$, the inner angular momentum is more relevant for the disc formation. Within $0.1\,R_{\rm vir}$ the direction of the gas angular momentum is consistent to better than 20° for all runs. The direction of angular momentum of molecular $\mathrm{H_2}$ gas generally follows that of the neutral gas, to better than 16°. The scale of $0.1\,R_{\rm vir}$ corresponds to $4-9$ kpc (Table 2.1); it varies from galaxy to galaxy because of the different redshift of the final available output.

Next we consider the principal axes of the shape tensor for neutral gas and stars, calculated with the $0.1\,R_{\rm vir}$ radius sphere. Their orientation can deviate from the angular momentum of neutral gas by as much as 36° in SFEturb run and 24° in

SFE50 run. For the other runs they are within $\approx 10°$. The moments of inertia calculated separately for the gas and stars generally agree with each other.

Although we first expected that the cold gas would settle into a thin disc and its angular momentum would be the best indicator of the rotation plane, the above comparison shows that the shape tensor gives a more consistent definition of the galaxy plane. The plane given by angular momentum has some deviations from the galaxy plane identified by eye. Therefore, we choose to use the gas shape tensor to define the galaxy plane and use this orientation throughout the paper. Similarly, Garrison-Kimmel et al. (2018) used the inertia tensor of stars to define the galaxy orientation in their analysis of FIRE-2 simulations.

Figure 2.1 shows examples of the density map of gas and stars projected along the three principle axes. The stellar distribution is strongly puffed up in the vertical direction and elongated in the galaxy plane. The axis ratios from the shape tensor show that, except for SFE50-SNR3 run, our simulated galaxies are not thin, axisymmetric discs. The axis ratios c/a and b/a for stars are generally around 0.4 and 0.6, respectively (for SFEturb run they are around 0.65 and 0.85), while the axis ratios for neutral gas are around 0.3 and 0.5 (for SFE200 run they are around 0.5 and 0.8). Axis ratios vary from galaxy to galaxy, but except for SFE50-SNR3 run (which has $b/a = 0.9$ for stars and $b/a = 0.8$ for molecular gas) they all indicate non-axisymmetric, triaxial shapes for both stars and molecular gas. This can be seen in the right panels of Figure 2.1, where the face-on view of the galaxy is not axisymmetric.

van der Wel et al. (2014) intuitively defined galaxy shapes to be oblate, prolate, or spheroidal based on the axis ratios. Here we adopt their definition and plot the axis ratios for the gas and stars in our simulated galaxies in Figure 2.2. Two runs stand out from the rest. The SFE50-SNR3 run has weaker feedback, which produces a more regular disc and shows features qualitatively different from all the other runs.

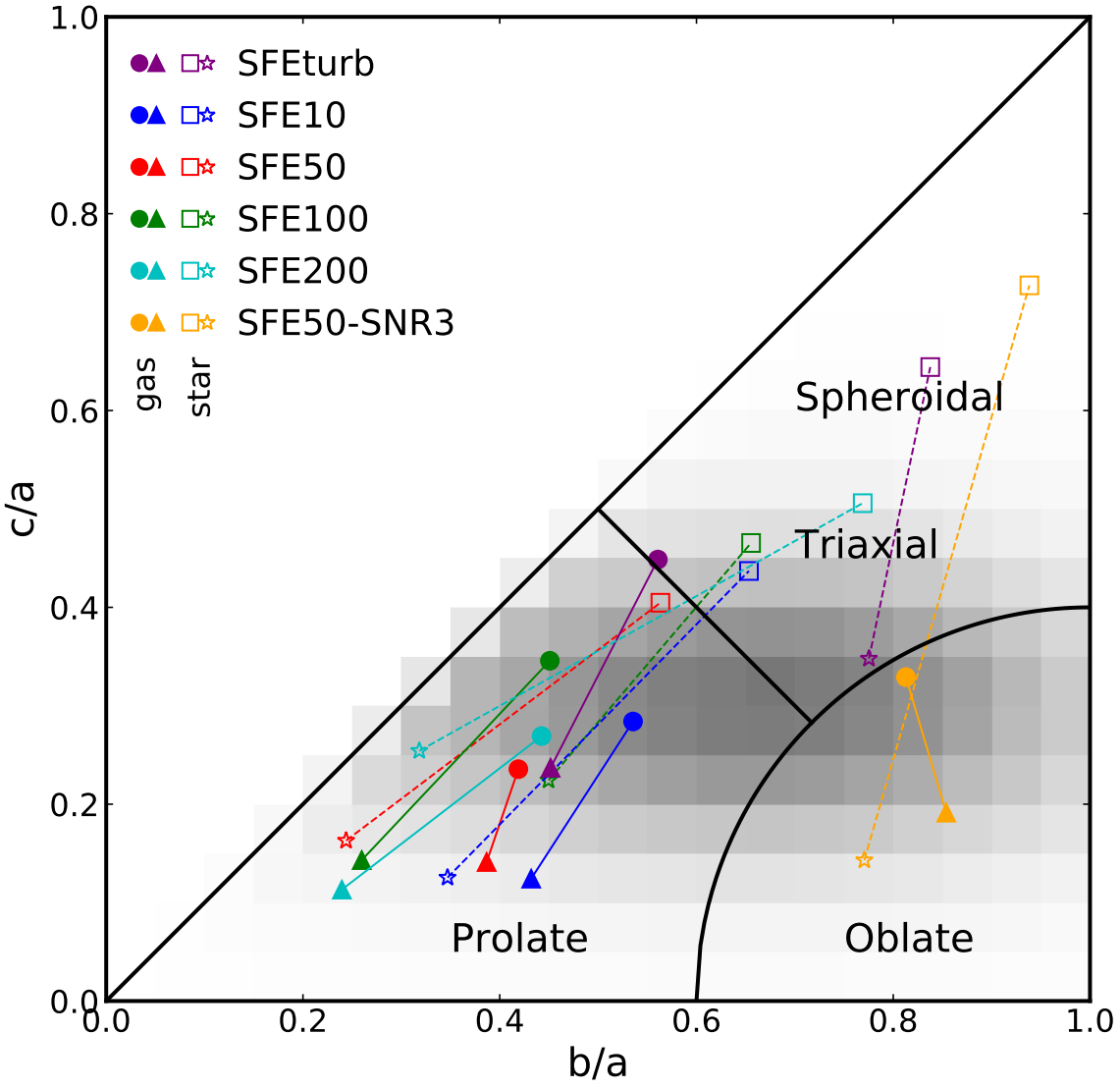

Figure 2.2 Axis ratios of all gas (filled circles) and molecular gas (filled triangles), and all stars (open squares) and young stars (asterisks) for different runs. The parameter space of the axis ratios b/a and c/a is divided into three parts to distinguish shapes of an ellipsoid, following the definition in Zhang et al. (2019) and van der Wel et al. (2014). The gray shades show the observed distribution of axis ratios for stars in the high redshift galaxies of similar mass range to our simulated galaxies (Zhang et al., 2019).

The SFEturb output is at an earlier epoch, which could be the cause of its different axis ratios of stars from the other runs with the fiducial feedback strength.

As pointed out by Zhang et al. (2019), the "spheroidal" regime also contains triaxial shapes and is better named as "spheroidal or triaxial". Zhang et al. (2019) assumed that the intrinsic shapes of galaxies are triaxial ellipsoids and fit ellipses with the same b/a axis ratios to isophote contours (instead of using the tensor of inertia). Then they used a model to reconstruct the three-dimensional shape based on the observed two-dimensional axis ratios for galaxies in the multi-wavelength CANDELS survey. They used data in wavelengths that are as close as possible to 4600Å in rest-frame, which better traces the distribution of young stars. We show in gray shades in Figure 2.2 their modeled axis ratio distribution for the redshift bin $1.5 < z < 2.0$ and mass bin $9.5 < \log M_*/\,M_\odot < 10$, which are closest to our simulated galaxies. Compared to the observations, the distribution of stars in our galaxies is typically rounder and thicker, extending from mildly prolate to triaxial to spheroidal regimes. The shape of the gas is significantly more flattened, falling into the prolate regime. Ceverino et al. (2015) and Tomassetti et al. (2016) also find that the shapes of $z \approx 1-3$ galaxies in their simulations tend to be triaxial or prolate.

However, the shapes of molecular gas and young stars are much more flattened ($c/a = 0.1 - 0.2$) and elongated ($b/a < 0.5$). Since the number of young star particles is not large and their distribution is clumpy (except for SFE50-SNR3 run) it is easy for young stars and molecular gas to display an elongated, prolate configuration. The shapes of young stars and molecular gas are similar, because the stars form out of molecular gas. As the galaxy evolves and interacts with other galaxies, these newly formed stars leave their formation sites and settle into a more spheroidal configuration.

Current observations suggest that galaxies transform from clumpy, thick discs at $z \gtrsim 1.5$ (e.g., Elmegreen et al., 2017) to regular, spiral structures at low redshift as galactic discs become less turbulent (Elmegreen & Elmegreen, 2014). During this

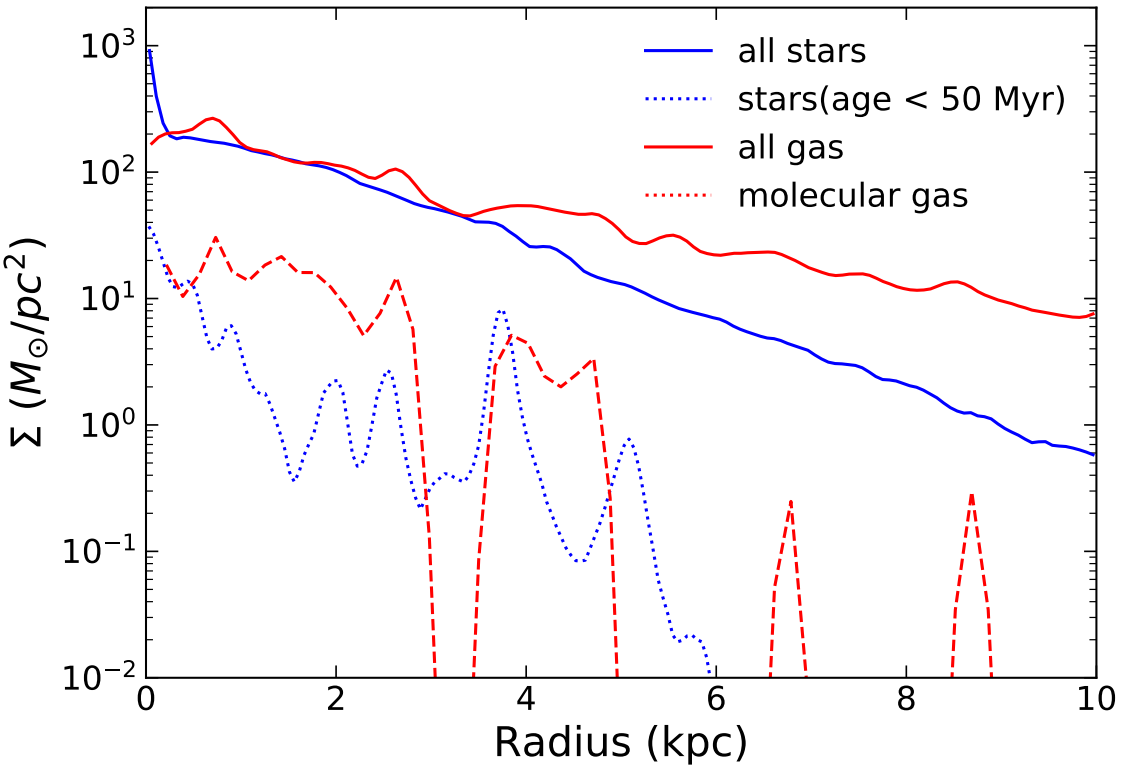

Figure 2.3 An example of the surface density profiles of gas and stars in cylindrical shells in run SFE50. The projection depth is ±5 kpc.

process galaxy shapes transit from prolate to oblate (Zhang et al., 2019), while the gas fraction decreases (Dessauges-Zavadsky et al., 2017a). Shibuya et al. (2016) found that the fraction of clumpy galaxies increases at $z \gtrsim 2$, peaks at $z \simeq 1-2$, and decreases at $z \lesssim 1$. This also indicates that galaxies transit from irregular morphology at high redshift to more regular, disky shapes at relatively low redshift.

Figure 2.3 shows an example of the surface density profile of stars and gas in one galaxy. The surface density profile is calculated in cylindrical shells in the galaxy plane defined previously, with projection depth being ±5 kpc. Stars dominate over gas near the centre but have similar density in the range of radii $0.4 < R < 3$ kpc. The gas density decreases less steeply with radius and begins to dominate over stars at larger radii. Since some of our galaxies have spheroidal shapes, we chose a large projection depth of 5 kpc to capture most of the stars and gas extending above the

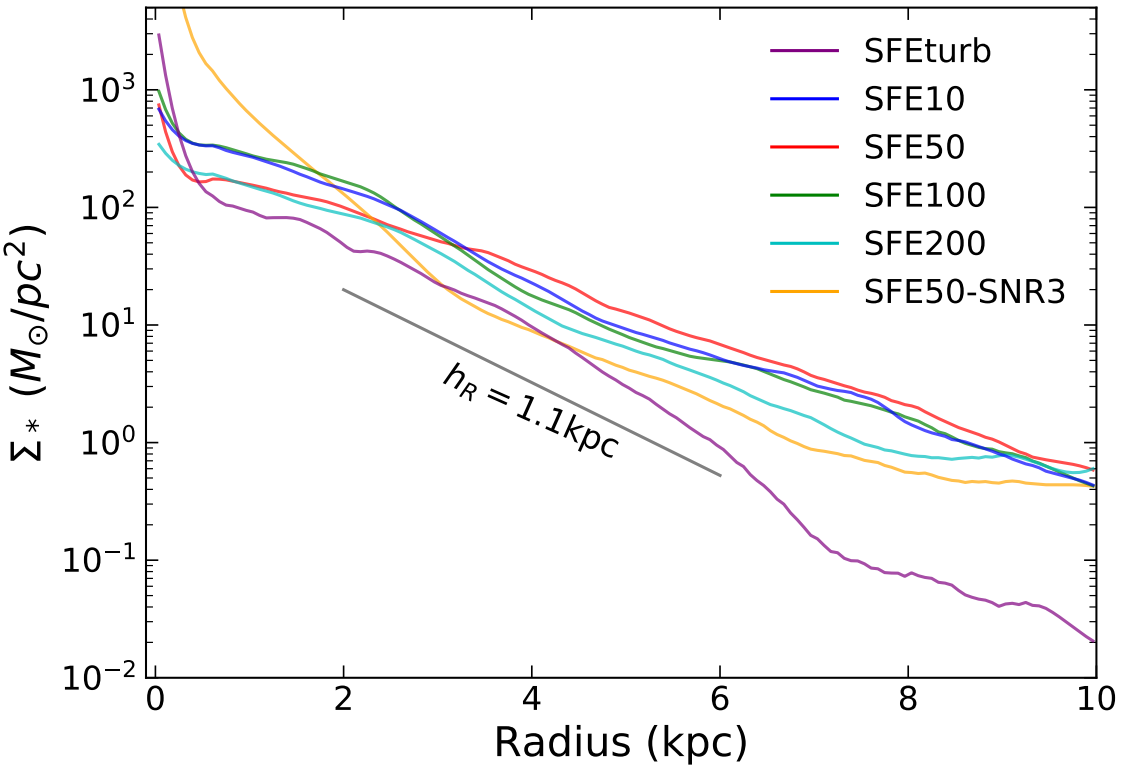

Figure 2.4 Comparison of the stellar surface density profiles for all six runs. Gray line corresponds to an exponential disc with the scale length of 1.1 kpc.

galaxy plane. Because of such a deep projection the molecular gas does not dominate in any cylindrical shell, but still reaches very high column density, $10-50\,M_\odot\,\mathrm{pc}^{-2}$, in the inner few kpc. Outside $R = 5$ kpc the H_2 density is low. When we calculate the Toomre Q parameter in later sections, we focus on the square region of ± 5 kpc from the centre of the galaxy in the galaxy plane with a certain thickness, which contains most of the star-forming molecular gas. The total gas and stellar surface densities in this region are comparable. The density profile of young stars formed within the last 50 Myr is quite irregular and does not directly follow the distribution of molecular gas. We return to this point in the discussion of density maps.

Figure 2.4 compares the stellar profiles of galaxies in different runs. Even though our galaxies are thick, all of them can still be fit reasonably well by an exponential disc model with the scale length $h_R = 1.1 \pm 0.1$ kpc, in the radial range $2-6$ kpc.

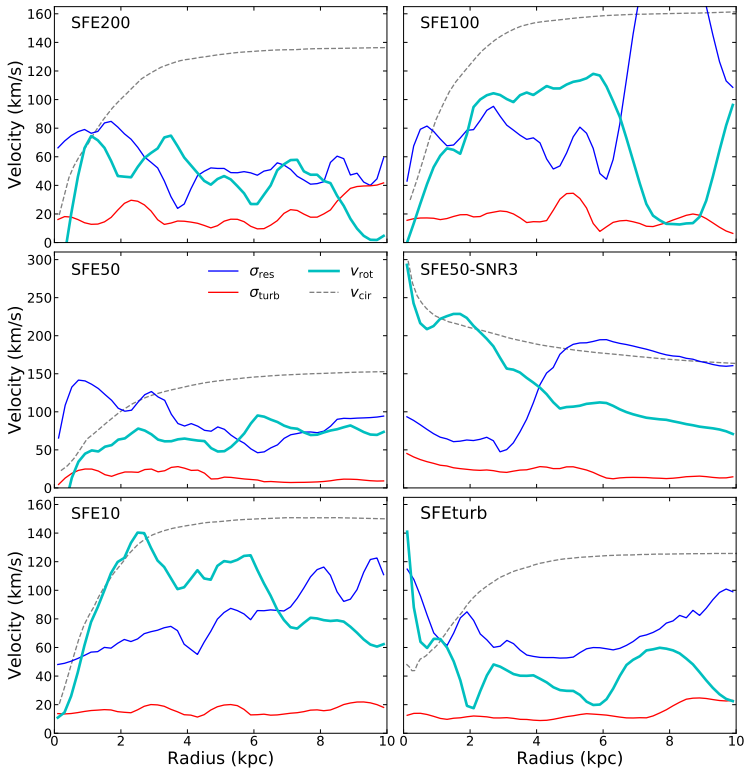

Figure 2.5 Kinematics of the molecular gas in the simulated galaxies, calculated in cylindrical shells: rotation velocity (cyan), subgrid-scale turbulence (red), resolved turbulence (blue). Grey dashed lines show the circular velocity profile.

The only exception is run SFE50 that has $h_R \approx 1.5$ kpc. Since the gas density profiles are shallower, formation of new stars over time would gradually increase the stellar scale length. We note that the exponential fits here exclude the central parts of the galaxies, which may harbor dense bulges.

Interestingly, observations of clumpy galaxies at $z \simeq 1-2$ (e.g., Shibuya et al., 2016) find surface brightness profiles with the Sérsic index of $n \simeq 1$, corresponding to an exponential disc. For example, massive ($M_* \sim 10^{9.6}\,M_\odot$) star-forming galaxies at $1 < z < 3$ in the ZFOURGE survey can be fit by $n \approx 1.2$ and an effective radius of $R_e \approx 2$ kpc (Papovich et al., 2015; Forrest et al., 2018). This is very close to the effective radii we obtain for our galaxies: $R_e \approx 1.68\,h_R \approx 1.8$ kpc. Thus the stellar density profiles of our simulated galaxies are typical of star-forming galaxies at these intermediate redshifts.

2.2.3 Velocity dispersion profiles

The kinematics of the interstellar medium in the simulated galaxies is shaped to a large extent by turbulent flows. The ART code models the injection of unresolved (subgrid scale) turbulence by SN explosions, as well as turbulent cascade from resolved scales. Even though these turbulent motions decay on a dynamical timescale, they still play an important role in distributing the SN energy over the neighbouring cells in actively star-forming galaxies.

Figure 2.5 shows kinematic profiles of the molecular gas in all six galaxies. Using the previously defined centre and orientation of the galaxy plane, the rotation curve is calculated in linearly spaced cylindrical shells of 200 pc in width. The SGS turbulence is averaged over the cells falling into the cylindrical shell. The resolved velocity dispersion $\sigma_{\rm res}$ is the residual motion of cells, after subtracting the mean rotation

velocity. Both dispersions are calculated in three dimensions, weighted by H_2 gas mass. When averaged over the cylindrical shells, the resolved dispersion dominates over the unresolved SGS turbulence. The resolved dispersion is comparable to the rotation velocity by amplitude and typically anticorrelates in radius.

Two runs stand out from the general trend. The weaker feedback run SFE50-SNR3 forms a massive bulge, which leads to the sharp rise of the circular velocity at the centre. The striking difference with the parallel run SFE50 with the same $\epsilon_{\rm ff}$, where the boost factor is only 5/3 times larger, may seem surprising. In the inner few kpc the rotation velocities differ by more than a factor of two. A likely cause of such a difference is in the unstable nature of formation of dense stellar clumps. If a dense bulge happens to begin forming, as was the case in run SFE50-SNR3, then stellar feedback is insufficient to halt its growth until most of the inner gas is converted into stars. In the other case when momentum feedback is always sufficiently strong, massive clumps do not form and subsequent feedback succeeds in regulating star formation. The other run that shows a rising rotation velocity towards the centre is SFEturb. The last available output for this run is at a higher redshift than the rest and the magnitude of the velocity increase is relatively small. It is difficult to predict whether this increase will persist to later epochs or conform to the other runs.

The turbulent flows of H_2 gas are highly supersonic. The 3D turbulent velocities reach $100 - 150$ km s^{-1} and violently stir the gas clouds in the galaxy plane. These flows are responsible for the thick structure of the gaseous discs of all our simulated galaxies. They also create significant vertical motion of the gas clouds.

2.3 Toomre analysis

The local stability criterion, first derived by Safronov (1960) and Toomre (1964) and reviewed in Toomre (1977), has been commonly used to identify star-forming regions in regular disc galaxies. For a gaseous disc, the Q parameter is defined as:

$$Q = \frac{\sigma\kappa}{\pi G\Sigma}, \tag{2.1}$$

where κ is the epicycle frequency, σ is the velocity dispersion, and Σ is the surface mass density. A self-gravitating disc is stable to axisymmetric perturbations on scale λ if

$$Q > 2\left(\frac{\lambda}{\lambda_{\rm crit}} - \frac{\lambda^2}{\lambda_{\rm crit}^2}\right)^{1/2}, \tag{2.2}$$

where $\lambda_{\rm crit}$ is the largest unstable wavelength for a zero-pressure ($\sigma = 0$) disc:

$$\lambda_{\rm crit} \equiv \frac{4\pi^2 G\Sigma}{\kappa^2}.$$

The disc with $Q > 1$ is stable on all perturbation scales λ, but even smaller Q can indicate stability for $\lambda < \frac{1}{2}\lambda_{\rm crit}$ (Binney & Tremaine, 2008).

The wavelength of the fastest growing perturbation is not $\lambda_{\rm crit}$ but instead is given by (e.g., Nelson, 2006)

$$\lambda_T \equiv \frac{2\sigma^2}{G\Sigma} = \lambda_{\rm crit}\frac{Q^2}{2}. \tag{2.3}$$

It is a two-dimensional analogue of the Jeans wavelength.

The amount of gas contained within a circle of diameter λ_T is an expected mass that would collapse into a self-gravitating object, called the Toomre mass:

$$M_T \equiv \frac{\pi}{4}\lambda_T^2\Sigma = \frac{\pi\sigma^4}{G^2\Sigma}. \tag{2.4}$$

In some recent work (e.g., Reina-Campos & Kruijssen, 2017; Pfeffer et al., 2018a),

the Toomre mass was defined alternatively as the mass within $\lambda_{\rm crit}$:

$$\tilde{M}_T \equiv \frac{\pi}{4}\lambda_{\rm crit}^2 \Sigma = \frac{4\pi^5 G^2 \Sigma^3}{\kappa^4}. \tag{2.5}$$

The two definitions are related by

$$M_T = \tilde{M}_T \frac{Q^4}{4}.$$

For dense clumps with $Q < 1$, the alternative $\tilde{M}_T$ can significantly overestimate the Toomre mass.

The original Toomre criterion applies to a fully gaseous, rotating, thin disc, under linear perturbations. Real galaxies, and our simulated galaxies, contain gas, stars, and dark matter. While the dark matter has much larger velocity dispersion and does not significantly affect disc stability, stars have velocity dispersion comparable to gas and contribute additional gravitational force. Different velocity dispersions, and the corresponding different scaleheights above the disc plane, of the stellar and gaseous components mean that they affect the stability criterion in a more complicated way. Rafikov (2001) presents a detailed analysis of the stability of multi-component discs and provides a modified criterion:

$$\frac{1}{Q_R} = \max_k \left\{ \frac{1}{Q_*} \frac{2\left(1 - e^{-q^2} I_0(q^2)\right)}{q} + \frac{1}{Q_g} \frac{2q\xi}{1 + q^2\xi^2} \right\} \tag{2.6}$$

where $q \equiv k\sigma_*/\kappa$, $\xi \equiv \sigma_g/\sigma_*$, I_0 is modified Bessel function of the first kind, and Q_* and Q_g are the separate parameters for stars and gas. The maximum is taken over values of the perturbation wavenumber $k = 2\pi/\lambda$. This criterion has been adopted in some simulations of galactic discs (e.g., Li et al., 2005).

Romeo & Wiegert (2011) and Romeo & Falstad (2013) suggest a simplified version of the modified Q parameter and show that it reasonably accurately reproduces

Rafikov's criterion. The modified parameter for N components of the matter distribution is

$$Q_N = \left(\sum_{i=1}^{N} Q_i^{-1} \frac{2\sigma_m \sigma_i}{\sigma_m^2 + \sigma_i^2} \right)^{-1} \tag{2.7}$$

where Q_i are the separate parameters for each component:

$$Q_i \equiv \frac{\sigma_i \kappa}{\pi G \Sigma_i} \tag{2.8}$$

and σ_m is the velocity dispersion of the component with the lowest Q_i. The ratio of dispersions in equation (2.7) acts as a weighting factor that reduces the contribution of the other components with $\sigma_i \neq \sigma_m$. They also suggest additional correction to account for the thickness of different components. This correction is small and we do not use it in our analysis for clarity, but we denote their full formalism for the Toomre parameter as Q_{RW}.

Romeo & Falstad (2013) show that their definition of Q_N with $N = 2$ can accurately approximate Rafikov's criterion. The definition of Q_2 was adopted, for example, in the model of Krumholz et al. (2018). To examine this approximation we calculated and compared the values of Q_R and Q_2 for our galaxies, and found that indeed they agree to better than 5%. The reason for such a close match can be understood as follows. If $Q_g < Q_*$, then the maximum of equation (2.6) is reached when $q\xi = 1$, and the weighting factors before Q_g are unity for both expressions (Q_R and Q_2) while the weighting factors before Q_* evaluate to be within 6% of each other. If instead $Q_g > Q_*$, then the maximum of equation (2.6) is reached when $q = 1$, and the weighting factors before Q_g are the same while the weighting factors before Q_* are within 7% of each other. Thus in general the two expressions Q_R and Q_2 differ by no more than a few percent.

For the calculation of Q in our simulations we use equation (2.7) and consider three components: molecular gas (H_2 multiplied by 1.32 to account for helium within

molecular clouds), atomic gas (HI plus HeI), and stars.

The wavelengths $\lambda_{\rm crit}$ and λ_T can be evaluated using the modified surface density:

$$\Sigma_{\rm eff} \equiv \sum_{i=1}^{N} \Sigma_i \frac{2\sigma_m^2}{\sigma_m^2 + \sigma_i^2} \tag{2.9}$$

as

$$\lambda_{\rm crit} = \frac{4\pi^2 G \Sigma_{\rm eff}}{\kappa^2}, \quad \lambda_T = \frac{2\sigma_m^2}{G\Sigma_{\rm eff}}. \tag{2.10}$$

If the velocity dispersions of any two components are equal, their surface densities can be effectively combined. For example, if the dispersions of atomic and molecular gas are the same, $\sigma_{\rm HI} = \sigma_{\rm H_2} = \sigma_g$, then their Q_i parameters can be combined into one

$$Q_g = (Q_{\rm HI}^{-1} + Q_{\rm H_2}^{-1})^{-1} = \frac{\sigma_g \kappa}{\pi G(\Sigma_{\rm HI} + \Sigma_{\rm H_2})}.$$

Analogously, if stars have the same velocity dispersion as the gas, their surface density can be added to the sum in the denominator. For example, Orr et al. (2018) included the sum of the gas and stars to obtain the total mass surface density, $\Sigma = \Sigma_{\rm g} + \Sigma_*$, in their analysis of the FIRE simulations.

Using the simulation outputs at the last available epoch, we calculate projected maps of Q-values in square patches of physical length $L = 0.2\,$kpc. The orientation of the patches is aligned with the galaxy plane, and the projection is along the direction perpendicular to the galaxy plane.

The size L is chosen to contain many gas cells for sufficient averaging, such that numerical discreteness does not affect the conclusions. The size of individual cells containing most of the neutral gas varies from about 50 to 100 pc at redshift $z = 1.5$. The mass-weighted size of individual cells at the redshift z=1.5 outputs varies from 36 to 54 pc when weighted by molecular gas mass, or from 92 to 112 pc when weighted by neutral gas mass. These cells are not at the highest level of refinement but dominate the gas mass. Therefore, we restrict the square patches to be no smaller than 200 pc.

Below we describe the calculation of the three variables determining the Q parameter: epicycle frequency, surface density, and velocity dispersion.

2.3.1 Epicycle frequency

For an axisymmetric disc with near-circular orbits (e.g., Binney & Tremaine, 2008)

$$\kappa^2 = \kappa_R^2 \equiv \frac{2\bar{V}^2}{R^2}\left(1 + \frac{d\ln\bar{V}}{d\ln R}\right) \tag{2.11}$$

where $\bar{V}$ is the average velocity of circular motion. We approximate it by a spherically-symmetric expression for the circular velocity:

$$V_{\rm c}(R) \equiv \left(\frac{GM(R)}{R}\right)^{1/2}.$$

The actual circular velocity for a razor-thin exponential disc is given by the Bessel functions, but the difference is only $\sim 10\%$, which is below other approximations necessary for our analysis.

In the limit of a flat rotation curve, the expression simplifies to

$$\kappa^2 = \kappa_c^2 \equiv \frac{2V_{\rm c}^2}{R^2}. \tag{2.12}$$

We define separate variables κ_R and κ_c to refer to these commonly used approximations. Below we use equation (2.12) as the definition of κ_c and use it to evaluate κ_c even when $V_{\rm c}$ is not constant.

Our simulated galaxies have much smaller and irregular rotation pattern because of additional pressure support provided by turbulent motions, as shown in Figure 2.5, and therefore we do not expect these approximations to hold. Since the epicycle expansion is done around the mean azimuthal velocity, a more appropriate quantity is the actual rotation velocity $\bar{V} = V_{\rm rot}$. To verify the accuracy of calculation of κ,

Table 2.2 Accuracy of calculation of κ in various approximations

Run	κ_c/κ_R 25-50-75% range	κ_ϕ/κ_R 25-50-75% range	κ_T/κ_R 25-50-75% range
SFE200	0.79 - 0.84 - 0.94	0.46 - 0.57 - 0.76	0.60 - 0.86 - 1.17
SFE100	0.78 - 0.82 - 0.92	0.57 - 0.66 - 0.79	0.61 - 0.84 - 1.13
SFE50	0.74 - 0.81 - 0.96	0.42 - 0.52 - 0.73	0.61 - 0.89 - 1.21
SFE10	0.75 - 0.86 - 0.96	0.61 - 0.89 - 1.07	0.60 - 0.90 - 1.26
SFEturb	0.75 - 0.85 - 1.03	0.27 - 0.62 - 1.14	0.58 - 0.86 - 1.19
SFE50-SNR3	1.03 - 1.09 - 1.13	0.58 - 0.74 - 0.96	0.49 - 0.76 - 1.11

we also evaluated the expression

$$\kappa^2 = \kappa_\phi^2 \equiv \frac{2V_{\rm rot}^2}{R^2}\left(1 + \frac{d\ln V_{\rm rot}}{d\ln R}\right). \tag{2.13}$$

Table 2.2 shows the distributions of ratios of κ given by equations (2.11-2.13). The epicycle frequencies are calculated in axisymmetric cylindrical shells of width $L = 200\,{\rm pc}$. As expected, the median κ_c is consistently lower than κ_R by $\sim 15\%$ because the circular velocity curves are still increasing in the range of radii we consider, $R < 5\,{\rm kpc}$. However, the difference is small.

The values calculated directly from the rotation velocity have a larger spread from κ_R, by $\sim 30\%$ in the median. The interquartile range of the distribution of κ_ϕ/κ_R is also wider, extending from 0.3 to 1.1. However, despite this significant scatter from shell to shell, the overall good accuracy of calculation of κ using the simple approximation of equation (2.11) is surprising. It validates our analysis of the Q parameter to better than a factor of two. For the distributions of Q parameter described below the epicycle frequency is calculated in cylindrical shells within the galaxy plane using equation (2.11), except in maps shown in Figure 2.7 where we use equation (2.12) to avoid visual artifacts.

Inoue et al. (2016) used their simulations of high-redshift galaxies to compare the calculation of κ using the circular velocity V_c and the actual rotation velocity $V_{\rm rot}$. In

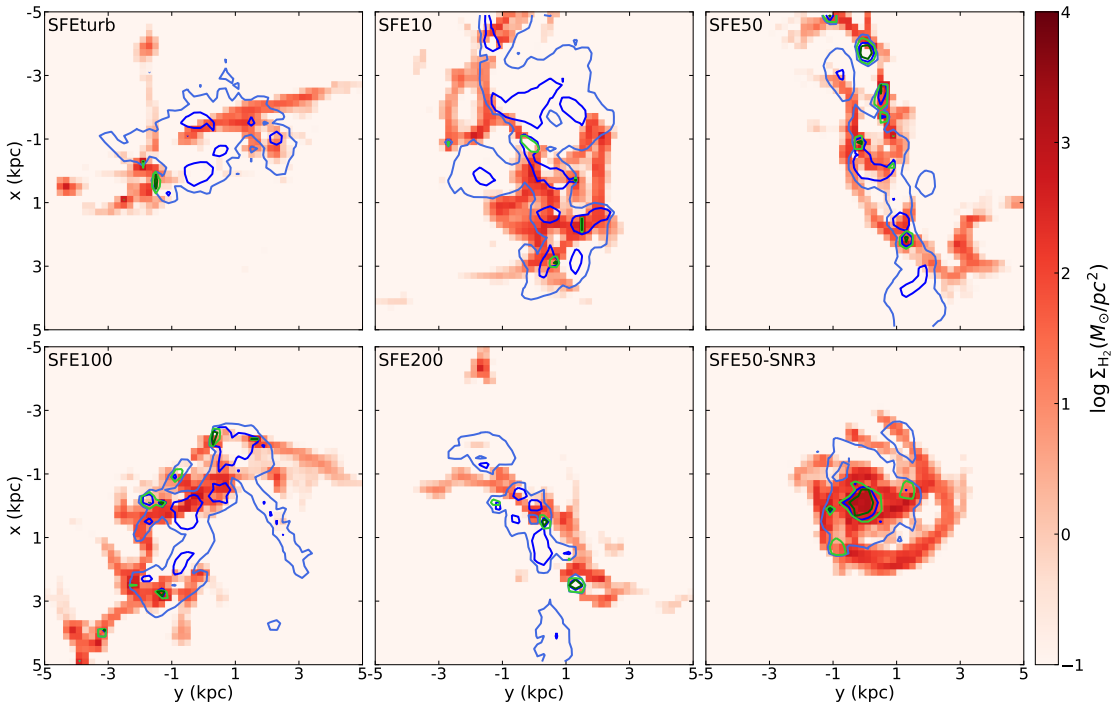

Figure 2.6 Map of H_2 surface density with projection thickness of ± 2 kpc. Blue contours enclose 80% and 99% of mass of young stars formed in last 50 Myr. Green contours enclose the same mass fractions of young stars formed only in last 10 Myr.

agreement with our results, they find that the latter gives somewhat lower values, but the difference is small. They also note that the rotation velocity can decrease with radius suddenly and lead to $\kappa^2 < 0$ in some shells in the outer parts of the galaxies.

In addition to the above approximations, we evaluate another expression for the epicycle frequency suggested by Pfeffer et al. (2018a):

$$\kappa^2 = \kappa_T^2 \equiv -\sum \lambda_i - \lambda_1, \tag{2.14}$$

where λ_i are the eigenvalues of the tidal tensor around the patch. The maximum eigenvalue is $\lambda_1 \geq \lambda_2, \lambda_3$. We calculate the tidal tensor on the uniform grid of 200 pc patches and evaluate the ratio of κ_T/κ_R on these patches, instead of the cylindrical shells as for the other axisymmetric definitions of κ. The ratio κ_T/κ_R is similar to κ_c/κ_R but with larger scatter. The absolute values of κ_R are similar to those in Fig. A1 of Pfeffer et al. (2018a): κ_R decreases from around $200\,\mathrm{Gyr}^{-1}$ in inner 1 kpc to $\sim 30\,\mathrm{Gyr}^{-1}$ at $R = 5$ kpc, for all runs except SFE50-SNR3.

2.3.2 Surface density

We calculate the surface densities of all components by projecting over a column of thickness $\pm 2\,\mathrm{kpc}$ around the galaxy plane. This thickness was chosen to contain most (about 80%) of the gas and stars. Only in SFE200 run such a column contains 70% of the mass. To account for 80% of the mass in that run we would need to integrate within $\pm 3\,\mathrm{kpc}$, but this difference is not essential to our analysis and we chose to keep the column thickness the same for all runs.

In most patches the gas dominates the surface density. The median ratio Σ_*/Σ_g is in the range $0.7 - 0.9$ for most runs. The two exceptions are SFE10 with $\Sigma_*/\Sigma_g \approx 1$ and SFE50-SNR3 with $\Sigma_*/\Sigma_g \approx 2$. Most of the gas near the galaxy plane is atomic and some is ionized. Below we investigate specifically the distribution of molecular

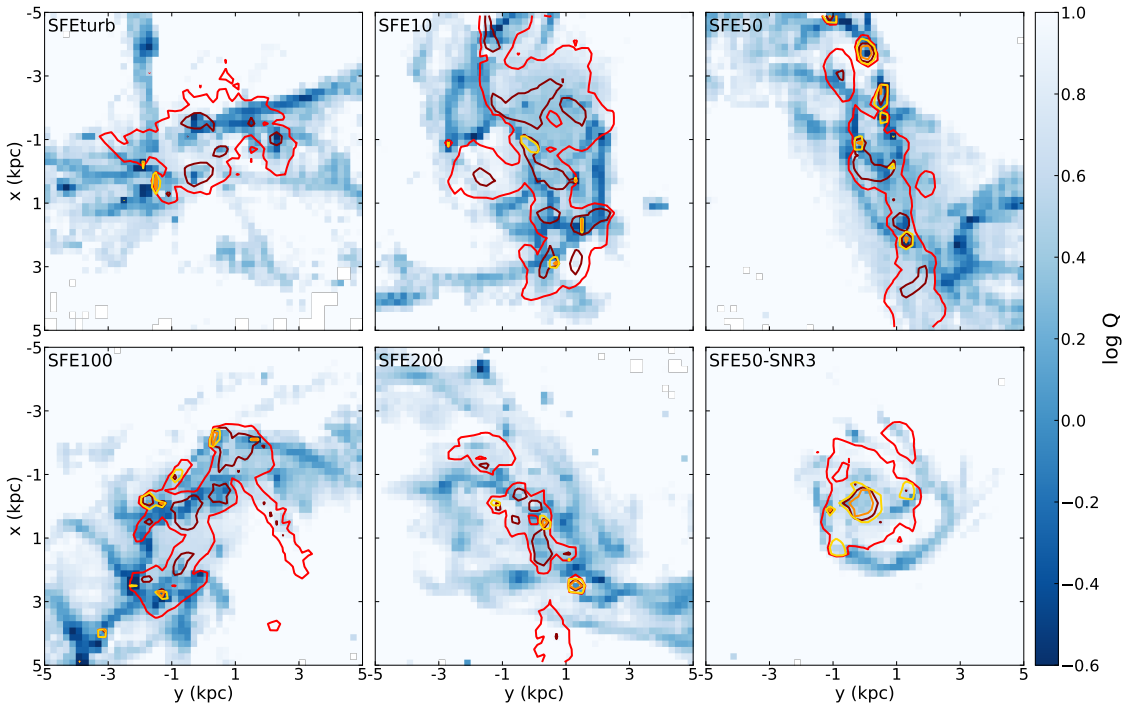

Figure 2.7 Map of the Q parameter in all six runs, at the last available output. Contours in red show the surface density of stars younger than 50 Myr enclosing 80% and 99% of the mass of these stars. Yellow contours enclose the same mass fractions of young stars formed only in last 10 Myr. The values of Q are calculated using the κ_c definition to avoid visual artifacts.

H_2 gas, which is directly linked to star formation.

Figure 2.6 compares the maps of molecular gas density with star formation rate density. To calculate the star formation rate in a large patch of a galaxy within a given interval of time, we need to take into account a finite duration of formation of cluster particles in our CCF algorithm (Li et al., 2018a). At the very beginning of the formation episode of each particle, the formation rate is low. Then as particle mass rises, the gas accretion rate increases and the star formation rate picks up, until stellar feedback halts the accretion. The variable $\tau_{\rm ave}$ approximately corresponds to the peak time of formation rate. Therefore, we take the star particle "zero age" to be the moment when it has gone through one $\tau_{\rm ave}$ after the creation. The typical values are $\tau_{\rm ave} \sim 1 - 2$ Myr.

Surprisingly, Figure 2.6 shows that for the majority of patches the regions of high H_2 density *do not* coincide with the regions of high SFR density. It is because strong stellar feedback quickly heats and removes the gas from star forming regions. This becomes an important point in the calculation of the depletion time in Section 2.4.

Observations of high-redshift galaxies (e.g., Daddi et al., 2010; Tacconi et al., 2013; Dessauges-Zavadsky et al., 2017a) provide an analogous comparison of CO and radio continuum maps with optical, UV, and IR images. Observed emission from molecular gas and young stars is generally in the same place, but they do not coincide exactly. Other simulation results also show decoupled gas and SFR maps. For example, in the FIRE simulations, which use $\epsilon_{\rm ff} = 100\%$, Oklopčić et al. (2017) find that gas clumps coincide with instantaneous SFR maps fairly well, but start to decouple from the SFR averaged over 10 Myr. The map of SFR averaged over 100 Myr shows that gas clumps do not trace the SFR peaks at all. This is similar to our results, where SF averaged over 10 Myr is located near the peaks of molecular gas, while SF averaged over 50 Myr correlates less well with the H_2 map. Most of the molecular gas is not participating in star formation at any given time.

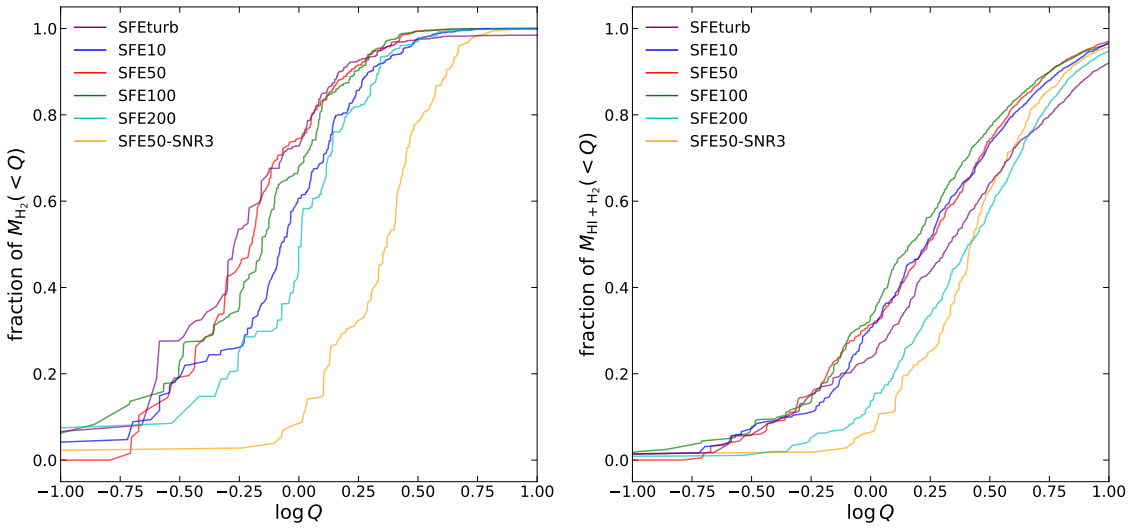

Figure 2.8 Cumulative distribution of the molecular H_2 (left) and neutral $HI + H_2$ (right) gas mass within ± 2 kpc as a function of Q. The velocity dispersion used to calculate Q for the right panel is weighted by the $HI + H_2$ mass for consistency. The epicyclic frequency is calculated using the κ_R definition.

2.3.3 Velocity dispersion

Since our galaxies are not thin, rotating discs with only velocity dispersion in radial direction, as in the original Toomre analysis, it is more reasonable to include all components of velocity dispersions as all of them resist gravitational collapse. The full velocity dispersion of gas includes three components: the sound speed within a cell c_s, the subgrid scale turbulence (SGS; unresolved velocity dispersion) within a cell σ_T, and velocity differences between neighbour cells (resolved velocity dispersion) $\sigma_{\rm cell}$:

$$\sigma_g^2 = c_s^2 + \sigma_T^2 + \sigma_{\rm cell}^2.$$

Using gas mass M_k in a cell k (see two versions below), we calculate the mass-weighted average of each component for the cells within a given patch:

$$\begin{aligned} c_s^2 &= \frac{\sum_k M_k\, c_{s,k}^2}{\sum_k M_k} \\ \sigma_T^2 &= \frac{2\sum_k E_{\mathrm{turb},k}}{\sum_k M_k} \\ \sigma_{\mathrm{cell}}^2 &= \sum_{i=r,\phi,z} \left[\frac{\sum_k {v_{i,k}}^2\, M_k}{\sum_k M_k} - \left(\frac{\sum_k v_{i,k}\, M_k}{\sum_k M_k} \right)^2 \right]. \end{aligned} \tag{2.15}$$

Here $\mathbf{v}_k$ is the velocity of cell k in the galacto-centric reference frame, and $E_{\mathrm{turb},k}$ is the energy in subgrid-scale turbulence in cell k, which is directly calculated in the simulation.

The goal of Toomre analysis is to identify regions of the ISM that are unstable to collapse and star formation. Which parts of the ISM are directly observable depends on the detection method. Radio observations of CO and HCN transitions probe dense molecular gas, while 21-cm line technique detects atomic hydrogen gas. To facilitate comparison with both types of observations, we consider two versions of the Q parameter, which differ only in the weighting of the velocities. In one version the sound speed and the resolved velocity dispersion are weighted by the mass of molecular gas, $M_k = M_{\mathrm{H_2}}$, while in the other they are weighted by the mass of total neutral gas, $M_k = M_{\mathrm{HI+H_2}}$. In the expression for the SGS turbulence, M_k is always the total gas mass. For most of our analysis we use the result of weighting by the $\mathrm{H_2}$ mass; it is therefore implied by default. Only when we discuss the distribution of neutral gas, we use the $\mathrm{HI} + \mathrm{H_2}$ mass weighting for consistency.

We take the sum over all cells in a given square patch, projected over $\pm 0.8\,\mathrm{kpc}$. We chose this smaller projection length than for the surface density to capture most of the molecular gas while avoiding the contamination by ionized gas that could skew the measurement of the sound speed. This choice also limits the resolved velocity

dispersion. The column $\pm 0.8\,\mathrm{kpc}$ contains over 90% of H_2 mass in all runs except SFEturb (which contains 80%).

Within patches of $L = 0.2\,\mathrm{kpc}$ the SGS turbulence dispersion is typically larger than the resolved dispersion. The median ratio $\sigma_{\mathrm{cell}}/\sigma_T \approx 0.6$, when σ_{cell} is H_2 mass weighted. This situation is different from the azimuthal averages shown in Figure 2.5, where σ_{res} is the tangential velocity dispersion of cells in a cylindrical shell. Here instead we compute σ_{cell} using the actual local mean velocity in a patch. This reduces the residual dispersion substantially.

The velocity dispersion of stars is calculated simply as inter-particle dispersion, analogously to σ_{cell} but weighted by stellar mass. In most patches (over 75%), the stellar dispersion dominates over the gas dispersion. The median ratio $\sigma_g/\sigma_* \approx 0.4$ for the molecular H_2 gas and 0.5 for the neutral $\mathrm{HI} + H_2$ gas.

2.3.4 Toomre mass

Using the epicycle frequency determined in Section 2.3.1 and the effective λ_{crit} and λ_T (Equation 2.10), we can now calculate the corresponding Toomre mass under the multi-component Q definition. We use patches of 200 pc, taking Σ to be the gas surface densitiy projected over $\pm\,2$ kpc. Then we average M_T and $\tilde{M}_T$ (equations 2.4–2.5) in linearly-spaced cylindrical shells. Our values of M_T generally increase with radius and vary from $10^{8.2}$ to $10^{10.6}\,M_\odot$ (interquartile range) in the patches with $M_{\mathrm{H_2}} > 10^6\,M_\odot$ (see Section 2.3.5 for justification of this threshold). However, $\tilde{M}_T$ is systematically higher; it ranges from $10^{8.6}$ to $10^{11}\,M_\odot$ for the same patches. This is several orders of magnitude larger than the values found by Pfeffer et al. (2018a), mainly because the gas surface density in our simulations is about an order of magnitude larger.

Note that Tamburello et al. (2015) account for non-linear growth of perturbations in collapsing clumps and find that the actual fragmentation mass is lower than M_T

by a factor of several. Subsequent dynamics of collapsed clumps may also affect their mass because of merging and agglomeration with other clumps. We do not investigate the distribution of clump masses in this paper but plan to do it in follow-up work.

2.3.5 Distribution of the Q parameter

Figure 2.7 shows projected maps of the Q parameter and compares them with the star formation rate density. The maps of Q generally follow the maps of gas density but pick out sharper features, such as spiral arms or filaments reaching to the galactic centre. The dynamic range of Q is significantly reduced relative to that of the molecular gas surface density, which makes Q a useful predictor of future star-forming regions.

An example of the weak feedback run (bottom right panel) illustrates that some high-density regions may not be unstable because of a steep potential well and high velocity dispersion. Thus a simple density threshold would not correctly pick gas patches that are unstable to gravitational collapse.

In Figure 2.8 we show the cumulative distribution of Q, weighted by the mass of molecular (left panel) and all neutral hydrogen (right panel). That is, the left panel shows the fraction

$$\frac{M_{\mathrm{H}_2}(< Q)}{M_{\mathrm{H}_2}},$$

and analogously for HI + H_2 in the right panel. The weaker feedback run SFE50-SNR3 shows a very different distribution, but all other runs with the same feedback strength show consistent results. The cumulative H_2 masses are rising sharply around the median values $Q \approx 0.5 - 1.0$. These medians are remarkably close to unity, given the many approximations in our calculation of the Q parameter and the irregular structure of these high-redshift galaxies. It also suggests that majority of the molecular gas is in the marginally stable dynamical state, which may indicate self-regulation of star formation by stellar feedback.

The distribution of neutral gas mass is shifted systematically towards higher values of $Q \approx 1.5 - 2.6$. The difference is mainly because the H_2 weighting selects regions of higher surface density and slightly lower velocity dispersion, both of which reduce Q.

Table 2.4 shows the interquartile ranges of the cumulative distribution of Q in patches of $L = 0.2\,\mathrm{kpc}$, weighted by H_2 surface density and $HI + H_2$ surface density, respectively.

The Toomre analysis indicates that there is a threshold at $Q \lesssim 1$ to distinguish the unstable regions of the disc. It is given by equation (2.2) that depends on the ratio of the perturbation scale to the largest unstable wavelength. To test the applicability of this analysis to the simulated galaxies, we calculated the largest unstable wavelength λ_{crit} on the patches that are capable of forming stellar particles. Because of the minimum adopted mass of stellar particles in the simulations ($\sim 10^3\,M_\odot$), patches with insufficient H_2 mass are unable to produce even a single particle. This is a numerical resolution limitation and therefore, such patches should not be included in our analysis. In fact, the patch size is much larger than the size of star-forming regions adopted in our runs ($\sim 5\,\mathrm{pc}$) and so the limiting mass should be significantly larger than the minimum particle mass. Accounting also for the low star formation efficiency in some of the simulations, we set the threshold H_2 mass at $10^6\,M_\odot$. Experimentation with lower values ($10^4 - 10^5\,M_\odot$) showed that the median λ_{crit} decreases as the threshold decreases, by $\approx 1 - 2$ kpc. However, we choose the larger threshold value because it leads to a more reliable calculation of the depletion time in the next section.

Table 2.3 shows the cumulative distributions of λ_{crit} for patches of 0.2 kpc on the side with H_2 mass above $10^6\,M_\odot$. The critical wavelength is quite large, between 4 and 8 kpc in the median for all stronger feedback runs. Such large values are caused mainly by the high surface density of gas and stars in these high-redshift galaxies.

The values of the Toomre wavelength λ_T are systematically lower and range between $1 - 3$ kpc in the median. SFE50-SNR3 run is again an exception because of its

Table 2.3 Distribution of $\lambda_{\rm crit}$ and λ_T in patches of $L = 0.2\,$kpc

Run	$\lambda_{\rm crit}$ (kpc) 25-50-75% range	λ_T (kpc) 25-50-75% range
SFE200	2.3 - 5.1 - 11.8	2.1 - 3.4 - 6.7
SFE100	3.4 - 6.4 - 16.0	1.5 - 2.5 - 4.7
SFE50	4.8 - 7.6 - 15.6	1.5 - 2.9 - 7.2
SFE10	2.8 - 4.2 - 6.2	1.4 - 2.2 - 5.1
SFEturb	3.1 - 4.8 - 10.2	0.9 - 1.3 - 2.6
SFE50-SNR3	1.1 - 1.6 - 2.5	3.0 - 5.9 - 11.7

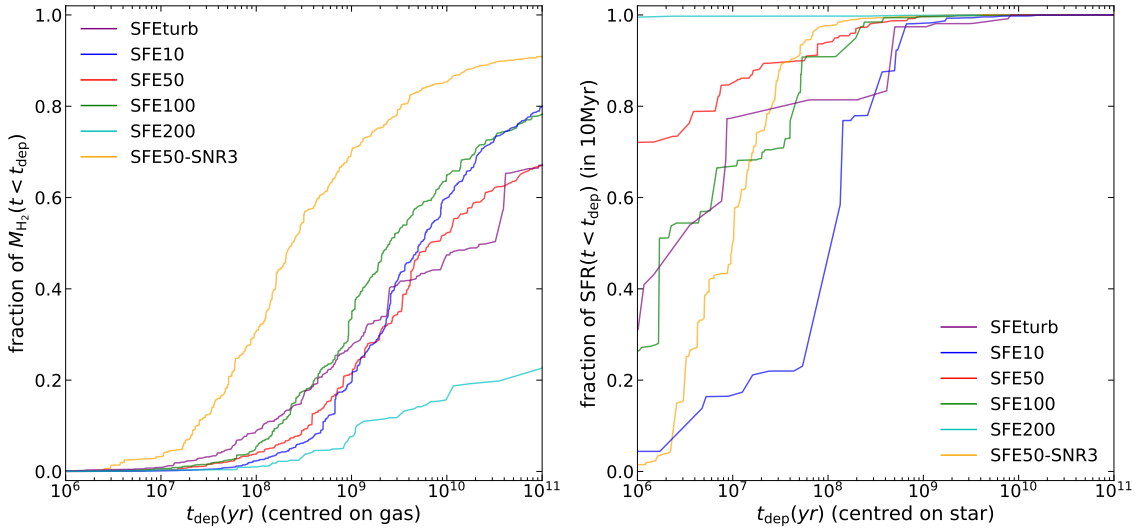

Figure 2.9 Cumulative distribution of the depletion time of molecular gas in patches of 100 pc in size. Left panel is for patches with $M_{\rm H_2} > 10^6\,M_\odot$ centered on gas density peaks, weighted by H_2 mass. Right panel is for patches with non-zero SFR within 10 Myr centered on SFR peaks, weighted by SFR.

higher Q values. It is straightforward to show, with $Q_N, \lambda_{\rm crit}, \lambda_T$ written in terms of σ_m and $\Sigma_{\rm eff}$ following Eq. (2.9) and (2.10), that the relation between the two wavelengths, given by equation (2.3), is still valid even for the multi-component definition of the Q parameter.

The condition $Q < 1$ could be used to select star-forming regions. The gas mass in patches that satisfy this condition is $\sim (1 - 5.5) \times 10^8\,M_\odot$, which generally accounts for 60-70% of the total H_2 mass (except for the weaker feedback run, which has less than 10% of the H_2 mass in $Q < 1$ patches). This mass is comparable to the mass

Table 2.4 Distribution of Q in patches of $L = 0.2\,\mathrm{kpc}$

Run	Q ($\Sigma_{\mathrm{H_2}}$ weighted) 25-50-75% range	Q ($\Sigma_{\mathrm{HI+H_2}}$ weighted) 25-50-75% range
SFE200	0.56 - 1.02 - 1.39	1.38 - 2.55 - 4.46
SFE100	0.33 - 0.70 - 1.18	0.76 - 1.47 - 2.95
SFE50	0.37 - 0.63 - 1.03	0.76 - 1.73 - 3.25
SFE10	0.47 - 0.84 - 1.36	0.87 - 1.71 - 3.35
SFEturb	0.26 - 0.53 - 1.02	1.04 - 2.07 - 4.12
SFE50-SNR3	1.35 - 2.35 - 2.94	1.75 - 2.64 - 4.17

converted to stars within 50 Myr. Except for SFE50-SNR3 run, the ratio of mass of young stars formed within 50 Myr to mass of H_2 in patches with $Q < 1$ varies between 0.34 and 3 for the different runs. Therefore, we can expect that most of the present molecular gas would be converted to stars on a timescale 50-100 Myr. This estimate is very approximate because Figure 2.7 shows that the star forming regions do not align with the unstable $Q < 1$ patches after 50 Myr. The match is better for very young star formation within only 10 Myr. In the next section we investigate the gas consumption timescale in detail, and consider different patch sizes and different ways of centering the search region.

2.4 Gas depletion time

An important measure of the global efficiency of star formation in galaxies is the gas depletion time. For our patches of size L, we define it as the ratio of the molecular gas surface density to the surface density of star formation:

$$t_{\mathrm{dep}}(L) = \frac{\Sigma_{\mathrm{H_2}}(L)}{\Sigma_{\mathrm{SFR}}(L)}. \tag{2.16}$$

Schruba et al. (2010), Feldmann et al. (2011), and Kruijssen & Longmore (2014) showed that the value of the depletion time depends on the spatial scale at which it is measured. It also depends on how the densities of gas and stars are calculated.

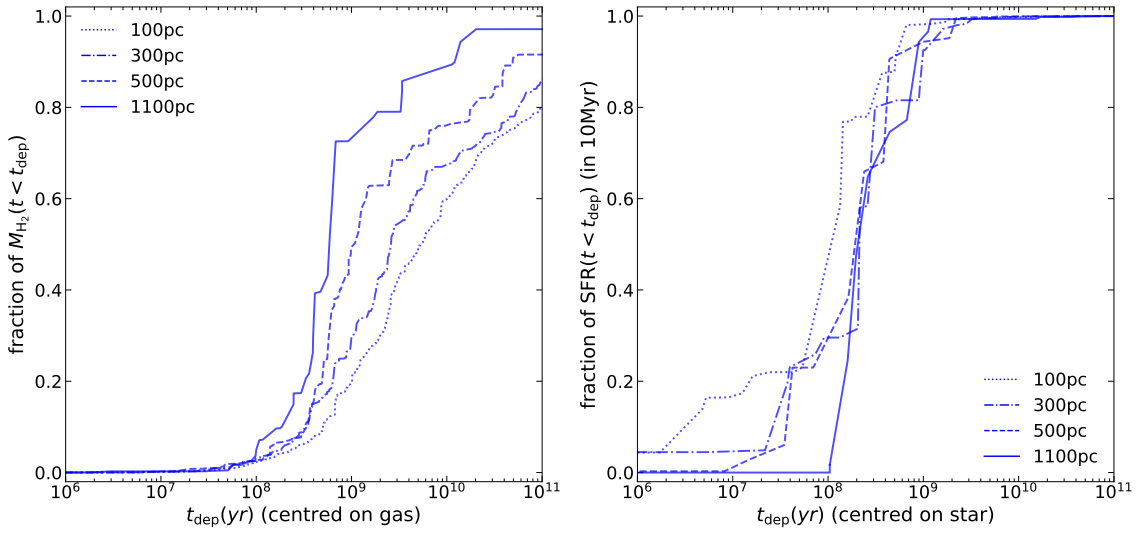

Figure 2.10 Cumulative distribution of H_2 depletion time for "centered-on-gas" (left panel) and "centered-on-stars" (right panel) versions for patches of different size. Here we show run SFE10 as an example. For "centered-on-gas" version, we use patches with $M_{H_2} > 10^6\,M_\odot$ and weight them by H_2 mass. For "centered-on-stars" version, we use patches with non-zero SFR and weight them by SFR. As patch size increases, the distribution shifts to smaller t_{dep} for "centered-on-gas" case and to larger t_{dep} for "centered-on-stars" case, similarly to the trend shown in Figure 2.11.

In most observations they are counted in circular apertures centred on peaks of SFR density. Such regions may not already contain most of the gas they had before the onset of the star formation episode, resulting in relatively short depletion timescales. On the other hand, if the apertures are centred on peaks of the current gas density, the depletion times appear systematically longer.

To compare these two approaches, we use the following algorithm to calculate the surface densities of gas and SFR in patches of 100, 300, 500, and 1100 pc. First, we cover the plane of the galaxy with a rectangular grid of 100×100 cells, each 100 pc wide and ±2kpc thick, and calculate Σ_{H_2} and Σ_{SFR} in each cell. To go to larger size of 300 pc, we search for maximum of H_2 mass (or SFR), record the position of that peak, group the peak and its surrounding $3^2 - 1 = 8$ cells into a larger patch and sum the H_2 mass and SFR in this group. We label these 9 cells as "counted" and repeat the loop for the yet uncounted cells. Now we have a list of patches of 300 pc. Then

we repeat this procedure for the chosen peaks and group 5^2 cell to obtain patches of 500 pc, etc.

Using this algorithm, we calculate the depletion time $t_{\rm dep}$ of H_2 gas in patches of different size from 100 pc to 1.1 kpc. The patches are chosen to be centred on peaks of H_2 mass or SFR. For the "centred-on-gas" version, we calculate the median of $t_{\rm dep}$ in patches with $M_{\rm H_2} > 10^6\,M_\odot$ to eliminate the low-density regions that would be unable to form stellar particles in our cluster formation algorithm, and count the SFR averaged over 50 Myr for better statistics. For the "centred-on-SFR" or "centred-on-young-stars" version (which we call "centred-on-stars" for brevity), we consider patches with any non-zero SFR, and count the SFR averaged over 10 Myr in order to approximate more closely the instantaneous star formation.

To show the full range of $t_{\rm dep}$ in both versions, in Figure 2.9 we plot the cumulative distribution of $t_{\rm dep}$ for patches of the smallest size (100 pc). In the "centred-on-gas" version (left panel), the cumulative distributions of $t_{\rm dep}$ weighted by H_2 mass do not reach 100%. This happens because some patches contain no stellar particles and therefore have formally infinite depletion time. The number of such patches is particularly large for SFE200 run, which has strongly misaligned molecular gas and young stars (see Figure 2.6). For the median H_2 mass the depletion time ranges between $10^{9.5}$ and $10^{10.5}$ yr in the strong-feedback runs.

In contrast, the right panel shows that the depletion time in the "centred-on-stars" version is significantly shorter, typically below 10^8 yr. Such discrepancy between the two counting methods is due to the gas-star misalignment which we emphasized above. Most of the star-forming sites have so little gas left within 100 pc that it would be exhausted in a relatively short interval of time.

Both versions of the gas depletion timescale vary strongly with the spatial scale on which they are calculated. To illustrate this dependence, we choose two representative runs (SFE50 and SFE10) and in Figure 2.10 we plot cumulative distributions of $t_{\rm dep}$

for different patch sizes. The distribution of $t_{\rm dep}$ in the "centered-on-gas" version shifts to smaller values as patch size increases, by an order of magnitude between 100 pc and 700 pc. This shift is monotonic with the patch size and similar for the two runs shown.

The "centered-on-stars" version shows the opposite trend: $t_{\rm dep}$ shifts to larger values. This change is roughly monotonic but differs between the two runs. In one case where $t_{\rm dep}$ was very short on 100 pc scale (SFE50) the increase is dramatic, by about two orders of magnitude. In the other case with larger $t_{\rm dep}$ (SFE10) the increase is only by a factor of two.

Adopting lower values of the threshold $M_{\rm H_2} > 10^4 - 10^6\,M_\odot$ increases the fraction of patches with infinite $t_{\rm dep}$, leading to larger scatter in $t_{\rm dep}$. If the threshold is taken to be $10^4\,M_\odot$, even the median value of $t_{\rm dep}$ in SFE200 run is infinite for all patch sizes. In the other runs, the median value of $t_{\rm dep}$ changes within 0.5 dex for the smallest patch size, and correspondingly less for larger patch sizes. On the other hand, increasing the threshold mass greatly reduces the available number of patches, so we do not set the threshold above $10^6\,M_\odot$.

Despite the above variations, the estimate of $t_{\rm dep}$ in both versions for all runs approaches a similar common range at the largest considered scale, $L = 1.1$ kpc. This convergence is illustrated in Figure 2.11. The median values over patches for different runs are all contained within $10^8 - 10^9$ yr. Unlike Figures 2.9 or 2.10, here the statistics of patches are not weighted by gas mass or SFR. The convergence is not strictly monotonic in all runs and there is large scatter from patch to patch. It is illustrated by shaded regions for one run; the amount of scatter is typical of all runs. In Section 2.5.2 we compare our results with the expectation of models of galactic star formation and available observational estimates.

The depletion times for gas-centered and star-centered patches do not match exactly on the largest scale, because we average the SFR over different timescales in the

two cases: 50 Myr for gas-centered patches and 10 Myr for star-centered patches. The reason for using different timescales is that young stars and molecular gas coincide little in our simulations due to strong feedback. Averaging the SFR in 50 Myr for star-centered $t_{\rm dep}$ would lead to values smaller by several orders of magnitude.

We have also checked how the estimate of the depletion time varies with time, by examining previous outputs of each run. For the weaker feedback run SFE50-SNR3, the difference on all scales is small. For the other runs, the depletion time on kpc scale generally changes by factor of a few, while on smaller scales the difference is larger. The variation of $t_{\rm dep}$ is larger for "center-on-stars" patches than for "center-on-gas" patches, because the value of SFR over 10 Myr is more stochastic, leading to larger scatter in $t_{\rm dep}$. Galaxies going through a merger show more divergent estimates of $t_{\rm dep}$ for both "center-on-gas" and "center-on-stars" versions on all scales, because of smaller overlap of SFR and molecular gas. The scatter of depletion time in previous outputs is similar to the scatter in snapshots shown in Figure 2.11.

2.4.1 Dependence of the depletion time on gas metallicity

We calculated the median values of the H_2 depletion time for patches of interest with metallicity higher and lower than the median metallicity. For almost all runs and all patch sizes, in both versions, the median $t_{\rm dep}$ of patches with high metallicity is smaller than that of patches with low metallicity. However, the whole metallicity distribution is contained to a very narrow range (the interquartile range is smaller than $0.1\,Z_{\odot}$ for all runs except SFE50-SNR3) such that the difference between the "high-metallicity" and "low-metallicity" values cannot be expected to lead to any substantial differences in physical properties of the gas. At the same time, the SFR density has a very large spread of several orders of magnitude in both cases. Therefore, we think that the dependence of $t_{\rm dep}$ on metallicity cannot be robustly determined with our data.

2.5 Discussion

2.5.1 Toomre analysis

The Toomre Q parameter has been measured in several observational studies, at high and low redshift. Genzel et al. (2011) mapped four $z \approx 2$ star-forming galaxies in gaseous Q, including a correction for multiple components. They find that H_α clumps marking young stellar systems are present at the locations of gravitationally unstable gas ($Q \lesssim 1$). However, this correspondence could result from their indirect inference of the gas density by using the same H_α flux coupled with the global star formation relation. Follow-up work by Genzel et al. (2014) find values of $Q < 1$ in the outer regions of the observed galaxies and an increase towards the center, which they associate with higher central mass concentration and larger κ.

Romeo & Wiegert (2011) and Romeo & Falstad (2013) used their Q_2 and Q_3 definitions to calculate the Toomre parameter for a sample of nearby spiral galaxies from the THINGS survey of Leroy et al. (2008). They find values of $Q \approx 2 - 5$, and no strong trend with galactocentric radius. In one third of the galaxies Q is dominated by H_2 in the inner parts and in the rest it is dominated by stars at all radii. Westfall et al. (2014) used integral field spectroscopy for 27 nearby face-on spiral galaxies from Martinsson et al. (2013) to calculate Q_{RW}, which is equivalent to Q_2 but includes corrections for disc thickness (Romeo & Wiegert, 2011). They find $Q \approx 1 - 3$, with some increase near the center due to rising κ_R. In two thirds of their galaxy sample, Q is dominated by the cold gas. Finally, Hitschfeld et al. (2009) estimate $Q \approx 2 - 4$ for the M51 galaxy, with smaller Q in spiral arms and in the outer disc. The total Q, calculated as a three-component sum assuming equal velocity dispersions, shows no obvious radial trend but the gaseous component $Q_{\rm g}$ alone increases towards the galaxy center.

In summary, observations of low-redshift galaxies indicate marginally stable discs,

with occasional collapsing regions due to spiral arms or other gravitational perturbations. In contrast, high-redshift galaxies show more unstable regions and higher star formation rates. Our simulated galaxies resemble these high-redshift observations, however, we do not find the values of Q increasing towards the galaxy center. Our distribution of Q is very patchy, and in general Q increases towards the outer parts with low gas density.

Various numerical simulations of galaxy formation have also investigated the Toomre stability criterion. For isolated disc galaxies, Li et al. (2005) used the full Rafikov (2001) definition of the Q parameter, and found that the star-formation timescale increases exponentially with Q_R. Li et al. (2006) further found an anti-correlation between the SFR and the minimum value of Q within the disc. This trend probably arises because both quantities depend on the gas density: SFR $\propto \Sigma$ and $Q_{\rm min} \propto \Sigma^{-1}$. Westfall et al. (2014) find a similar anti-correlation in their sample between the SFR surface density and $Q_{\rm min}$, although with large scatter. For our galaxies, the lowest value of Q in patches is not representative of all star formation, but we checked that the H_2 mass-weighted median value of Q does not correlate with the SFR.

In simulations of high-redshift ($z \simeq 2.3$) galaxies, Ceverino et al. (2010) calculated a two-component Q and found unstable regions in spiral arms and dense clumps. More recently, Inoue et al. (2016) calculated the Q_2 parameter for their high-redshift clumpy disc galaxies after removing the bulge and treating stars younger than 100 Myr as a gas component. They found relatively high values $Q \gtrsim 2-3$ in interclump regions and $Q < 1$ only in very dense clumps. They also found that clumps begin forming with a high value of Q, which then decreases as clumps become denser. Such stable discs could be a consequence of high mass concentration and low gas density, resulting from insufficiently strong stellar feedback. We find similarly high Q values, low gas fraction, and high SFR$\gtrsim 10\,M_\odot\,{\rm yr}^{-1}$ in our weak feedback SFR50-SNR3 run.

Analysis of the FIRE simulations presented in Oklopčić et al. (2017) has a very similar setup to ours: a 10 kpc square grid, with a 50 pc cell size smoothed to ~ 120 pc to identify stellar clumps. They calculate the gaseous parameter Q_g by approximating $\sigma = \sigma_z$ and $\kappa = \Omega$, and find that many gas clumps at the $z \approx 2$ output overlap with regions of $Q < 1$, but do not match exactly. This is similar to our results that low values of Q trace high-gas-density regions. We are also in agreement that the spatial coincidence between gas density peaks and SFR peaks washes out with increasing age of the stars.

To summarize our analysis of disc stability, we can ask: How well does the $Q < 1$ criterion work to predict the location and amount of star formation in these high-redshift galaxies? Is it a better criterion than a simple threshold on the H_2 density? We think the answer is yes.

Most of the neutral gas in our galaxies has values of Q greater than one. However, even in this case turbulent discs may be unstable to gravitational collapse on small scales, below λ_T (e.g., Romeo et al., 2010; Hoffmann & Romeo, 2012). One way to evaluate the correspondence of the Q criterion to star formation is to compare the amount of gas mass contained within a given threshold of Q and SFR.

The amount of H_2 currently contained within the contours enclosing 99% of SFR, averaged over 50 Myr, is similar to that in patches with $Q < 1$. However, when we look only at regions containing 99% of SFR within the shorter timescale of 10 Myr, the H_2 mass shows much more variation among the different runs (with the same stronger-feedback prescription): from 0.04 to 90 times the mass of stars formed in this period. Even if we select the very narrow part of the disc, calculating the column density of gas and SFR within only ± 0.2 kpc, the range of variation still extends from 0.3 to 38 times. In contrast, the H_2 mass selected by the $Q < 1$ criterion varies only between 5.5 and 220 times the young star mass. This shorter range of variation makes the Q criterion more useful for predicting future star formation.

Another way to make this comparison is to define a threshold in H_2 surface density such that the mass contained in patches above that threshold matches the mass in patches with $Q < 1$. We find that this threshold would vary from 48 to 110 $M_\odot\,\mathrm{pc}^{-2}$ for the different strong-feedback runs, that is, by more than factor of two. For HI + H_2 gas the corresponding thresholds would be even higher: $140 - 250\,M_\odot\,\mathrm{pc}^{-2}$. (For the weak-feedback run the thresholds are another order of magnitude larger.) These densities significantly exceed the prediction of Schaye (2004) model, in which transition to star-forming gas in present-day galaxies is expected to happen at $3 - 10\,M_\odot\,\mathrm{pc}^{-2}$. This may be an evidence for denser and more compact ISM in galaxies at $z \approx 1 - 2$.

2.5.2 Depletion time

Utomo et al. (2017) and Colombo et al. (2018) measured the depletion time of molecular gas on kpc scale in the EDGE-CALIFA survey of nearby galaxies. They found $t_{\mathrm{dep}} \approx 2.4$ Gyr, with large scatter of about 0.5 dex. The depletion time decreases near the center in some of the galaxies, especially at lowest masses $M_* \lesssim 10^{10}\,M_\odot$.

For high-redshift galaxies the depletion time of molecular gas appears to be shorter. According to Genzel et al. (2010), the depletion time for normal star-forming galaxies decreases from 1.5 Gyr at $z \approx 0$ to 0.5 Gyr at $z \approx 2$. Tacconi et al. (2010, 2013) measured $t_{\mathrm{dep}} \approx 0.7$ Gyr, with a dispersion of 0.24 dex, in a survey of $z \approx 1 - 3$ galaxies with $M_* > 2.5 \times 10^{10}\,M_\odot$ and SFR$\gtrsim 30\,M_\odot\,\mathrm{yr}^{-1}$. However, it is important to note the large scatter associated with this mean trend – a small selection of galaxies may deviate significantly. For example, Tadaki et al. (2018) find a very short depletion time of 10^8 yr in a sub-mm starburst galaxy at $z = 4.3$. This galaxy shows giant kpc-scale molecular gas clumps with low $Q \approx 0.3$, due to the very high gas density. Forrest et al. (2018) also find spread of two orders of magnitude for the specific SFR of $1 < z < 4$ galaxies in the large ZFOURGE survey.

Our simulated galaxies, except in the weaker feedback SFE50-SNR3 run and the

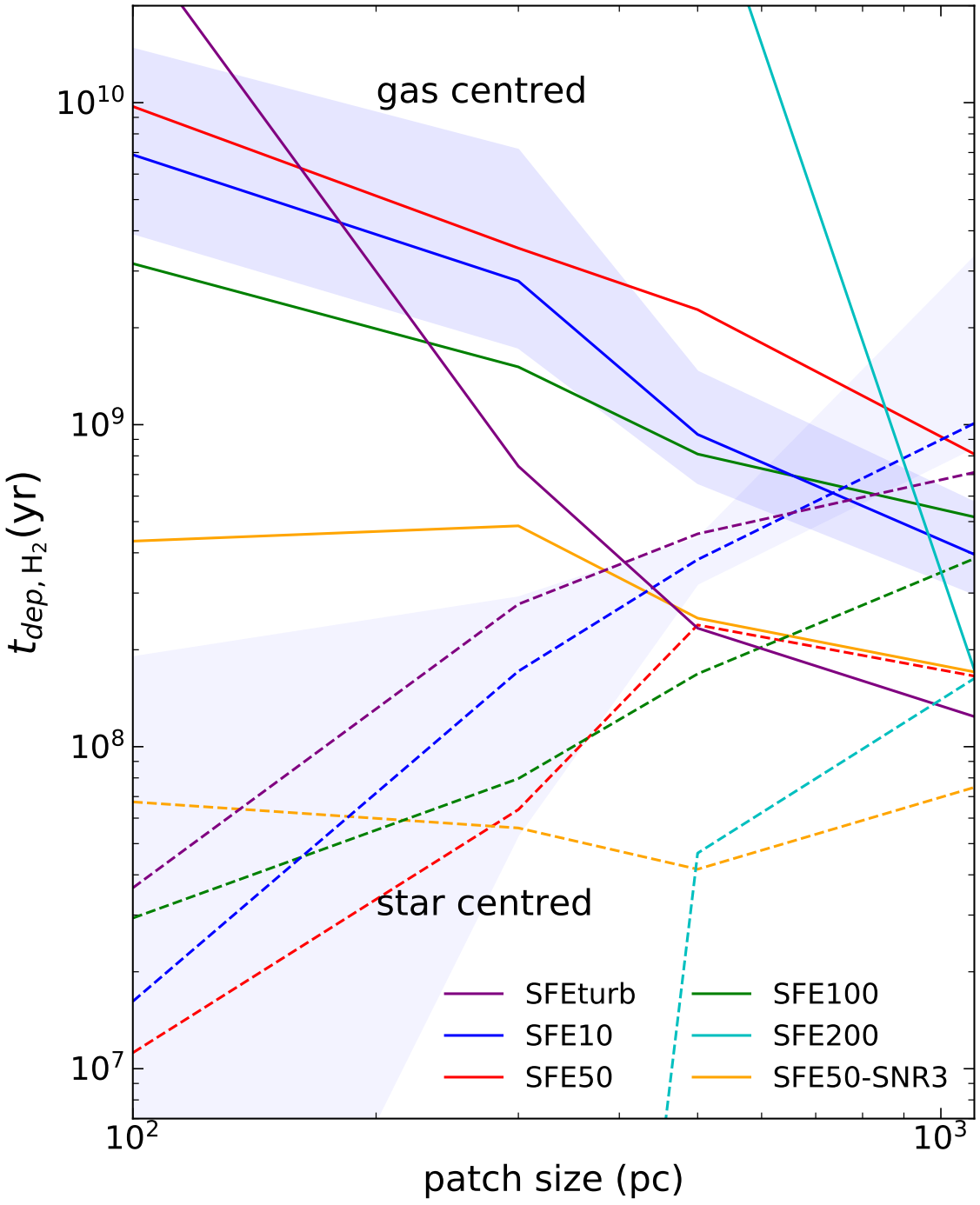

Figure 2.11 Depletion time of H_2 for "centered-on-gas" (solid lines) and "centered-on-stars" versions (dashed lines) as a function of patch size. For "centered-on-gas" version, we take the median of $t_{\rm dep}$ in patches with $M_{\rm H_2} > 10^6\,M_\odot$. For "centered-on-stars" version, we take the median of $t_{\rm dep}$ in patches with non-zero SFR. Shaded regions show the 40%-60% range of the values of $t_{\rm dep}$ for SFE10 run, to illustrate the typical wide spread of the distribution.

higher-redshift SFEturb output, show comparable values of $t_{\rm dep}$ to the observations of $z \approx 1-2$ galaxies. When averaged in cylindrical shells, the depletion time roughly follows the trend of decreasing towards the center but given the strong patch-to-patch variation we do not study it further.

Recent observations by Rebolledo et al. (2015) and Leroy et al. (2017) have been able to probe individual star-forming regions on scales below 100 pc in nearby galaxies. They reveal a wide scatter of SFR density by three orders of magnitude at H_2 densities $10-10^3\,M_\odot\,{\rm pc}^{-2}$. Our simulations show a correspondingly large scatter in the local depletion times. Taking the patches above our adopted threshold on the H_2 mass, which corresponds to the surface density of $10^6\,M_\odot/(200\,{\rm pc})^2 = 25\,M_\odot\,{\rm pc}^{-2}$, we find few individual patches with $t_{\rm dep} < 10^8$ yr in the stronger-feedback runs. This corresponds to the lowest bound derived by Rebolledo et al. (2015). We find some patches with $t_{\rm dep} > 10^{10}$ yr, which fall above the observed upper limit. However, such regions are more likely to escape detection because of their lower SFR. There is no systematic trend with $\epsilon_{\rm ff}$ used in the simulation. Also, as with most of our results, the weaker-feedback run is an exception, as it contains patches with the depletion times as short as 10^7 yr.

The high-redshift galaxies appear to have a high molecular gas fraction: Tacconi et al. (2013) measured $f_{\rm mol} = M_{\rm mol}/(M_{\rm mol}+M_*) \sim 50\%$. After correcting for incompleteness, $f_{\rm mol}$ lowers to $30-40\%$. That fraction is larger than what we find for our simulated galaxies (about 15%), even though our galaxies are less massive and therefore should be more gas-rich. Daddi et al. (2010) also found the molecular gas fraction of galaxies at $z \approx 1.5$ to be $\sim 50-65\%$. At very high redshift, the molecular gas fraction is even larger: Dessauges-Zavadsky et al. (2017a) measured $f_{\rm mol} \approx 60-79\%$ for a lensed $M_* \sim 5\times10^9\,M_\odot$ galaxy at $z \approx 3.6$. In our galaxies the fraction of neutral gas reaches about 50%, but the fraction of molecular gas stays low regardless of the value of local star formation efficiency adopted in the simulation

(see Table 2.1).

Theoretical models predict the scaling of the depletion time with properties of star-forming regions. Semenov et al. (2017) constructed an analytical model based on the mass conservation and the physical picture of rapid gas evolution between star-forming and non-star-forming states to study $t_{\rm dep}$ and the fraction of gas that participates in star formation. They tested this model with a suite of L_*-sized galaxy simulations (Semenov et al., 2018) with different values of $\epsilon_{\rm ff}$, feedback strength $f_{\rm boost}$ (b in their notation), and star formation threshold. According to their model, gas regulation in galaxies is divided into two regimes: the self-regulation regime where feedback is strong or $\epsilon_{\rm ff}$ is large enough, and the dynamics-regulation regime where feedback is weak or $\epsilon_{\rm ff}$ is small. In the dynamics-regulation regime, the supply of star-forming gas is balanced by dispersal due to dynamical processes such as turbulent shear, differential rotation, etc. The depletion time is inversely proportional to $\epsilon_{\rm ff}$, and the star-forming gas fraction is insensitive to $\epsilon_{\rm ff}$ or feedback strength. In the self-regulation regime, gas spends most of the time in non-star-forming stages, and gas regulation is mainly controlled by star formation and feedback. The depletion time scales with feedback strength, but is insensitive to $\epsilon_{\rm ff}$. The star-forming gas fraction is small and scales inversely with $f_{\rm boost}$ and $\epsilon_{\rm ff}$. Although the model is formulated for the depletion time of all gas on kpc scales, the behavior of $t_{\rm dep,H_2}$ and H_2 fraction with different $\epsilon_{\rm ff}$ and feedback strength is similar.

Our simulated galaxies fall in the self-regulation regime. Consistent with the Semenov et al. (2017) model, our $t_{\rm dep,H_2}$ on the largest scale shows a slightly decreasing trend with $\epsilon_{\rm ff}$, although non-monotonic and with large scatter. It may be mainly due to the fraction of gas in molecular phase, $M_{\rm H_2}/M_{\rm HI+H_2}$, generally falling with $\epsilon_{\rm ff}$. The weaker feedback run has slightly smaller $t_{\rm dep,H_2}$ and higher H_2 fraction than the other runs, which is also consistent with their model. The scale dependence of $t_{\rm dep}$ is also similar, but our results are less regular and show significant scatter.

Galactic star formation relations are expected to break down below a certain spatial scale due to incomplete sampling of star formation or gas tracers, and relative motion of dense gas and young stars because of stellar feedback. The scatter of star formation relations, coming from the discreteness and stochasticity of star formation, and drifting of young stars, becomes more significant on smaller scales (Feldmann et al., 2011, 2012). Kruijssen & Longmore (2014) constructed a model to describe this breakdown of SF relations based on the concept that a galaxy consists of many independent star-forming regions separated by some length scale, and that these regions are going through the SF process during which gas and/or young stars can be observed by some tracer. The timescale of the whole SF process is a combination of the epoch when gas is visible ($t_{\rm gas}$) and when stars are visible ($t_{\rm star}$), with some overlap time ($t_{\rm over}$). The scatter in $t_{\rm dep}$ increases from ~ 0.1 dex on kpc scale to ~ 1 dex on tens of pc scale for randomly positioned apertures. Centering apertures on gas or stellar peaks systematically biases $t_{\rm dep}$ – centering on gas peaks overestimates $t_{\rm dep}$ and centering on stellar peaks underestimates $t_{\rm dep}$ – making the "tuning fork" diagram, as shown in our Figure 2.11. The relative durations of the various phases of SF process ultimately determine the excess or deficit of $t_{\rm dep}$ on small scales.

Kruijssen et al. (2018) provide a detailed method to reconstruct the timescales of star formation and feedback ($t_{\rm gas}$, $t_{\rm star}$, $t_{\rm over}$) from the maps of gas and stellar flux. Using their method requires that the SFR for both gas-centered and star-centered apertures is averaged over the same time span. We find the ratio of the gas-centered to star-centered $t_{\rm dep}$ about a factor of 100 at the smallest scale of 100 pc, with a very large variation between the different runs; these numbers exceed even the largest expected spread shown in Kruijssen et al. (2018). From our Figure 2.11 we can at least see that our $t_{\rm gas}$ is much larger than $t_{\rm star}$ (taken to be 10 Myr here). From Figure 2.6 we can infer that our $t_{\rm over}$ is definitely smaller than 50 Myr, and probably close to 10 Myr, which explains why we have so many infinities in the gas-centered

determination of $t_{\rm dep}$. The depletion time is so short on small scales when centered on stars that dense gas does not coincide with young stars, and causes formally infinite $t_{\rm dep}$.

2.6 Summary

We have investigated the structure of high-redshift ($z \approx 1.5 - 2$) galaxies in a suite of cosmological simulations with different star formation efficiency and feedback strength. Our main results are summarized below:

- Unlike the regular appearance of low-redshift disc galaxies, the galaxies in our simulations have thick stellar components with irregular, prolate shapes. The kinematics are dominated by turbulent motions and not by rotation. The stellar surface density profiles are approximately exponential, with the scale length of about 1 kpc.

- Although the vertical scale heights for all gas and all stars are large, cold molecular gas is concentrated to a relatively thin plane. Young stars, which form from the molecular gas, likewise have the distribution with axis ratios $c/a = 0.1-0.2$.

- Spatial correlation between the peaks of gas density and SFR deteriorates with the age of stellar population and almost disappears after ~ 50 Myr, because of stellar feedback dispersing old gas clouds around star-forming regions.

- We calculate the maps of Toomre Q parameter in patches of 200 pc, combining three components with different velocity dispersions: stars, molecular gas, and atomic gas. The median value of Q weighted by H_2 mass is in the range $0.5-1$, surprisingly close to unity given the irregular structure of the galaxies.

- The median value weighted by neural HI + H_2 mass is higher: $Q \approx 1.5 - 2.6$. The Q parameter in the weaker feedback run SFE50-SNR3 is systematically

larger than in the other runs, because of the low gas density and high central mass concentration.

- The dynamic range of Q maps is much smaller than that of the H_2 surface density maps, making the Toomre Q parameter a better indicator of unstable regions that would collapse and form stars. The Q parameter also depends on the spatial scale over which it is calculated: enlarging the averaging scale increases the value of Q.

- The depletion time of molecular gas in our galaxies is around 1 Gyr on the kpc scale, with large scatter from run to run. On smaller scales, $t_{\rm dep}$ splits to systematically larger or smaller values when centering the aperture on gas peaks or stellar peaks, respectively.

CHAPTER III

Origin of giant stellar clumps in high-redshift galaxies

This chapter was published as: ***Meng, X.*** *& Gnedin, O. Y. 2020, MNRAS, 494, 1263*

Abstract

We examine the nature of kpc-scale clumps seen in high-redshift galaxies using a suite of cosmological simulations of galaxy formation. We identify rest-frame UV clumps in mock HST images smoothed to 500 pc resolution, and compare them with the intrinsic 3D clumps of young stars identified in the simulations with 100 pc resolution. According to this comparison for the progenitors of Milky Way-sized galaxies probed by our simulations, we expect that the stellar masses of the observed clumps are overestimated by as much as an order of magnitude, and that the sizes of these clumps are also overestimated by factor of several, due to a combination of spatial resolution and projection. The masses of young stars contributing most of the UV emission can also be overestimated by factor of a few. We find that most clumps of young stars present in a simulation at one time dissolve on a timescale shorter than $\sim$150 Myr. Some clumps with dense cores can last longer but eventually disperse.

Most of the clumps are not bound structures, with virial parameter $\alpha_{\rm vir} > 1$. We find similar results for clumps identified in mock maps of Hα emission measure. We examine the predictions for effective clump sizes from the linear theory of gravitational perturbations and conclude that they are inconsistent with being formed by global disc instabilities. Instead, the observed clumps represent random projections of multiple compact star-forming regions.

3.1 Introduction

While most nearby L* galaxies present disc-like morphologies, deep observations with the *Hubble Space Telescope* (HST) reveal irregular and clumpy shapes of high-redshift galaxies at rest-frame ultraviolet (UV) and optical wavelengths (Elmegreen et al., 2007; Overzier et al., 2010; Swinbank et al., 2010a). Integral field spectroscopic surveys of these galaxies show both rotation and turbulent motions (Förster Schreiber et al., 2006; Shapiro et al., 2008; Wisnioski et al., 2015). The fraction of galaxies that are clumpy at rest-frame UV evolves with time and varies with galaxy mass. The clumpy fraction for star-forming galaxies increases from $z \simeq 8$ to $z \simeq 1-3$, reaches a peak, and subsequently decreases until $z \simeq 0$ (Murata et al., 2014; Guo et al., 2015; Shibuya et al., 2016). This trend is similar to the evolution of the cosmic star formation rate (SFR) density (Madau & Dickinson, 2014). Moreover, the clumpy fraction tends to increase with SFR at $z \simeq 0-2$ (Shibuya et al., 2016), suggesting a correlation with star formation activity. The clumpy fraction decreases with stellar mass in galaxies at $z \simeq 0.8-2$ (Tadaki et al., 2014; Guo et al., 2015).

Some of the earliest found clumpy galaxies were characterized as "chain", "tadpole", and "clump cluster" galaxies (Elmegreen et al., 2004b, 2005). There are typically 2-8 clumps per galaxy, with estimated stellar mass $\sim 10^7 - 10^9\,M_\odot$ (Förster Schreiber et al., 2011; Guo et al., 2012; Soto et al., 2017). The clumps typically have high SFR, resembling mini-starbursts in their galaxies (Bournaud et al., 2015;

Zanella et al., 2015). While the clumps contribute only a few percent individually and ∼20% altogether to their host galaxy's total luminosity, their contribution to SFR is larger, ∼10-50% (Förster Schreiber et al., 2011; Guo et al., 2012; Wuyts et al., 2012). The SFR of individual clumps varies from $10^{-1}\,M_{\odot}\,\mathrm{yr}^{-1}$ (Soto et al., 2017) to $1-10\,M_{\odot}\,\mathrm{yr}^{-1}$ (Guo et al., 2012). The inferred ages of the clump stars range from 10^{6} to $10^{10}\,\mathrm{yr}$ (Soto et al., 2017). There is also wavelength dependence: clumps identified at different wavebands do not fully overlap (Förster Schreiber et al., 2011).

The clumpy structure is also observed in Hα (e.g. Livermore et al., 2012; Mieda et al., 2016) and CO maps (e.g. Swinbank et al., 2010b; Dessauges-Zavadsky et al., 2017b). Kinematic studies of high-redshift massive clumpy galaxies ($M \sim 10^{10.6}\,M_{\odot}$) in H$\alpha$ emission using SINFONI/VLT show that these galaxies are turbulent and rotation dominated. Some galaxies have a massive stellar bulge (Genzel et al., 2008, 2011).

Typical sizes of clumps in HST images ∼1 kpc are at the limit of angular resolution at high redshift. Gravitational lensing has afforded us a magnified view of these galaxies. Adamo et al. (2013) identified 30 clumps of $10^{6}-10^{9}\,M_{\odot}$ in a lensed spiral galaxy Sp 1149 at redshift 1.5 with spatial resolution ∼100 pc. Girard et al. (2018) observed a lensed rotating galaxy at $z = 1.59$ and identified three Hα clumps, which together contribute ∼40% of total SFR inferred from the Hα flux. The SFR density in these clumps is ∼100 times higher than in nearby HII regions. Livermore et al. (2012) obtained the luminosity function of clumps with median source plane spatial resolution ∼360 pc and compared it with the luminosity function of HII regions in galaxies at $z \approx$1-1.5. They conclude that high-redshift clumps are HII regions that are larger and brighter than local HII regions. The clump sizes in lensed galaxies are smaller than those in unlensed galaxies. Jones et al. (2010) found clumps with diameter 300 pc-1 kpc in lensed galaxies at z=1.7-3.1 with spatial resolution achieving ∼100 pc. Livermore et al. (2015) extracted 50 star-forming Hα and Hβ clumps with

sizes in the range 60 pc - 1 kpc in 17 lensed galaxies at $1< z <4$. Wuyts et al. (2014) found clumps of diameter ∼300-600 pc in a highly magnified lensed galaxy at $z = 1.70$ and found a radial gradient of their rest-frame UV colour. Johnson et al. (2017) found star-forming clumps of radius smaller than 100 pc in a lensed galaxy at $z = 2.5$. Olmstead et al. (2014) quantified relative stellar-to-nebular extinction in two z=0.91 galaxies with ∼0.3 kpc resolution. They found that the integrated extinction measurements agree with other studies in that the ionized gas is more obscured than stars. However, when examining on a clump-by-clump basis, they show that the hypothesis that stars and ionized gas experience identical extinction cannot be ruled out.

Unlike the clumpy structure observed in rest-frame UV, ALMA observations of cold dust of massive ($M_* \sim 10^{11}\,M_\odot$) star forming galaxies at $z \sim 3$ with 200 pc resolution show smooth, disc-like morphology (Rujopakarn et al., 2019).

The puzzling appearance of giant clumps has inspired many theoretical studies that have investigated the formation and evolution of these high-redshift clumps. They include isolated disk simulations (e.g. Tamburello et al., 2015; Inoue & Yoshida, 2018) and cosmological zoom-in simulations (e.g. Ceverino et al., 2010; Oklopčić et al., 2017). The clumpy fraction in simulations decreases from high redshift ($z \approx 2$) to low redshift (e.g. Buck et al., 2017; Mandelker et al., 2017). These studies identify clumps using a variety of methods: in projected gas density maps (e.g. Oklopčić et al., 2017; Benincasa et al., 2019), projected stellar density maps (e.g. Mayer et al., 2016), mock observational maps (e.g. Hopkins et al., 2012; Moody et al., 2014; Buck et al., 2017), in 3D gas or stellar density distributions (e.g. Mandelker et al., 2014, 2017), or as gravitationally bound objects (e.g. Tamburello et al., 2015; Benincasa et al., 2019). Clumps identified in one type of maps do not necessarily correspond to clumps found in other maps (Moody et al., 2014). In stellar maps, galaxies are only clumpy in UV light, but not in projected stellar mass density (Buck et al., 2017).

The origin of these high-redshift clumps is yet unclear. There are two general scenarios for clump formation: one in which clumps grow through gravitational instability within galactic discs, the other in which clumps are caused by external perturbations, such as mergers. The internal scenario, including violent disc instability and spiral arm instability, is supported by many simulations (e.g. Ceverino et al., 2010; Genel et al., 2012; Inoue & Yoshida, 2018) and observations (e.g. Elmegreen et al., 2007; Genzel et al., 2008; Guo et al., 2012; Zanella et al., 2015). Studies of nearby turbulent disc galaxies that resemble the high-redshift clumpy galaxies (Fisher et al., 2017b) show that the clump sizes are consistent with the results of instabilities in self-gravitating gas-rich discs (Fisher et al., 2017a). If clumps form from fragmentation driven by turbulence, the clump stellar mass function should follow a power law of slope close to -2. Dessauges-Zavadsky & Adamo (2018) derived the mass function of star-forming clumps at $z \sim 1 - 3.5$ and found the power-law slope ≈ -1.7 at $M_* > 2 \times 10^7\, M_\odot$, in agreement with the turbulence-driven scenario.

On the other hand, some simulation results indicate that *ex situ* mergers contribute to at least a portion of the clumps (Mandelker et al., 2017), and that massive clumps could form from minor galactic mergers (Mandelker et al., 2014) and clump-clump mergers (Tamburello et al., 2015). Observations of merging galaxies appear to support such merger-driven clump formation (Puech et al., 2009; Puech, 2010; Guo et al., 2015; Ribeiro et al., 2017).

The final fate of the clumps is also currently under debate. One alternative is that these clumps would migrate to the galactic center and potentially contribute to the galactic bulge (Ceverino et al., 2010; Inoue & Saitoh, 2012; Perez et al., 2013; Bournaud et al., 2014). Evidence for this scenario is built on the observed radial gradient of clump's colour or age, such that clumps closer to the galaxy center are older (Noguchi, 1999; Genzel et al., 2008; Adamo et al., 2013). This scenario requires clump lifetime to be longer than a few orbital times. Another alternative is that

these clumps dissolve in a relatively short time and may contribute to the thick disc, whereas the colour gradient could instead be a result of the inside-out disc growth (Murray et al., 2010; Genel et al., 2012; Hopkins et al., 2012; Buck et al., 2017; Oklopčić et al., 2017). Some studies advocate both scenarios: low-mass clumps get disrupted in a short time, while more massive clumps survive and migrate to the center (Genzel et al., 2011; Mandelker et al., 2017).

The interpretation of clump origins is complicated by possible overestimation of the clump masses and sizes due to limited angular resolution and sensitivity (e.g., Tamburello et al., 2015, 2017; Cava et al., 2018). The observed kpc-scale clumps may also be clusters of clumps or blending of smaller structures (Behrendt et al., 2016). Due to limited sensitivity, the observed clumps may be biased against low-mass structures (Dessauges-Zavadsky et al., 2017a). These effects need to be taken into consideration when measuring clump properties.

In this work, we revisit the nature of giant clumps using our state-of-the-art simulations of galaxy formation. We explore how resolution and sensitivity affect the inferred clump mass and size, and investigate the final state of these clumps. Our high-resolution cosmological simulations include novel and most realistic modeling of star formation and stellar feedback, which allows us to produce galaxy structures that closely resemble observed high-redshift galaxies. In Section 3.2 we describe our simulations and clump identification method, both in 3D and in 2D projection. We present the properties of real 3D clumps and mock 2D clumps in our simulations in Section 5.3. In Section 5.4 we discuss how simulated clumps compare with the observations and other simulations. We present our conclusions in Section 5.5.

3.2 Identification of clumps

3.2.1 Simulation suite

We use a suite of cosmological simulations performed with the Adaptive Refinement Tree (ART) code (Kravtsov et al., 1997; Kravtsov, 1999, 2003; Rudd et al., 2008) and described in Li et al. (2018b) and Meng et al. (2019). All runs start with the same initial conditions in a periodic box of 4 comoving Mpc, so that the main halo has total mass $M_{200} \approx 10^{12}\,M_{\odot}$ at $z = 0$, similar to that of the Milky Way. The ART code uses adaptive mesh refinement to increase spatial resolution in dense regions. There are 128^3 root grid cells, setting the dark matter particle mass $m_{\rm DM} \approx 10^6\,M_{\odot}$. The finest refinement level is adjusted in runtime to keep the physical size of gas cells at that level between 3 and 6 pc. Because of strong stellar feedback, few cells reach the finest refinement level and the typical spatial resolution of molecular gas is 36–63 pc. We calculated the halo spin parameter $\lambda = J/(\sqrt{2}MVR)$ (Bullock et al., 2001) for our dark matter halos, where J is the total angular momentum, M is the halo mass, V and R are virial velocity and virial radius of the dark matter halo. The halo spin parameter varies in the range of $\sim$0.01-0.05, depending on the redshift. Here we list the halo spin parameters of our dark matter halos at the analyzed snapshots: $\lambda \sim$0.05 for SFE50 and SFE100 runs at z=1.50, $\lambda \sim$0.01 for SFE200 and SFE10 at z=1.78, and $\lambda \sim$0.04 for SFEturb run at z=1.98.

The simulations include three-dimensional radiative transfer (Gnedin & Abel, 2001) of ionizing and UV radiation from stars (Gnedin, 2014) and extragalactic UV background (Haardt & Madau, 2001), non-equilibrium chemical network of ionization states of hydrogen and helium, and phenomenological molecular hydrogen formation and destruction (Gnedin & Kravtsov, 2011). The simulations incorporate a subgrid-scale (SGS) model for unresolved gas turbulence (Schmidt et al., 2014; Semenov et al., 2016b). Star formation is implemented with the continuous cluster formation (CCF)

algorithm (Li et al., 2017b) where each star particle represents a star cluster that forms at a local density peak and grows mass via accretion of gas until the feedback from its own young stars terminates the accretion. The feedback recipe includes early radiative and stellar wind feedback, as well as a supernova (SN) remnant feedback model (Martizzi et al., 2015; Semenov et al., 2016b). The momentum feedback of the SN remnant model is boosted by a factor $f_{\rm boost} = 5$ to compensate for numerical underestimation and to match the star formation history expected from abundance matching. To explore the variation of results with the speed of star formation, we ran a suite of simulations with different value of the local star formation efficiency (SFE) per free-fall time, $\epsilon_{\rm ff}$. For full description of star formation and feedback recipe, see Li et al. (2017b, 2018b).

Our simulation suite with $f_{\rm boost} = 5$ produces the star formation history (SFH) expected for $10^{12}\,M_{\odot}$ halos from abundance matching. The initial distribution of star clusters also matches observations of young clusters in nearby galaxies. The cluster initial mass function can be described by a Schechter function, the slope of which is close to the observed value -2 for $\epsilon_{\rm ff} = 0.5 - 1.0$. The fraction of clustered star formation correlates with the SFR density, consistent with observations for $\epsilon_{\rm ff} = 0.5 - 2.0$. The formation timescales of clusters in runs with $\epsilon_{\rm ff} \geqslant 0.1$ are shorter than 3 Myr, within the range of the observed age spread of young star clusters.

In this paper we analyze several runs with different value of $\epsilon_{\rm ff}$. The number after "SFE" in the run names corresponds to the percentage of local $\epsilon_{\rm ff}$. In SFEturb run $\epsilon_{\rm ff}$ is variable and turbulence-dependent (as implemented by Semenov et al. 2016b). The typical values are 3%, with a log-normal scatter of about 0.3 dex (Li et al., 2018b). For each run we focus on the main galaxy in the simulation box. In all the runs the galaxies have a similar star formation history, reproducing the expectation of abundance matching. The galaxies also have similar axisymmetric stellar surface brightness profile (Meng et al., 2019). The SFE has systematic effects on small

spatial and temporal scales: higher SFE leads to shorter formation timescales for star clusters and more concentrated stellar feedback. This results in more bursty star formation rate and lower fraction of star-forming gas in galaxies with higher SFE. Our simulated galaxies have stellar masses $(2.4-8.2)\times10^9\,M_\odot$ and neutral gas mass $(2.7-8.4)\times10^9\,M_\odot$ at the analyzed snapshots ($z\approx1.5-3$, depending on the run). Other global properties of the simulated galaxies are listed in our previous paper (Meng et al., 2019).

3.2.2 Identification criteria

To match the procedures for finding clumps in rest-frame UV images of high-redshift galaxies, we made mock observations of our simulated galaxies and identified clumps in the mock images. We use the *Flexible Stellar Population Synthesis* (FSPS) model (Conroy et al., 2009; Conroy & Gunn, 2010) to generate spectral energy distributions of all star particles and shift them to their corresponding redshifts in the simulation outputs.

To focus on rest-frame near-UV (∼2800Å) light, Guo et al. (2015) detected clumps in HST/ACS F435W band for the galaxy redshift range $0.5\leqslant z<1.0$, F606W band for $1.0\leqslant z<2.0$, and F775W band for $2.0\leqslant z<3.0$. Note that the CANDELS galaxies used in Guo et al. (2018) have stellar masses $10^9-10^{11}\,M_\odot$ at z=1.5-3, covering the range of our galaxy masses but extending to higher values.We follow their choice and calculate the simulated galaxy fluxes in the latter two bands, because our outputs lie in the range $1<z<3$. First we bin the mock observational images in two-dimensional (2D) patches of 100 pc side (physical, not comoving) to avoid uneven resolution of cells at different refinement level. Then we degrade the images to patches of 500 pc to approximate the HST resolution at these redshifts. Galaxy orientation is chosen to be face-on, where the galaxy plane is determined using the shape tensor of neutral gas (Meng et al., 2019).

We do not take into account the effect of dust because dust is not modeled explicitly in our simulations. Including dust attenuation could change the degree of clumpiness of the mock images, suppressing intrinsically bright clumps and enhancing less luminous ones (Buck et al., 2017).

We use the `python` package `astrodendro`[1] to identify clumps in our mock observations. This package computes dendrograms, which are tree diagrams particularly useful for identifying hierarchical structures. A dendrogram contains *leaves*, which have the highest values and no substructure, and combines them into *branches*, which merge hierarchically into the largest branch, the tree *trunk*. We consider only leaves in the dendrogram of mock observational maps, since structures larger than leaves contain more than one clump. We focus our analysis on the brightest clumps, by setting a lower limit on the fractional contribution of each clump to the galaxy UV luminosity: $f_{\rm LUV} \equiv L^{\rm UV}_{\rm clump}/L^{\rm UV}_{\rm galaxy} > 3\%$, following Guo et al. (2018).

There are three parameters in the `astrodendro` setup: `min-value`, which is the minimum value of the surface brightness for a structure to be identified; `min-delta`, which is the minimum difference between adjacent structures; and `min-npix`, which is the minimum number of pixels required to form a structure. Experimentation showed that the results depend largely on `min-npix`, but not the other two parameters. We set the value of `min-npix` according to the expected observational resolution, but then vary this parameter to find the sensitivity of the results.

We set the parameter `min-value` to be the typical surface brightness in the outer parts of the galaxy, about 26 mag arcsec^{-2}, to include most of it in the dendrogram tree. Raising `min-value` would make the smallest clumps disappear, most of which are excluded from our sample anyway due to the $f_{\rm LUV}$ cut. It would also make some of the largest clumps to be restricted only to their brightest parts.

The parameter `min-delta` determines whether a peak is considered a clump or

[1] `http://dendrograms.org/`

noise. We found `min-delta` not as important as `min-npix` in clump identification, as long as it is smaller than the mean difference among pixels that are in the region of clumps. We set the value of `min-delta` at 0.5 mag arcsec^{-2}.

Since high-redshift galaxies do not have well-defined thin discs, intrinsic clumps could be overlapping with each other, even when the galaxies are viewed face-on. Identifying clumps in 3D uncovers this projection effect. We also identify clumps in three dimensions (3D), using dendrograms of stellar 3D density. We calculate the mass density of all stars and of stars younger than 100 Myr on a uniform 3D grid of cubes of 100 pc side. The Z-direction of this uniform 3D grid is perpendicular to the galaxy plane, so that a column of 3D cubes coincides with one 2D patch. This way we can associate each 2D clump with one or several 3D clumps in the same projected area. The 3D dendrogram builds a similar tree-like structure starting from regions of highest density. We identify 3D clumps using the density of young stars because young stars are the main contributor to UV luminosity, and the distribution of old stars does not appear clumpy.

The choice of parameter `min-value` and `min-delta` for 3D clump identification is similar to that in 2D. We choose the parameter `min-value` by examining the 3D light and young stellar density profile of the faintest clumps and take the value at the boundary where the profile dissolves into the background. `min-delta` is set to be small enough that we do not artificially combine peaks due to this parameter. We take `min-value`=$10^{-1.6}\,M_{\odot}\mathrm{pc}^{-3}$ and `min-delta`=0.2 dex for 3D clump identification.

3.3 Results

3.3.1 Clump properties in simulations

Using the clump finding method described above, we identified clumps in 2D surface brightness maps and in 3D density grid of young stars. In a given snapshot

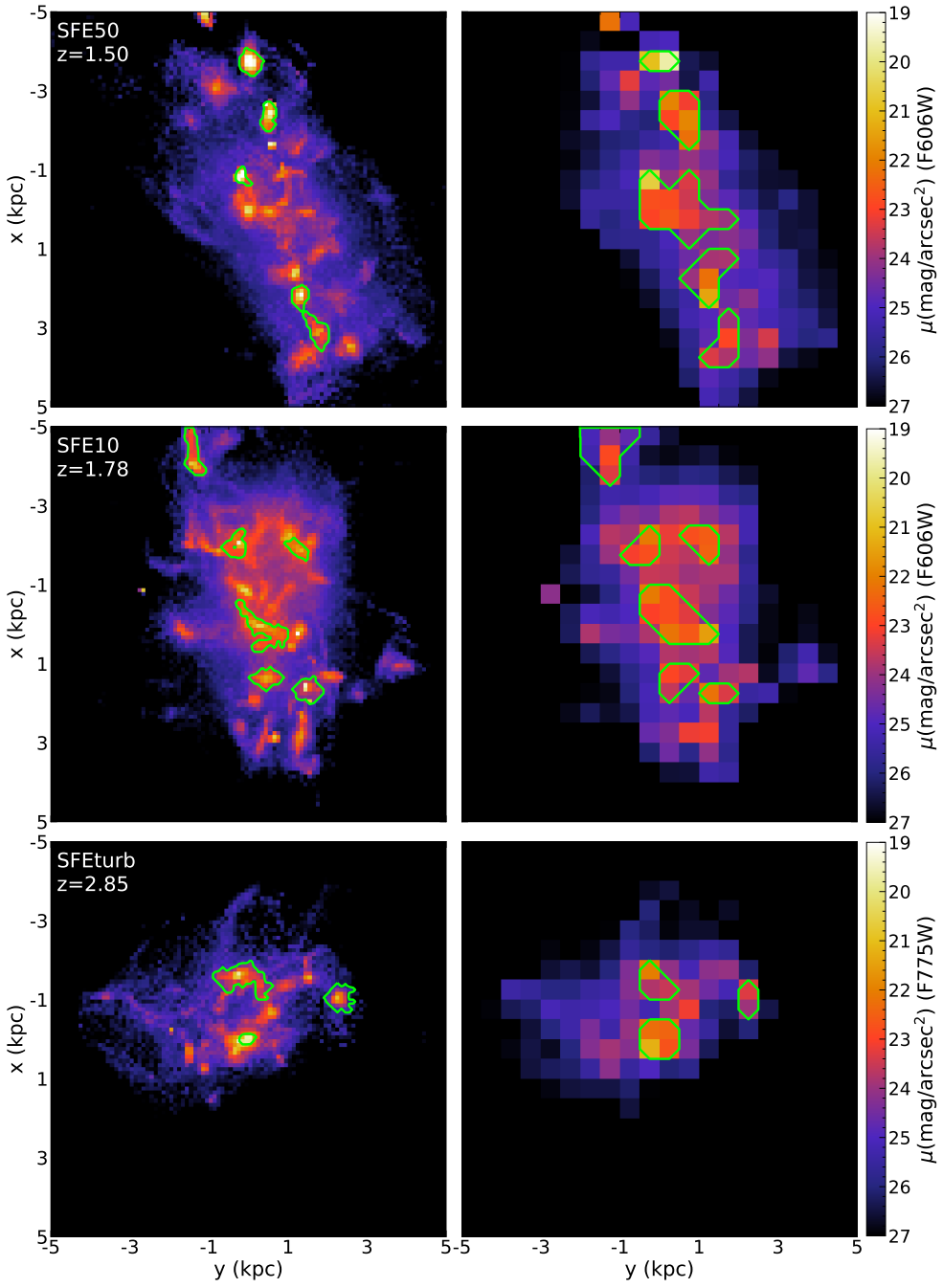

Figure 3.1 Comparison of 2D and 3D clumps in mock images of simulated galaxies. Colour maps show the surface brightness in HST/ACS F606W band (except lower panels for SFEturb which are in the F775W band because of higher redshift of the simulation output) and do not include internal extinction. Left panels have pixels of 100 pc, right panels are degraded to HST resolution (500 pc) at these redshifts. Contours in the right panels show 2D clumps identified in the degraded maps using `astrodendro` with `min-npix`=2. Only clumps with fractional luminosity $> 3\%$ are shown. Contours in the left panels show the most luminous 3D clump of young stars within the projected area of the corresponding 2D clump, with `min-npix`=14.

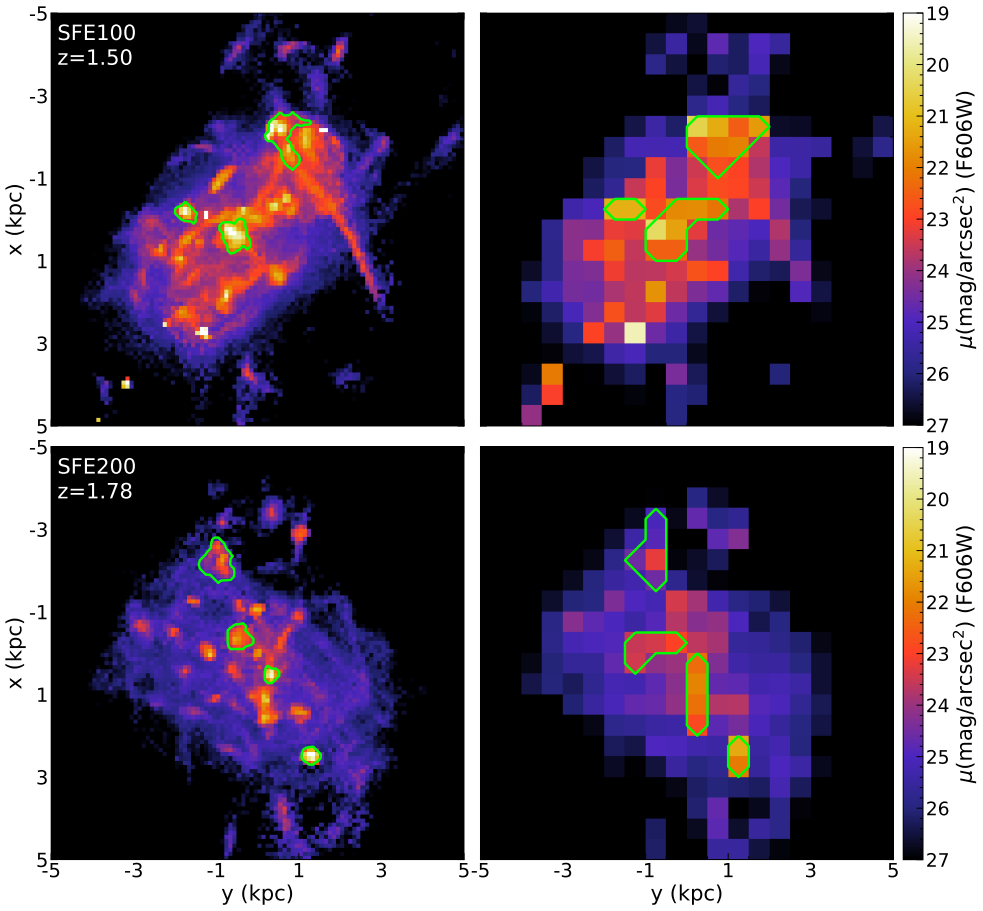

Figure 3.1 (continued)

there are several 2D clumps in the main galaxy, and 10–40 3D clumps, depending on the parameter `min-npix`.

To explore the effects of parameter choice, we tried different values of `min-npix` for the identification of 3D clumps. We aim to find separate 3D clumps of young stars that are large enough to account for most of the UV luminosity from the corresponding projected 2D clump. Our 3D patch size is fixed at 100 pc, so that larger `min-npix` means larger minimum required clump volume. We concluded that `min-npix`=14 is a good choice for 3D identification: lower value of `min-npix` would only produce more small 3D clumps, while higher value of `min-npix` would artificially lump together clumps that are already large enough. Also, if we use larger value of `min-npix`, we would miss one very small and compact clump in SFE50 that accounts for 48% of the rest-frame UV luminosity of its 2D counterpart. Based on these tests, we think `min-npix`=14 gives the most stable result for 3D clump finding.

For 2D clump identification in degraded mock observations with patches of 500 pc, we take `min-npix`=2, since it best matches the 2D clumps identified by eye. This choice also corresponds to the roughly kpc scale of the observed clumps.

Therefore, we adopt the values `min-npix`=2 for 2D and `min-npix`=14 for 3D clump finding. This choice implies a minimum size of the identified clumps. We define the effective clump size as the radius of a circle (in 2D) or a sphere (in 3D) with the same area/volume as the sum of patches in the clump, via `Clump Area` $\equiv \pi R_{\rm 2D}^2$ and `Clump Volume` $\equiv \frac{4\pi}{3} R_{\rm 3D}^3$. Our choice of `min-npix` corresponds to the minimum effective radius of 400 pc in 2D and 150 pc in 3D. Note that Oklopčić et al. (2017) uses similar minimum effective radius (125 pc) for 2D clump identification in FIRE simulations.

Figure 3.1 shows the identified 2D clumps and 3D clumps in five simulation runs with different SFE. Colour maps show the surface brightness of each galaxy in HST/ACS F606W (for $1 \leq z < 2$) and F775W (for $2 \leq z < 3$) bands. The spatial

resolution for the left panels is a typical simulation patch of 100 pc, while the right panels are degraded to 500 pc to approximate the HST resolution at $z > 1.5$. Contours in the right panels show 2D clumps identified in the degraded maps. For each of these clumps, in the corresponding left panel we show the projected contours of the 3D clump that contributes the most to the rest-frame UV luminosity within the area of that 2D clump. Reduced resolution tends to mix smaller clumps into one big peak, resulting in the kpc-scale clumps seen in the observations.

Table 3.1: Clump masses and sizes identified in 2D and 3D

Run	L_{2D}/L_{gal}[1]	Flux[2] (μJy)	M_{2D}[3] ($10^7\,M_\odot$)	R_{2D}[4] (pc)	$M_{3D,tot}$[5] ($10^7\,M_\odot$)	$M_{3D,max}$[6] ($10^7\,M_\odot$)	R_{3D}[7] (pc)	$L_{3D,max}/L_{2D}$[8]	$SFR_{3D,max}/SFR_{2D}$[9]	$L_{3D,H\alpha}/L_{2D,H\alpha}$[10]
SFE200	0.195	0.060	22.9	560	5.6	0.7	160	0.42	0.10	0.11
	0.176	0.054	4.9	400	1.1	1.1	200	0.95	0.97	-
	0.074	0.023	15.4	560	2.1	0.8	230	0.37	0.56	0.88
	0.041	0.013	6.1	630	2.1	2.1	330	0.74	0.90	-
SFE100	0.198	0.213	34.9	800	9.8	6.4	300	0.57	0.52	0.43
	0.198	0.213	68.7	750	13.6	6.5	300	0.50	0.48	0.14
	0.073	0.078	12.4	400	1.3	1.3	220	0.46	0.65	-
SFE50	0.320	0.211	6.3	400	4.7	4.7	260	0.99	0.97	0.73
	0.201	0.132	58.2	940	6.8	0.7	150	0.48	0.09	0.82
	0.111	0.073	10.3	630	0.7	0.7	190	0.69	0.58	0.70
	0.090	0.059	24.3	630	3.9	0.8	210	0.51	0.17	0.19
	0.032	0.021	9.6	560	1.8	1.5	270	0.38	0.61	-
SFE10	0.236	0.108	82.9	800	8.3	3.7	280	0.21	0.41	0.26
	0.076	0.035	21.6	490	3.9	3.1	220	0.52	0.57	-
	0.074	0.034	26.0	490	4.2	1.1	180	0.23	0.16	-
	0.047	0.022	5.0	400	1.0	1.0	220	0.89	0.81	-
	0.046	0.021	4.5	690	0.8	0.8	260	0.65	0.60	0.49
	0.042	0.019	14.6	490	1.2	1.2	230	0.56	0.83	0.35
SFEturb	0.372	0.091	35.6	560	20.3	8.0	160	0.31	0.29	0.28
	0.132	0.032	15.6	490	11.7	9.8	350	0.82	0.84	-
	0.043	0.011	4.9	400	3.1	3.1	290	0.91	0.94	-

[1] Fraction of the galaxy's luminosity contributed by the 2D clump. Only clumps with fractional luminosity $> 3\%$ are listed.
[2] Clump's flux in HST/ACS F606W waveband, except for SFEturb run which is in F775W because of higher output redshift.
[3] Stellar mass of 2D clumps, defined using surface density as: $M_{\rm 2D} = \Sigma_* \times$ ClumpArea.
[4] Effective radius of 2D clumps, defined as: Clump Area $= \pi R_{\rm 2D}^2$.
[5] Sum of stellar mass of all 3D clumps within the projected area of the 2D clump.
[6] Stellar mass of the most luminous 3D clump within the projected area of the 2D clump.
[7] Effective radius of 3D clumps, defined as: Clump Volume $= \frac{4\pi}{3} R_{\rm 3D}^3$.
[8] Ratio of luminosity of 3D clumps to luminosity of 2D clumps in each galaxy's corresponding waveband.
[9] Ratio of star formation rates in 3D clumps and 2D clumps, averaged over 100 Myr.
[10] Ratio of Hα luminosity in 3D clumps and 2D clumps, for corresponding 2D UV clumps (matched by eye). "-" means there is no corresponding Hα 2D clump for a 2D UV clump.

The SFE50 run has five 2D clumps. The top one is the brightest while the middle one is the largest. The large clumps are massive and extended and most of them include multiple 3D clumps in projection. This leads to significant overestimation of both mass and size. The most luminous corresponding 3D clump is bright and compact, but there is also another 3D clump that accounts for 15% of the 2D luminosity, and four other clumps, each contributing a few percent of the luminosity. At the top edge of the galaxy there is another structure which produces a single 2D pixel, but since we require at least two pixels for 2D clumps, this single pixel is not identified as a clump.

The SFE10 run has a big 2D clump in the center, which is also a complex of multiple 3D clumps. Some of its smaller 3D clumps show elongated structure. The last snapshot of SFEturb run is at $z = 2.85$, much earlier than the other runs, so the main galaxy in SFEturb run is smaller. It has only three 2D clumps; one of the corresponding 3D clumps is compact, and the other two are more extended. The SFE100 run has three 2D clumps, two of which are large and extended, and so are the corresponding 3D clumps. The SFE200 run has four 2D clumps, all of which look elongated to some extent. The corresponding 3D clumps are small and compact, with many young stars located between the clumps, further leading to overestimation of mass and size of the 2D clumps.

Table 3.1 lists the properties of 2D and 3D clumps. The masses of the 2D clumps are in the range $10^7 - 10^9\,M_\odot$, in agreement with clump masses in Guo et al. (2018). While the effective sizes of 2D clumps are of order kpc, as in the observations (Förster Schreiber et al., 2011; Guo et al., 2012; Soto et al., 2017), the sizes of 3D clumps are significantly smaller. The projected area of some 2D clumps may contain multiple 3D clumps. To demonstrate the effects of mixing and projection, we include two estimates of 3D mass: the total mass of all 3D clumps that contribute more than 32 magnitude in the corresponding HST band (but not necessarily centered) within

the 2D clump column, and the mass of only one clump that contributes the highest rest-frame UV luminosity. Although mixing of multiple 3D clumps increases the mass by a factor of a few in some cases, inclusion of stars that are not in any identified 3D clump contributes much more to the overestimation of 2D clump masses.

The flux of the 2D clumps in their corresponding wavebands is similar to that in the observations of Guo et al. (2018). The flux of their clumps at $1.5 < z < 3$ in F606W and F775W bands is mostly between 0.02 to 0.3 μJy. Similarly, the largest of our simulated 2D clumps can contain between 20% and 37% of their host galaxy UV luminosity.

In the remaining columns we show the ratio of luminosities and SFRs of the most luminous 3D clump and the corresponding 2D clump. Although the masses of 2D clumps are significantly overestimated, the difference in luminosity is smaller. This is because we identify 3D clumps using young stars and they are the main contributor to the rest-frame UV luminosity. Similar to the luminosity ratio, the SFR in 2D clumps is also less overestimated than mass and size, because UV luminosity traces young stars.

The total contribution of 3D clumps to SFR in 100 Myr is 50-80%. Each individual 3D clump typically contributes a few percent to total SFR, up to ∼30%.The contribution of individual 2D clumps to the total SFR typically varies from a few percent to ∼15%, and can be up to 60%. This is generally consistent with the observational results of Guo et al. (2015) who find that the clumps typically contain a few percent of the total SFR, and of Soto et al. (2017) who find median clump contribution of 5% to the total SFR.

3.3.2 Comparison of 2D and 3D masses and sizes

The first obvious result of our comparison of clumps identified in 2D and 3D is the differences in mass and size. We compare the properties of 2D and 3D clumps in

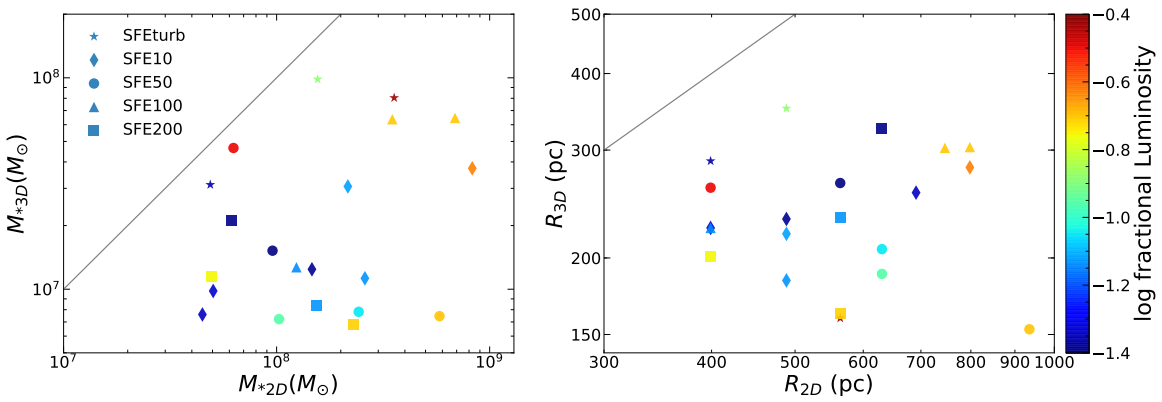

Figure 3.2 Stellar mass (left panel) and effective radius (right panel) of clumps identified in 2D degraded mock observation map (resolution 500 pc) with `min-npix`=2 and 3D density of young stars (younger than 100 Myr) identified with `min-npix`=14. Here stellar mass in 2D is total stellar mass within the projected column, and stellar mass in 3D is total stellar mass of the most luminous 3D clump within that 2D clump. Effective radius is the radius of a circle (in 2D) or a sphere (in 3D) that has the same area/volume as the corresponding clump. The colour of each point is the fractional luminosity of the 2D clump, $L_{\rm 2D}/L_{\rm gal}$. Gray thick solid lines are 1:1.

Figure 3.2. Each point corresponds to the 3D clump that contributes the most to the luminosity of the 2D clump when projected onto the galaxy plane. The stellar mass of the identified 2D clumps range from several times $10^7\,M_\odot$ to $10^9\,M_\odot$, while few of the corresponding 3D clump masses exceed $10^8\,M_\odot$. All the points are below the 1:1 line, which means that 2D clump masses and sizes are overestimated.

The overestimation of the 2D clump mass can be as large as an order of magnitude. Fractional luminosity of 2D clumps is not predictive of the extent of mass overestimation. The most overestimated 2D clumps are also more likely to contain multiple 3D clumps of comparable mass. Combining multiple snapshots, the most massive clump mass in both 2D and 3D seems to increase with decreasing SFE. The most overestimated clumps tend to have large fractional luminosity, since they are more likely to be combination of multiple clumps.

Similarly, the effective sizes of 2D clumps are also overestimated, especially for those that are made up of multiple 3D clumps. The ratio of 2D to 3D effective radii

can be a factor of several. We do not see any correlation between clump sizes and SFE.

All of the 3D clumps have intrinsic sizes below the HST resolution of ~ 500 pc, even after combining simulation grid cells to uniform patches of 100 pc. Therefore, the sizes of clumps are *always overestimated* in observations, unless they benefit from the magnification by gravitational lensing.

3.3.3 Clump longevity

Important questions for the interpretation of these giant clumps are: Are they gravitationally bound? How long do they remain identifiable as distinct clumps? One way to address this is to look at how the average distance between pairs of clump stars varies over time. If majority of stars are self-bound, the average distance should not change much. On the other hand, if the stars that appear to be in a clump are unbound after the initial gas clearance, then their average distance would increase steadily over time. The maximum rate of expansion is set by the velocity dispersion of all stars in that region of the galaxy. Below we conduct several tests of the boundedness and longevity of the clumps in our simulations.

Figure 3.3 shows the evolution of the average distance between pairs of young stars in 3D clumps, over the last three available simulation snapshots. After identifying a clump in the first snapshot, we calculate the mass-weighted average pair distance of the same stars in the following two snapshots, about 150 Myr and 300 Myr later. This can be used as a proxy for characteristic size of the clump. We can see that the average sizes of most clumps increase significantly after only 150 Myr, which means they dissolve and spread out. There are some less massive clumps that spread out slowly after 150 Myr, but they eventually dissolve after 300 Myr. We note that this analysis is limited by the time between saved simulation snapshots, and that clumps could dissolve in shorter than 150 Myr.

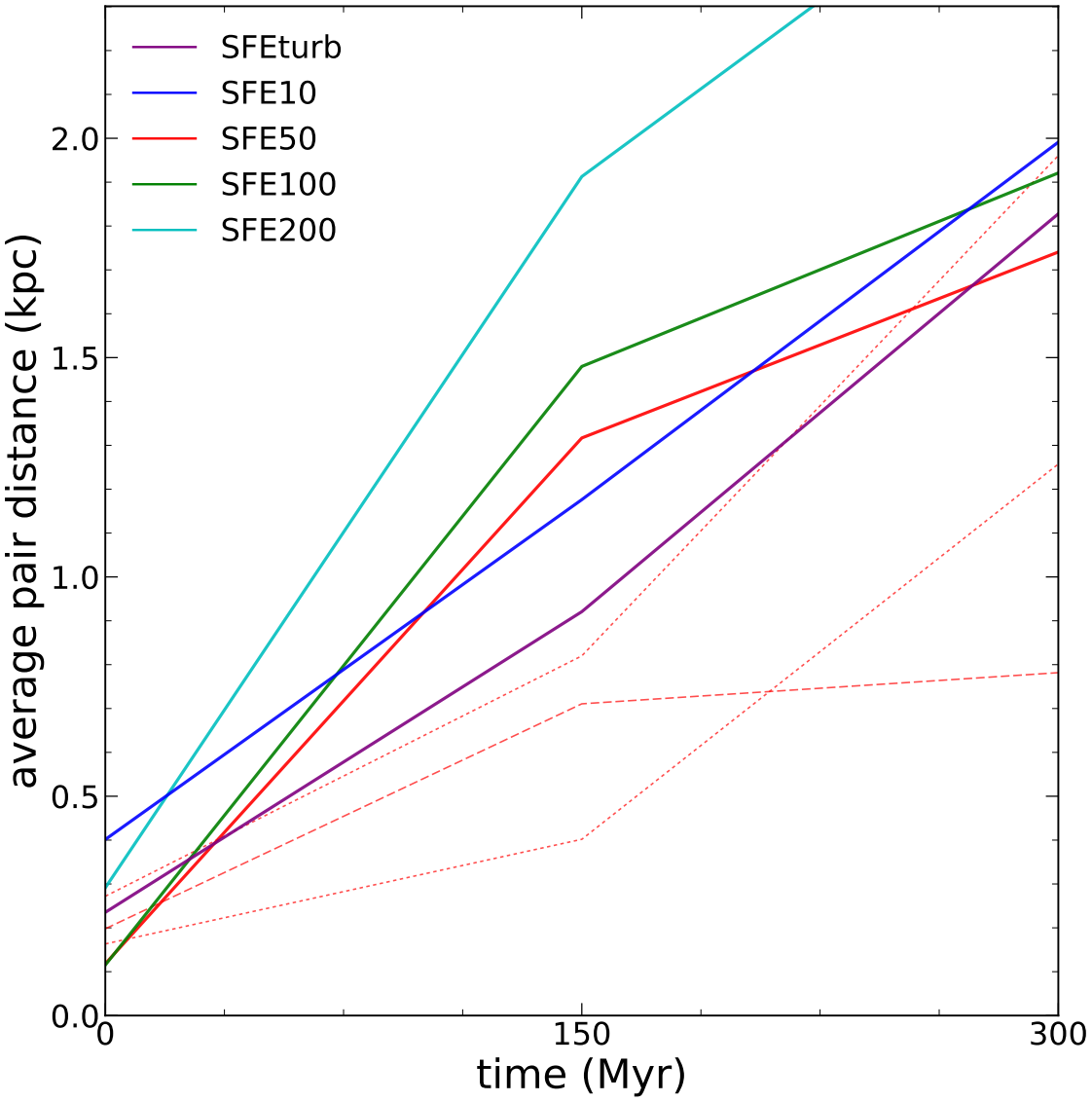

Figure 3.3 Evolution of mass-weighted average pair distance of young stars in 3D clumps over three consecutive snapshots. The time interval between snapshots is about 150 Myr. Solid lines show the most massive clump in each run. For SFE50 we show also the second most massive clump (dashed line) and the less massive ones (dotted lines). The upper limit of the y-axis corresponds to the distance traveled at 15 km/s over 150 Myr.

To estimate whether the 3D clumps are bound, we calculate the virial parameter

$$\alpha_{\rm vir} = a\frac{\sigma^2 R}{GM}, \tag{3.1}$$

where M is clump mass, R is the effective radius, and σ is the 3D velocity dispersion of young stars. We take the parameter $a = 5/3$ for a constant density sphere (Bertoldi & McKee, 1992). Values of $\alpha_{\rm vir} \lesssim 1$ would indicate the clump is bound, although the precise value depends on the structure of the stellar distribution. Note that 1 is not a strict boundary, so the assumption of constant density sphere is valid.

To quantify the evolution of the mass and size of a clump of young stars identified in one simulation snapshot, we look for clumps in the same group of stars in the next two consecutive snapshots. We keep the same density threshold for clump identification, so that stellar particles that dissolve into the background are not counted in the successor clumps. Then we calculate the fraction of mass of the original stars that remain in the most massive successor clump. We also take the mass-weighted average distance between pairs of the original stars remaining in the successor clump as a measure of size of the successor clump.

We find a weak correlation between $\alpha_{\rm vir}$ and the mass fraction of young stars that can still be found in the successor clump: clumps with high remaining mass fraction have small $\alpha_{\rm vir} \lesssim 1$. For example, all (20) but two clumps with remaining mass fraction above 70% have $\alpha_{\rm vir} < 1$. Equivalently, clumps with small $\alpha_{\rm vir}$ tend to retain more of their original mass: the median remaining mass fraction for clumps with $\alpha_{\rm vir} < 1$ is 61%, while the median for clumps with $\alpha_{\rm vir} > 1$ is only 7%. Mandelker et al. (2017) similarly found that clumps with shorter lifetime have larger $\alpha_{\rm vir}$.

The overall evolution of clump mass and size, relative to the moment it was first identified, is shown in Figure 3.4. Every clump begins at point (1,1) and then appears as a circle one snapshot later, with an arrow pointing to the second snapshot later.

If there is only a circle without an arrow, it means that the successor was not found after two snapshots. Since we use the average pair distance as a proxy for size of the surviving clump, we can illustrate decrease in the average density with lines of constant M/r^3. Most circles (one snapshot later) lie below the dashed line, which means that the remaining clumps are less dense than the original clumps. Many arrows point in the direction of decreasing density, indicating that the remaining part of clumps keeps dissolving over time. After 300 Myr the density of most clumps decreases by an order of magnitude, after which they would no longer be interpreted as clumps.

There is one clump in SFEturb run whose average density increases between the second and third snapshots, as the the mass continues to decrease. In this case the less bound outer part of the clump dissolves, leaving the remaining inner part more dense. At the time it was first identified, the clump was gravitationally bound, with the virial parameter $\alpha_{\rm vir} = 0.37$. After 125 Myr, the inner part of this clump still remains bound with $\alpha_{\rm vir} = 0.19$. However, after 250 Myr, half of this clump is dissolved and spread out all over the galaxy, while core now has $\alpha_{\rm vir} = 4.9$, with a significant increase in the velocity dispersion. We can expect that this clump will dissolve soon afterwards.

To examine one possible fate of the clumps, that they migrate to the galaxy center due to dynamical friction and contribute to the bulge (Ceverino et al., 2010), we calculated the evolution of the average galactocentric distance of the young stars in clumps. We found no net inward or outward migration of clump stars, similar to the conclusion of Buck et al. (2017). Instead of massive bulges, our galaxies contain nuclear star clusters that are consistent with being formed by in-situ star formation (Brown et al., 2018).

It has been argued that feedback recipe affects the longevity of clumps in simulations. Simulations with only thermal feedback from supernovae often produce

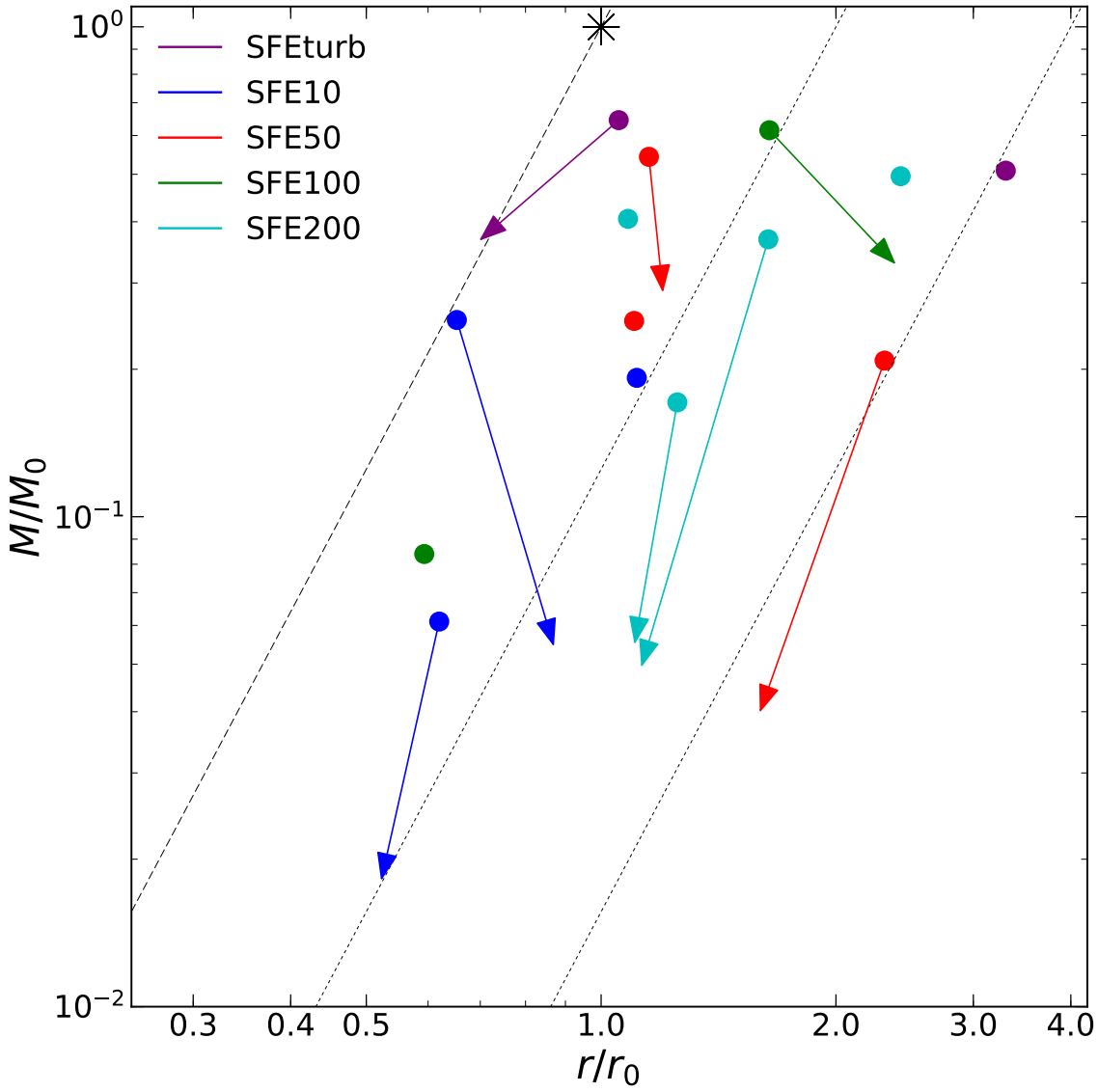

Figure 3.4 Evolution of the fraction of mass and mean separation of stars that remain in the most massive successor of clumps identified two snapshots before last available. The separation r is calculated as the mass-weighted average distance between all pairs of stars. By construction every clump starts at (1,1), marked by the starry symbol. Circles show the ratios one snapshot later (∼150 Myr), and arrows point to the values two snapshots later (∼300 Myr). The dashed line marks the initial average density of the clumps, while dotted lines mark constant density 8 and 64 times lower. We note that, since r is a proxy for size, "density" here is also a proxy for the actual density.

long-lived clumps which then migrate to the galaxy center (e.g. Ceverino et al., 2010, 2012; Mandelker et al., 2014). In simulations with strong momentum feedback, clumps usually dissolve in a short time $\lesssim$100 Myr (e.g. Oklopčić et al., 2017). The results of Mandelker et al. (2017) with radiation pressure feedback fall into an intermediate category, where massive clumps survive and low-mass clumps disrupt. Our simulation results are consistent with the strong feedback regime: most clumps dissolve in a short time and do not migrate to the center. As shown by Li et al. (2018b) this strong feedback is required to match the star formation history of Milky Way-sized galaxies inferred from abundance matching, and therefore, we favour the conclusions obtained with our simulations.

We have available one weaker feedback run, which produced too high star formation rate and metallicity. The distribution of young stars in that run is smoother and the disc appears regular even at high redshift. If we run 3D clump identification algorithm over that galaxy, we get 3D clumps that are similar to clumps in other galaxies at face value, except that they are slightly larger. However, these 3D clumps would be less prominent than clumps in other galaxies, given the smoother distribution of young stars. Thus, these clumps should not be viewed as the same "clumps" expected in a clumpy galaxy.

Some observed clumps appear to be old: e.g., Guo et al. (2012) found clump ages in the range $10^8 - 10^9$ yr, Soto et al. (2017) in the range $10^6 - 10^{10}$ yr, Zanella et al. (2019) in the range $10^6 - 10^9$ yr. The oldest end of these intervals is used to argue the longevity of the clumps. Some clumps even appear older than the underlying disc stellar population. However, inferred age can be severely affected by contamination by disc stars, as well as measurement systematics. When we calculate the mass-weighted average age of all stars in the clump regions in our simulations, we also find rather large ages 0.1-1 Gyr, because of the large amount of old stars that do not contribute much to the rest-frame UV light. The true age of stars producing most of

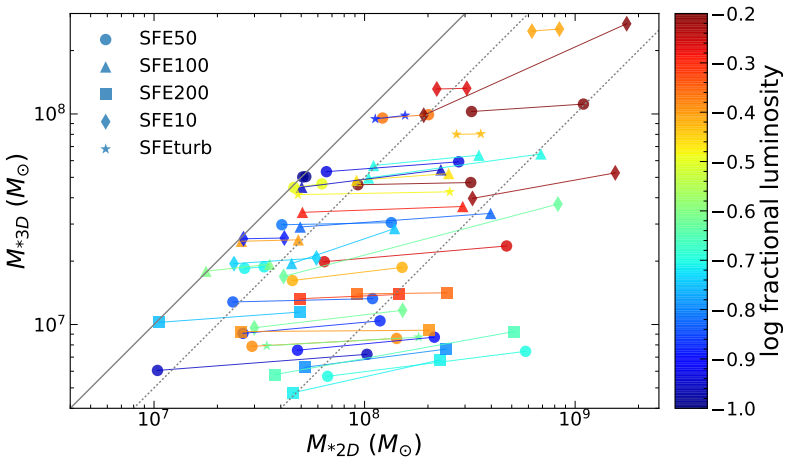

Figure 3.5 Masses of 2D clumps and the most luminous 3D clumps within them in different snapshots at redshifts from 1.5 to 3 (between 2 and 8 snapshots per run). Lines between symbols connect the mass of young stars (age less than 100 Myr) to the mass of all stars. To reduce the number of plotted points, we include only 2D clumps that contribute fractional luminosity $> 10\%$. The solid line shows 1:1 ratio, and the dotted lines show 1:2 and 1:10 ratios.

the UV light is of course below 100 Myr and cannot be used to set constraints on the dynamical longevity of the clumps.

3.4 Discussion

3.4.1 Effects of angular resolution

In our simulations, masses of the 2D clumps are overestimated by about an order of magnitude. Comparison of clump masses in high-redshift field galaxies with those magnified by gravitational lensing points to a similar overestimation due to insufficient resolution. The observed clump stellar mass increases from around $10^7\,M_\odot$ for lensed galaxies to $\sim 10^9\,M_\odot$ for field galaxies (Dessauges-Zavadsky et al., 2017a; Cava et al., 2018). Cava et al. (2018) observed the strongly lensed Cosmic Snake galaxy with the spatial resolution down to ∼30 pc, as well as the less strongly lensed Counterimage of the same galaxy with the resolution ∼300 pc. They found rest-frame UV clumps in the Counterimage with masses $\gtrsim 10^8\,M_\odot$, while the corresponding clumps in the

Cosmic Snake image have masses down to $\sim 10^7\,M_\odot$. Our 3D clumps and 2D clumps show similar overestimation in mass. Thus angular resolution is critical for correct characterization of giant clumps in high-redshift galaxies.

Clump sizes in lensed galaxies range from several hundred pc to kpc (Adamo et al., 2013; Wuyts et al., 2014; Livermore et al., 2015), smaller than clumps found in unlensed galaxies. Tamburello et al. (2017) tested the resolution effect using the Hα map smoothed with different Gaussian FWHM and found that clump sizes increase from ∼120 pc to ∼800-900 pc when FWHM increases from 100 pc to 1 kpc.

Two effects can increase the observed mass of 2D clumps: mixing of multiple 3D clumps within the projected column and inclusion of stars that are not in clumps. From Table 3.1 we can see that when clump masses are dramatically overestimated, the mass of intra-clump stars usually contributes more to the total mass than additional smaller 3D clumps. Including the mass of all 3D clumps usually increases the mass only by a factor of few, while including intra-clump stars can inflate the clump mass by over an order of magnitude.

The amount of discrepancy between the 2D and 3D masses is much larger when counting stars of all ages, relative to counting only young stars. Figure 3.5 illustrates both ways of calculating the mass: lines connect the mass of young stars (lower left) to the mass of all stars (upper right) within the same clump. Counting all stars instead of only young stars is roughly equivalent to overestimating the SFR. From the plot we can see that SFR in 2D clumps could be overestimated by a factor of 3 or more. Similar result is found in Fisher et al. (2017b), where blurred Hα clumps in the DYNAMO-HST sample of low-redshift turbulent disc galaxies have $\sim 2-3$ times higher SFR than the associated full-resolution clumps. The overestimation of SFR is larger for more massive 2D clumps, because they may contain multiple distinct 3D clumps.

However, we do not find systematic dependence of the mass ratios on the output

redshift. We plot masses of clumps identified at several epochs in all five runs: 8 outputs for SFE50 and 5 each for SFE100, SFE200, and SFE10 at redshift between 1.5 and 2, and 2 outputs for SFEturb at $z \approx 3$. The distributions of points all these epochs are comparably broad.

A commonly used observational proxy for SFR is the luminosity L_ν in the wavelength range 1500-2800Å (Kennicutt, 1998a):

$$\mathrm{SFR}\ (\,M_\odot\,\mathrm{yr}^{-1}) = 1.4 \times 10^{-28} L_\nu\ (\mathrm{erg\,s^{-1}Hz^{-1}}). \tag{3.2}$$

This relation holds for the Salpeter stellar initial mass function (IMF). In this study we adopt the Kroupa (2001) IMF, which requires a correction factor of 0.64 (Kennicutt & Evans, 2012). Using this relation, we compared the SFR calculated from the near-UV L_ν in our mock images to the actual SFR in the simulation, averaged over 100 Myr. The SFR inferred from L_ν is systematically 0.13 dex higher than the actual SFR, but with a large scatter of about 0.4 dex. The scatter mainly comes from the difference in age and metallicity of young stars. Stars older than 100 Myr could also contribute to the UV luminosity but are not counted in the SFR, which adds even more scatter.

The ratio of the UV luminosities and SFRs of the 3D and 2D clumps are listed in Table 3.1. The difference in the two ratios is caused by the scatter in the SFR$-L_\nu$ relation discussed above. Most clumps have similar values of $L_{\mathrm{3D,max}}/L_{\mathrm{2D}}$ and $\mathrm{SFR_{3D,max}}/\mathrm{SFR_{2D}}$. Some clumps, however, show large discrepancies. For example, the first clump in SFE200 run and the second and fourth clumps in SFE50 run contain fairly large luminosity fractions, but smaller SFRs by a factor 3-5. This occurs because these 2D clumps combine multiple 3D clumps, and the selected most luminous 3D clump happens to be much younger than the other 3D clumps. For instance, in the second 2D clump in SFE50 run, the average age of the selected 3D clump is 5.9 Myr, while all other overlapping 3D clumps are about 50 Myr old.

3.4.2 Effects of numerical resolution

We find that clumps identified at one snapshot in our simulations disperse and disappear in subsequent snapshots. To check that the dissolution of clumps is not caused artificially by the numerical relaxation effects, we calculated the half-mass relaxation time (Spitzer, 1969) for stellar particles making up the identified clumps:

$$\begin{aligned} t_{\rm rh} &= \frac{0.17\,N}{\ln(\lambda N)}\sqrt{\frac{r_{\rm h}^3}{GM}} \\ &= \frac{0.78\,{\rm Gyr}}{\ln(\lambda N)}\frac{1\,M_\odot}{m}\left(\frac{M}{10^5\,M_\odot}\right)^{1/2}\left(\frac{r_{\rm h}}{1{\rm pc}}\right)^{3/2}. \end{aligned} \tag{3.3}$$

Here m, M, and N are the average mass, total mass, and number of the stellar particles, respectively, and r_h is the half-mass radius. We take $\lambda = 0.2$ following Binney & Tremaine (2008). Using another suggested value $\lambda = 0.1$ does not change the results.

The relaxation times for our clumps are typically several hundred Myr, with the smallest ones being about 100 Myr. Since it usually takes several relaxation times for a stellar system to dissolve, which is much longer than the time between two snapshots (about 150 Myr), we can conclude that the dissolution of the 3D clumps in our simulations is not due to numerical relaxation.

3.4.3 Gaseous clumps in line emission

While we mainly focus on clumps of young stars in this paper, angular resolution also affects inferred sizes and masses of analogous clumps of ionized gas, found via their emission lines. For example, studies of Hα emission from lensed high-redshift galaxies ($z = 1-4$) find clumps with sizes ranging from $\sim$100 pc to 1 kpc (Jones et al., 2010; Livermore et al., 2012, 2015). The clump luminosity and SFR density increase with redshift. Measuring shifts of the Hα provides also kinematic information on the high-redshift clumps (e.g., Mieda et al., 2016).

Hα emission traces even younger stars (age less than $\sim$10 Myr) than the UV light, because youngest stars contribute most of the ionizing radiation. To make more direct comparison with Hα observations, we created mock maps of the Hα emission measure. Similar to the UV clumps, we identify Hα clumps in both 2D and 3D.

The emission measure is defined as $\int n^2\, ds$, where n is the density of ionized hydrogen (here we use ionized hydrogen in cells with $T < 20,000$ K, which contribute most to the cross-section), and s is the length in the line of sight. In a simulation output the ionized hydrogen density is given on the adaptive mesh, with some cells on low refinement levels that are larger than our preferred uniform grid of 100 pc. Therefore, we resample all cells in the galaxy region to level 9, which corresponds to physical size 15-25 pc, and then map the gas density on a uniform grid of 100 pc. We calculate the emission measure on this grid and then identify 3D Hα clumps using `astrodendro`. The parameters we use here are: `min-pix`=14 and `min-delta`=0.4 dex. The parameter `min-value` is chosen using a similar approach of matching the background value as we did for 3D UV clumps. We choose different `min-value` for different runs because these galaxies are at different redshift and have different $n_{\rm HII}$: `min-value`=$\log n^2/\mathrm{cm}^{-6}$=1 for SFE50 and SFE100 runs, $\log n^2/\mathrm{cm}^{-6}$=0.2 for SFE10 run, $\log n^2/\mathrm{cm}^{-6}$=0.5 for SFE200, and $\log n^2/\mathrm{cm}^{-6}$=0 for SFEturb run.

To obtain a 2D map of the emission measure, we integrate it in the vertical direction and degrade it to 500 pc patches. Then we identify 2D Hα clumps using the following `astrodendro` parameters: `min-npix`=2, `min-delta`=0.1 dex, `min-value`=200 pc cm^{-6}.

The Hα clumps do not necessarily correspond to rest-frame UV clumps since they trace star formation on different timescales. For those UV and Hα clumps that match spatially, we show the ratios of Hα luminosity in 3D and 2D clumps in the last column of Table 3.1. For the UV clumps that do not have a corresponding Hα clump, we leave dashes in the table.

The first clump in SFE200 run has $L_{\rm 3D,H\alpha}/L_{\rm 2D,H\alpha}$ of only 11%. The emission

measure distribution in this galaxy is relatively smooth, thus producing this very small 3D Hα clump.

We can compare the 3D/2D ratio of Hα luminosities, which trace the SFR over $\sim$10 Myr, with the corresponding ratio of SFR averaged over 100 Myr. For many clumps they are close in value, and in these cases the 3D and 2D Hα and UV clumps correspond to each other fairly well. Some clumps have large discrepancies, generally because the UV clump and the Hα clump differ significantly in geometry.

The Hα light is more clumpy than UV light distribution, since it traces younger stellar population. The total fractional luminosity of 3D Hα clumps ranges from $\sim$10%-95%, while the contribution to Hα luminosity from individual 3D clumps ranges from less than 1% to $\sim$90%. This large range results from the short timescale of star formation traced by Hα and thus the clumpiness of Hα emission. Most individual 3D clumps contribute less than 10% to total Hα luminosity, while only one or two largest clumps contribute most of the Hα flux. The gas in Hα clumps is not bound, with virial parameters (Equation 3.1)) larger than 1 by orders of magnitude.

3.4.4 Origins of giant stellar clumps

Gravitational instability can induce fragmentation of the galactic disc, thus leading to growth of clumps. Stability criterion to linear axisymmetric perturbations for gaseous discs can be described by the Toomre parameter (Safronov, 1960; Toomre, 1964):

$$Q = \frac{\sigma\kappa}{\pi G\Sigma}, \tag{3.4}$$

where κ is the epicycle frequency, σ is the velocity dispersion, and Σ is the mass surface density. A disk is unstable if $Q \lesssim 1$. Observations (e.g. Puech, 2010; Girard et al., 2018) found that clumpy galaxies are marginally stable, with Toomre $Q \sim 1$. Most Hα clumps in high-redshift galaxies are located in regions where Q is low (Genzel et al., 2011; Wisnioski et al., 2012; Mieda et al., 2016). In galaxy formation simulations

(Inoue et al., 2016; Oklopčić et al., 2017) gas clumps also coincide with regions of $Q < 1$, albeit with large scatter of Q value. Inoue et al. (2016) also found that stellar clumps coincide with regions of low stellar Q.

In Meng et al. (2019) we studied the distribution of Q in our simulations, accounting for different phases of the gas and stars. Such multi-component Q more accurately describes the linear stability criterion in realistic galaxies. We found that strong feedback from young stars disperses gas around them, leading to spatial *anti-correlation* of dense gas and stars up to 50 Myr old. Here we analogously find that locations of 2D UV clumps do not coincide with the regions of low Q. This is because low Q regions closely trace high gas density, while rest-frame UV clumps trace young stars. The average value of Q in 2D UV clump regions ranges from 2.1 to 3.6, which is larger than the median value of $Q = 0.5 - 1.0$ weighted by molecular gas mass in Meng et al. (2019). For Hα clumps $Q = 1.4 - 3.3$ is closer to the gas values, as expected for very young stars. The discrepancy in location of gas and young stars is also seen in the simulations of Oklopčić et al. (2017), where they find that the gas clumps coincide with the location of instantaneous star formation, but not of the SFR averaged over more than 10 Myr.

If clumps form out of instability of a self-gravitating disc, the wavelength of the fastest growing perturbation is

$$\lambda_T = \frac{2\sigma^2}{G\Sigma}. \tag{3.5}$$

The effective radius of a fully-formed clump may not exactly correspond to this scale, but it can still be used as a rough guide. If we take $\lambda_T/2$ for $R_{\rm clump}$, and approximate the disc rotation velocity profile as flat, then $\kappa \approx \sqrt{2}V_{\rm rot}/R_{\rm disc}$, and we obtain

$$\frac{R_{\rm clump}}{R_{\rm disc}} \approx \frac{\pi Q}{\sqrt{2}} \left(\frac{\sigma}{V_{\rm rot}} \right). \tag{3.6}$$

Here $R_{\rm disc}$ is a characteristic size of the gas distribution. When the gaseous disc is

marginally stable, i.e., $Q \sim 1$, we expect a linear relation between $R_{\rm clump}/R_{\rm disc}$ and $\sigma/V_{\rm rot}$.

Fisher et al. (2017a) found the relation $R_{\rm clump}/R_{\rm disc} = (0.38\pm0.02)\,\sigma/V_{\rm rot}$ in nearby turbulent discs with properties closely resembling $z \sim 2$ star-forming galaxies. They take twice the half-light radius of Hα light as $R_{\rm disc}$ and identify clumps in Hα + [NII] map. They use flux-weighted $\sigma(\rm H_\alpha)$ and the modeled rotation velocity $V_{\rm rot}$ at 2.2 disc scalelengths. Note that the proportionality coefficient 0.38 is a factor of 6 smaller than that expected from the linear perturbation theory above. This already serves as a warning that the observed clump sizes may not be related to global disc instabilities.

We examined whether the 2D or 3D clumps in our simulations obey a similar relation. We take $R_{\rm disc}$ to be twice the half-light radius in our mock Hα emission measure map. We use $\sigma_{\rm HII}$ in the brightest pixel and the circular velocity at galactocentric radius of 10 kpc as σ and $V_{\rm rot}$. In the 25 snapshots of our 5 runs at different redshifts, most galaxies have the value of $\sigma/V_{\rm rot}$ between 0.2 and 1.2. The ratio of sizes $R_{\rm clump}/R_{\rm disc}$ for 2D clumps is in the range 0.1 to 0.4; for 3D clumps in the range 0.02 to 0.2. We do not find any correlation between the ratios $R_{\rm clump}/R_{\rm disc}$ and $\sigma/V_{\rm rot}$, either for 2D or 3D clumps. We also examined this relation for the projected UV clumps and did not find any correlation. This shows that our clump sizes are not set by gravitational instability in thin axisymmetric discs. In fact, if the relation (3.6) were to hold, we would expect clump sizes of the order λ_T, which is several kpc (Meng et al., 2019), and therefore significantly larger than the effective radii of our identified clumps. We also calculated the radius $(M_f/\pi\Sigma)^{1/2}$ corresponding to the "fragmentation mass" $M_f = 2\lambda_T\Sigma c_s/(\kappa f_g)$ defined by Tamburello et al. (2015) through their equation 10, and found that it still overestimates the clump sizes, by factor of a few.

We also investigated the dependence of clump properties on the specific SFR (sSFR). Shibuya et al. (2016) found that the fraction of high-redshift clumpy galaxies

increases with sSFR. Fisher et al. (2017b) found a positive relation between sSFR and the maximum fractional luminosity of low-redshift Hα clumps, both blurred and with full resolution ($\sim$100 pc). Our Hα clumps also show a positive correlation between the maximum fractional luminosity and sSFR. For most galaxies with sSFR$>$ $2 \times 10^{-10}\,\mathrm{yr}^{-1}$, the maximum fractional luminosity of 3D Hα clumps is larger than 10%, while the maximum fractional luminosity of 2D Hα clumps (resembling their blurred clumps) is larger than 20%. These results agree with those of Fisher et al. (2017b).

3.4.5 Comparison with other simulation studies

A number of studies have investigated the formation of giant clumps in galaxy formation simulations. Some simulation analyses identify 2D clumps in surface density maps. Oklopčić et al. (2017) identified clumps using gas surface density in a massive galaxy in the FIRE simulation at $z = 1 - 2$. They use a similar setup to our study: bin the surface density in patches of 50 pc and use `astrodendro` with a minimum of 20 patches. Thus their clump sizes are several hundred pc, with the baryon masses $10^{6.5-9.5}\,M_\odot$. Moody et al. (2014) identified clumps in projected stellar mass maps and obtained clump masses $\sim 10^{6.5} - 10^{9}\,M_\odot$. Our 2D clump masses are similar to the high-mass end of these clumps, but we miss the low-mass end possibly because of the adopted cut on fractional luminosity.

Other simulation studies identify clumps using 3D density, which is similar to our clump identification in 3D. Mandelker et al. (2017) identified clumps in the 3D distribution of both the cold gas component and stellar component in the VELA simulation galaxies, with halo masses $10^{11} - 10^{12}\,M_\odot$ at $z = 2$. They identified clumps as connected regions, which is similar to our hierarchical tree construction, containing at least 8 cells of (70 pc)^3 each and found clumps of baryon mass $\sim 10^{7} - 10^{9}\,M_\odot$. Tamburello et al. (2015) and Mayer et al. (2016) used the `SKID` algorithm to identify

bound structures in their simulations. Their typical clump gas masses and stellar masses are $\sim 10^7 - 10^8\,M_\odot$. Our 3D clumps are similar to these clumps in mass.

Some simulations investigate the effect of spatial resolution on identified clump masses using maps smoothed to different resolution. Tamburello et al. (2017) identified clumps in mock Hα images of their simulations, convolved with the 2D Gaussian aperture with full width at half maximum FWHM = 1 kpc and 100 pc. They found clump gas mass to be $\sim 10^8 - 10^9\,M_\odot$ for FWHM=1 kpc, and $\sim 10^{6.5} - 10^{8.5}\,M_\odot$ for FWHM=100 pc. The clumps in the full-resolution Hα maps are similar to the 3D bound structures found in Tamburello et al. (2015), but clumps found with FWHM=1 kpc have masses overestimated by an order of magnitude. Our results support this order-of-magnitude overestimation. We also agree on the amount of overestimate of the clump size: Tamburello et al. (2017) find a median intrinsic radius ∼120 pc vs. ∼800 pc in the blurred images, while our 3D and 2D median clump sizes are ∼200 pc and ∼630 pc, respectively. How much clump sizes are overestimated depends largely on the spatial resolution, and this comparison shows that the clump sizes in observations of unlensed galaxies are likely severely overestimated.

Behrendt et al. (2016) also identified clumps in mock Hα maps. They convolved the surface density map with a 2D Gaussian of FWHM=1.6 kpc to mimic the instrumental response, and found clump baryon masses to be $(1.5 - 3) \times 10^9\,M_\odot$. Buck et al. (2017) identified clumps in luminosity maps in the NIHAO galaxy sample, using both intrinsic clumps in non-dust-attenuated rest-frame U band and clumps in HST bands with dust taken into account. Their clump masses range from a few times $10^6\,M_\odot$ to $10^9\,M_\odot$, and sizes are $\sim 300 - 900$pc. The high-mass end is similar to our 2D rest-frame UV clumps in mock observation maps. Benincasa et al. (2019) identified clumps in isolated disc galaxy simulations both in 3D using `SKID` and in 2D gas surface density map using `astrodendro`. They used 2D resolution of 10 pc and 100 pc, and found only a factor of a few difference between the 2D and 3D clump

masses. This result guarantees that our 3D clumps identified on grid of 100 pc are not much overestimated compared to the intrinsic clumps.

3.5 Conclusions

We investigated the nature of giant kpc-scale clumps in high-redshift progenitors of the Milky Way-sized galaxies. We identified both 2D clumps in rest-frame UV mock HST observation maps with 500 pc resolution and intrinsic clumps in 3D density of young stars with 100 pc resolution (Figure 3.1). The 2D clumps are chosen to resemble observed giant clumps. The main results of our comparison are summarized below:

- The masses and sizes of 2D clumps are overestimated due to limited angular resolution and projection of several 3D clumps along the line of sight. The overestimate of mass can be as large as an order of magnitude, and the overestimate of size can be a factor 2–3, compared to the most luminous corresponding 3D clump (Figure 3.2 and Table 3.1). The intrinsic sizes of clumps (150–300 pc) are below the HST resolution at $z > 1$, unless the source galaxy is strongly lensed.

- Most clumps of young stars in our simulated galaxies dissolve on a timescale shorter than ∼150 Myr. The average pair distance between young stars in a clump increases dramatically after 150 Myr (Figure 3.3), and the density of the remaining parts of these clumps continues to decrease over time (Figure 3.4).

- Most of the 3D clumps are not gravitationally bound structures, with the virial parameter $\alpha_{\rm vir} > 1$. However, a few clumps with $\alpha_{\rm vir} < 1$ are more likely to have a fraction of mass remain bound in the next simulation snapshot, after ∼150 Myr.

- Although total stellar masses of 2D clumps are significantly overestimated, the masses of young stars are only overestimated by factor of a few relative to the most luminous corresponding 3D clump (Figure 3.5).

- We created mock images of Hα emission measure and compared the sizes of Hα clumps with the expectation of the linear theory of global disc instabilities. We do not find the expected correlation between $R_{\rm clump}/R_{\rm disc}$ and $\sigma/V_{\rm rot}$, primarily because the clump sizes are much smaller than expected. Therefore, we conclude that the observed clumps are not the result of gravitational instabilities in thin axisymmetric discs, at least in Milky Way-sized galaxies.

CHAPTER IV

Evolution of disc thickness in simulated high-redshift galaxies

This chapter was published as: ***Meng, X.*** *& Gnedin, O. Y. 2021, MNRAS, 502, 1433*

Abstract

We study the growth of stellar discs of Milky Way-sized galaxies using a suite of cosmological simulations. We calculate the half-mass axis lengths and axis ratios of stellar populations split by age in galaxies with stellar mass $M_* = 10^7 - 10^{10}\,M_\odot$ at redshifts $z > 1.5$. We find that in our simulations stars always form in relatively thin discs, and at ages below 100 Myr are contained within half-mass height $z_{1/2}$ ∼0.1 kpc and short-to-long axis ratio $z_{1/2}/x_{1/2}$ ∼0.15. Disc thickness increases with the age of stellar population, reaching median $z_{1/2}$ ∼0.8 kpc and $z_{1/2}/x_{1/2}$ ∼0.6 for stars older than 500 Myr. We trace the same group of stars over the simulation snapshots and show explicitly that their intrinsic shape grows more spheroidal over time. We identify a new mechanism that contributes to the observed disc thickness: rapid changes in the orientation of the galactic plane mix the configuration of young stars. The frequently mentioned "upside-down" formation scenario of galactic discs, which

posits that young stars form in already thick discs at high redshift, may be missing this additional mechanism of quick disc inflation. The actual formation of stars within a fairly thin plane is consistent with the correspondingly flat configuration of dense molecular gas that fuels star formation.

4.1 Introduction

The stellar disc of the Milky Way (MW) galaxy shows two distinct geometric components, the thin disc and the thick disc, with characteristic scale heights ∼0.3 kpc and 0.9 kpc, respectively (Jurić et al., 2008). Similar thin disc plus thick disc structure is also observed in other nearby galaxies (e.g. Reddy et al., 2006; Yoachim & Dalcanton, 2008b,a). The MW thin and thick discs differ in chemical abundances (Bensby et al., 2005) – the thin disc being more metal rich with lower [α/Fe], and the thick disc being more metal poor with higher [α/Fe] – suggesting that the thin disc is mainly composed of relatively young stars, while the thick disc mainly of old stars. The typically quoted axis ratios of the thin and thick discs are $\sim$ 0.1 and 0.4, respectively (e.g., Bovy & Rix, 2013; Bensby et al., 2011). Results of massive spectroscopic surveys such as APOGEE suggested that instead of a clear separation between the thin and thick discs, the MW structure may be better described by a superposition of many mono-abundance populations, each with a single exponential scale height and scale length (Bovy et al., 2012, 2016). For older populations, indicated by lower metallicity and enhanced [α/Fe], the scale heights monotonically increase with age (from ≈200 pc to 1 kpc), while the scale lengths decrease (from >4.5 kpc to 2 kpc).

The disc structure has also been studied at high redshift. Unlike the regular thin discs seen at low redshift, high-redshift galaxies show thick discs which are often clumpy (e.g. Elmegreen et al., 2007; Overzier et al., 2010; Swinbank et al., 2010a). Chain and spiral galaxies in the *Hubble Space Telescope (HST)* Ultra Deep Field have the ratio of scale height to radial scale length $\sim$ 1/3 (Elmegreen & Elmegreen, 2006).

Elmegreen et al. (2017) found the scale heights of galaxies in the *HST* Frontier Fields Parallels to increase with galaxy mass and decrease with redshift. By measuring the scale heights at differt locations of edge-on disks, they noted that clump regions have smaller scale height, and that the thick disc is observed best between the clumps, with a larger scale height. The overall evolution of galaxy shape appears to transform galaxies from prolate and spheroidal shapes at high redshift to thin discs at low redshift (Law et al., 2012; van der Wel et al., 2014; Zhang et al., 2019).

Current cosmological simulations successfully reproduce the observations that the present day distribution of young stars is thinner than that of old stars, and that galaxy discs are thicker at high redshift (e.g. Bird et al., 2013; Pillepich et al., 2019; Buck et al., 2020). However, the origin of such a transition from thick to thin discs, and how stars form in thick discs at high redshift, is still actively debated.

There are several scenarios that attempt to explain the formation of thick discs. One posits that the old stellar populations formed thick and kinematically hot at high redshift (e.g. Tutukov et al., 2000; Bird et al., 2013; Stinson et al., 2013), as mono-age or mono-abundance populations in simulations show thicker and shorter discs at birth. Other scenarios suggest that stars always form in thin discs and are heated to larger heights because of scattering by massive clumps (Bournaud et al., 2009; Comerón et al., 2011, 2014; Beraldo e Silva et al., 2020) or radial migration (Schönrich & Binney, 2009; Loebman et al., 2011). The third type of studies suggest that thick discs are related to galaxy mergers, including heating of a thin disc in a merger, direct accretion of stars from satellite galaxies, and star formation in merging gas-rich systems (e.g. Reddy et al., 2006). Considering that mergers are more frequent at high redshift, these scenarios may lead to thicker discs of old stellar populations.

In this paper, we revisit the evolution of galactic disc thickness with stellar age, taking advantage of a recent suite of cosmological simulations of MW-sized galaxies with ultrahigh resolution. We test whether young stars actually form in thin discs

and how quickly the scale height grows with stellar age. In section 4.2 we describe the simulations and how we calculate the shape of the simulated galaxies. We present the evolution of thickness of our simulated galaxies with stellar age in section 5.3. We compare our results with observations and other simulations and discuss the factors that contribute to the thickening of galactic discs in section 5.4, and present our conclusions in section 5.5.

4.2 Shape of High-Redshift Galaxies

4.2.1 Simulation Suite

We use a suite of cosmological simulations run with the Adaptive Refinement Tree (ART) code (Kravtsov et al., 1997; Kravtsov, 1999, 2003; Rudd et al., 2008) and described in Li et al. (2018b) and Meng et al. (2019). All runs start with the same initial conditions in a periodic box of 4 comoving Mpc, producing a main halo with total mass $M_{200} \sim 10^{12}\,M_{\odot}$ at $z = 0$, similar to that of the MW. The ART code uses adaptive mesh refinement to reach higher spatial resolution in dense regions. The lowest resolution is set by the root grid, which has 128^3 cells. This sets the dark matter particle mass $m_{\rm DM} \approx 10^6\,M_{\odot}$. The finest refinement level is adjusted in runtime to keep the physical size of gas cells on that level between 3 and 6 pc. Because of strong stellar feedback, few cells remain at this finest refinement level and the typical spatial resolution of molecular gas is 36-63 pc.

The simulations include three-dimensional radiative transfer using the Optically Thin Variable Eddington Tensor approximation (Gnedin & Abel, 2001) of ionizing and ultraviolet (UV) radiation from stars (Gnedin, 2014) and the extragalactic UV background (Haardt & Madau, 2001), non-equilibrium chemical network that calculates the ionization states of hydrogen and helium, and phenomenological molecular hydrogen formation and destruction (Gnedin & Kravtsov, 2011). The simulations in-

corporate a subgrid-scale model for unresolved gas turbulence (Schmidt et al., 2014; Semenov et al., 2016b). Star formation is implemented with the continuous cluster formation (CCF) algorithm (Li et al., 2017b), where each stellar particle represents a star cluster that forms at a local density peak and grows its mass via accretion of gas and star formation until feedback of its own stars terminates the growth. The feedback recipe includes early radiative and stellar wind feedback, as well as a supernova (SN) remnant feedback model (Martizzi et al., 2015; Semenov et al., 2016b). The momentum feedback of the SN remnant model is boosted by a factor $f_{\rm boost} = 5$ to compensate for the underestimation due to numerical discreteness of the isolated SN model and to match the star formation history expected from abundance matching. The simulations include several runs with different values of the local star formation efficiency (SFE) per free-fall time, which ranges from $\epsilon_{\rm ff} = 10\%$ to 200%. For a full description of the star formation and feedback recipe, see Li et al. (2017b, 2018b) and Li & Gnedin (2019).

The simulations are run to $z \approx 1.5$. The high-redshift galaxies in our simulations are turbulent and irregular. They show prolate shape and contain clumps of young stars, consistent with the observed high-redshift clumpy galaxies. We remind the reader that these galaxies are unlike the regular galaxies at low redshift with a thin rotating disc and central bulge. Instead, the high-redshift galaxies are rather more irregular and dominated by turbulence.

Apart from the fiducial runs with $f_{\rm boost} = 5$, our simulation suite also includes a weaker feedback run with $f_{\rm boost} = 3$. The weaker feedback run produces a much higher stellar mass than expected from abundance matching and has a thin rotating disc even at high redshift. We do not consider that run as realistic and exclude it from analysis in this paper. More details about the structure of our simulated galaxies are given in Meng et al. (2019) and Meng & Gnedin (2020)

In this paper we analyze four runs that used different values of $\epsilon_{\rm ff}$. The number

after "SFE" in the run name corresponds to the percentage of local $\epsilon_{\rm ff}$. At redshifts $z \approx 2-5$, there is one main galaxy and several smaller galaxies. We investigate the structure of the main galaxy and other isolated galaxies with stellar mass above $10^7\,M_\odot$.

4.2.2 Measurements of galaxy shape

To measure galaxy shape, first we need to define the extent of a central galaxy and eliminate possible effects of satellites and galaxy mergers. For this purpose we calculate galaxy properties in a sphere centered on the galaxy center with radius 10 kpc for $z \leqslant 5$ galaxies, and 5 kpc for $z > 5$ galaxies. This covers most of the stellar distribution of our galaxies.

We examined the robustness of this choice by evaluating several alternative radial cuts. We determined the stellar extent of the galaxies by identifying a local minimum $r_{\rm M,min}$ in the spherically-averaged mass profile $dM(r)/dr$. If the profile decreases with r and then begins to increase again, it means there is another stellar structure such as a satellite galaxy. We use $r_{\rm M,min}$ as one of alternative radial cuts. The median of these radii is about 10 kpc for $z \leqslant 5$ galaxies, and 5 kpc for $z > 5$ galaxies, which informs our final choice. The other alternative we tried is 20 kpc for $z \leqslant 5$ galaxies, and 10 kpc for $z > 5$ galaxies, which is close to the extent of the largest galaxies in high-redshift observations (e.g. Elmegreen et al., 2017).

We found that for most of our simulated galaxies, choosing different maximum radius has little effect on the determination of galaxy orientation or stellar scale lengths along all axes. Using these alternative radial cuts results in the median fractional difference in the half-mass axis lengths smaller than 0.1 dex, and the orientation of the majority of galaxies changes by less than 1 degree.

We study all isolated (also referred to as central to their dark matter halo) galaxies, which are located outside of virial radii of other galaxies. The number of stellar

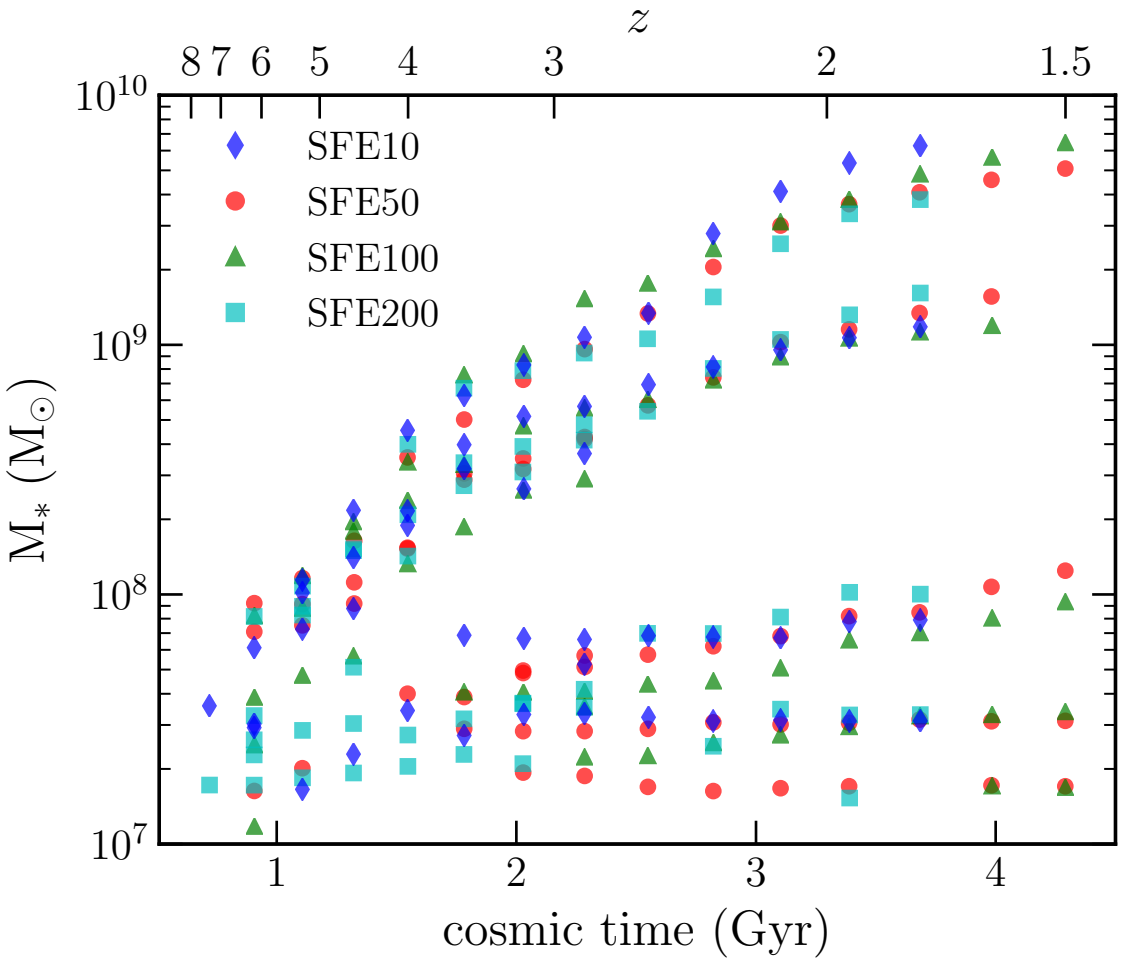

Figure 4.1 Stellar mass of simulated galaxies vs. cosmic time and redshift in the simulation snapshots selected for analysis. In addition to the main galaxy, each snapshot contains several smaller isolated galaxies.

particles in a galaxy ranges from a few hundred for galaxies with $M_* \sim 10^7\,M_\odot$ to several million for galaxies with $M_* \sim 10^{10}\,M_\odot$. We require galaxies to have $M_* > 10^7\,M_\odot$ so that they contain at least 100 stellar particles for the shape analysis to be reliable.

For the purpose of comparison with observational samples, different stages in the evolution of a given simulated galaxy can be considered uncorrelated if they are separated by a long enough time. We choose the separation between simulation snapshots based on the dynamical time calculated at the typical largest extent of stellar distribution, which is $R/V_{\rm circ} \sim 60 - 100$ Myr. We wish to use the snapshots separated by at least two dynamical times. The time interval between our snapshots is ∼150 Myr at $z < 5$ and ∼100 Myr at higher redshift, thus we select one in every two snapshots. This results in separation of $200 - 300$ Myr between the analyzed epochs.

We show the galaxy stellar mass and cosmic time of our sample in Figure 4.1. Each point corresponds to an isolated galaxy at a selected snapshot. One can see two relatively large galaxies and several smaller galaxies at these epochs.

High-redshift galaxies often show irregular shapes unlike axisymmetric discs observed at low redshift, and they are often dominated by turbulent motions instead of rotation. Therefore, following Meng et al. (2019) we use the shape tensor (Zemp et al., 2011) to determine the orientation of our simulated galaxies:

$$\mathbf{I} \equiv \sum_{k,i,j} M_k\, r_{k,i}\, r_{k,j}\ \mathbf{e}_i \otimes \mathbf{e}_j,$$

where M_k is the mass of k-th stellar particle, $r_{k,i}$ are its coordinates in the galactocentric reference frame ($i = 1, 2, 3$), and $\mathbf{e}_i$ are the three unit vectors of the coordinate axes. The tensor can be diagonalized by a rotation matrix to obtain the principal moments of inertia $I_1 \geqslant I_2 \geqslant I_3$. From these we calculate the axis ratios $b/a \equiv \sqrt{I_2/I_1}$

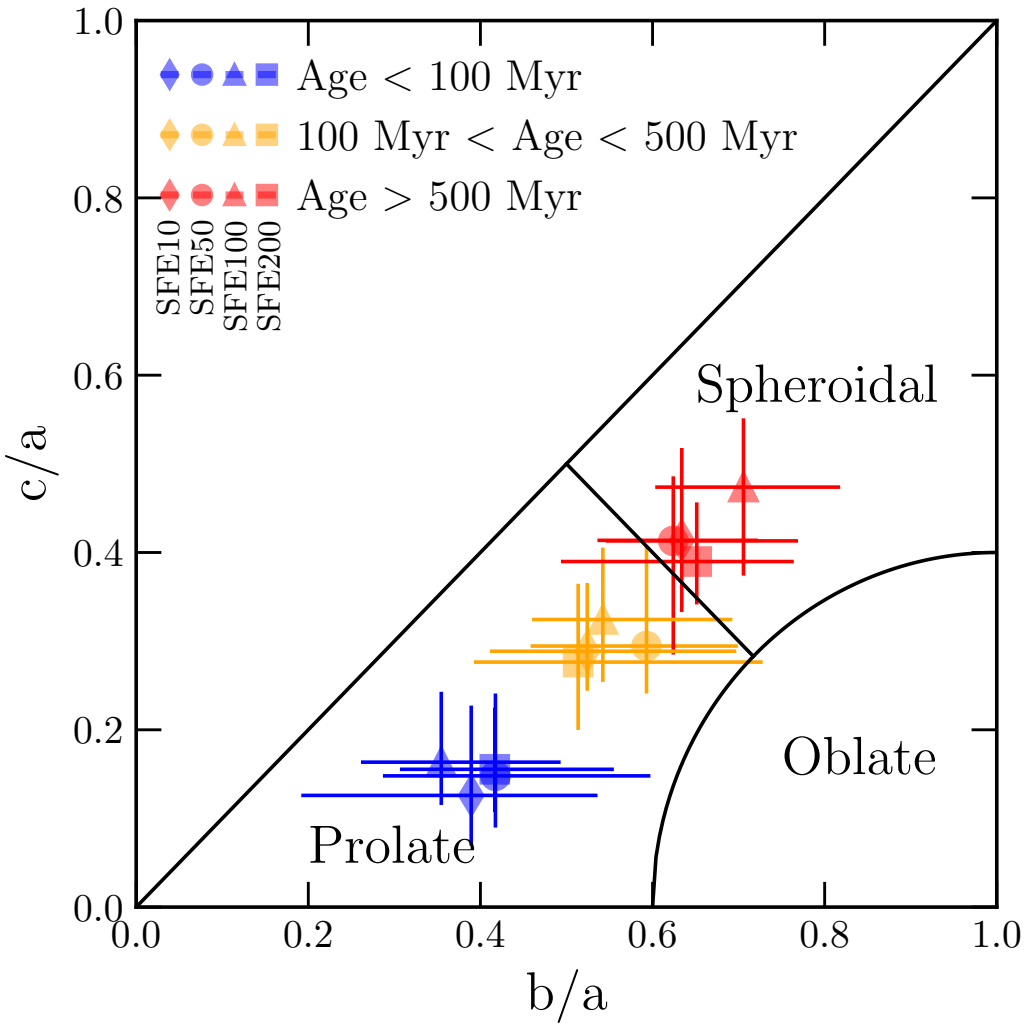

Figure 4.2 Intrinsic axis ratios of stellar populations of different age. Points with error bars show the median and interquartile ranges of stars in a given age bin for all galaxies in our sample. For most galaxies, the stellar distribution transitions from prolate to more spheroidal shape as stellar age increases.

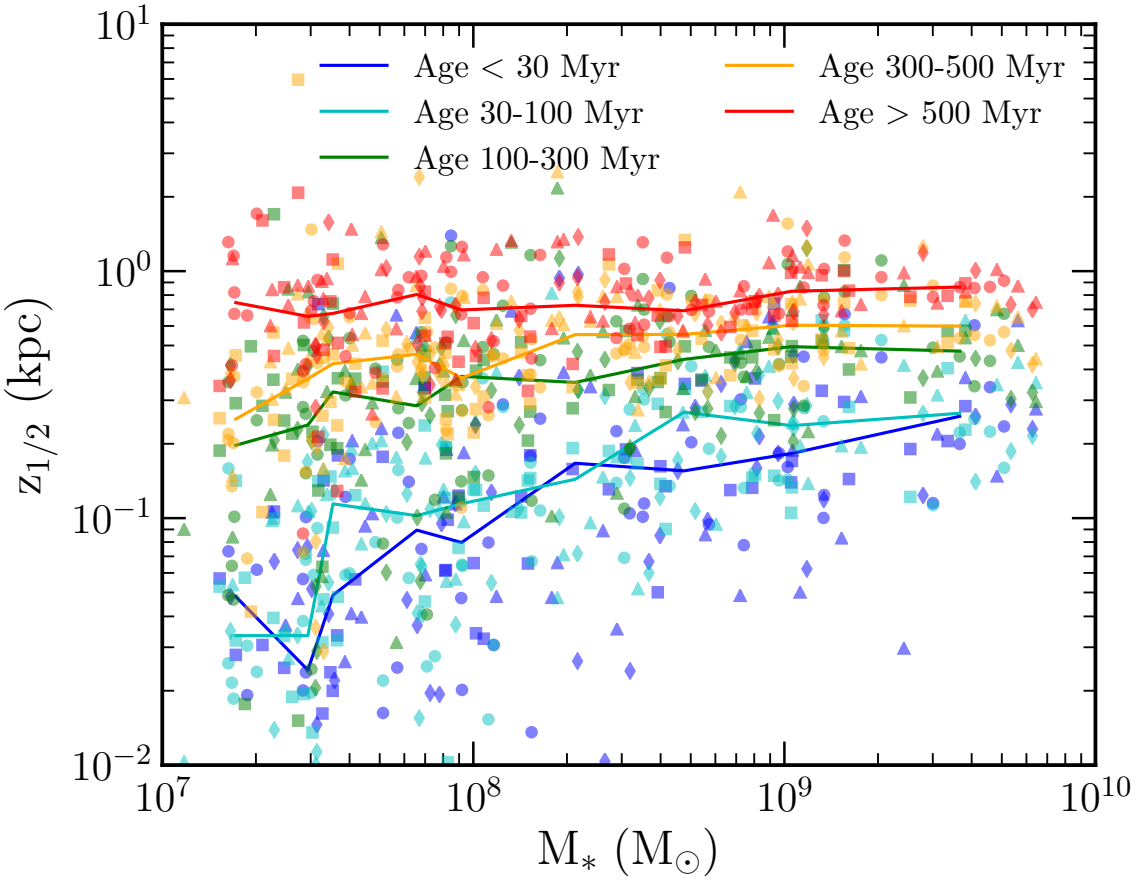

Figure 4.3 Disc half-mass height of stellar populations split by age. Galaxies in the four runs are shown by the same symbols as in Figure 4.1. Lines show median values of $z_{1/2}$ in bins of galaxy mass. For the youngest stars the disc height increases with galaxy mass, but at ages above 100 Myr the median height is independent of mass.

and $c/a \equiv \sqrt{I_3/I_1}$, which describe the shape of the mass distribution. The orientation of the galaxy plane is given by the eigenvector corresponding to the smallest eigenvalue. Following van der Wel et al. (2014), we divide the parameter space of the axis ratios b/a and c/a into three distinct shapes of an ellipsoid: oblate, spheroidal, and prolate (elongated).

In Figure 4.2 we show the evolution of the axis ratios with stellar age. We separate stellar populations in three groups: less than 100 Myr, between 100 and 500 Myr, and older than 500 Myr. Each point shows the median axis ratios of stars in an age group for all galaxies in our sample, and the errorbars show the interquartile ranges of the

c/a and b/a distributions. In most galaxies, c/a and b/a both increase with age, and the morphology of the stellar populations transitions from prolate to more spheroidal or triaxial shape. Note that for this plot we calculated the shape tensor of each age group separately, so that the intrinsic orientation is possibly different for each group. Our goal here was to show that older stars have more spheroidal distribution even when the coordinate axes are chosen to minimize the short axis ratio c/a.

For all subsequent analysis we fix a single orientation for a given galaxy, as it would be assigned in observation. We choose the plane orientation based on the shape tensor of stars younger than 100 Myr, because they contribute most of the observable rest-frame UV light. We then transform the coordinates of stellar particles into the new coordinates given by the adopted orientation.

We use half-mass axis lengths to describe the thickness of different components of our galaxies. The half-mass axis length is calculated as the length that contains half the cumulative mass along a given direction: $M(|x| < x_{1/2}) \equiv \frac{1}{2}M_{\rm tot}$, where x, y and z correspond to the long, middle, and short axes. We use $z_{1/2}/x_{1/2}$ as a proxy for the ratio of scale height to scale length, and refer to it as "disc thickness". Note here we use "disc thickness" only to describe the thickening of stellar distributions as in observations. We do not distinguish between a flattened stellar distribution or a rotationally-supported disc.

In previous papers (Meng et al., 2019; Meng & Gnedin, 2020) where we studied the relation between the star formation rate and interstellar medium, we defined galaxy plane as the orientation of the short axis of the shape tensor of neutral gas. Here we use the definition based on stars to match the observations of high-redshift stellar populations. We checked the difference in these two definitions of galaxy orientation: the median difference is about 30° for all galaxies in our sample. For the most massive galaxies, the difference is smaller, with a median of 16°.

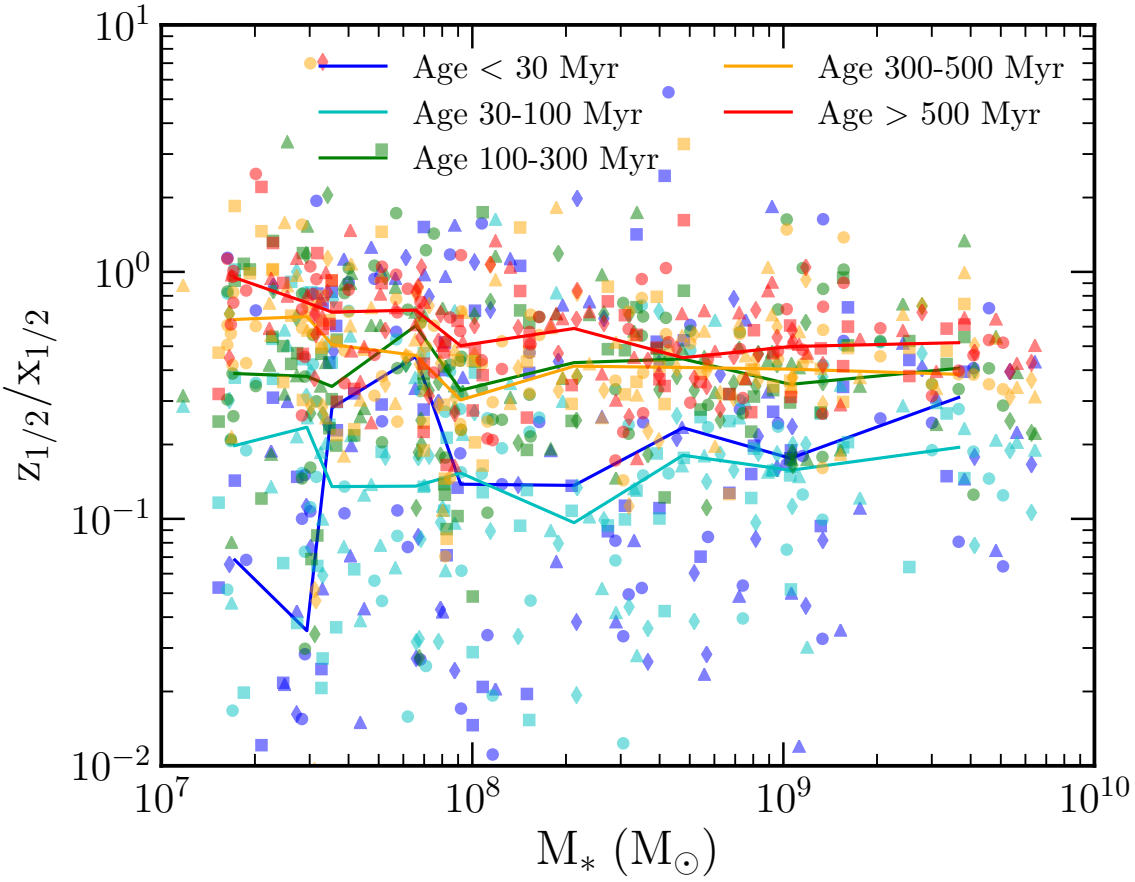

Figure 4.4 Ratio of the short to long axes, split by stellar age. Symbols are as in Figure 4.3. The median disc thickness shows little variation with galaxy mass.

4.3 Evolution of disc thickness

For a more detailed analysis of the thickness of stellar populations, we group stars into more fine age bins: <30 Myr, 30–100 Myr, 100–300 Myr, 300–500 Myr, >500 Myr. We then calculate the half-mass length for each age bin in the three established coordinate directions.

First we explore whether the thickness of stellar discs depends on galaxy mass, which itself increases over time. Figure 4.3 shows the half-mass height $z_{1/2}$ for stellar populations of different age. Each galaxy is represented by five small points corresponding to the heights of its five age populations. The shape of the symbols indicates from which run they were selected, but there is no systematic dependence in any of the results on the SFE adopted in the simulations.

Big diamonds show median values of the small points in galaxy mass bins, to allow better inspection of any systematic trends. The clearest trend is the half-mass height monotonically increasing with age of the stellar population. We describe it in more detail in the following plots. Any dependence on galaxy stellar mass is much less obvious. There may be a slight trend for the height of the youngest stars to increase with mass, but not for stars older than 100 Myr. Considering that the low-mass galaxies contain fewer than 10^3 stellar particles, and even less in a given age bin, this trend may not be significant.

Figure 4.4 shows disc thickness as the ratio of the short to long axis lengths $z_{1/2}/x_{1/2}$. Similar to the half-mass height, the thickness also increases with the age of the stellar populations. The only possible exception is with the two youngest age bins, which show more scatter and variation at the low mass end. This is because the orientation we choose to define the disc plane is based on the shape of stars younger than 100 Myr, which includes both age bins. This definition automatically minimizes the thickness of both stellar populations. Depending on stellar mass and specific star formation history, either the $<$ 30 Myr bin or the 30-100 Myr bin may display a

thinner disc. Therefore, we find little difference in all of our results between these two bins.

A small fraction of points lie above the line $z_{1/2} = x_{1/2}$. As our disc orientation minimizes the thickness of stars younger than 100 Myr, some older stellar populations that formed in a very different orientation may show $z_{1/2} > x_{1/2}$. This includes even one of the younger age bins (<30 Myr or 30–100 Myr) for a few galaxies. Our definition of the disc orientation guarantees that at least one of the younger bins will have $z_{1/2}/x_{1/2} < 1$, and for majority of the galaxies both bins do. However, a few cases with rapidly precessing disc plane show the other young bin with $z_{1/2} > x_{1/2}$. The rapid precession of the plane of young stars is illustrated later in Figure 4.8. Such precession may be caused by the clumping of star forming regions and the cycles of gas infall and expansion due to strong stellar feedback.

For the oldest bin with age >500 Myr the ratio $z_{1/2}/x_{1/2}$ decreases slightly with increasing stellar mass, from about 0.75 at $M_* \sim 10^7\,M_\odot$ to about 0.5 at $M_* > 10^9\,M_\odot$. A linear fit between $\log(z_{1/2}/x_{1/2})$ and $\log M_*$ has a slope -0.102±0.023. Thus disc thickening appears to be more pronounced for the older stars in dwarf galaxies. Still the trend is weak and does not appear to be related to the velocity dispersion of gas from which the stars form. We find the median ratio of the velocity dispersion to galaxy circular velocity is roughly constant $\sigma_g/v_{\rm circ} \approx 0.6$ at all galaxy masses. The more likely cause of higher $z_{1/2}/x_{1/2}$ in dwarf galaxies is just the small number of young stellar particles, which leads to a more stochastic galaxy orientation.

Both the disc height and thickness increase with stellar age. Figure 4.5 illustrates this explicitly for the half-mass height. Instead of individual points for each galaxy's populations, vertical errorbars indicate the interquartile range of the galaxy distribution. Diamonds show the median values for all galaxies in our sample in a given age bin. We can see that the disc thickness clearly increases from ∼0.1 kpc at formation to ∼0.8 kpc after 1 Gyr. The scatter in the galaxy sample is approximately constant

at 0.15-0.2 kpc, although it appears larger for younger ages due to the logarithmic scale.

Similarly, Figure 4.6 shows that the half-mass ratio $z_{1/2}/x_{1/2}$ steadily increases with stellar age. Young stars always form in relatively thin discs with $z_{1/2}/x_{1/2} \lesssim 0.2$, while after several hundred Myr the stars occupy significantly thicker shapes with $z_{1/2}/x_{1/2} \approx 0.6$. Some runs have $z_{1/2}/x_{1/2}$ larger in the first age bin than in the second, again because the galaxy orientation is based on all stars younger than 100 Myr. Overall, both young stellar populations appear to occupy a similarly thin disc. As in the previous plots, there is no clear difference for the simulations with different adopted SFE.

We provide the medians and interquartile ranges of $z_{1/2}/x_{1/2}$ for stellar populations of different age in Table 4.1.

Table 4.1: Medians and interquartile ranges of the half-mass short-to-long axis ratios of stellar populations in age bins.

	25–50–75% range of $z_{1/2}/x_{1/2}$				
Run	Age < 30 Myr	Age 30–100 Myr	Age 100–300 Myr	Age 300–500 Myr	Age > 500 Myr
SFE10	0.04 – 0.16 – 0.41	0.04 – 0.14 – 0.27	0.20 – 0.33 – 0.64	0.31 – 0.40 – 0.56	0.43 – 0.54 – 0.69
SFE50	0.05 – 0.14 – 0.48	0.08 – 0.19 – 0.30	0.26 – 0.41 – 0.69	0.36 – 0.48 – 0.61	0.43 – 0.57 – 0.84
SFE100	0.05 – 0.20 – 0.64	0.10 – 0.17 – 0.29	0.32 – 0.43 – 0.68	0.34 – 0.53 – 0.67	0.48 – 0.61 – 0.77
SFE200	0.11 – 0.25 – 0.52	0.07 – 0.14 – 0.26	0.22 – 0.41 – 0.54	0.30 – 0.43 – 0.77	0.35 – 0.49 – 0.66

To study the evolution of thickness of individual stellar populations, we trace the shape of the same group of stars through time. We choose all stars younger than 100 Myr within 10 kpc from galaxy center, for galaxies in the redshift range $z = 2 - 4.6$. We follow these same stars in all consecutive snapshots and calculate the intrinsic half-mass axis ratio, in the coordinate frame defined by their own shape tensor. Note that this is different from the previous plots where the orientation of all stars is defined by the shape tensor of stars younger than 100 Myr at that snapshot. We choose the intrinsic orientation for this calculation because the galaxy orientation changes over time, and simple misalignment with current plane may lead to thicker inferred shape of the tracked stars. We stop tracking the stars when their host galaxies merge into larger halos, since mergers significantly alter the original shape of the stellar distribution.

Figure 4.7 shows the resulting distribution for all selected populations. In total we have stellar groups selected from 118 independent galaxy snapshots. We only include groups with at least 100 stellar particles to be able to calculate the shape reliably. The stellar populations start from median age of $\sim$50 Myr and are traced in every consecutive snapshot after selection. The median value of $z_{1/2}/x_{1/2}$ increases from 0.15 when the stars are youngest to about 0.5 after 2 Gyr. The thickening of their distribution is unambiguous. This process saturates after 1-2 Gyr and therefore it is easy to miss in studies that have coarser time resolution. We do not find any dependence in this evolution on galaxy mass. Note that in Figure 4.7 the ratio $z_{1/2}/x_{1/2}$ is measured in the coordinate system defined by the shape tensor of the same stars that we trace. If instead $z_{1/2}/x_{1/2}$ is measured in the orientation of the galaxy defined by young stars in each snapshot, the median of $z_{1/2}/x_{1/2}$ will increase to 0.5 in 1 Gyr, and further to 0.58 in 2 Gyr. This is because different stellar populations do not always align. We describe the orientation of different stellar populations in more detail in Figure 4.8.

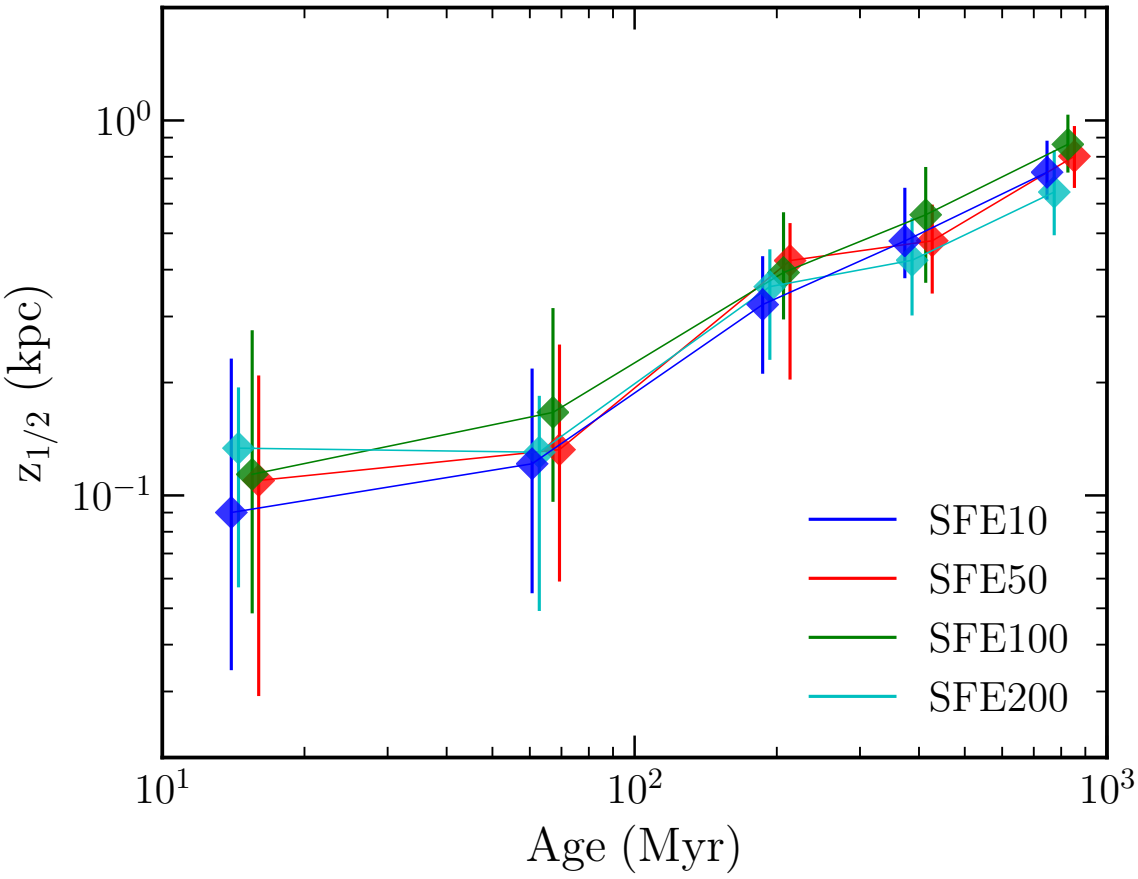

Figure 4.5 Evolution of the disc half-mass height with stellar age. Median $z_{1/2}$ of stellar populations in a given age bin for all galaxies are shown as big diamonds and connected with solid lines for clarity, analogously to Figure 4.3. Instead of individual points, vertical errorbars show the interquartile range. The value of age is taken to be the median in a bin: about 15, 65, 200, 400, 800 Myr, respectively. A small offset in age is added to distinguish multiple points. Disc scale height monotonically increases with stellar age. There is no clear trend with the value of SFE used in the simulations.

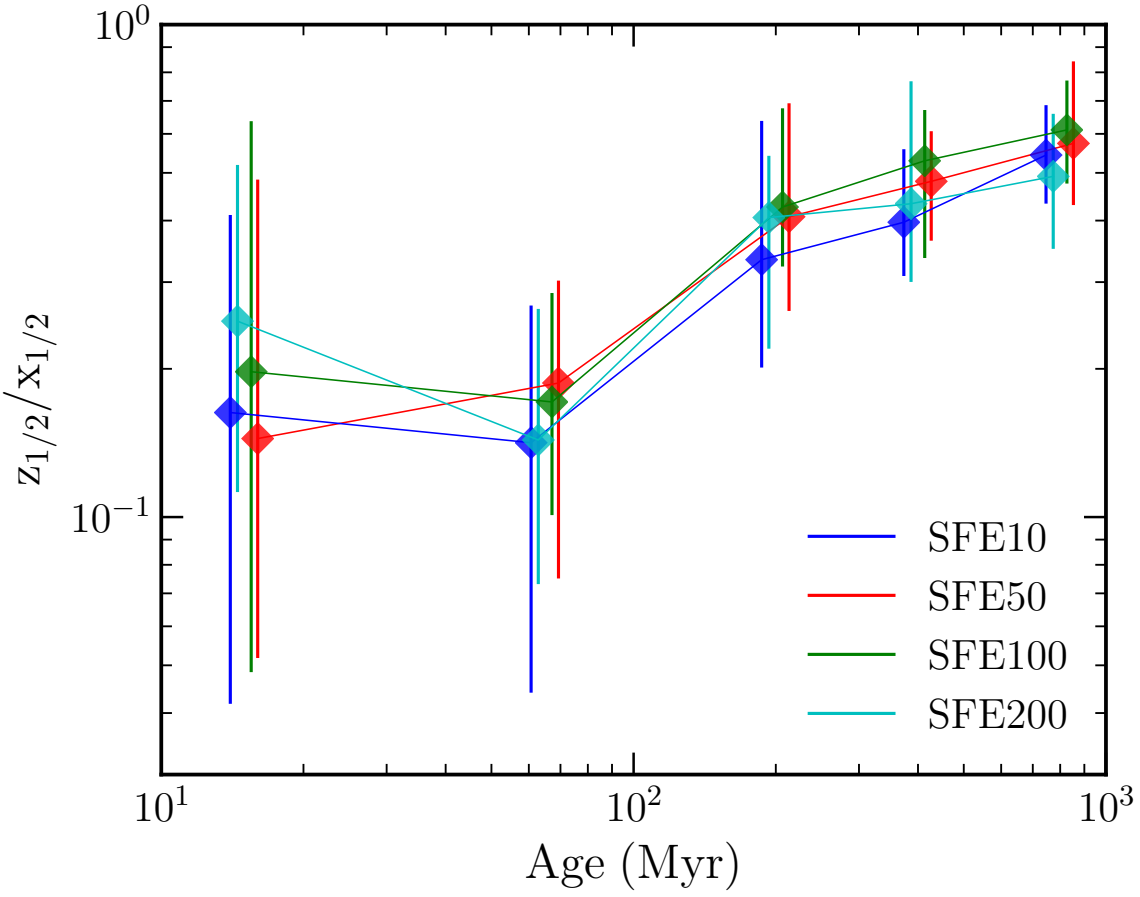

Figure 4.6 Evolution of the half-mass short-to-long axis ratio $z_{1/2}/x_{1/2}$ with stellar age. Symbols are as in Figure 4.5. The median disc thickness clearly increases with age.

4.4 Discussion

4.4.1 Comparison with observations of high-redshift galaxies

When galaxies are observed edge-on, their vertical heights can be directly measured. At $z > 1.5$, optical and near infrared wavebands correspond to the rest-frame UV, which is contributed mostly by young stars. Elmegreen & Elmegreen (2006) fitted the vertical profiles of galaxies in *HST* UDF with the functional form $\mathrm{sech}^2(z/z_0) = 4/[\exp(z/z_0) + \exp(-z/z_0)]^2$, where the half-light height $z_{1/2}$ is related to the scale height z_0 by $z_{1/2} = 0.55\, z_0$. Their results show that the half-light height of clumpy galaxies in the F850LP band to be $z_{1/2} = 0.5 \pm 0.2$ kpc. This is larger than the thickness of the youngest stellar population of our galaxies: $0.1 - 0.2$ kpc. Similarly, Elmegreen et al. (2017) measured the sech^2 scale height of the high-redshift galaxy discs in the *HST* Frontier Fields Parallels. They divided the galaxies into spiral type and clumpy (or "chain") type, and found that spiral galaxies are only present at $z < 1.5$. Since our simulated galaxies are at $z \geq 1.5$ and appear clumpy, we compare our results to the clumpy/chain galaxies. For these systems Elmegreen et al. (2017) find that the height is smaller for the star-forming clumps and larger in the interclump regions. They interpret it as the clumps representing the thin disc, while the interclump regions being more similar to the thick disc. Therefore, the average scale height quantifies the thin disc and the maximum scale height in each galaxy quantifies the thick disc. The median of the average half-light height of the clumpy galaxies in the redshift range 1.5-2.5 is $z_{1/2} = 0.35 \pm 0.13$ kpc. This is closer to, but still larger than, the median half-mass height of the youngest stars in our simulated galaxies (see Figure 4.3). Note that the observational sample includes more massive galaxies than in our simulations, from $10^9\, M_\odot$ up to $10^{10.7}\, M_\odot$, whereas our sample is below $10^{9.8}\, M_\odot$. In the Elmegreen et al. (2017) sample the disc height strongly increases with galaxy mass, which may partly account for the larger values in the

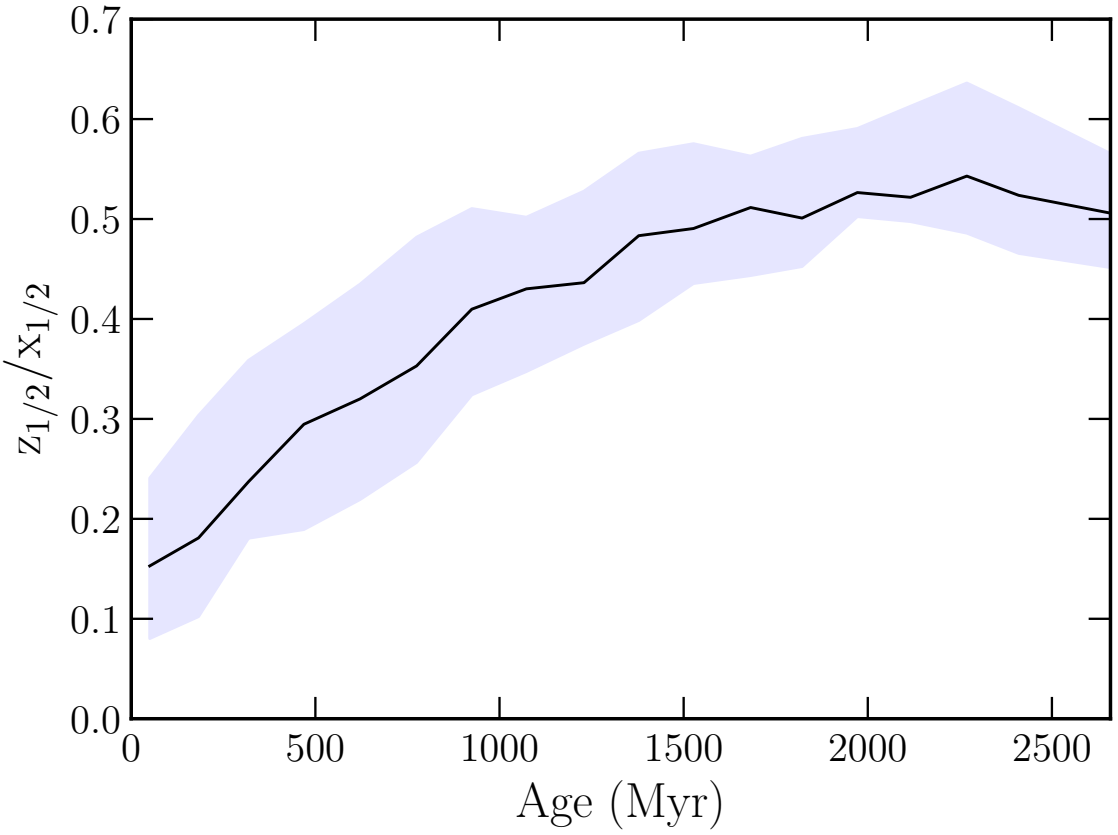

Figure 4.7 Evolution of the intrinsic short-to-long axis ratio of the same group of stars. Each group is selected as stars younger than 100 Myr in independent snapshots and traced over all subsequent snapshots until the galaxy mergers with a larger galaxy or reaches the end of the simulation. The line shows the median value for all combined groups in age bins, while the shaded region shows the interquartile range.

observational sample. In contrast, the median of the maximum heights is even higher at 0.58 ± 0.24 kpc, well above our results.

Although stars form in relative thin configuration in our simulated high-redshift galaxies, this configuration is still somewhat thicker than local star-forming discs, which typically have axis ratios below 0.15. As a comparison, in our simulations the median short-to-long axis ratio of young stars is 0.15-0.2. High-redshift observations directly support thicker star-forming discs. For example, Reshetnikov et al. (2003) find that in the Hubble Deep Field galaxies the ratio of disc scale height to scale length increases by a factor of at least 1.5 from low redshift to $z > 1$. Elmegreen

& Elmegreen (2006) quote the mean ratio of the radial half-length between clumps to sech^2 scale height to be 3.4±1.6, which corresponds to $z_{1/2}/x_{1/2} \gtrsim 0.16$, although exact conversion is not possible. This ratio is similar to that for young stars in our simulated galaxies.

Besides the direct way of measuring the thickness of edge-on discs, there are also studies of shapes of high-redshift galaxies with other orientations. If galaxies are randomly projected on the sky, the observed shape distribution can be modeled to infer the intrinsic 3D shape distribution. Elmegreen et al. (2005) found the intrinsic thickness of spiral galaxy discs in the *HST* UDF to be ≈0.2-0.3. Other studies divide high-redshift galaxies into mass and redshift bins and reconstruct the shape distribution for each bin to analyse the evolution of intrinsic galaxy shapes. These studies found that high-redshift galaxies are more prolate or spheroidal than oblate (e.g. Law et al., 2012), and that the prolate fraction decreases with decreasing redshift and increasing galaxy mass (e.g. van der Wel et al., 2014; Zhang et al., 2019). For the mass and redshift range that best matches our simulated galaxies, $M_* = 10^9 - 10^{11}\,M_\odot$ and $z = 1.5 - 3.6$, Law et al. (2012) found galaxy minor/major and intermediate/major axis ratios $c/a \sim 0.3$ and $b/a \sim 0.7$. Zhang et al. (2019) give the average ratios $c/a \sim 0.3$ and $b/a \sim 0.5$ for galaxies with mass $10^{9.0-9.5}\,M_\odot$ at $1.5 < z < 2.5$. This puts the intrinsic shape of observed high-redshift galaxies in the prolate regime, similar to our simulated galaxies.

Elmegreen et al. (2017) also find the scale height of the most massive ($M_* = 10^{9.5} - 10^{10.5}\,M_\odot$) clumpy galaxies to increase with galaxy mass and decreasing redshift. This trend is present for the youngest stars in our galaxies (Figure 4.3).

4.4.2 Comparison with simulations

Numerical simulations have also investigated the evolution of disc thickness. For example, the Illustris TNG50 simulation suite has a much larger number of galaxies

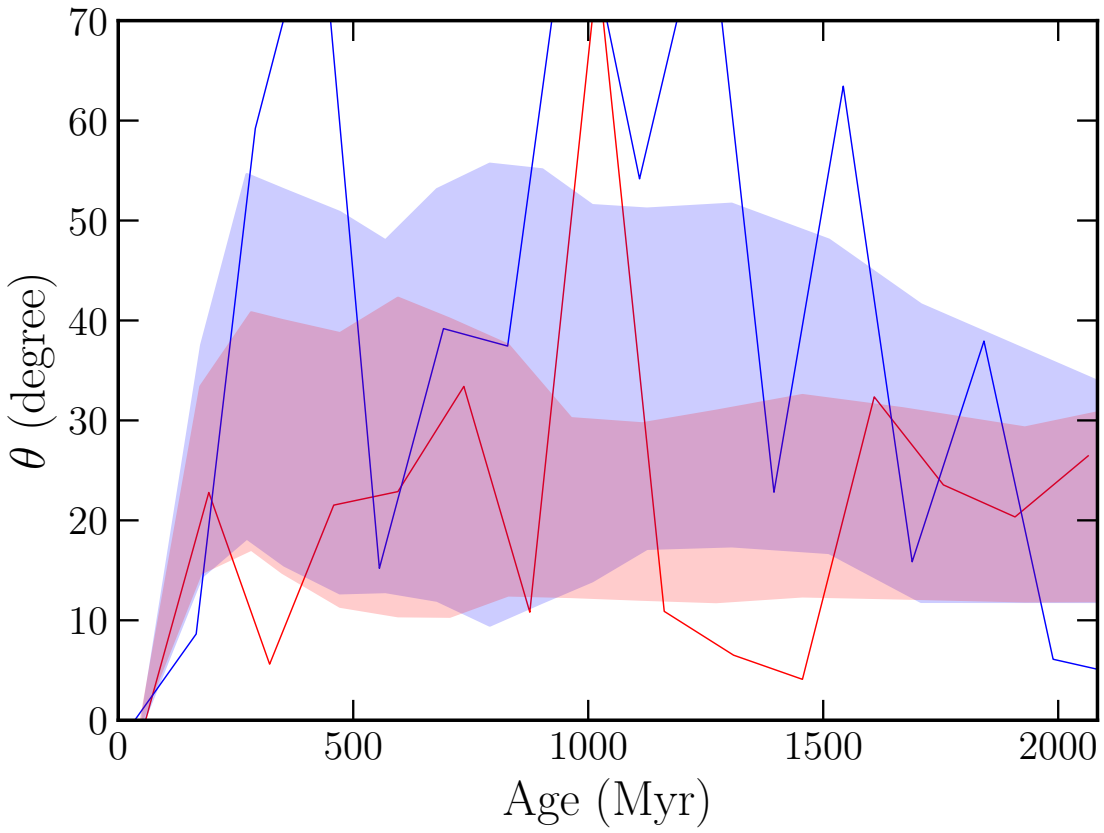

Figure 4.8 Angle between the orientation (normal) of young stars in a galaxy and the orientation of the group of stars chosen from this galaxy for Figure 4.7, as a function of the age of the tracked stars. This angle is between 0° and 90° and always starts from 0° in the snapshot when the chosen group itself defines the orientation of the galaxy. The red/blue shaded regions show the interquartile range for galaxies with stellar mass larger/smaller than $10^8\,M_\odot$ (roughly dividing the main progenitor branch and future satellites – see Figure 4.1). Thin red and blue lines trace individual example galaxies in each mass bin. Age bins are selected to contain equal number of points, and are different from Figure 4.7 because splitting by mass reduces the number of points.

but lower spatial resolution (100-300 pc) than our simulations. In TNG50 the half-light height of star-forming galaxies with $M_* = 10^7 - 10^{10}\,M_\odot$ in the rest-frame V band at $z = 2$ is 200-400 pc (Pillepich et al., 2019). Since the V band light is produced by stars of a wide range of age, this result should be compared to our age bins up to 500 Myr. Indeed, it agrees with the height of 100-500 Myr old stars in our galaxies. Pillepich et al. (2019) define disc thickness somewhat differently from us, as the ratio of the half-mass height to the 2D circularized half-mass radius, $z_{1/2}/r_{1/2}$, but the difference should be limited to tens of percent. For galaxies in the mass range $10^8 - 10^9\,M_\odot$ at $z = 1.5 - 2$, the ratio $z_{1/2}/r_{1/2}$ ranges from 0.15 to 0.35. More massive galaxies, with $10^9 - 10^{10}\,M_\odot$, are significantly less thick: $z_{1/2}/r_{1/2} \approx 0.1 - 0.2$. While these values are comparable to the average thickness of the 100-300 Myr old populations in our galaxies, we do not find as strong a dependence on galaxy mass. This difference may provide a useful observational test of the models.

The TNG50 results were not split by stellar age, but they provide another measure of height for the Hα emitting gas, which is a proxy for star formation rate and may indicate future location of very young stars. In all the TNG50 galaxies the height of the Hα gas is consistently smaller than that of the V-band stars, which indicates the age trend that we described above for our results.

In addition to measuring the disc height and length, Pillepich et al. (2019) studied the intrinsic shape of the stellar distribution in elliptical shells. They find that elongated (or prolate) galaxies are more common at higher redshift and lower galaxy mass: at $z > 2$ about half of the galaxies with $M_* = 10^7 - 10^9\,M_\odot$ have prolate shapes while the other half show spheroidal shapes. It is similar to our results (Figure 4.2) for the >100 Myr old stars. None of the TNG50 stellar distributions are as prolate as our young stars (<100 Myr old). The only comparably small values of b/a are shown by the distribution of Hα emitting gas, which may be a proxy for the very young stars. Therefore it is possible that in the TNG50 simulation stars also form only in

thin discs. A more direct investigation of this trend would be desirable.

In other suite of Vela simulations, Ceverino et al. (2015) investigated the shape of high-redshift galaxies by fitting an ellipsoid to the 3D isodensity surface around the half-mass radius. They show one example of a galaxy at z=2.2 with $M_* \approx 10^9\,M_\odot$ that has the axis lengths of the stellar distribution equal to $(a, b, c) = (2.2, 0.76, 0.67)$ kpc. Both the short axis length and the short-to-long axis ratio are similar to our results (Figure 4.2).

A number of simulation studies favor the "upside-down" formation scenario for galactic discs, wherein stars form in thick discs at high redshift with a kinematically "hot" configuration, and then in progressively thinner discs at lower redshift. For example, Buck et al. (2020) used the NIHAO-UHD simulations of isolated MW-sized galaxies to investigate the evolution of height of stellar populations split in age bins of Δt =2 Gyr. They find that the exponential scale heights at birth decrease with stellar age, i.e. the "upside-down" formation. They also find that the scale heights at present day are larger than those at birth. This suggests that the disc thickness today is set partly at the time of formation and partly by subsequent secular heating. Note that unlike our globally measured $z_{1/2}$, Buck et al. (2020) excluded the central galaxy and calculated the exponential scale height in the radial range $2 < R < 25$ kpc. The numbers they find for ∼10 Gyr old stars (i.e. formation at $z \approx 1-2$) are $z_d = 0.8 - 1.5$ kpc, which correspond to the half-mass height $z_{1/2} = z_d \ln 2 \approx 0.5 - 1.0$ kpc. This is within the range of thickness of the oldest stars in our galaxies. However, these estimates are likely to be incorrectly classified as "at birth". Buck et al. (2020) choose wide age bins of Δt =2 Gyr, during which the disc orientation can evolve significantly as shown by our results, and therefore the stellar thickness may already be artificially inflated. This is supported by inspecting their Figs. 13 and 14, which show 3 of 5 galaxies with larger scale height than scale length "at birth" for stars older than 10 Gyr. This confirms the disc precession which hides

the actual thickness of the star-forming disc. A finer time resolution is required to uncover the trend indicated by our results.

Other simulation studies that favor the "upside-down" disc formation scenario also take wide 1-2 Gyr age bins for mono-age stellar populations (e.g. Bird et al., 2013; Stinson et al., 2013; Ma et al., 2017; Navarro et al., 2018). They consistently find that old stars form in structures that are radially compact and relatively thick, while younger stars form in progressively larger, thinner, and colder configurations. We show that this cannot be used as confirmation of the "upside-down" scenario. The stars could either form in actually thicker discs or form in thin discs and get quickly displaced from the changing midplane. This is especially important at high redshift where more frequent mergers may change the galaxy orientation more often. Only tracing the stellar population from the very young stars can distinguish between these two cases. For example, a recent paper by Bird et al. (2020) analysed the upside-down formation scenario in simulated galaxies with smaller time intervals (50 Myr), illustrating both effects of the disc forming kinematically hotter at high redshift, and the stars being heated after birth. Since their analysis covers later epochs ($z < 1.5$) than our simulations and quantifies only the velocity dispersion instead of disc thickness, we cannot make a direct comparison with their results.

4.4.3 On the cause of thickening of galactic discs

There are many physical mechanisms that lead to real thickening of galactic discs with fixed orientation. Classic papers by Spitzer & Schwarzschild (1951) and Spitzer & Schwarzschild (1953) pointed out that gravitational scattering by massive gas clouds could increase the velocity dispersion of stars in the galactic plane, especially considering differential galactic rotation. Lacey (1984) generalized the calculation by considering the vertical motion of stars and found that the velocity dispersion increases steadily while the shape of the velocity ellipsoid remains fixed. More recently,

Bournaud et al. (2009) found that secular internal processes, such as gravitational instabilities and scattering by large clumps, lead to increased but constant disc height over galactocentric radius, while galaxy mergers lead to flaring of discs at large radii. Grand et al. (2016) showed the galactic bar and tidal perturbations from satellites to be responsible for most of the disc heating at $z \lesssim 1$ in the Auriga simulations.

In addition to these real processes, we have identified another important, and possibly dominant, mechanism leading to apparently larger axis ratios of high-redshift galaxies. It is the rapidly changing orientation of the disc midplane. For example, the intrinsic axis ratio for the same stellar population 1 Gyr since birth is only $\sim$0.4 (Figure 4.7), while the observable axis ratio for the oldest stars with the median age 800 Myr is already $\sim$0.6 (Figure 4.6). The difference comes from the different definitions of the galaxy orientation. In observations, only one orientation exists in the plane of the sky, and because of redshifting of starlight, it is defined by the distribution of blue young stars. Because of rapid midplane precession, this orientation may not align with the intrinsic orientation of older populations and artificially inflate the measured thickness.

Figure 4.8 illustrates the precession of the disc plane directly. The angle between the short axis of the configuration of tracked stars for Figure 4.7 and the short axis defined by the current young stars rises up to $\sim$20°-40° in less than 200 Myr and then oscillates around those values. The path of individual galaxies has a much wider variation than the interquartile trend, as illustrated by two cases in the figure.

The median angle offset is noticeably larger for low-mass galaxies, $M_* < 10^8\,M_\odot$. Less massive galaxies are more easily affected by accretion and mergers which can change the disc orientation. Also, the determination of the plane is more stochastic for low-mass galaxies simply because of the smaller number of stellar particles.

Overall, the plane precession is a significant contributor to the thicker appearance of high-redshift discs. Investigation of such offset at lower redshifts will be important

for disentangling this effect from the real disc expansion by secular and external heating. Unfortunately, our current simulation suite does not extend beyond $z \approx 1.5$. We plan to investigate this issue with new simulations in progress and encourage other groups with available simulations to look into it as well.

4.5 Conclusions

We investigated the thickening of high-redshift galaxy discs using a suite of ultrahigh-resolution cosmological simulations. We selected a galaxy sample with stellar mass $M_* = 10^7 - 10^{10}\,M_\odot$ at redshifts $z \geq 1.5$. We calculated the half-mass disc height and the axis ratio for stellar populations split into narrow age bins, from < 30 Myr to > 500 Myr. The main results are summarized below:

- Using the shape tensor, we find the intrinsic 3D shape of high-redshift galaxies to be prolate or elongated, unlike the local axisymmetric discs. Stars younger than 100 Myr are confined to a more prolate and thinner configuration, while older stars gradually transition to a spheroidal shape (Figure 4.2).

- Young stars in our simulated galaxies always form in thin discs, with half-mass height $\sim$0.1 kpc. The disc height increases from $\sim$0.04 kpc in low-mass galaxies to $\sim$0.2 kpc in higher mass galaxies. The disc height gradually increases with the age of stellar population, from $\sim$0.1 kpc to $\sim$0.8 kpc (Figure 4.5). There may be a weak trend for the disc height of young stars to increase with galaxy mass, but no systematic trend for stars older than 100 Myr (Figure 4.3).

- The short to long axis ratio $z_{1/2}/x_{1/2}$ also increases with the age of stellar population, from $\sim$0.15 to $\sim$0.6 (Figure 4.6). There is a slight trend for the axis ratio of the oldest stars to decrease with galaxy mass (Figure 4.4).

- We trace the same group of stars over consecutive simulation snapshots and

calculate the evolution of their intrinsic axis ratio (Figure 4.7). We confirm that stars form in thin discs $z_{1/2}/x_{1/2} \sim 0.15$ and then rapidly expand away from the current disc plane.

- However, in addition to the real kinematic heating of stars, we find a new effect contributing to thicker appearance of galactic disks. The disc plane in observations of high-redshift galaxies is defined by rest-frame UV light dominated by young stars. The orientation of this plane rapidly varies by $\sim$20°-40° (Figure 4.8), which mixes and artificially inflates the configuration of older stars. The plane continues to precess after $\sim$200 Myr.

In future studies it will be desirable to quantify the effect of disc precession using other simulations and at lower redshift. Since the mixing of stellar populations is rapid, with the timescale less than 200 Myr, this will require correspondingly high temporal resolution of the analyzed stellar populations. Previous studies advocating the "upside-down" disc formation scenario lacked sufficient time resolution to detect the plane precession effect. We plan to address these issues with upcoming galaxy formation simulations.

CHAPTER V

Tidal disruption of star clusters in galaxy formation simulations

This chapter was published in MNRAS with authors ***Meng, X.*** *& Gnedin, O. Y., 2022, MNRAS, 515, 1065*

Abstract

We investigate the evolution of the tidal field experienced by massive star clusters using cosmological simulations of Milky Way-sized galaxies. Clusters in our simulations experience the strongest tidal force in the first few hundred Myr after formation, when the maximum eigenvalue of the tidal tensor reaches several times 10^4 Gyr^{-2}. After about 1 Gyr the tidal field plateaus at a lower value, with the median $\lambda_{\rm m} \sim 3 \times 10^3$ Gyr^{-2}. The fraction of time clusters spend in high tidal strength ($\lambda_{\rm m} > 3 \times 10^4$ Gyr^{-2}) regions also decreases with their age from $\sim$20% immediately after formation to less than 1% after 1 Gyr. At early ages both the *in situ* and *ex situ* clusters experience similar tidal fields, while at older ages the *in situ* clusters in general experience stronger tidal field due to their lower orbits in host galaxy. This difference is reflected in the survival of clusters: we looked into cluster disruption calculated in simulation runtime and found that *ex situ* star clusters of the same ini-

tial mass typically end up with higher bound fraction at the last available simulation snapshot than the *in situ* ones.

5.1 Introduction

All massive galaxies ($M_* \gtrsim 10^9\,M_\odot$) in the local universe host globular cluster (GC) populations (Brodie & Strader, 2006). The overlapping metallicity, density, and mass distributions between GCs and young star clusters suggest that they share similar formation mechanisms, except that they differ in their cosmological history (Krumholz et al., 2019). Observations show that the combined mass of the GC system scales almost linearly with the host galaxy halo mass (Spitler & Forbes, 2009; Hudson et al., 2014; Harris et al., 2017; Forbes et al., 2018). This suggests that GC systems are tightly linked to the assembly history of their host halos. Therefore, studying the formation and evolution of GC systems can help us understand galaxy evolution better.

From a newly formed population of star clusters to old GCs there is a significant decrease in the fraction of low-mass clusters. To understand how the power-law initial mass function of young clusters transforms into the lognormal mass function of GCs at present (e.g. Kruijssen, 2009; Muratov & Gnedin, 2010; Li & Gnedin, 2014), we must understand the dynamical evolution of star clusters. There are internal processes, including stellar evolution and relaxation, as well as external processes, namely tidal disruption, that affect cluster properties. The mass loss due to stellar evolution for single stellar populations (Leitherer et al., 1999) and the two-body relaxation for isolated clusters (Spitzer, 1987) are well studied. However, since tidal disruption involves changing gravitational potential over the cosmic time, complicated orbits of clusters, encounters with massive objects, interactions with the dense disc etc., modelling the disruption effects on star clusters is not straightforward.

Analytical and semi-analytical models of tidal disruption usually assume a fixed

average tidal field without considering the position information (e.g., Boylan-Kolchin, 2017; Choksi et al., 2018; Choksi & Gnedin, 2019; El-Badry et al., 2019) or an analytical tidal field along idealised elliptical orbits in the host galaxy (e.g., Prieto & Gnedin, 2008; Webb et al., 2014; Rossi et al., 2016). Either assumption does not capture the cosmic evolution or the granularity of the gravitational potential. There have also been several studies that used numerical simulations of galaxy formation and galaxy mergers to obtain more realistic tidal fields for modelling of GCs disruption. These studies select a subset of stellar particles to represent GCs orbits. For example, Renaud et al. (2017) selected particles older than 10 Gyr in a cosmological zoom-in simulation of a MW-like galaxy and found that the tidal field grows stronger with cosmic time and that the *in situ* clusters experience significantly stronger tides than the accreted clusters. Halbesma et al. (2020) applied the same selection to the Auriga simulation suite but found that it cannot match the observed Milky Way (MW) and M31 GC populations. The E-MOSAICS simulations explored cluster formation times, mass and metallicity distributions, orbits and assembly of the GC system (e.g. Reina-Campos et al., 2019, 2020; Bastian et al., 2020; Keller et al., 2020; Pfeffer et al., 2020; Trujillo-Gomez et al., 2021). These relations can, in turn, be used to infer the merger history of the MW and possibly other galaxies (Kruijssen et al., 2019b, 2020). There are also merger and interacting galaxy simulations that explored the effects of merger environment on the formation and disruption of star clusters (e.g. Kruijssen et al., 2012; Kim et al., 2018). These efforts have greatly contributed to our understanding of GC evolution.

The previous work can still be improved with more realistic modeling of the formation of massive star clusters. The history of the tidal field experienced by real clusters may be significantly different from that of all old stellar particles. In this work, we present a study of the tidal disruption processes in a suite of cosmological simulations that directly model the formation of star clusters. We track the posi-

tion and tidal field of massive star clusters throughout cosmic time from their birth until redshift $z \approx 1.5$. We introduce our simulations and tidal field calculations in Section 5.2. In Section 5.3 we describe how the tidal field as well as the location of massive clusters evolve with time, and how that affects the bound fraction of the star clusters. In Section 5.4 we discuss caveats in our analysis and compare with other studies. We also discuss possible implications of our results for building more realistic analytical models, as well as recovering the assembly history of the MW. We present our conclusions in Section 5.5.

5.2 Simulations

We use a suite of cosmological simulations run with the Adaptive Refinement Tree (ART) code (Kravtsov et al., 1997; Kravtsov, 1999, 2003; Rudd et al., 2008) and described in Li et al. (2018b) and Meng et al. (2019). All runs start with the same initial conditions in a periodic box of 4 comoving Mpc, producing a main halo with total mass $M_{200} \sim 10^{12}\,M_{\odot}$ at $z = 0$, similar to that of the Milky Way. The ART code uses adaptive mesh refinement to achieve better spatial resolution in denser regions. The lowest resolution level is set by the root grid, which is 128^3 cells. This sets the dark matter (DM) particle mass $m_{\rm DM} = 1.05 \times 10^6\,M_{\odot}$. The finest refinement level is set to be kept between 3 and 6 physical pc. The ART code uses adaptive mesh refinement, and each particle (stellar particle and DM particle) contributes to the gravitational potential and feels the potential via the cell that it is in. The gravitational softening length is essentially the cell size, which is adaptive according to a combination of the Lagrangian and the Jeans refinement criteria. For details of the refinement criteria we refer the readers to Section 2 of Li et al. (2017b). The maximum refinement level of DM particles is 4 levels above the maximum refinement level of gas cells.

The simulations include three-dimensional radiative transfer using the Optically

Thin Variable Eddington Tensor approximation (Gnedin & Abel, 2001) of ionizing and ultraviolet radiation from stars (Gnedin, 2014) and the extragalactic UV background (Haardt & Madau, 2001), non-equilibrium chemical network that deals with ionization stars of hydrogen and helium, and phenomenological molecular hydrogen formation and destruction (Gnedin & Kravtsov, 2011). The simulations incorporate a subgrid-scale model for unresolved gas turbulence (Schmidt et al., 2014; Semenov et al., 2016b).

A unique advantage of these simulations for our study is direct modeling of bound star clusters, rather than generic stellar particles. Star formation is implemented with the continuous cluster formation (CCF) algorithm (Li et al., 2017b), where each stellar particle represents a star cluster that forms at a local density peak and grows its mass via accretion of gas until its own feedback terminates its growth. The feedback recipe includes early radiative and stellar wind feedback, as well as a supernova (SN) remnant feedback model (Martizzi et al., 2015; Semenov et al., 2016b). The momentum feedback of the SN remnant model is boosted by a factor $f_{\rm boost} = 5$ to compensate for numerical underestimation due to limited resolution and to match the star formation history expected from abundance matching. The simulations have been run with different values of local star formation efficiency (SFE) per free-fall time, $\epsilon_{\rm ff}$. For a full description of the star formation and feedback recipe, see Li et al. (2017b, 2018b).

For this analysis we use five runs with different local efficiency: SFE10, SFE50, SFE100, SFE200, and SFE200w. The number after 'SFE' in their names is $\epsilon_{\rm ff}$ in percent. There is also run SFE200w with more frequent output of the detailed tidal information for all massive stellar particles ("tidal writeout"). It is a rerun of the SFE200 run, but because of the rather explosive nature of star formation with high SFE, they are not exactly the same (Li & Gnedin, 2019).

We restrict our analysis to massive clusters because their disruption timescale is

longer under the same tidal field, giving them greater probability to survive to the present and become GCs. We take the threshold of the initial mass of the massive stellar particles to be $M_i > 3 \times 10^5\,M_\odot$ in all runs except SFE10, where we take $M_i > 2 \times 10^5\,M_\odot$ because that run has few massive stellar particles due to its low SFE. This selection resulted in 40, 73, 46, 121 and 133 star clusters chosen from the SFE10, SFE50, SFE100, SFE200 and SFE200w runs, respectively.

New massive stellar particles are identified in the halo of the main galaxy in each snapshot and then traced throughout the simulation. The center of the galaxy is defined as the location of maximum stellar density, found iteratively using smoothing kernels of decreasing size as in Brown et al. (2018). The galaxy plane is defined using the shape tensor and its principle axes for neutral gas within 10 kpc from the galaxy center. For full description of the determination of the galaxy plane see Meng et al. (2019). The cylindrical radius from the galaxy center (R) and the height above the disc plane (z) are calculated for all massive stellar particles.

The ambient matter density ρ is calculated on the scale of 100 pc as a sum of the gas density in the cell containing the stellar particle, and the average density of DM particles and stellar particles: $\rho = \rho_{\rm gas} + \rho_{\rm DM} + \rho_*$. The latter two are calculated by representing DM and stellar particles are constant density spheres of radius 100 pc, and summing them within a 100 pc radius sphere around the stellar particle.

The tidal tensor around massive clusters is calculated in post-processing of simulation snapshots as the second derivative of the potential:

$$T_{ij}(\mathbf{x}_0, t) \equiv - \left. \frac{\partial^2 \Phi(\mathbf{x}, t)}{\partial x_i \partial x_j} \right|_{\mathbf{x}=\mathbf{x}_0} \tag{5.1}$$

where i and j are two orthogonal directions in the Cartesian coordinate frame. We calculate the tidal tensor using the second-order finite difference across a $3 \times 3 \times 3$ cell cube centered on the stellar particle. We then calculate the three eigenvalues of

	Cell size (kpc)			$\lambda_{\rm m} > 10^4$ Gyr^{-2}	age<150 Myr
Run	25%	50%	75%	50%	50%
SFE10	0.13	0.19	0.28	0.10	0.17
SFE50	0.14	0.19	0.31	0.10	0.19
SFE100	0.14	0.21	0.34	0.10	0.17
SFE200	0.20	0.38	0.71	0.11	0.17
SFE200w	0.22	0.31	0.52	0.13	0.17

Table 5.1 The median and interquartile range of the sizes of cells containing massive star clusters in each run, combined over all analysed snapshots. The last two columns show the median size for a subset of cells with strong tidal field or young stellar particles.

the tidal tensor, $\lambda_1, \lambda_2, \lambda_3$, and use

$$\lambda_{\rm m} \equiv \max|\lambda_i|$$

as an upper limit of the tidal strength, following Li & Gnedin (2019). In the rest of this paper we use $\lambda_{\rm m}$ to refer to the tidal strength. According to the Poisson equation, the trace of the tidal tensor $\lambda_1 + \lambda_2 + \lambda_3$ equals $-4\pi G\rho$.

The gravitational potential calculated in a cosmological simulation contains a cosmological term due the universal expansion (e.g., Martel & Shapiro, 1998). We confirmed that this term is much smaller than the Newtonian potential we wish to evaluate around star clusters. It contributes less than 0.4 Gyr^{-2} to T_{ij} at redshifts $z < 9$, while the typical Newtonian values are above 10^3 Gyr^{-2}. Therefore, we can differentiate the values of Φ taken directly from the simulation snapshots.

We chose the averaging scale of the ambient density of 100 pc because it matches best the scale on which we calculate the tidal tensor. The size of cells containing the traced stellar particles at each snapshot ranges from tens of pc to kpc scale, as they appear in various locations in the halo. We include the median value and the interquartile range of the cell sizes for each run in Table 5.1. The median of the total distribution is $200 - 300$ pc, while the cells containing clusters younger than 150 Myr

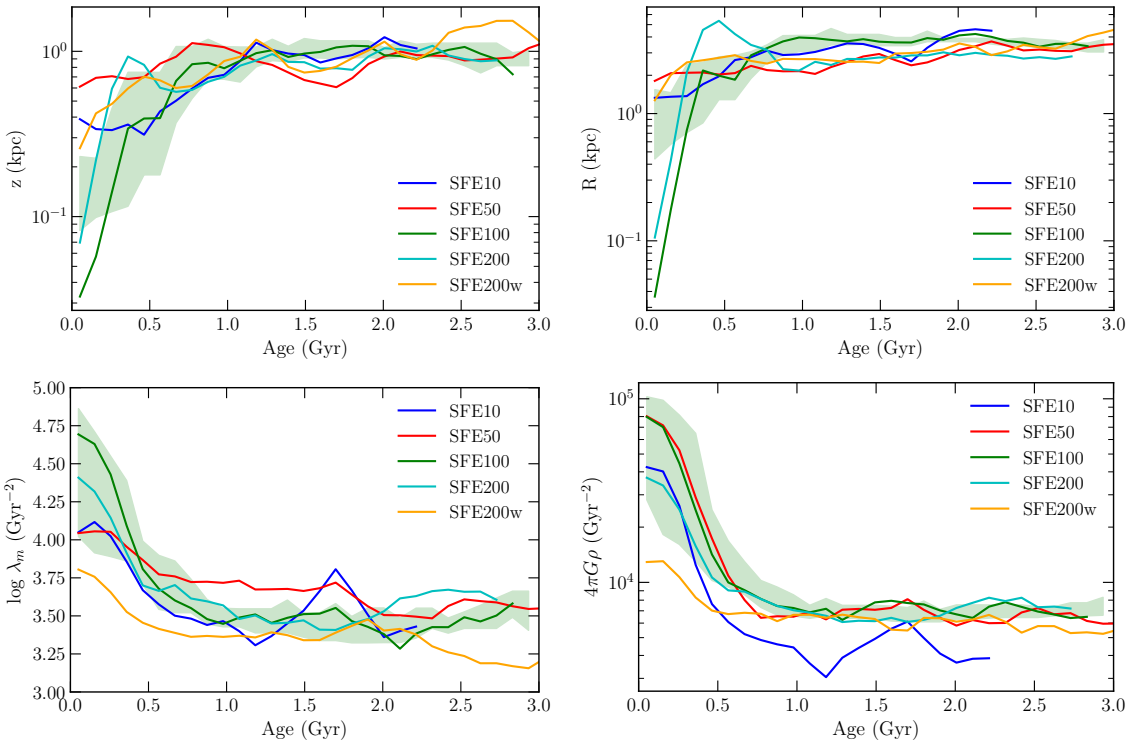

Figure 5.1 Median values of height z, radius R, maximum eigenvalue of the tidal tensor $\lambda_{\rm m}$, and ambient density ρ in age bins from the simulation snapshots. Only star clusters born in the main galaxy are included (first found within 7 kpc from the center and younger than 150 Myr). The lines are smoothed with a Savitzky-Golay filter with window size of 900 Myr. Shaded regions are the interquartile range (25-75%) in a moving window of 700 Myr for the SFE100 run.

have median size about 170 pc. In regions with strongest tidal field ($\lambda_{\rm m} > 10^4$ Gyr^{-2}) the cells are 100–130 pc. Therefore, the averaging scale of the ambient density does not need to be much smaller than that. The 75% percentile for the SFE200 run is larger than for the other runs because it has a group of stellar particles in a low density region far from the galaxy center. This group biases the overall distribution of cell sizes. We will discuss the scale for the tidal tensor calculation in more detail in Section 5.4.

5.3 Results

5.3.1 Overall evolution of cluster location, tidal field, and ambient density

First we focus on *in situ* star clusters, which remain in the main halo at all times and do not transfer to another halo during mergers. Our criterion for a cluster to be identified as formed *in situ* is to be first found within 7 kpc from the main galaxy center and younger than 150 Myr. We chose the 7 kpc threshold to guarantee that the clusters formed in the main galaxy. If we change this criterion to 5 kpc, it would only change the identified origin of one cluster. We chose the 150 Myr age threshold because the largest time interval between two consecutive snapshots is about 150 Myr, and we want to ensure that the clusters we found were not formed earlier in other places and then brought in to the main galaxy.

We exclude a group of clusters in the SFE200 run that are likely born in a nearby satellite galaxy at $z \approx 5.2$, although they fit the other *in situ* criteria. This satellite is at ~ 2.5 kpc from the main galaxy center at that snapshot, so it would be challenging to revise the criterion to avoid this group of clusters automatically. However, this group has highly elliptical orbits around the main galaxy and never gets as close to the center again, so we do not consider them as born *in situ* and exclude them from Figure 5.1.

We trace the evolution of the four variables (z, R, ρ, $\lambda_{\rm m}$) for these massive clusters in the simulation snapshots. We first describe the overall evolution of these quantities with cluster age. In Figure 5.1 we show the median values in age bins for clusters born in the main galaxy. This plot shows that *in situ* star clusters tend to move away from the center and out of the galaxy disc, into the regions of lower density and weaker tidal field. We can also see that the four tidal variables change the most in the first 500–1000 Myr after cluster formation, and subsequently reach almost constant plateaus. Note that the lines show only the average trend – most of *in situ* clusters (about 70%) move outward, but some ($\sim$30%) move inward. Some clusters have elliptical orbits that are on average farther away from the galaxy center than at formation. Since the snapshots are taken at arbitrary times, the clusters can be found at random phases of their orbits. The outward-going trend is thus an average effect of all massive star clusters that are born in the main galaxy.

Note that the position (z, R) and tidal field related properties ($\lambda_{\rm m}$, ρ) are not independent of each other. Density decreases with distance from the galaxy center and from the disc plane, and tidal strength is positively correlated with density. In Figure 5.2 we show a correlation between the galactocentric distance and the tidal field at a given point. Here we include all star clusters (not only *in situ* ones) because this relation describes the overall potential of the main galaxy. The distribution extends far into the lower right corner because several clusters move to large galactocentric radii, and thus small $\lambda_{\rm m}$ regions, and some of the *ex situ* clusters come from large radii. These star clusters can bias the whole distribution due to a limited number of massive star clusters we trace. The farther away from the center, the weaker is the tidal field and the lower is the ambient density. A similar negative correlation holds for $\lambda_{\rm m}$ and the height above the disc plane z, but with a different slope from the relation in Figure 5.2 and larger scatter. The values of $\lambda_{\rm m}$ closely correlate with $4\pi G\rho$. This means the decrease in the tidal strength experienced by clusters is usually

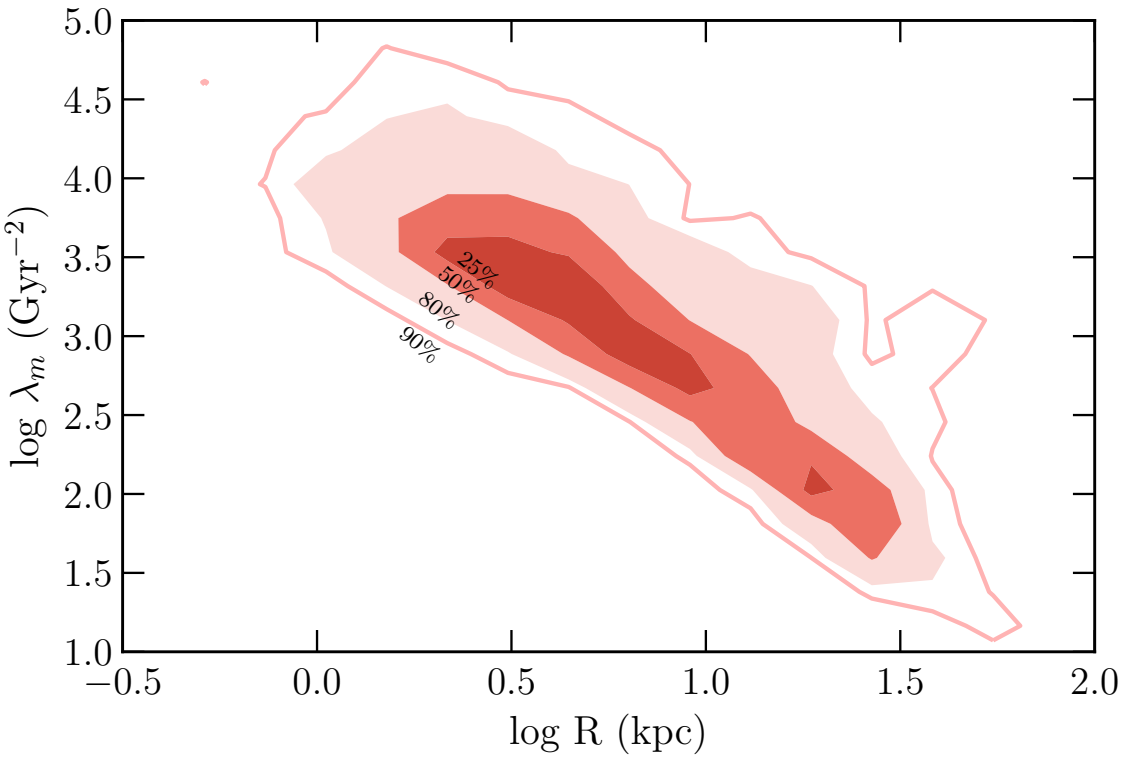

Figure 5.2 Distribution of $\lambda_{\rm m}$ and R shown as 25%, 50%, 80%, and 90% contours of all traced massive star clusters. All five simulation runs and all epochs are combined. There is a negative correlation between $\lambda_{\rm m}$ and R with a power law slope of -1.23.

correlated with the clusters going away from the disc plane, out of the galaxy, and into the lower density regions.

As we can see from Figure 5.1, the position and tidal field of the star clusters do not change linearly with time. We can calculate a slope of this evolution, but must be careful on what timescale to calculate it. The slope is steeper and has larger scatter among different clusters shortly after cluster formation. Therefore, for the following analysis we discuss slopes of the quantities over the first 1 Gyr after formation, since this is the interval when their position and tidal field change the most.

5.3.2 Time spent in strong tidal fields

The disruption processes of clusters are integrated effects over time. Both the intensity of the tidal field and the duration of strong tidal events matter the most

in cluster disruption. However, not only the finite spatial resolution makes us underestimate the tidal strength, but the time between our simulation snapshots is also relatively long ($\sim$100–150 Myr). We only have access to the potential at these snapshots, which are taken at random phases of the orbit and likely miss the short periods when the clusters experience strong tidal field.

We consider two ways to get around this limitation. We can either reconstruct the orbits between two snapshots using the information at the snapshots, or re-run the simulation and output the tidal field at shorter intervals. Figure 5.3 shows the former, an example of calculating one cluster orbit in the SFE50 run assuming the potential is fixed between two consecutive snapshots. We deduct the bulk velocity of the galaxy to account for the moving galaxy center. Every block of gray curves is one orbit calculated between two snapshots. The blocks are not smoothly connected because the calculated orbits cannot guarantee to match the points of the next snapshot. This figure illustrates that z, R, and $\lambda_{\rm m}$ vary rapidly and the values taken at the simulation snapshots may not be representative of the actual orbit.

We compared the tidal field values of our calculated detailed orbit with that from the tidal writeout. The values are close within $\sim$100 Myr from the calculation start point, but our calculated orbit could not capture the tidal field if there are encounters with high-density structures. We also tried interpolating the potential between the start and end epochs and it does not seem to improve the accuracy of our calculated orbits.

We compare the slopes calculated from separate snapshots and detailed orbits and found the slopes from snapshots can basically match slopes from detailed orbits. We fit the median of the values and median of the time of these red lines. We also connect the values at each snapshot by blue lines, so that each section of the gray curves start on the blue lines, and fit a slope to these data. We compared the slopes of these two linear fits and found the two slopes are basically consistent but with

a relatively large scatter. This shows that, although using snapshots loosely spaced in time loses information of detailed orbital evolution, it captures the overall low-frequency evolution of position and tial field on a larger timescale, as presented in the analysis in the previous section.

Since the disruption of a star cluster is related to the integrated time it spends in tidal field, we calculate the fraction of the time that a given cluster experienced in tidal field with $\lambda_{\rm m} > 3 \times 10^4$ Gyr^{-2}. The choice of this threshold is based on the disruption timescale as well as the evolution of $\lambda_{\rm m}$ with age – most star clusters have $\lambda_{\rm m}$ below several times 10^4 Gyr^{-2}. In Figure 1 of Li & Gnedin (2019), based on the same simulation suite, 10^4 Gyr^{-2} is also a fairly high value that is reached in a very small portion of the time a few hundred Myr after the example cluster is formed. We then average this time fraction for all *in situ* star clusters in the SFE50 run and plot the result in Figure 5.4. In this figure we can see that immediately after formation clusters spend more time in regions of high tidal strength. After they grow older and move away from the galaxy center they spend less time in high tidal field regions and experience less disruption. This is consistent with the previously mentioned outward motion of the majority of the *in situ* clusters, and the decreasing tidal strength with cluster age.

Our chosen threshold value for $\lambda_{\rm m}$ is somewhat arbitrary, and therefore we also show results for lower threshold values. As expected, decreasing the threshold increases the time fraction. When the threshold value decreases to 5×10^3 Gyr^{-2}, the time fraction rises to ∼40–50% in the beginning. This indicates that this threshold value 5×10^3 Gyr^{-2} is a typical value of the tidal strength in the first Gyr after the formation of these massive star clusters. Rather than using the arbitrary time and location given by the widely separated snapshots, with the detailed orbits we can see that even immediately after formation the clusters do not spend all of their time in strong tidal field regions, but only ∼10% of the time. As they grow older, this

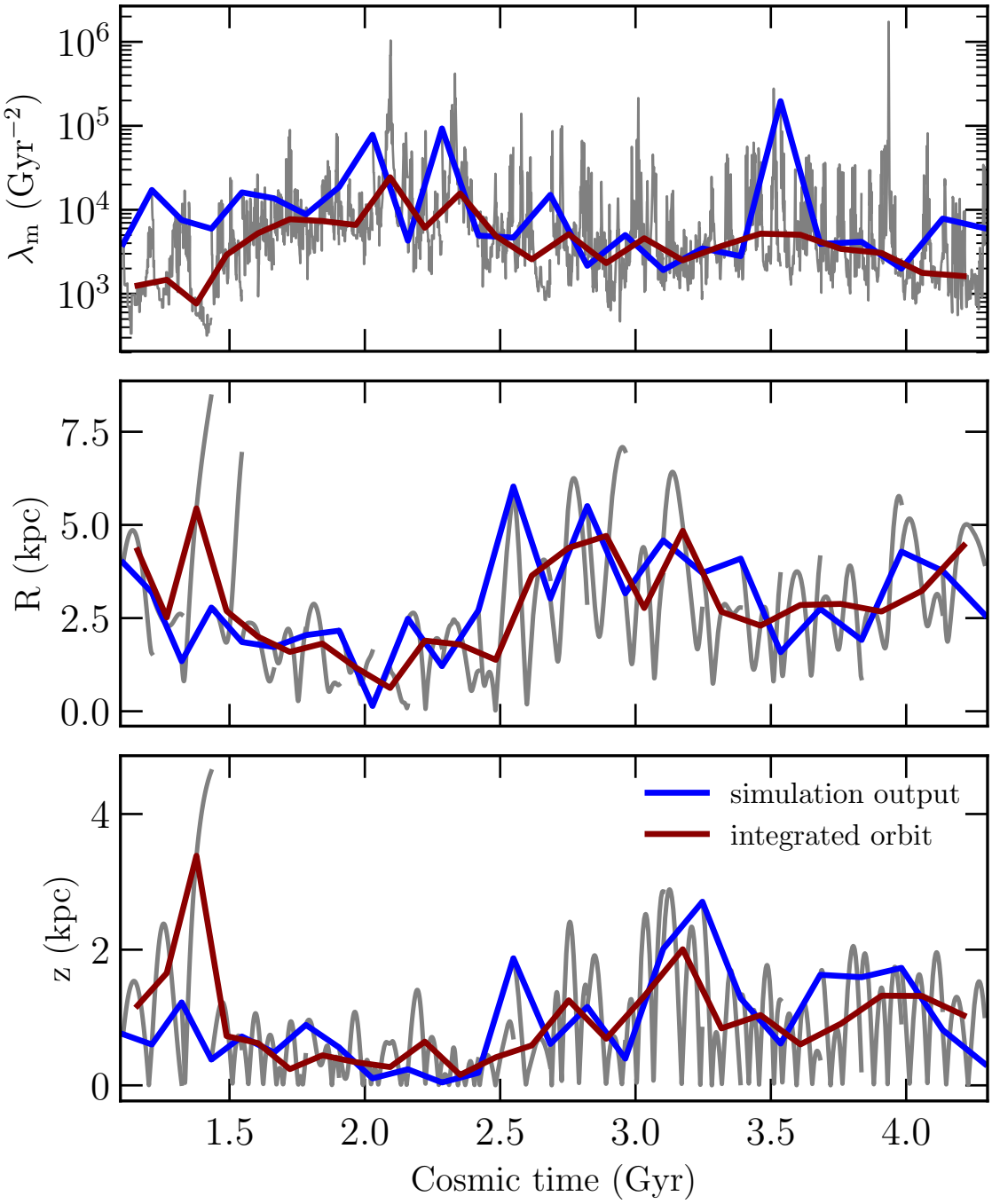

Figure 5.3 Calculated orbit for one star cluster (gray lines) assuming the potential is not changing between simulation snapshots. Blue lines connect values output at the snapshots. The actual orbit goes through a wide range of distances from the galaxy center and tidal field strengths, so that the snapshot values may not be representative of the tidal history (although the orbit post-processing is definitely inaccurate). Red lines are the medians of the post-processed quantities between the snapshots.

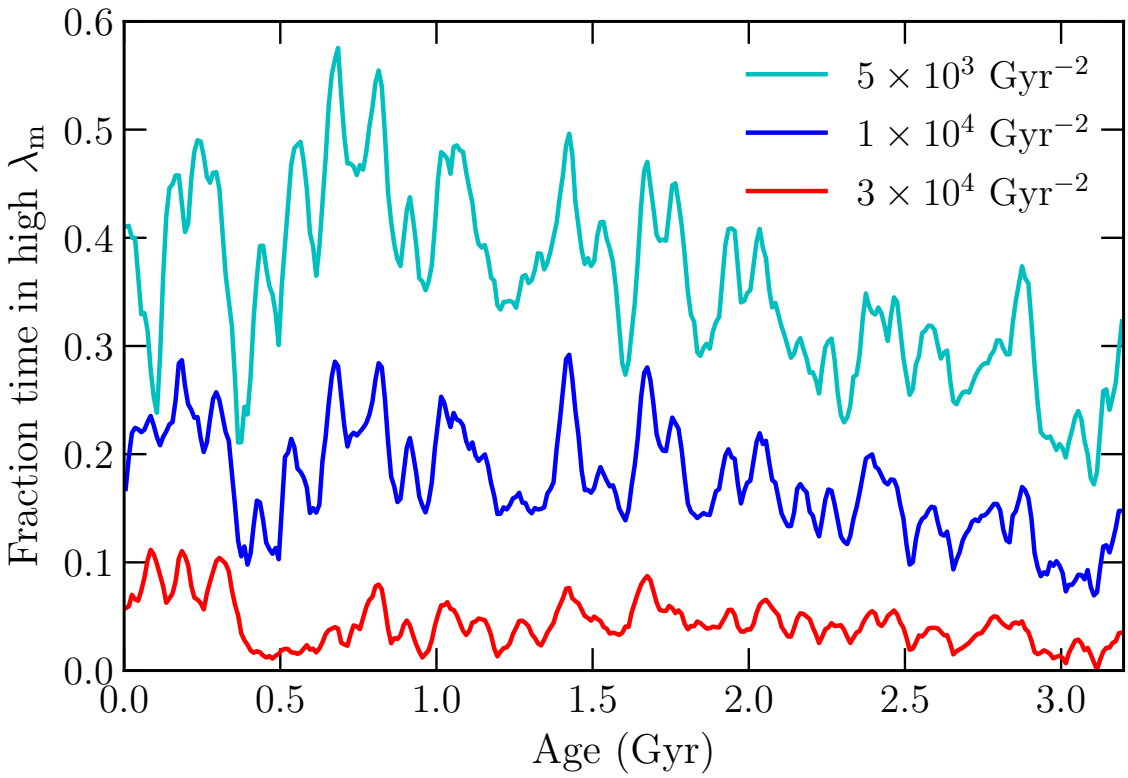

Figure 5.4 Average fraction of time in strong tidal field above three threshold values ($\lambda_m > 3 \times 10^4$ Gyr^{-2}, 1×10^4 Gyr^{-2}, and 5×10^3 Gyr^{-2}) for 38 star clusters in SFE50 run. The fraction is calculated in bins of 10 Myr from detailed orbits integrated assuming fixed potential between the snapshots. The time fraction in strong tidal field regions is larger at early ages.

fraction drops to a few percent.

As an alternative to reconstructing cluster orbits, we have the SFE200w run with detailed output of all massive star clusters at densely spaced intervals of 0.1 Myr. We also compared the R and λ_m slopes from snapshots and from tidal output for this run, to examine if using tidal field and position from snapshots is acceptable. It turns out that the slopes are consistent.

We could also calculate the average time in strong tidal field regions for this one run with detailed tidal output. Using this SFE200w run, in Figure 5.5 we plot the average time fraction in strong tidal fields for *in situ* and *ex situ* star clusters. We do not include star clusters that have not yet merged into the main halo by the last

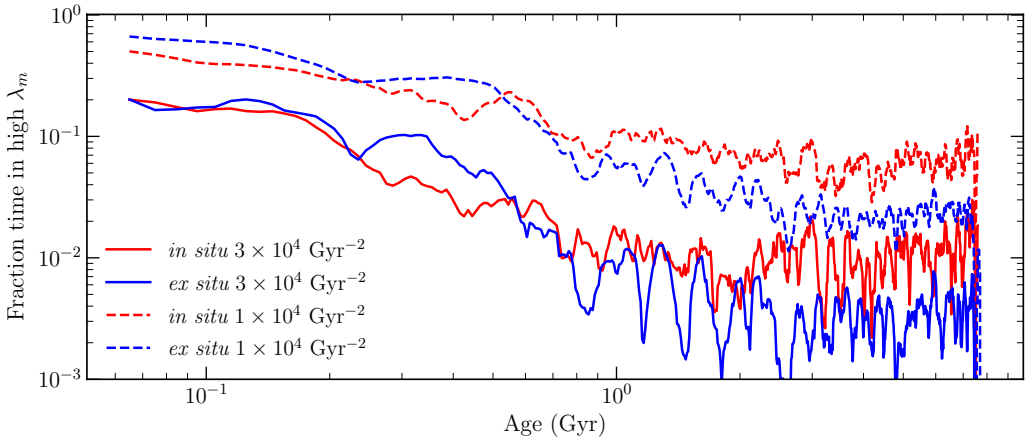

Figure 5.5 Average fraction of time in tidal fields with $\lambda_{\rm m} > 3\times10^4$ Gyr^{-2} (solid lines) and $\lambda_{\rm m} > 1\times10^4$ Gyr^{-2} (dashed lines) in SFE200w run. The fraction is calculated in 10 Myr bins and smoothed with the Savitzky-Golay filter. The average value for all *in situ* clusters is shown as red lines, the average for all *ex situ* clusters as blue lines. Again, the time fraction in strong tidal field is higher at early ages and decreases with time. *Ex situ* star clusters end up farther from the center, in weaker tidal fields.

available snapshot ($z \approx 0.79$).

The evolution of the time fraction in the SFE200w run is similar to the SFE50 run shown in Figure 5.4: $\sim$ 10–20% in the beginning and then a drop to a few percent at old ages. The average time fractions in strong tidal field regions for the *in situ* and *ex situ* clusters are similar in the first few million years after formation but differ significantly at old ages. When their host galaxies merge with the main galaxy, the *ex situ* clusters have larger total energy than the *in situ* clusters, and therefore end up farther away from the galaxy center and in weaker tidal field regions. We discuss this more in the next section.

5.3.3 Starting and ending locations

Star clusters have different tidal histories depending on where they were born. This difference could reflect in different expectations of final bound fraction for star clusters formed in the main galaxy or satellites, and also scatter in the bound fraction for clusters from the same origin. Inspired by Figure 5.5, we examine the evolution

of cluster location in Figure 5.6. If the initial distance from the galaxy center $R_{\rm init}$ is small, then the cluster is most likely born *in situ*, and vice versa. We divide all star clusters into the inner ($R_{\rm init} \leq 7$ kpc) and outer ($R_{\rm init} > 7$ kpc) groups to roughly distinguish *in situ* and *ex situ*. Note that unlike Section 5.3.1, here we do not exclude the group of clusters that formed in a satellite galaxy near the pericenter of its orbit around the main galaxy.

We find that about 80% of the *in situ* clusters move outwards from their birth locations. At the final snapshot, most *in situ* clusters still end up within ∼7-8 kpc of the galaxy center. As expected, most of the *ex situ* clusters end up much farther from the galaxy center. *Ex situ* clusters coming from the same origin most likely form a stream and have similar orbits in the main halo. In a given snapshot, they may be at different orbital phases but all of them can reach the furthest point of its group.

In Figure 5.7 we show examples of the orbits of two *in situ* clusters in the SFE200w run, corresponding to the orange stars with black edges in Figure 5.6. One cluster stays deep in the main halo throughout the whole simulation time. The other escapes the main galaxy on a highly elliptical orbit after the energy jumps which usually happen during galactic mergers. The orbits are color-coded by the tidal strength $\lambda_{\rm m}$. The inner cluster remains in the regions of relatively strong tidal field throughout the simulation time, while the escaping cluster experiences weaker tidal field most of the time. Note that the times of output snapshots do not coincide with the timing of strong tidal field, especially for highly elliptical orbits as clusters spend most of the time at larger radii. From this example we can see that even star clusters formed in the same galaxy could have different orbits and therefore different tidal histories, and the final survival of these clusters.

We can also examine the difference in the tidal field experienced by *in situ* and *ex situ* clusters by looking at the host galaxy mass. In Figure 5.8 we not only include clusters that end up in the main halo, but also take a few massive star clusters in all

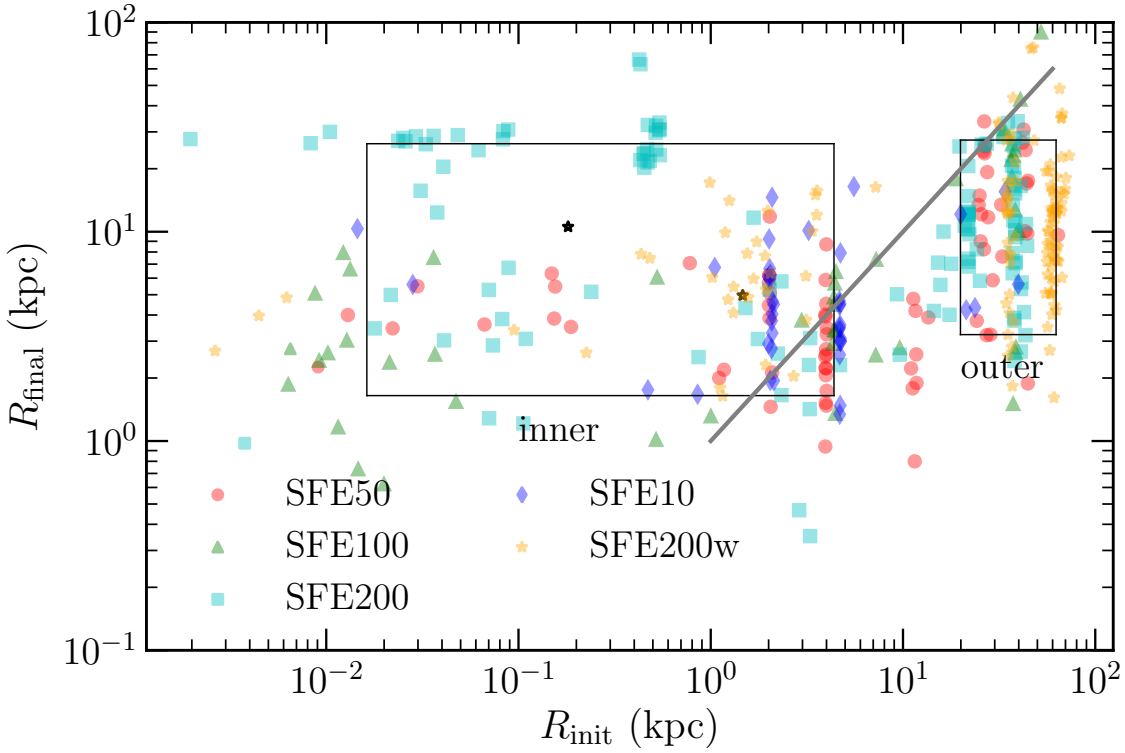

Figure 5.6 Radius from the center of the main galaxy in the final snapshot vs. radius when the clusters initially appear. Rectangular frames mark the 20%−80% percentiles of the inner ($R_{\rm init} \leq 7$ kpc) and outer ($R_{\rm init} > 7$ kpc) groups of clusters. The gray line marks 1:1 relation. Majority of clusters born in the main galaxy (∼80%) move outward. (This fraction is different from the 70% we quote in Section 5.3.1 because the criterion for *in situ* is not exactly the same and we are comparing starting and ending position instead of in 1 Gyr.) On the other hand, most of clusters brought in by mergers (outer group) move inward. Nevertheless, majority of the *in situ* clusters still end up closer to the galaxy center then the *ex situ* clusters.

halos and record their host halo mass at the time of formation. Then we compare the tidal field strengths before the star clusters are 100 Myr old and after they are 2 Gyr old, which we call 'initial' and 'final' values. Unlike the previous plots, here we do not use any threshold on $\lambda_{\rm m}$. The initial tidal strength does not differ much for clusters formed in smaller or larger host galaxies, but there does seem to be a difference in the final tidal field. This is consistent with what we found before – clusters that form in smaller galaxies (which will potentially merge with the main galaxy) end up in less dense regions, and farther away from the galaxy center.

5.3.4 The bound fraction

We can now look at the bound fraction of the star clusters we trace. Since star clusters lose mass due to a variety of processes, we write the fraction of the initial cluster particle mass that remains gravitationally bound at age t as

$$f_{\rm bound}(t) = f_i \, f_{\rm se}(t) \, f_{\rm dyn}(t) \tag{5.2}$$

where f_i is the initial bound fraction accounting for gas expulsion during cluster formation stage, $f_{\rm se}$ is accounting for mass loss due to stellar evolution, and $f_{\rm dyn}$ is accounting for tidal stripping of stars. The initial bound fraction f_i the fraction of mass in a star-forming complex that remains bound in the early phase, calculated as

$$f_i \equiv \min\left(\frac{\epsilon_{\rm int}}{\epsilon_{\rm core}}, 1\right)$$

where $\epsilon_{\rm core} = 0.5$ is the correction factor for mass loss due to protostellar outflows, and $\epsilon_{\rm int}$ is the integral star formation efficiency – the mass of the active stellar particle divided by the maximum baryon mass of the GMC throughout the whole course of cluster accretion (see Li et al. (2018b) for more details). The fraction $f_{\rm dyn}$ is calculated in run time using the current tidal tensor for each star cluster (Li &

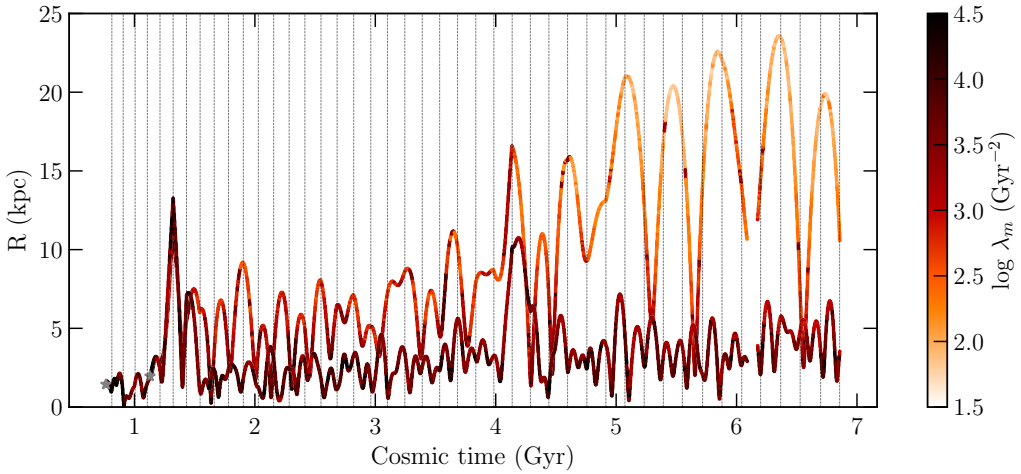

Figure 5.7 Examples of the orbits and $\lambda_{\rm m}$ of two clusters in the SFE200w run. One stays in the main galaxy, one goes to a higher energy orbit. The gray stars are the starting points (around 15 Myr of age) of the two curves.

Gnedin, 2019). Since here we focus on disruption by tidal process, i.e. $f_{\rm dyn}$, for simplicity we use the phrase "bound fraction" to represent $f_{\rm dyn}$ in the following text.

The timescale for tidal disruption can be approximated as

$$t_{\rm tid}(M,\,\Omega_{\rm tid}) = 10\,{\rm Gyr}\left(\frac{M(t)}{2\times10^5\,M_\odot}\right)^{2/3}\frac{100\,{\rm Gyr}^{-1}}{\Omega_{\rm tid}(t)}. \tag{5.3}$$

where $M(t)$ is the current cluster mass, and the frequency $\Omega_{\rm tid}(t)$ is related to the maximum eigenvalue of the tidal tensor along the cluster trajectory as

$$\Omega_{\rm tid}^2 = \frac{1}{3}\lambda_{\rm m}$$

(Li & Gnedin, 2019). This is a simplification of the actual disruption calculation that must include the full tidal field and the structure of the cluster. We use it here fore convenience to obtain first estimates of the cluster mass loss. We compare the results with an alternative simplified expression for $t_{\rm tid}$ in Appendix **??**. We find general consistency between the two methods, however note that neither expression can give an accurate value for the actual mass loss.

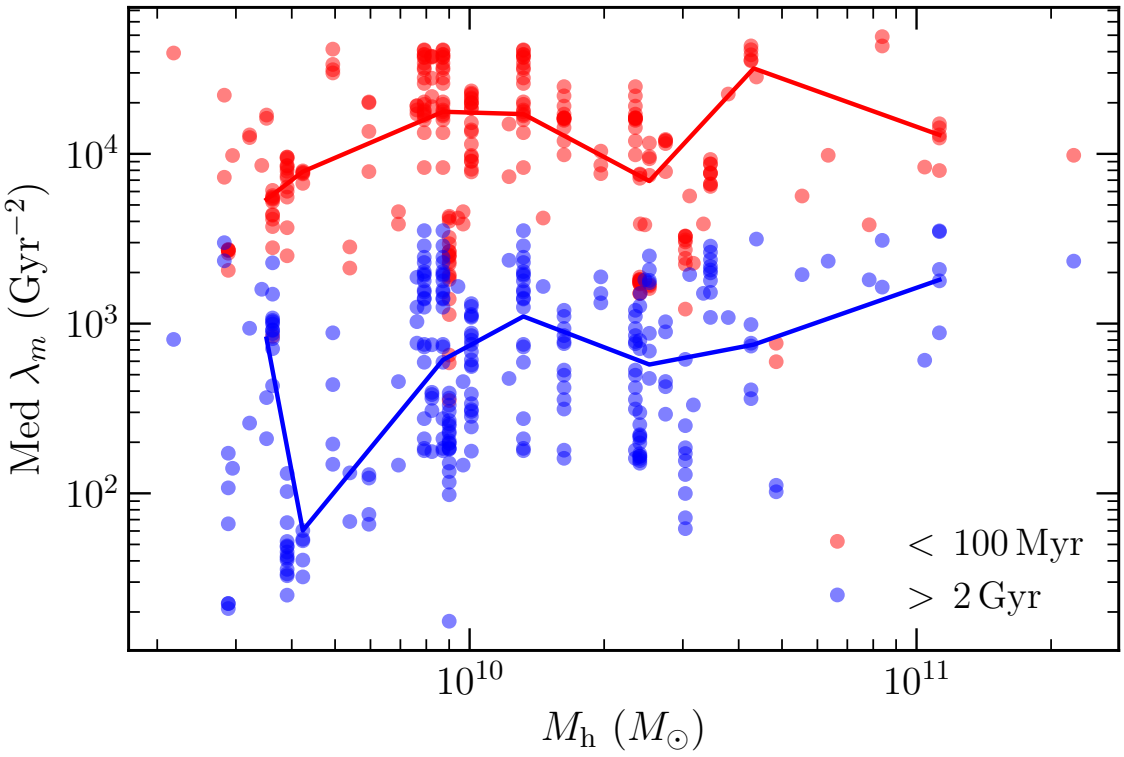

Figure 5.8 Median values of $\lambda_{\rm m}$ over two time periods: right after cluster formation (red) and at late times (blue) for star clusters in the SFE200w run, vs. host halo mass at formation. Each point represents one cluster. The lines are the median of the corresponding points. There is no noticeable dependence of the tidal field at birth on the host mass. In contrast, the final tidal field strength increases with host mass. Clusters born in smaller halos tend to end up in weaker tidal field regions, possibly because when they merge into the main halo they end up farther from the center.

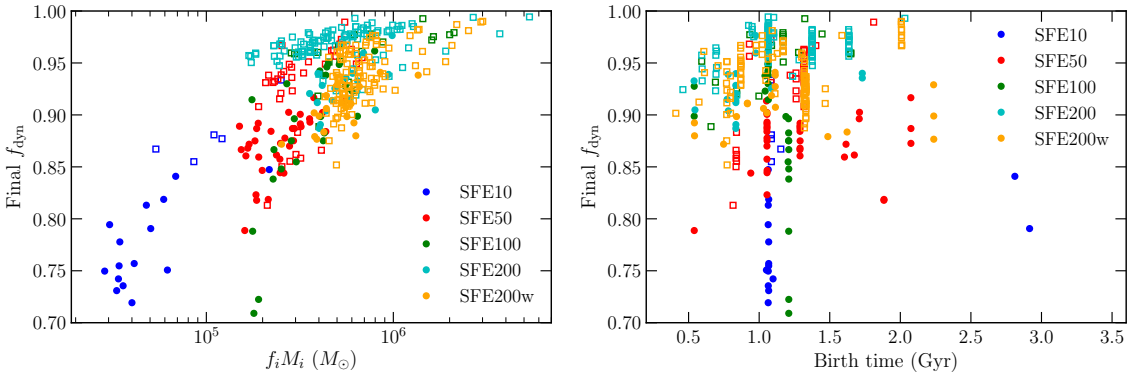

Figure 5.9 Relation between the bound fraction of clusters at last available snapshot and their initial mass (*left panel*) or birth time (*right panel*). Different colours represent different runs. Solid circles are for *in situ* clusters, open squares are for *ex situ* clusters. The final bound fraction positively correlates with both the initial mass and birth time. Star clusters born in the main galaxy generally end up with smaller bound fractions than those born in satellite galaxies.

During one time step from time t_n to $t_n + \mathrm{d}t$, the bound fraction $f_{\rm dyn}$ was calculated as

$$f_{n+1} = \exp\left(-\frac{\mathrm{d}t}{t_{\rm tid}^{\rm out}}\right) f_n.$$

However the disruption timescale $t_{\rm tid}^{\rm out}$ used in these simulations had an old normalization factor of $41.4\,\mathrm{Gyr}^{-1}$ instead of $100\,\mathrm{Gyr}^{-1}$, and also because of a coding mistake, the value of $\lambda_{\rm m}$ was overestimated by the expansion scale factor $a_{\rm box}(t)$. We correct for this mistake using a method described in Appendix A.

5.3.5 Testing the tidal disruption timescale

Tidal disruption is an important component of semi-analytical models of globular cluster evolution. Analytical models usually use a single parameter to describe tidal strength. For example, Choksi et al. (2018) introduced a parameter to normalize the strength of average tidal field along cluster trajectories. Choksi et al. (2018) found that the average tidal field $\Omega_{\rm tid} = 200\,\mathrm{Gyr}^{-1}$ matches best the observed mass function of globular clusters systems in nearby galaxies. Our results allow us to calibrate the value of the average tidal field.

From the models of tidal disruption, we expect the cluster bound fraction to depend on cluster age t and initial mass $M_{\rm init}$. If a cluster experiences constant tidal field throughout its lifetime, then the bound fraction can be derived by integrating Eq. 5.3. The expected value is

$$\tilde{f}_{\rm dyn}(t) = \left[1 - \frac{2}{3}\frac{t}{10\,{\rm Gyr}}\frac{\Omega_{\rm tid}}{100{\rm Gyr}^{-1}}\left(\frac{M_{\rm init}}{2\times10^5\,M_\odot}\right)^{-2/3}\right]^{3/2}. \tag{5.4}$$

According to this solution, clusters with larger initial mass will have larger bound fraction after the same time. Our higher SFE runs form star clusters with larger initial mass than lower SFE runs, and therefore should have more clusters with large final bound fraction. Apart from $M_{\rm init}$, tidal history also plays a role: clusters born in the main galaxy should have smaller final bound fraction than clusters born in satellite galaxies and brought to the main galaxy by mergers.

Figure 5.9 shows that these general trends are present in our simulations. From the left panel we can see that star clusters with larger initial mass tend to have larger final bound fraction, while from the right panel we can see that clusters formed later tend to have larger bound fraction. Since the cluster initial mass correlates with SFE, there is also a correlation between SFE and the final bound fraction: star clusters in runs with lower SFE have lower $f_{\rm dyn}$.

We also distinguish between *in situ* and *ex situ* clusters, using the same criteria as in Section 5.3.1. We consider the group of clusters that were likely born in a close satellite at $z \approx 5.2$ in the SFE200 run as *ex situ*, just as in Figure 5.1. We can see that the *ex situ* clusters have higher final bound fraction than the *in situ* ones, indicating that they experience different tidal history, and that the *ex situ* clusters on average experience weaker tidal field. This can be seen more clearly in the right panel, where is a quite obvious distinction between the *in situ* and *ex situ* samples.

Different tidal history for clusters from different origins (host galaxies) has a no-

ticeable effect on their final bound fraction. Although the tidal strength immediately after formation is similarly high for clusters born in halos of all mass, the tidal field differs more at late times and affects the final bound fraction. Clusters formed in smaller satellite galaxies, which tend to end up farther away from the main galaxy, will have larger final bound fraction than clusters formed in larger galaxies.

Figure 5.10 shows the evolution of the bound fraction with age for clusters from different origins. Different initial masses of the clusters result in a spread of values of $f_{\rm dyn}$. In general, the bound fraction for clusters from satellites decreases more slowly than for clusters formed in the main halo.

The rate of decrease of $f_{\rm dyn}$ generally slows down with cluster age. This differs from the expectation of Equation 5.4 shown as a triangular frame, because the tidal field strength along cluster orbits cannot be treated as a constant. To relate our results to the formulation of Equation 5.4 we define the 'effective' value of $\Omega_{\rm tid}$ that would produce the actual bound fraction $f_{\rm dyn}$ of the simulated clusters:

$$\Omega_{\rm tid,eff} = \frac{3}{2}\frac{10\,{\rm Gyr}}{t}\left(\frac{M_{\rm init}}{2\times10^5\,M_\odot}\right)^{2/3}(1-f_{\rm dyn}^{2/3})\times100\,{\rm Gyr}^{-1} \tag{5.5}$$

and plot $\Omega_{\rm tid,eff}$ in Figure 5.11. The figure shows significant scatter of $\Omega_{\rm tid,eff}$ for all host galaxies, which reflect a variety of orbital histories in every galaxy. We split the figure into two panels for *in situ* and *ex situ* clusters. *In situ* clusters in general experience stronger tidal field than *ex situ* clusters. At the same time, we can see that the average $\Omega_{\rm eff,tid}$ value for *ex situ* clusters increases with the host mass, which means that star clusters formed in more massive halos tend to end up with smaller bound fraction than those formed in less massive halos.

We compare our results with the GC formation model in Chen & Gnedin (2022), which is based on the outputs of the TNG50 simulation. In their model, star particles or dark matter particles in the TNG50 simulation are selected to represent star

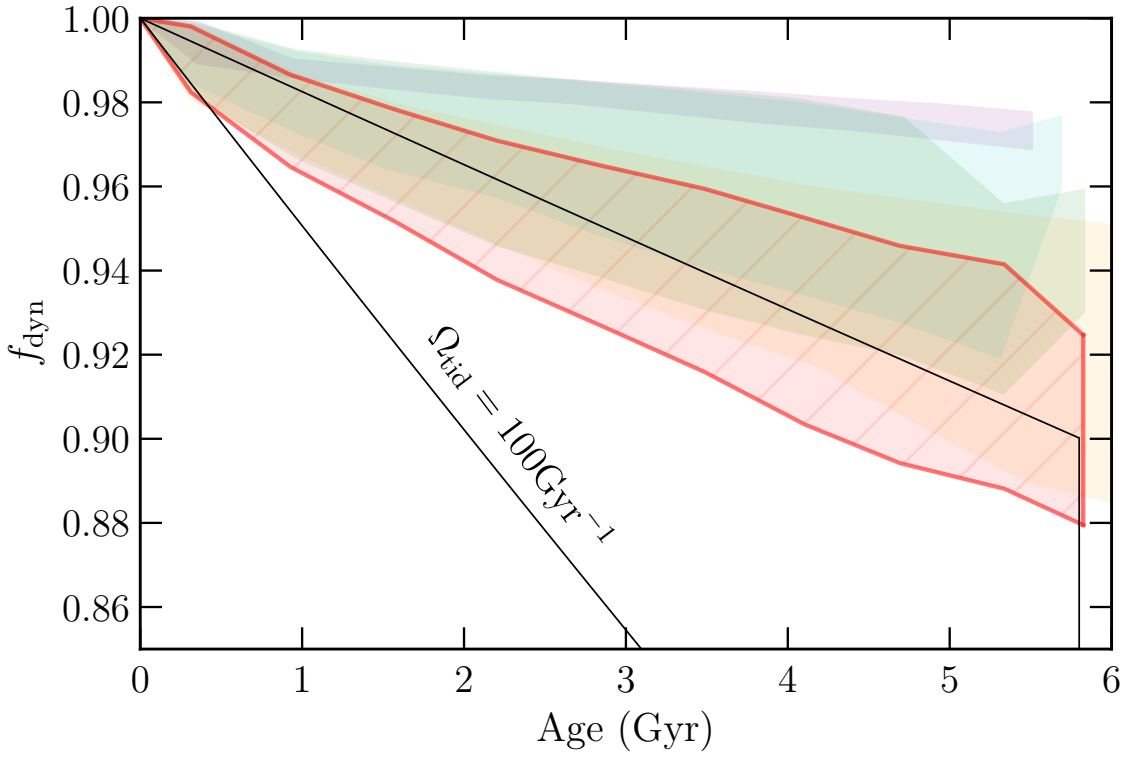

Figure 5.10 Another illustration of the effect of different tidal history on the final bound fraction. Each shaded region shows the 10-90% percentiles of the distribution of star cluster masses formed in one galaxy. Red colour is for the main galaxy, emphasized by the red frame; the orange, cyan, and green colours are for different satellite galaxies that merge into the main galaxy, ranked from early to late times. The purple band is for two clusters that originally formed in a satellite of the host galaxy of the cyan clusters and merged into the main galaxy along with the cyan clusters. The bound fraction for clusters from satellites usually decreases more slowly than for clusters formed in the main halo. For comparison, we include a triangular frame that corresponds to the evolution of the bound fraction in a constant tidal field $\Omega_{\rm tid} = 100\,{\rm Gyr}^{-1}$, for the same range of initial cluster masses.

clusters that would form GC systems. They calculated the tidal field surrounding these clusters using the simulation outputs. Here we calculated the $\Omega_{\rm tid,eff}$ for each star cluster with the same definition as Equation 5.5 and plotted the IQR as a gray shaded region in Figure 5.11. We used only clusters with initial mass above $2\times10^5\,M_\odot$ to make the cluster sample consistent with our clusters in f_iM_i. The $\Omega_{\rm tid,eff}$ values in the TNG50 outputs are similar to those in our simulations, and also show a decreasing trend with the host halo mass at birth, consistent with our results.

5.4 Discussion

5.4.1 Spatial resolution and tidal strength calculation

The calculated values of the tidal tensor depend on the spatial scale (cell size). Tidal field calculated on larger scale is typically weaker, as the limited resolution of the simulations may underestimate the actual tidal strength. A typical scale on which we calculate the tidal field is several hundred pc. The median size of cells where the traced star clusters are located is between 100 and 300 pc (see Table 5.1), and the differentiating scale is twice that. The whole range of cells containing the star clusters extends from as small as ∼20 pc to as large as 1 kpc in size, depending on the location in the galaxy. This is significantly larger than the typical cluster size, so the actual tidal field experienced by the star clusters is likely underestimated.

We cannot increase the simulation resolution but we can reduce it. We checked that if we double the scale on which we calculate $\lambda_{\rm m}$ by combining two cells, the resulting value of $\lambda_{\rm m}$ can decrease. The decrease is ∼10% if the original scale is ∼kpc, but it can be as much as a factor of 10 if the original scale is ∼100 pc. The amount of decrease scales with the local density because the runtime cell refinement is done according to the matter density. Thus regions of higher density are more affected by the loss of resolution.

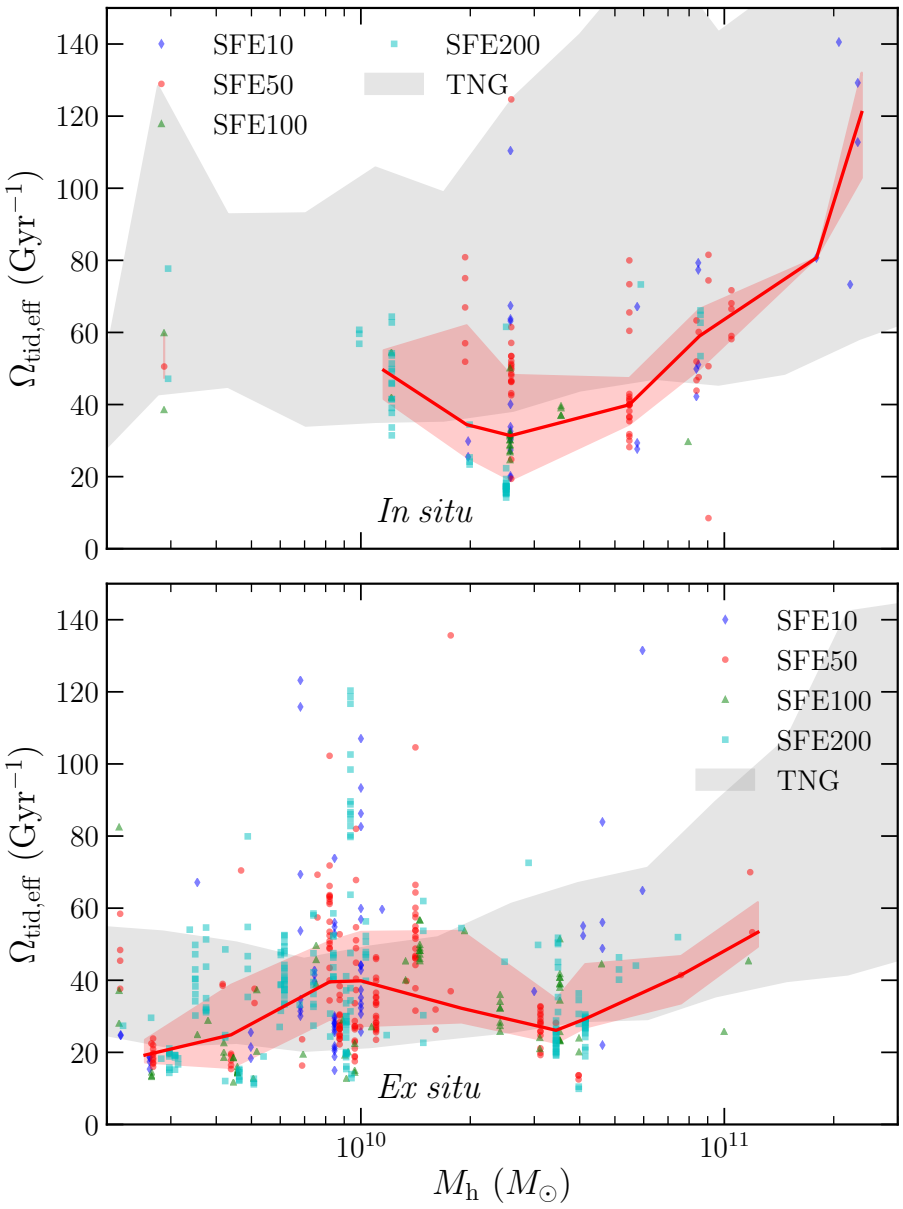

Figure 5.11 The expected effective $\Omega_{\rm tid,eff}$ value for each cluster versus host halo mass at time of formation, split into *in situ* and *ex situ* clusters. Different colours represent different runs. The red line and red shaded region are the median and IQR of all the points in bins of $\log M_{\rm h}$. *In situ* clusters in general experience stronger tidal field than *ex situ* clusters. There is a increasing trend with the host halo mass at birth for *ex situ* clusters, which means that on average clusters formed in more massive halos experience stronger tidal field. The gray shaded region is the IQR of the $\Omega_{\rm tid,eff}$ value before $z = 1.5$ for star clusters with initial mass larger than $2 \times 10^5\,M_\odot$ modeled by Chen & Gnedin (2022) based on the TNG50 simulation outputs. The $\Omega_{\rm tid,eff}$ value in the TNG50 simulation has a increasing trend with host halo mass at birth, consistent with our results.

5.4.2 Comparison with other studies

Renaud et al. (2017) performed a cosmological zoom-in simulation of a Milky Way-like galaxy and selected particles older than 10 Gyr to represent globular clusters. They calculated the tidal tensor and its eigenvalues along the orbits since redshift $z \gtrsim 1.8$ until $z = 0.5$. In agreement with our results, they found that the tidal field experienced by accreted clusters is weaker than that of the *in situ* clusters. Also in agreement, before accretion clusters experience generally similar tides regardless of in which galaxy they formed. However, most of their *ex situ* clusters are accreted after $z = 1.2$, later than the epochs we can probe with our simulations ($z > 1.5$). Renaud et al. (2017) do not distinguish between *in situ* and *ex situ* clusters that merged into the main galaxy during the "rapid increase phase" at redshifts $2 \lesssim z \lesssim 5$. Although these early accreted clusters are likely to mix better with the *in situ* ones than the later accreted clusters, at the end of our simulations they are still located farther from the center and experience weaker tidal field than the true *in situ* clusters.

Renaud et al. (2017) also found that the tidal field strength, characterized by the maximum eigenvalue λ_1 of the tidal tensor, increases with cosmic time after $z \approx 1.5$. This result applies to the later evolution of the overall tidal field and does not conflict with our conclusion that youngest clusters experience the largest $\lambda_{\rm m}$ near their birthplaces. Because of the different time intervals and different measures of the tidal field, the two studies are thus complimentary to each other. Although direct quantitative comparison is not possible, we tried to compare our measures of the tidal field around $z \approx 1.5$. Their average maximum eigenvalue $\lambda_1 \sim 320$ Gyr^{-2} is smaller than our typical $\lambda_{\rm m} > 10^3$ Gyr^{-2} around star clusters at late times. However, these two eigenvalues are likely different: λ_1 is the maximum of the three eigenvalues (and always positive), while $\lambda_{\rm m}$ is the maximum of the absolute value of the eigenvalues and may sometimes take value of $-\lambda_3$. Indeed, Figure 12 of Renaud et al. (2017) shows the average of the other two eigenvalues λ_2 and λ_3 is negative and 1-2 times

larger than λ_1. Even taking this into consideration, their average λ_1 is an order of magnitude smaller than our $\lambda_{\rm m}$. Our simulations have slightly better numerical resolution – a factor of 2 smaller dark matter particle mass, which may partly explain the difference but not all of it. It is possible that accurate force resolution close to the time of cluster formation is required to capture the earliest and most violent part of their tidal history.

Kruijssen et al. (2011) and Kruijssen et al. (2012) explored star cluster formation and destruction in the context of an isolated galaxy as well as galaxy merger. They found that dynamical heating of stellar clusters by tidal shocks can be an order of magnitude higher in interacting galaxies than in isolated galaxies. Although our high-redshift galaxies are more complicated and turbulent than isolated galaxies, we similarly found increased tidal strength experienced by star clusters during galaxy mergers. These authors were the first to propose that young clusters that evolve into old globular clusters are those that escape the merging system toward large distances early. They also found low-mass clusters being disrupted preferentially, which is consistent with our work and other current models (e.g. Elmegreen, 2010; Gieles & Renaud, 2016). Renaud & Gieles (2013) found in N-body simulations of star clusters that the galaxy major merger only indirectly affects the evolution of clusters by modifying their orbits in or around the galaxies, which is similar in essense to the difference in the bound fraction of *in situ* and *ex situ* clusters.

5.4.3 On recovering the halo assembly history

Recent kinematic studies of the Galactic GC population, coupled with the improved measurements of the age-metallicity distribution, have been used to uncover the assembly history of the Milky Way (e.g. Massari et al., 2019; Kruijssen et al., 2019b). Piatti (2019) pointed out differences in orbits and kinematics between the *in situ* and accreted Galactic GCs. Armstrong et al. (2021) found a clear kinematic

distinction between the metal-rich and metal-poor GCs using Gaia data. Although our simulations extend only to $z = 1.5$ and cannot be directly used to reconstruct the assembly history, we also found distinct orbital energy distributions for the *in situ* and *ex situ* clusters at the last available snapshot. Combined with the weaker tidal field experienced by *ex situ* clusters after accretion onto the main galaxy, we can expect that a larger fraction of the *ex situ* star clusters would survive until the present than of the *in situ* clusters.

Other simulations find similar results. Keller et al. (2020) found that more massive halos subject their proto-GCs to stronger disruption in the E-MOSAICS simulations. Their *ex situ* proto-GCs have higher survival fraction than the *in situ* ones, in agreement with our conclusions. They also found that GCs formed in major mergers are more likely to survive than those formed in minor mergers. They argue for the importance of major mergers because they are effective in ejecting clusters to orbits of higher energy and weaker tidal field.

When considering the different tidal field experienced by *in situ* and *ex situ* clusters in models of GC formation and evolution, we note that the normalization of the disruption time ($\Omega_{\rm tid,eff}$) in our Figure 5.11 is averaged not to $z = 0$, but only up to $z \sim 1.5$. A better calibration of $\Omega_{\rm tid,eff}$ would require simulations that run to redshift zero.

The orbital information of the GCs can potentially be used to reconstruct the assembly history of the MW. Pfeffer et al. (2020) pointed out that the orbits of GCs are sensitive to the masses and merger timing of their host satellite galaxies. Combined with the results from Keller et al. (2020) and our results that *ex situ* clusters are usually on higher orbits and more likely to survive than *in situ* ones, it is possible to use this information to recover the origins of GCs in the MW. Kruijssen et al. (2019b, 2020) used the metallicity plus orbital information to recover the merger history of the MW and added a Kraken galaxy corresponding to the 'low-energy'

GCs, merged prior to the previously found *Gaia*-Enceladus, the Helmi streams, the Sequoia galaxy, and the Sagittarius dwarf galaxy. We caution that using *only* the orbital information to reconstruct the merging history may not be effective, because the chaotic merging could mix spatial and kinematic distribution of GCs (Keller et al., 2020). For example, the integrals of motion for our star clusters at the last available snapshot cannot easily distinguish between clusters from different progenitors. In addition, Garrow et al. (2020) argued that the presence of dwarf galaxies can alter the orbital energies and actions of GCs, and that outer clusters are more strongly affected than inner clusters.

5.4.4 On parametrization of tidal strength

Recent literature has several approximations to parametrize strength of the tidal field. One is the largest absolute value of the eigenvalues of the tidal tensor, $\lambda_{\rm m}$, which we used in our previous work. Another is the largest effective eigenvalue $\lambda_{1,\rm e} = \lambda_1 - 0.5(\lambda_2 + \lambda_3)$ proposed by Pfeffer et al. (2018b) based on numerical experiments with a Plummer sphere potential, and later adopted by citetRodriguez:2022aa. We repeated our analysis with $\lambda_{1,\rm e}$ as an alternative parametrization of the tidal strength to compare the conclusions with our default choice $\lambda_{\rm m}$. We found the value of $\lambda_{1,\rm e}$ to be within a factor of 2 from $\lambda_{\rm m}$. Clusters experience stronger tidal field and end up with smaller bound fraction using this alternative definition, but our conclusions that tidal strength overall decreases with cluster age and that *ex situ* clusters usually end up with larger bound fraction than *in situ* clusters remain qualitatively the same. We describe the detailed comparison in Appendix B.

5.5 Conclusions

We use a suite of cosmological simulations with the continuous cluster formation algorithm to investigate the evolution of the tidal field around massive star clusters.

We calculate the tidal tensor for the simulation snapshots and reconstruct cluster orbits between the snapshots. Our main results are summarized below:

- Massive star clusters experience the strongest tidal field in the first 1 Gyr after formation in high-density regions close to their birthplaces. The maximum eigenvalue of the tidal tensor can reach $10^4 - 10^5$ Gyr^{-2}. After about 1 Gyr the median value plateaus at $\lambda_{\rm m} \sim 3 \times 10^3$ Gyr^{-2}.

- Most of the *in situ* clusters in our simulations move out of the disc plane and to larger distances from the galaxy center (Figure 5.1).

- The tidal field and cluster positions always correlate. The farther from the center and the galactic disc, the weaker the tidal field (Figure 5.2). A typical relation is $\lambda_{\rm m} \propto R^{-1.2}$.

- Cluster orbits are typically eccentric, alternating between regions of high and low density. Simulation snapshots record only random phases of the orbit, which may not capture the regions of strongest tidal field. Accurate calculation of tidal disruption requires detailed cluster orbits at time intervals $\lesssim 1$ Myr.

- As clusters age, they spend a smaller fraction of time in regions of high tidal strength. For example, in regions with $\lambda_{\rm m} > 3 \times 10^4$ Gyr^{-2}, this fraction decreases from $\sim$20% immediately after formation to less than 1% after 1 Gyr (Figure 5.5).

- *Ex situ* star clusters are brought in to the main galaxy by accretion of satellite galaxies. They experience similar tidal strength to the *in situ* clusters in the first 100 Myr after formation, independent of their host galaxy mass (Figure 5.8). After accretion *ex situ* clusters experience typically weaker tidal field than the *in situ* ones (Figure 5.11). They also tend to end up in higher energy orbits than the *in situ* clusters (Figure 5.6).

- Most *in situ* clusters remain near the galaxy center, but some are ejected to radii as large as 20 kpc on highly eccentric orbits, where they experience weaker tidal field (Figure 5.7).

- The cluster bound fraction at the last available snapshot increases with initial cluster mass, meaning that lower mass clusters are more easily disrupted. At the same initial mass, the final bound fraction is usually higher for *ex situ* clusters than *in situ* clusters (Figure 5.9).

- Clusters formed in more massive halos on average experience stronger tidal field. The difference is larger for *ex situ* clusters. For these clusters, the average normalization of the disruption rate increases by a factor of $\sim$3 as host halo mass increases from 3×10^9 to $10^{11}\,M_\odot$ (Figure 5.11).

- We also use an alternative definition for the tidal strength $\Omega^2_{\rm tid} = \lambda_{1,\rm e}$ and compare with our default choice $\lambda_{\rm m}$. Clusters experience stronger tidal field and end up with smaller bound fraction using this alternative, but our conclusions remain qualitatively the same (Appendix B). Note that both methods provide only rough estimates for the actual cluster mass loss and may be quantitatively inaccurate. Better understanding and parametrization of cluster disruption is urgently needed.

These results are important for calibrating the disruption rates of globular cluster in semi-analytical modeling of their evolution. Even the highest-resolution current cosmological simulations are unable to capture the detailed dynamics affecting cluster disruption. Therefore, such models still require additional boost of the disruption rates. Chen & Gnedin (2022) present an example of such a calibration that allows it to match observed properties of globular cluster systems.

CHAPTER VI

Summary and future work

Finally, I summarize my work in this chapter and provide potential future directions.

In Chapter II, I introduce the measurements of the shape of the high-redshift galaxies in our simulations defined by the shape tensor, for different components. The distribution of gas and young stars show prolate shapes, while the distribution of all the stars are somewhat spheroidal. This is consistent with observations of high-redshift galaxies, which, by reconstructing 3D shape from 2D images, also show mostly prolate shape. Our galaxies are also turbulent and thick, similar to observed high-redshift galaxies. I then applied Toomre analysis, which is originally applicable for thin, axisymmetric, and single-component disks only, to our simulated thick, irregular and turbulent galaxies. I found that a revised version of Toomre analysis for multiple components still works for our simulated galaxies even when the galaxies are far from thin, axisymmetric disks. We can still use the close to unity Toomre Q value to identify unstable regions that would collapse and form stars. Moreover, I showed the spatial non-coincidence of molecular gas and young stars due to the strong stellar feedback implemented in the simulations. This is consistent with observations, which use the 'fork diagram' to show the difference in depletion time when centering small apertures on cold gas and young stellar density peaks.

As I have compared part of our simulated galaxies' properties with high-redshift galaxy observations in Chapter II, in Chapter III I compare simulations to observations in a more realistic and one-to-one way. I created mock observation maps in *HST* bands with *HST* resolution for the simulated galaxies. This way we can understand better how observational constraints would affect the appearance of high-redshift galaxies. I found that the simulated galaxies do show kpc-scale stellar clumps, just like in observations. However, by comparing the clumps in the mock observation maps with the intrinsic clumps, I found that the masses and sizes of the clumps could be largely overestimated and thus affect our interpretation of them, and that the clumps are not one single object, and would dissolve in a relatively short time. This work provided potential insights for interpreting observational data.

In Chapter IV I extended my work on high-redshift galaxy structure and looked into the evolution of stellar disk thickness at high redshift. I found that the observed large thickness of high-redshift galaxy disks is due to rapid thickening of young stellar populations, and not because the disks are formed that thick. The rapidly changing disk orientation also plays a role in the observed large disk thickness. By looking into the simulations, I clarified that the 'upside-down' disk formation scenario inspired by the observed thickness of high-redshift galaxies may be missing the rapid disk inflation in the first few hundred Myr.

Finally, in Chapter V, I look into the formation of globular clusters, which are the fossils from the young universe and contain information about the assembly history of their host galaxies. Specifically, I investigated the tidal field evolution of the massive star clusters formed at high redshift, which could become globular clusters at present day. I found that the tidal strength around massive star clusters is strongest immediately after they formed and quickly decreases to a lower value after several hundred Myr. This is associated with the massive star clusters going out of the disk and away from the galaxy center after their formation. I also found that *ex*

situ star clusters experience weaker tidal field in their late time evolution than *in situ* clusters, probably because of their higher orbit when they merge into the main galaxy. Therefore, *ex situ* clusters have a better chance to survive to present day than *in situ* clusters. With this information we get better insights into recovering the assembly history with current GC populations.

In this dissertation work I only started making close comparison of simulations and observations in terms of galaxy structure. Future work calls for closer comparison in more aspects of galaxy properties, for more cosmic epochs, with more up-to-date telescopes and instruments (e.g. JWST and ELT), and with gravitational lensing. It would be great if kinematic information is also compared for simulated and observed galaxies through mock IFU observations. Even more interesting would be to put simulated galaxies through gravitational lensing to mimic the boosted resolution seen in observations of lensed galaxies. In this way we can understand in greater details how much simulated galaxies are consistent with observed ones. It would be very helpful to have simulations with different star formation and feedback models and see whether current observations can distinguish between these models. The work can also be extended to larger mass range, not only the MW-like galaxies.

In terms of GC formation and evolution, simulations with star cluster formation and evolution models should be run to $z = 0$ with more efficient simulation code to see if the GC population is consistent with observations. It would be especially useful to construct a model to use spatial, kinematic, and metallicity information of the current GC populations to find out the galaxy assembly history – some groups have already started working on this. This would make the most use of simulation works to interpret observations.

APPENDICES

APPENDIX A

Bound fraction correction

Becuase of a coding mistake, the disruption timescale $t_{\rm tid}^{\rm out}$ used in these simulations had an old normalization factor of $41.4\,{\rm Gyr}^{-1}$ instead of $100\,{\rm Gyr}^{-1}$, and the value of $\lambda_{\rm m}$ was overestimated by the expansion scale factor $a_{\rm box}(t)$. We describe how we correct for these two factors in this appendix.

The disruption timescale used during the simulation differs from the true one by a factor

$$\frac{t_{\rm tid}^{\rm out}}{t_{\rm tid}^{\rm true}} \equiv \omega = 0.414\, a_{\rm box}.$$

The analysis presented in Li & Gnedin (2019) corrected for this overestimate of the disruption rate by recalculating the bound fraction in post-processing of the simulation outputs. They calculated the tidal tensor for each cluster at a given snapshot n and applied the true timescale $t_{\rm tid}^{\rm true}(M_n, \Omega_n)$ to obtain a corrected value of the bound fraction at the next snapshot.

Since we showed that the tidal field along cluster trajectories varies rapidly between the snapshots separated by a relatively large interval of ~ 150 Myr, we may expect the above correction to be inaccurate. The largest uncertainty comes from the timing of snapshots: the tidal field at a snapshot is at a random phase of the cluster orbit, and probably in a weaker tidal field region because clusters on eccentric orbits spend more time farther from the galaxy center.

We can check the accuracy of this correction using the SFE200w run with very finely spaced output of the tidal field. That output matches the frequency of update of the bound fraction in the simulation runtime, once every global simulation timestep on the root grid level. If the number of timesteps between two snapshots is m, then we can write the calculation of $f_{\rm dyn}$ from one snapshot to the next as

$$\frac{f_{n+m}}{f_n} = \exp\left(-\sum_{i=n+1}^{n+m} \frac{{\rm d}t_i}{t_{\rm tid}(M_i, \Omega_i)}\right).$$

We corrected the tidal tensor history from the SFE200w run by the factor ω and calculated the true bound fraction of clusters in that run. Then we repeated the method used by Li & Gnedin (2019), which also includes the factor ω but applies it over the whole duration between the snapshots at once, $\Delta t_n = \sum_i {\rm d}t_i$, instead of at m individual steps. We found that using the coarser steps usually overestimates the bound fraction by $\sim$3% for final $f_{\rm dyn} > 0.9$ and up to 10% for final $f_{\rm dyn} \lesssim 0.8$. The more mass clusters lose, the more inaccurate this recalculation becomes.

Here we derive an alternative correction for the bound fraction in the main runs, using directly the output values $f_{\rm dyn}^{\rm out}$ at the snapshots. The disruption timescale used in the runtime differs from the true one in two aspects: the normalization of the disruption rate was overestimated, and in turn it lead to underestimating current bound cluster mass M_i. In terms of the ratio of the output cluster mass to its true value, $\mu = f^{\rm out}/f^{\rm true}$, the disruption time used in the simulations can be written as

$$t_{\rm tid}^{\rm out} = t_{\rm tid}^{\rm true}(M\mu, \Omega/\omega) = \omega\,\mu^{2/3} t_{\rm tid}^{\rm true}(M, \Omega).$$

Then the logarithmic difference of the bound fractions in the outputs is

$$\log\left(\frac{f_{n+m}}{f_n}\right)^{\rm out} = -\sum_{i=n+1}^{n+m}\frac{{\rm d}t_i}{\omega_i\,\mu_i^{2/3}t_{\rm tid}(M_i,\Omega_i)}$$
$$\approx -\frac{1}{\bar{\omega}\,\mu_n^{2/3}}\sum_{i=n+1}^{n+m}\frac{{\rm d}t_i}{t_{\rm tid}(M_i,\Omega_i)} = \frac{1}{\bar{\omega}\,\mu_n^{2/3}}\log\left(\frac{f_{n+m}}{f_n}\right)^{\rm true}$$

where $\bar{\omega}$ is taken as a simple average between the two snapshots $\bar{\omega} = 0.414(a_{{\rm box},n} + a_{{\rm box},n+m})/2$, and $\mu_n = f_n^{\rm out}/f_n^{\rm true}$ is calculated from the quantities already known at snapshot n. Therefore, we correct the output fractions as

$$\log\left(\frac{f_{n+m}}{f_n}\right)^{\rm true} = \bar{\omega}\,\mu_n^{2/3}\log\left(\frac{f_{n+m}}{f_n}\right)^{\rm out}.$$

For the first snapshot where a new cluster appears, the corrected and uncorrected bound fractions $f_n^{\rm true}$ and $f_n^{\rm out}$ are both taken to be 1.

Essentially this new method assumes an average $a_{\rm box}$ and an average mass correction factor. Although still an approximation, it encodes the evolution of the tidal field between snapshots via the available output values of the bound fraction. We checked this method on the SFE200w run and found very good agreement within 1%.

APPENDIX B

An alternative definition of tidal strength

As an alternative, we introduce another parametrization of the tidal strength $\lambda_{1,\mathrm{e}} = \lambda_1 - 0.5(\lambda_2 + \lambda_3)$ to describe the tidal strength (Renaud et al., 2011; Rodriguez et al., 2022). Here λ_1, λ_2 and λ_3 are the three eigenvalues of the tidal tensor from the largest to the smallest. The tidal field strength that sets the tidal radius of a cluster on a circular orbit is $\partial^2\Phi/\partial r^2 + \Omega^2$ (Renaud et al., 2011). Pfeffer et al. (2018b) shows in their Appendix C that for a Plummer sphere the tidal strength is well approximated by $\lambda_1 - 0.5(\lambda_2 + \lambda_3)$. Thus, we take this $\lambda_{1,\mathrm{e}}$ as our alternative tidal strength, as is done in Rodriguez et al. (2022). We find the value of $\lambda_{1,\mathrm{e}}$ to be always positive and within a factor of 2 of the value of λ_m. Using this new $\lambda_{1,\mathrm{e}}$, we take $\Omega^2_\mathrm{tid} = \lambda_{1,\mathrm{e}}$ instead of the original $\Omega^2_\mathrm{tid} = \lambda_\mathrm{m}/3$ in the disruption calculation.

We show a comparison between the median evolution with cluster age of the two in Figure B.1. In general, the median evolution of $\lambda_{1,\mathrm{e}}$ is about 2-3 times larger than $\lambda_\mathrm{m}/3$. The evolutionary trend of $\lambda_{1,\mathrm{e}}$ is similar to that of λ_m: it has higher value at young ages and then plateaus to a lower value at late times. The time fraction that clusters experience high $\lambda_{1,\mathrm{e}}$ also goes down with cluster age. Since with this alternative definition of tidal strength clusters experience stronger disruption, they end up with smaller bound fraction in the end. We show the evolution of bound fraction for clusters from different origins in the SFE200w run, similar to Figure 5.10,

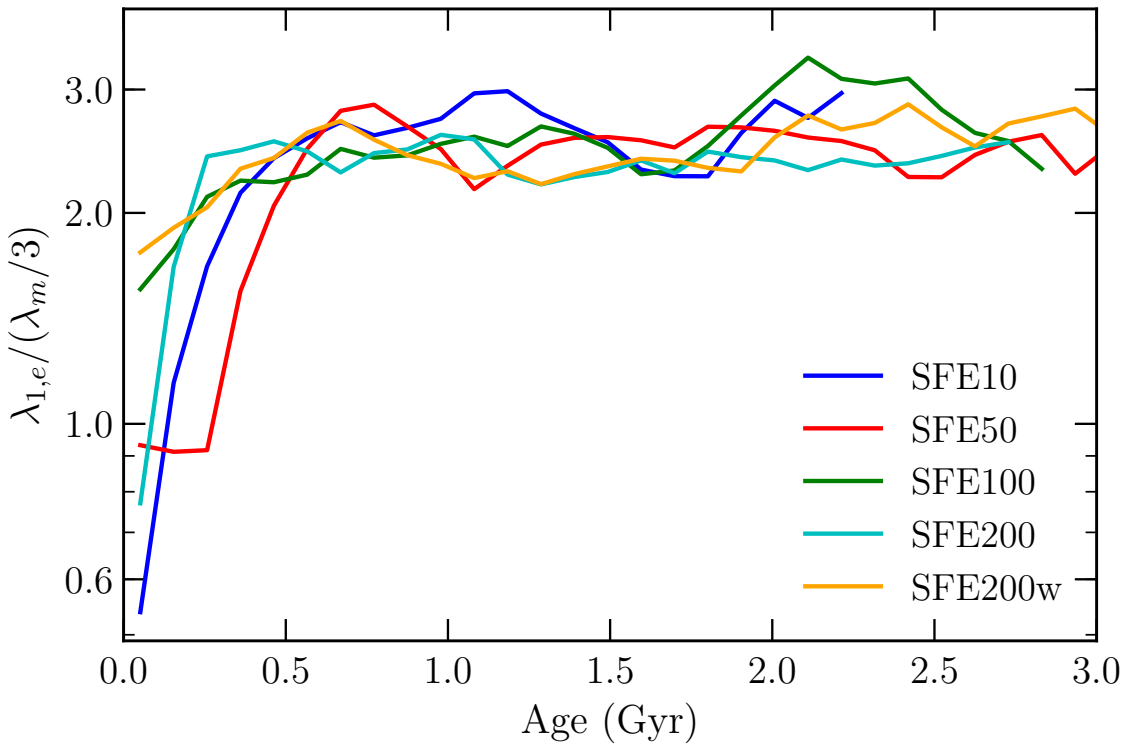

Figure B.1 Comparison of $\lambda_{1,e}$ with $\lambda_m/3$. The y axis is the ratio of the median of the two λ in age bins, i.e. ratios of the tidal strength evolution as in the bottom left panel of Figure 5.1. The median evolution of $\lambda_{1,e}$ is about 2-3 times larger than $\lambda_m/3$. Note that although the ratio between $\lambda_{1,e}$ and $\lambda_m/3$ is smaller when clusters are young, $\lambda_{1,e}$ is still larger at the beginning than it is at later times.

but calculated with $\Omega^2_{\rm tid} = \lambda_{1,e}$ instead of the original $\Omega^2_{\rm tid} = \lambda_m/3$ in Figure B.2. Our conclusion that the bound fraction for clusters from satellites usually decreases more slowly than for clusters formed in the main halo, remains unchanged with this alternative.

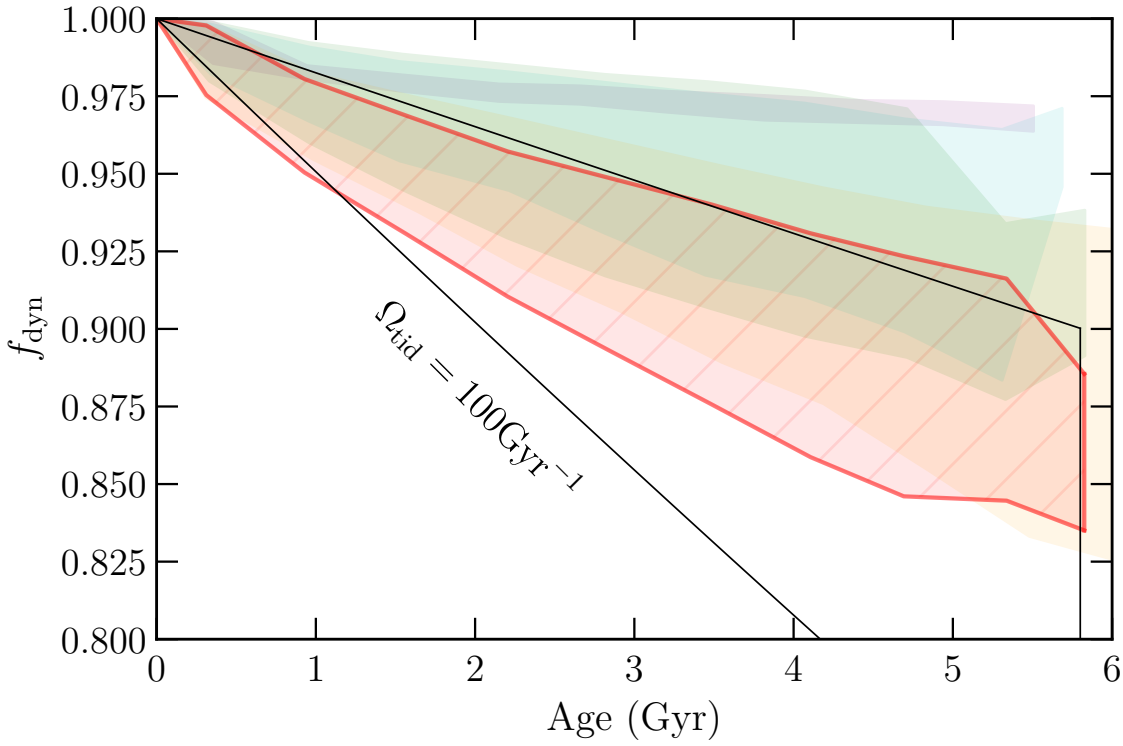

Figure B.2 A reproduction of Figure 5.10, but with $\Omega^2_{\rm tid} = \lambda_{1,\rm e}$. The bound fraction goes down with this alternative, but the distinction for clusters from different origins remains similar.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adamo, A., Östlin, G., Bastian, N., et al. 2013, *ApJ*, 766, 105

Agertz, O., Kravtsov, A. V., Leitner, S. N., & Gnedin, N. Y. 2013, *ApJ*, 770, 25

Armstrong, B. M., Bekki, K., & Ludlow, A. D. 2021, *MNRAS*, 500, 2937

Bastian, N., Pfeffer, J., Kruijssen, J. M. D., et al. 2020, *MNRAS*, 498, 1050

Behrendt, M., Burkert, A., & Schartmann, M. 2016, *ApJ*, 819, L2

Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, *ApJ*, 770, 57

Benincasa, S. M., Wadsley, J. W., Couchman, H. M. P., Pettitt, A. R., & Tasker, E. J. 2019, *MNRAS*, 486, 5022

Bensby, T., Alves-Brito, A., Oey, M. S., Yong, D., & Meléndez, J. 2011, *ApJ*, 735, L46

Bensby, T., Feltzing, S., Lundström, I., & Ilyin, I. 2005, *A&A*, 433, 185

Beraldo e Silva, L., Debattista, V. P., Khachaturyants, T., & Nidever, D. 2020, *MNRAS*, 492, 4716

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

Binney, J., & Tremaine, S. 2008, Galactic Dynamics (Princeton, NJ: Princeton Univ. Press)

Bird, J. C., Kazantzidis, S., Weinberg, D. H., et al. 2013, *ApJ*, 773, 43

Bird, J. C., Loebman, S. R., Weinberg, D. H., et al. 2020, arXiv e-prints, arXiv:2005.12948

Blanton, M. R., & Moustakas, J. 2009, *ARA&A*, 47, 159

Bournaud, F., Daddi, E., Weiß, A., et al. 2015, *A&A*, 575, A56

Bournaud, F., Elmegreen, B. G., & Martig, M. 2009, *ApJ*, 707, L1

Bournaud, F., Perret, V., Renaud, F., et al. 2014, *ApJ*, 780, 57

Bovy, J., & Rix, H.-W. 2013, *ApJ*, 779, 115

Bovy, J., Rix, H.-W., Liu, C., et al. 2012, *ApJ*, 753, 148

Bovy, J., Rix, H.-W., Schlafly, E. F., et al. 2016, *ApJ*, 823, 30

Boylan-Kolchin, M. 2017, *MNRAS*, 472, 3120

Bozek, B., Fitts, A., Boylan-Kolchin, M., et al. 2019, *MNRAS*, 483, 4086

Brodie, J. P., & Strader, J. 2006, *ARA&A*, 44, 193

Brown, G., Gnedin, O. Y., & Li, H. 2018, *ApJ*, 864, 94

Buck, T., Macciò, A. V., Obreja, A., et al. 2017, *MNRAS*, 468, 3628

Buck, T., Obreja, A., Macciò, A. V., et al. 2020, *MNRAS*, 491, 3461

Bullock, J. S., Kravtsov, A. V., & Weinberg, D. H. 2001, *ApJ*, 548, 33

Cava, A., Schaerer, D., Richard, J., et al. 2018, Nature Astronomy, 2, 76

Ceverino, D., Dekel, A., & Bournaud, F. 2010, *MNRAS*, 404, 2151

Ceverino, D., Dekel, A., Mandelker, N., et al. 2012, *MNRAS*, 420, 3490

Ceverino, D., Primack, J., & Dekel, A. 2015, *MNRAS*, 453, 408

Chabrier, G. 2001, *ApJ*, 554, 1274

Chen, Y., & Gnedin, O. Y. 2022, arXiv e-prints, arXiv:2203.00599

Choksi, N., & Gnedin, O. Y. 2019, *MNRAS*, 488, 5409

Choksi, N., Gnedin, O. Y., & Li, H. 2018, *MNRAS*, 480, 2343

Colombo, D., Kalinova, V., Utomo, D., et al. 2018, *MNRAS*, 475, 1791

Comerón, S., Elmegreen, B. G., Salo, H., et al. 2014, *A&A*, 571, A58

Comerón, S., Elmegreen, B. G., Knapen, J. H., et al. 2011, *ApJ*, 741, 28

Conroy, C., & Gunn, J. E. 2010, *ApJ*, 712, 833

Conroy, C., Gunn, J. E., & White, M. 2009, *ApJ*, 699, 486

Conselice, C. J. 2014, *ARA&A*, 52, 291

Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, *MNRAS*, 450, 1937

Daddi, E., Bournaud, F., Walter, F., et al. 2010, *ApJ*, 713, 686

Dekel, A., Sari, R., & Ceverino, D. 2009, *ApJ*, 703, 785

Dessauges-Zavadsky, M., & Adamo, A. 2018, *MNRAS*, 479, L118

Dessauges-Zavadsky, M., Schaerer, D., Cava, A., Mayer, L., & Tamburello, V. 2017a, *ApJ*, 836, L22

Dessauges-Zavadsky, M., Zamojski, M., Rujopakarn, W., et al. 2017b, *A&A*, 605, A81

Diesing, R., & Caprioli, D. 2018, Physical Review Letters, 121, 091101

El-Badry, K., Quataert, E., Weisz, D. R., Choksi, N., & Boylan-Kolchin, M. 2019, *MNRAS*, 482, 4528

Elmegreen, B. G. 2010, *ApJ*, 712, L184

Elmegreen, B. G., & Elmegreen, D. M. 2006, *ApJ*, 650, 644

Elmegreen, B. G., Elmegreen, D. M., Tompkins, B., & Jenks, L. G. 2017, *ApJ*, 847, 14

Elmegreen, D. M., & Elmegreen, B. G. 2014, *ApJ*, 781, 11

Elmegreen, D. M., Elmegreen, B. G., & Hirst, A. C. 2004a, *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

Elmegreen, D. M., Elmegreen, B. G., & Sheets, C. M. 2004b, *ApJ*, 603, 74

Feldmann, R., Gnedin, N. Y., & Kravtsov, A. V. 2011, *ApJ*, 732, 115

—. 2012, *ApJ*, 758, 127

Fisher, D. B., Glazebrook, K., Abraham, R. G., et al. 2017a, *ApJ*, 839, L5

Fisher, D. B., Glazebrook, K., Damjanov, I., et al. 2017b, *MNRAS*, 464, 491

Forbes, D. A., Read, J. I., Gieles, M., & Collins, M. L. M. 2018, *MNRAS*, 481, 5592

Forrest, B., Tran, K.-V. H., Broussard, A., et al. 2018, *ApJ*, 863, 131

Förster Schreiber, N. M., & Wuyts, S. 2020, *ARA&A*, 58, 661

Förster Schreiber, N. M., Genzel, R., Lehnert, M. D., et al. 2006, *ApJ*, 645, 1062

Förster Schreiber, N. M., Shapley, A. E., Genzel, R., et al. 2011, *ApJ*, 739, 45

Garrison-Kimmel, S., Hopkins, P. F., Wetzel, A., et al. 2018, *MNRAS*, 481, 4133

Garrow, T., Webb, J. J., & Bovy, J. 2020, *MNRAS*, 499, 804

Genel, S., Naab, T., Genzel, R., et al. 2012, *ApJ*, 745, 11

Genzel, R., Burkert, A., Bouché, N., et al. 2008, *ApJ*, 687, 59

Genzel, R., Tacconi, L. J., Gracia-Carpio, J., et al. 2010, *MNRAS*, 407, 2091

Genzel, R., Newman, S., Jones, T., et al. 2011, *ApJ*, 733, 101

Genzel, R., Förster Schreiber, N. M., Lang, P., et al. 2014, *ApJ*, 785, 75

Gieles, M., & Renaud, F. 2016, *MNRAS*, 463, L103

Girard, M., Dessauges-Zavadsky, M., Schaerer, D., et al. 2018, *A&A*, 619, A15

Gnedin, N. Y. 2014, *ApJ*, 793, 29

Gnedin, N. Y., & Abel, T. 2001, NewA, 6, 437

Gnedin, N. Y., & Kravtsov, A. V. 2011, *ApJ*, 728, 88

Grand, R. J. J., Springel, V., Gómez, F. A., et al. 2016, *MNRAS*, 459, 199

Griffiths, R. E., Casertano, S., Ratnatunga, K. U., et al. 1994, *ApJ*, 435, L19

Guo, Y., Giavalisco, M., Ferguson, H. C., Cassata, P., & Koekemoer, A. M. 2012, *ApJ*, 757, 120

Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, *ApJ*, 800, 39

Guo, Y., Rafelski, M., Bell, E. F., et al. 2018, *ApJ*, 853, 108

Haardt, F., & Madau, P. 2001, Clusters of Galaxies and the High Redshift Universe Observed in X-rays, 64

Halbesma, T. L. R., Grand, R. J. J., Gómez, F. A., et al. 2020, *MNRAS*, 496, 638

Harris, W. E., Blakeslee, J. P., & Harris, G. L. H. 2017, *ApJ*, 836, 67

Harrison, E. R. 1970, *Phys. Rev. D*, 1, 2726

Hawking, S. W. 1982, Physics Letters B, 115, 295

Hitschfeld, M., Kramer, C., Schuster, K. F., Garcia-Burillo, S., & Stutzki, J. 2009, *A&A*, 495, 795

Hockney, R. W., & Eastwood, J. W. 1981, Computer Simulation Using Particles

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

Holmberg, E. 1941, *ApJ*, 94, 385

Hopkins, P. F. 2015, *MNRAS*, 450, 53

Hopkins, P. F., Kereš, D., Murray, N., Quataert, E., & Hernquist, L. 2012, *MNRAS*, 427, 968

Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, *MNRAS*, 480, 800

Hudson, M. J., Harris, G. L., & Harris, W. E. 2014, *ApJ*, 787, L5

Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, *ApJS*, 221, 8

Inoue, S., Dekel, A., Mandelker, N., et al. 2016, *MNRAS*, 456, 2052

Inoue, S., & Saitoh, T. R. 2012, *MNRAS*, 422, 1902

Inoue, S., & Yoshida, N. 2018, *MNRAS*, 474, 3466

Johnson, T. L., Rigby, J. R., Sharon, K., et al. 2017, *ApJ*, 843, L21

Jones, T. A., Swinbank, A. M., Ellis, R. S., Richard, J., & Stark, D. P. 2010, *MNRAS*, 404, 1247

Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, *ApJ*, 673, 864

Katz, H., & Ricotti, M. 2014, *MNRAS*, 444, 2377

Keller, B. W., Kruijssen, J. M. D., Pfeffer, J., et al. 2020, *MNRAS*, 495, 4248

Kennicutt, Robert C., J. 1998a, *ARA&A*, 36, 189

—. 1998b, *ApJ*, 498, 541

Kennicutt, R. C., & Evans, N. J. 2012, *ARA&A*, 50, 531

Kim, J.-h., Ma, X., Grudić, M. Y., et al. 2018, *MNRAS*, 474, 4232

Kravtsov, A. V. 1999, PhD thesis, NEW MEXICO STATE UNIVERSITY

—. 2003, *ApJ*, 590, L1

Kravtsov, A. V., Klypin, A. A., & Khokhlov, A. M. 1997, *ApJS*, 111, 73

Kroupa, P. 2001, *MNRAS*, 322, 231

Kruijssen, J. M. D. 2009, *A&A*, 507, 1409

—. 2015, *MNRAS*, 454, 1658

Kruijssen, J. M. D., & Longmore, S. N. 2014, *MNRAS*, 439, 3239

Kruijssen, J. M. D., Pelupessy, F. I., Lamers, H. J. G. L. M., et al. 2012, *MNRAS*, 421, 1927

Kruijssen, J. M. D., Pelupessy, F. I., Lamers, H. J. G. L. M., Portegies Zwart, S. F., & Icke, V. 2011, *MNRAS*, 414, 1339

Kruijssen, J. M. D., Pfeffer, J. L., Crain, R. A., & Bastian, N. 2019a, *MNRAS*, 486, 3134

Kruijssen, J. M. D., Pfeffer, J. L., Reina-Campos, M., Crain, R. A., & Bastian, N. 2019b, *MNRAS*, 486, 3180

Kruijssen, J. M. D., Schruba, A., Hygate, A. P. S., et al. 2018, *MNRAS*, arXiv:1805.00012

Kruijssen, J. M. D., Pfeffer, J. L., Chevance, M., et al. 2020, *MNRAS*, 498, 2472

Krumholz, M. R., Burkhart, B., Forbes, J. C., & Crocker, R. M. 2018, *MNRAS*, 477, 2716

Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, *ARA&A*, 57, 227

Lacey, C. G. 1984, *MNRAS*, 208, 687

Law, D. R., Steidel, C. C., Shapley, A. E., et al. 2012, *ApJ*, 745, 85

Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, *ApJS*, 123, 3

Leroy, A. K., Walter, F., Brinks, E., et al. 2008, *AJ*, 136, 2782

Leroy, A. K., Schinnerer, E., Hughes, A., et al. 2017, *ApJ*, 846, 71

Li, H., & Gnedin, O. Y. 2014, *ApJ*, 796, 10

—. 2019, *MNRAS*, 486, 4030

Li, H., Gnedin, O. Y., & Gnedin, N. Y. 2018a, *ApJ*, 861, 107

—. 2018b, *ApJ*, 861, 107

Li, H., Gnedin, O. Y., Gnedin, N. Y., et al. 2017a, *ApJ*, 834, 69

—. 2017b, *ApJ*, 834, 69

Li, Y., Mac Low, M.-M., & Klessen, R. S. 2005, *ApJ*, 626, 823

—. 2006, *ApJ*, 639, 879

Livermore, R. C., Jones, T., Richard, J., et al. 2012, *MNRAS*, 427, 688

Livermore, R. C., Jones, T. A., Richard, J., et al. 2015, *MNRAS*, 450, 1812

Loebman, S. R., Roškar, R., Debattista, V. P., et al. 2011, *ApJ*, 737, 8

Ma, X., Hopkins, P. F., Wetzel, A. R., et al. 2017, *MNRAS*, 467, 2430

Madau, P., & Dickinson, M. 2014, *ARA&A*, 52, 415

Mandelker, N., Dekel, A., Ceverino, D., et al. 2017, *MNRAS*, 464, 635

—. 2014, *MNRAS*, 443, 3675

Martel, H., & Shapiro, P. R. 1998, *MNRAS*, 297, 467

Martinsson, T. P. K., Verheijen, M. A. W., Westfall, K. B., et al. 2013, *A&A*, 557, A130

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

Massari, D., Koppelman, H. H., & Helmi, A. 2019, *A&A*, 630, L4

Mayer, L., Tamburello, V., Lupi, A., et al. 2016, *ApJ*, 830, L13

Meng, X., & Gnedin, O. Y. 2020, *MNRAS*, 494, 1263

—. 2021, *MNRAS*, 502, 1433

—. 2022, *MNRAS*, 515, 1065

Meng, X., Gnedin, O. Y., & Li, H. 2019, *MNRAS*, 486, 1574

Mieda, E., Wright, S. A., Larkin, J. E., et al. 2016, *ApJ*, 831, 78

Moody, C. E., Guo, Y., Mandelker, N., et al. 2014, *MNRAS*, 444, 1389

Murata, K. L., Kajisawa, M., Taniguchi, Y., et al. 2014, *ApJ*, 786, 15

Muratov, A. L., & Gnedin, O. Y. 2010, *ApJ*, 718, 1266

Murray, N., Quataert, E., & Thompson, T. A. 2010, *ApJ*, 709, 191

Naab, T., & Ostriker, J. P. 2017, *ARA&A*, 55, 59

Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, 477, 1206

Navarro, J. F., Yozin, C., Loewen, N., et al. 2018, *MNRAS*, 476, 3648

Nelson, A. F. 2006, *MNRAS*, 373, 1039

Nelson, D., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, 475, 624

Noguchi, M. 1999, *ApJ*, 514, 77

Oklopčić, A., Hopkins, P. F., Feldmann, R., et al. 2017, *MNRAS*, 465, 952

Olmstead, A., Rigby, J. R., Swinbank, M., & Veilleux, S. 2014, *AJ*, 148, 65

Orr, M. E., Hayward, C. C., Hopkins, P. F., et al. 2018, *MNRAS*, 478, 3653

Overzier, R. A., Heckman, T. M., Schiminovich, D., et al. 2010, *ApJ*, 710, 979

Papovich, C., Labbé, I., Quadri, R., et al. 2015, *ApJ*, 803, 26

Peebles, P. J. E., & Dicke, R. H. 1968, *ApJ*, 154, 891

Perez, J., Valenzuela, O., Tissera, P. B., & Michel-Dansac, L. 2013, *MNRAS*, 436, 259

Pfeffer, J., Kruijssen, J. M. D., Crain, R. A., & Bastian, N. 2018a, *MNRAS*, 475, 4309

—. 2018b, *MNRAS*, 475, 4309

Pfeffer, J. L., Trujillo-Gomez, S., Kruijssen, J. M. D., et al. 2020, *MNRAS*, 499, 4863

Piatti, A. E. 2019, *ApJ*, 882, 98

Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, *MNRAS*, 475, 648

Pillepich, A., Nelson, D., Springel, V., et al. 2019, *MNRAS*, 490, 3196

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, 594, A13

Prieto, J. L., & Gnedin, O. Y. 2008, *ApJ*, 689, 919

Puech, M. 2010, *MNRAS*, 406, 535

Puech, M., Hammer, F., Flores, H., Neichel, B., & Yang, Y. 2009, *A&A*, 493, 899

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

Rebolledo, D., Wong, T., Xue, R., et al. 2015, *ApJ*, 808, 99

Reddy, B. E., Lambert, D. L., & Allende Prieto, C. 2006, *MNRAS*, 367, 1329

Reina-Campos, M., Hughes, M. E., Kruijssen, J. M. D., et al. 2020, *MNRAS*, 493, 3422

Reina-Campos, M., & Kruijssen, J. M. D. 2017, *MNRAS*, 469, 1282

Reina-Campos, M., Kruijssen, J. M. D., Pfeffer, J. L., Bastian, N., & Crain, R. A. 2019, *MNRAS*, 486, 5838

Renaud, F., Agertz, O., & Gieles, M. 2017, *MNRAS*, 465, 3622

Renaud, F., Gieles, M., & Boily, C. M. 2011, *MNRAS*, 418, 759

Reshetnikov, V. P., Dettmar, R. J., & Combes, F. 2003, *A&A*, 399, 879

Ribeiro, B., Le Fèvre, O., Cassata, P., et al. 2017, *A&A*, 608, A16

Rodriguez, C. L., Hafen, Z., Grudić, M. Y., et al. 2022, arXiv e-prints, arXiv:2203.16547

Romeo, A. B., Burkert, A., & Agertz, O. 2010, *MNRAS*, 407, 1223

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

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

Rossi, L. J., Bekki, K., & Hurley, J. R. 2016, *MNRAS*, 462, 2861

Rudd, D. H., Zentner, A. R., & Kravtsov, A. V. 2008, *ApJ*, 672, 19

Rujopakarn, W., Daddi, E., Rieke, G. H., et al. 2019, *ApJ*, 882, 107

Safronov, V. S. 1960, Annales d'Astrophysique, 23, 979

Salpeter, E. E. 1955, *ApJ*, 121, 161

Schaye, J. 2004, *ApJ*, 609, 667

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, 446, 521

Schechter, P. 1976, *ApJ*, 203, 297

Schmidt, W., Almgren, A. S., Braun, H., et al. 2014, *MNRAS*, 440, 3051

Schönrich, R., & Binney, J. 2009, *MNRAS*, 399, 1145

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

Semenov, V. A., Kravtsov, A. V., & Gnedin, N. Y. 2016a, *ApJ*, 826, 200

—. 2016b, *ApJ*, 826, 200

—. 2017, *ApJ*, 845, 133

—. 2018, *ApJ*, 861, 4

Shapiro, K. L., Genzel, R., & Förster Schreiber, N. M. 2010, *MNRAS*, 403, L36

Shapiro, K. L., Genzel, R., Förster Schreiber, N. M., et al. 2008, *ApJ*, 682, 231

Shibuya, T., Ouchi, M., Kubo, M., & Harikane, Y. 2016, *ApJ*, 821, 72

Soto, E., de Mello, D. F., Rafelski, M., et al. 2017, *ApJ*, 837, 6

Spitler, L. R., & Forbes, D. A. 2009, *MNRAS*, 392, L1

Spitzer, Lyman, J., & Schwarzschild, M. 1951, *ApJ*, 114, 385

—. 1953, *ApJ*, 118, 106

Spitzer, L. 1987, Dynamical Evolution of Globular Clusters (Princeton: Princeton University Press)

Spitzer, Jr., L. 1969, *ApJ*, 158, L139

Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *MNRAS*, 475, 676

Starobinsky, A. A. 1982, Physics Letters B, 117, 175

Stinson, G. S., Bovy, J., Rix, H. W., et al. 2013, *MNRAS*, 436, 625

Swinbank, A. M., Smail, I., Chapman, S. C., et al. 2010a, *MNRAS*, 405, 234

Swinbank, A. M., Smail, I., Longmore, S., et al. 2010b, *Nature*, 464, 733

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

Tadaki, K., Iono, D., Yun, M. S., et al. 2018, *Nature*, 560, 613

Tadaki, K.-i., Kodama, T., Tanaka, I., et al. 2014, *ApJ*, 780, 77

Tamburello, V., Mayer, L., Shen, S., & Wadsley, J. 2015, *MNRAS*, 453, 2490

Tamburello, V., Rahmati, A., Mayer, L., et al. 2017, *MNRAS*, 468, 4792

Tomassetti, M., Dekel, A., Mandelker, N., et al. 2016, *MNRAS*, 458, 4477

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

—. 1977, *ARA&A*, 15, 437

Trujillo-Gomez, S., Kruijssen, J. M. D., Reina-Campos, M., et al. 2021, *MNRAS*, 503, 31

Tutukov, A. V., Shustov, B. M., & Wiebe, D. S. 2000, Astronomy Reports, 44, 711

Utomo, D., Bolatto, A. D., Wong, T., et al. 2017, *ApJ*, 849, 26

van der Wel, A., Chang, Y.-Y., Bell, E. F., et al. 2014, *ApJ*, 792, L6

Vogelsberger, M., Marinacci, F., Torrey, P., & Puchwein, E. 2020, Nature Reviews Physics, 2, 42

Vollmer, B., & Leroy, A. K. 2011, *AJ*, 141, 24

von Hoerner, S. 1960, *ZAp*, 50, 184

Webb, J. J., Leigh, N., Sills, A., Harris, W. E., & Hurley, J. R. 2014, *MNRAS*, 442, 1569

Weinberger, R., Springel, V., & Pakmor, R. 2020, *ApJS*, 248, 32

Westfall, K. B., Andersen, D. R., Bershady, M. A., et al. 2014, *ApJ*, 785, 43

White, S. D. M., & Rees, M. J. 1978, *MNRAS*, 183, 341

Windhorst, R. A., Fomalont, E. B., Kellermann, K. I., et al. 1995, *Nature*, 375, 471

Wisnioski, E., Glazebrook, K., Blake, C., et al. 2012, *MNRAS*, 422, 3339

Wisnioski, E., Förster Schreiber, N. M., Wuyts, S., et al. 2015, *ApJ*, 799, 209

Wuyts, E., Rigby, J. R., Gladders, M. D., & Sharon, K. 2014, *ApJ*, 781, 61

Wuyts, S., Förster Schreiber, N. M., Genzel, R., et al. 2012, *ApJ*, 753, 114

Yoachim, P., & Dalcanton, J. J. 2008a, *ApJ*, 683, 707

—. 2008b, *ApJ*, 682, 1004

Zanella, A., Daddi, E., Le Floc'h, E., et al. 2015, *Nature*, 521, 54

Zanella, A., Le Floc'h, E., Harrison, C. M., et al. 2019, *MNRAS*, 489, 2792

Zeldovich, Y. B. 1972, *MNRAS*, 160, 1P

Zemp, M., Gnedin, O. Y., Gnedin, N. Y., & Kravtsov, A. V. 2011, *ApJS*, 197, 30

Zhang, H., Primack, J. R., Faber, S. M., et al. 2019, *MNRAS*, 484, 5170