Star Cluster Formation in Cosmological Simulations

by

Hui Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Astronomy and Astrophysics)
in The University of Michigan
2017

Doctoral Committee:

Associate Professor Oleg Y. Gnedin, Chair
Professor Eric F. Bell
Professor August E. Evrard
Associate Professor Christopher J. Miller

hliastro@umich.edu
ORCID ID:0000-0002-1253-2763

To my beloved wife, my parents, my teachers and friends.

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Prof. Oleg Gnedin, for his guidance during the last five years with patience, encouragement, and comprehensive knowledge. Whenever I recall the scene that he and I discussed in his office, I remember clearly how he evaluated the priorities of different projects, how he divided knotty problems into simple pieces, and how he solved physical problems that struggle me days with only a few seconds using his strong physical intuition. Besides science, he spent a vast amount of time helping me improve my scientific writing and oral presentation. I keep all paper drafts that are filled with heavy annotations of his corrections and comments. I also remember all the moments that he helped me go over the slides before my talks. Oleg to me is not only an advisor, but also an English instructor, a collegue, and a mentor. My only hope is to maintain the style of thinking I learned from him in my future career.

I also owe my deep gratitude to Prof. Nick Gnedin for his generous support on the numerical developments of my simulations. I cannot count how many emails I sent him to ask questions about numerical methods in general and the ART code in details. He even spent days digging into my codes helping me fix the bugs that I created by myself. Without him, I cannot imagine how slow my progress will be.

I thank all thesis committee members, Gus Evrard, Eric Bell, and Chris Miller, for their numerous suggestions during every committee meetings. Many of the suggestions are game-changers that are critical to the direction of my thesis. I also thank my colleagues Marcel Zemp, Vadim Semenov, Andrey Kravtsov for their supports

and suggestions on my projects and manuscripts.

I thank my classmates, Vivienne Baldassare, Traci Johnson, Marina Kounkel, Kamber Schwarz, Meghin Spencer, for their company in the department. I still remember the time when we solved problems in homeworks together and killed many weekends to prepare the prelim exam.

I would like to thank Prof. Mario Mateo to allow me join his cycling trips during the last several years so that I could continue my cycling career in the US, even though I am a slow rider. Hopefully, we can climb more mountains in the future.

I thank the whole Astronomy department for providing me such a relaxed and friendly environment. I appreciate the timely supports from the IT and admin stuff. Especially, I would like to thank Brian Cox, whom I constantly bother for various immigration issues.

I thank all my friends I made in Ann Arbor: the cycling teams, the badminton club, and many others from the Chinese community. Thanks for sharing with me stories outside astronomy and bringing me happiness during the tough Ph.D. period.

I thank my wife, Letian Yan, for her endurance on our long-distance relation for the last five years. I know it was extremely hard for her to live without my company in the New York city. There are some ups and downs but luckily we survive. I appreciate every moments that we are together, whether on biking or hiking, or other things. I believe future is brighter for us.

Finally, I would like to thank my parents for their love and support. Confucius said, “While a father or mother are alive, a son should not travel far. If he travel, he must have a meaning destination”. Hopefully, my journey to the U.S. can later be demonstrated to be a “meaning destination”.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

Using cosmological hydrodynamic simulations, researchers are making rapid progress in reproducing different types of galaxies and various global scaling relations. One important step has been to recognize the importance of feedback mechanisms that suppress the excess of star formation activities. Although many feedback processes have been explored, the star formation prescription has remained unchanged for over two decades.

To model star formation in a more realistic way, in my thesis, I develop a new implementation in cosmological simulations, continuous cluster formation (CCF) , by considering star clusters as a unit of star formation, inspired by observations that most stars form in clusters. In CCF, a cluster particle grows its mass through gas accretion within a star-forming sphere. The accretion is terminated by its own feedback, thus the final mass is set self-consistently. I also introduce the initial bound fraction, f_i, to estimate the mass fraction that is remind bound to the cluster when it emerges from the giant molecular clouds (GMCs). I implement CCF in the Adaptive Refinement Tree code and perform a series of simulations of Milky Way-sized galaxies.

I find that the global star formation history (SFH) of the main galaxy is sensitive to the feedback parameter. Varying the star formation efficiency per free-fall time $\epsilon_{\rm ff}$, on the other hand, has no systematic effect on SFH.However, $\epsilon_{\rm ff}$ has a dramatic effect on the properties of modeled star clusters, which can be used to calibrate the star formation and feedback models on a scale that is compatible with the size of GMCs.

I find that the cluster initial mass function is best described by the Schechter function. The cutoff mass scales with the star formation rate of the host galaxies,

suggesting that cluster formation depends strongly on galactic environments. I find f_i increases strongly with cluster masses, irrespective to the global galactic environment. However, f_i is very sensitive to the choice of $\epsilon_{\rm ff}$: the higher the $\epsilon_{\rm ff}$, the larger the f_i. This trend also leads to a positive correlation between the maximum cluster mass and $\epsilon_{\rm ff}$. I measure the integrated cluster formation efficiency and find it correlates with star formation rate surface density. Moreover, I find a clear trend that cluster formation timescale is shorter with higher $\epsilon_{\rm ff}$. Future observations of this timescale in the nearby star formation regions can be used as another powerful diagnostic to constrain $\epsilon_{\rm ff}$.

Based on CCF, I implement a new algorithm to model the tidal disruption of clusters along their orbits around the galaxies. I find that various disruption processes significantly changes the shape of the mass function from Schechter-like to log-normal, which suggests young massive clusters formed at high-z are promising candidates of the progenitor of globular clusters (GCs). To better understanding the formation and evolution of GCs, I construct a semi-analytical model onto the halo merger trees in Millennium-II simulations. This model successfully reproduces the observed multi-modal metallicity distribution of GCs in a wide range of host galaxy masses in the Virgo cluster.

CHAPTER I

Introduction

How did the Universe evolve from an extremely hot and dense fireball to the current complex structure with galaxies, stars, planets, and even human beings living with it? This multifaceted question has puzzled our ancestors for centuries and many of the best minds in history have contributed their whole lives to pursue the true nature of the Universe. From the ancient Chinese view of a square earth and a canopy-like heaven (天圆地方), to the current widely accepted Λ-CDM cosmology, studies of the Universe have been transformed from pure philosophical speculations to accurate scientific measurements. It is not surprising that *Time* magazine ranked "the origin of the Universe" as one of the top ten most important problems for the 21st century. Fortunately, we are living in the so-called "golden age" of cosmology, where the understanding of the origin and evolution of the Universe are developing rapidly, thanks to state-of-the-art astronomical observational techniques as well as ever-increasing computing power. However, the road leading to our current achievements has been riddled with false starts and dead ends, and is marked by several important paradigm shifts.

In this introduction, I provide a short overview of the general consensus of the current model of cosmic evolution and describe the use of numerical simulations as the most important tool to tackle the problems of galaxy formation. After highlighting

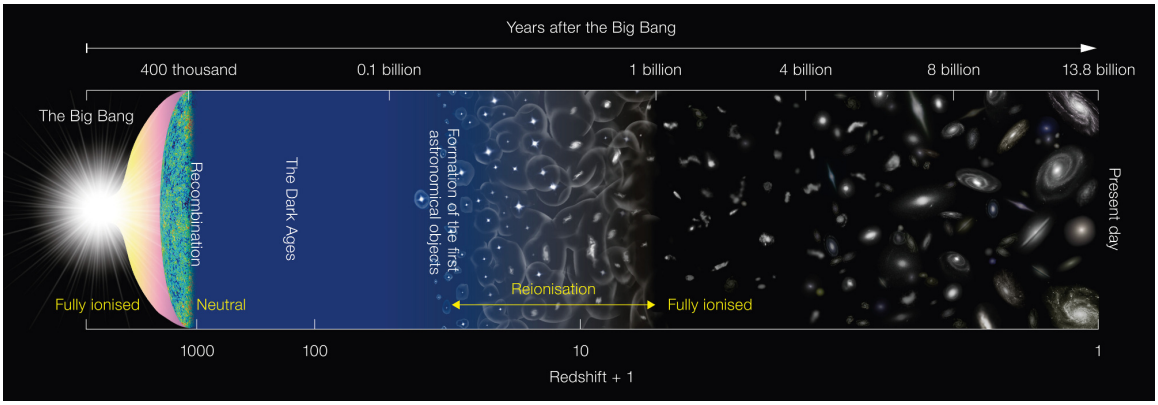

Figure 1.1: Pictorial sketch of the evolution of the Universe along cosmic time with some highlights to monumental epochs, such as recombination and the formation of Pop III stars.

some theoretical challenges that we currently face, I will demonstrate how my thesis work contributes to the quickly evolving field of galaxy formation.

1.1 A brief history of the Universe

And God said, "Let there be light," and there was light.

Genesis 1:3

The commonly accepted picture of the evolution of the universe is elegantly summarized in Figure 1.1[1]. From the leftmost singularity epoch in this figure, our cosmic voyage starts with an extremely fast expansion ~ 13.7 billion years ago, in a state of enormous density and pressure. During the first picosecond, the universe experiences 60 e-folds of expansion (Kolb & Turner, 1990) and the initial quantum fluctuations are stretched to larger scales (Guth & Pi, 1982). Mathematically, the stretching of space can be described by a time-dependent scale factor, $a(t)$, so that the position measurements $r(t_1)$ at time t_1 and $r(t_2)$ at time t_2 follow the relation $R = r(t_2)/t_2 = r(t_1)/t_1$, where R is the comoving distance.

As the Universe expands, both the density and temperature start to decline. As soon as the temperature cools to < 13.6 eV, hot electrons and protons start to bind

[1]Credit: NAOJ; `https://www.eso.org/public/images/eso1620a/`

together to create Hydrogen atoms, causing the ionization fraction of the Universe to drop dramatically. At the same time, with the decline of the density of charged particles, the optical depth of Thomson scattering also drops and photons decouple from baryonic matter. Detailed calculations suggest that the decoupling temperature is around $T_{\rm dec} \approx 0.26$ eV (see Mo et al., 2010). This represents the last scattering surface of photons, where the Cosmic Microwave Background (CMB) emerges.

In the late 1940s, George Gamow first predicted that the Universe should be filled with the relic of the CMB as homogeneous and isotropic blackbody radiation (Gamow, 1946). Later in 1950, based on some early results of cosmic nucleosynthesis, Ralpha Alpher and Robert Herman estimated the temperature of this relic radiation to be 5 K (Alpher & Herman, 1950). Unfortunately, these science-fiction like predictions have been largely forgotten since the then far-fetched idea of the Big Bang theory was not accepted by the mainstream scientists at that time - at least until Penzias & Wilson (1965) discovered an isotropic background radiation at 7.5 cm wavelength with antenna temperature 3.5 ± 1.0 K, which helped the two researchers to win the 1978 Nobel Prize.

Later satellite observations confirmed this emission as the most perfect blackbody spectrum with $T = 2.73$ K with a NEARLY isotropic spatial distribution (Mather et al., 1990). A more careful examination revealed a slight anisotropy on the order of 10^{-5} (Smoot et al., 1992), which reflects the density fluctuation of the Universe at the last scattering surface. Why is this miniature anomaly worth another Nobel Prize in 2006 [2]? It is this tiny deviation from perfect that provides the seeds of the hierarchical structures that were later developed to create galaxies and stars.

If we put aside the tiny anisotropy for a while, another important implication from the CMB observations is that our Universe acts in the simplest way possible: it

[2] "The Nobel Prize in Physics 2006 was awarded jointly to John C. Mather and George F. Smoot for their discovery of the blackbody form and anisotropy of the cosmic microwave background radiation"

satisfies the so-called "cosmological principle" (Coles & Lucchin, 2002) which states that the universe is homogeneous and isotropic on large scales with a flat geometry. It is quite surprising that the "cosmological principle" that was once criticized for making "our lack of knowledge a principle of knowing something" [3] actually reflects the true nature of the Universe.

The theoretical description of an expanding, homogeneous, and isotropic Universe with a flat geometry is straightforward. With the cosmological principle and general theory of relativity, the evolution of the Universe is described by the Friedman equation (Friedmann, 1922)

$$\left(\frac{\dot{a}}{a}\right) \equiv H^2(t) = \left(\Omega_m a^{-3} + \Omega_\gamma a^{-4} + \Omega_\Lambda\right) H_0^2, \tag{1.1}$$

where $H_0 = \sqrt{8\pi G \rho_{\rm crit,0}/3}$ is the Hubble's Constant at present, Ω_m, Ω_γ, and Ω_Λ are the density fractions of matter, radiation, and dark energy components with respect to the critical density, respectively. The expansion history of the Universe can be solved for a given set of cosmological parameters described above. Most recent observations of CMB, supernova (SN) Ia, and clustering of galaxies jointly give a very good constraint on these cosmological parameters with high precision. This is why we are in the era of precise cosmology. In this thesis, I adopt a Λ-CDM cosmology with $\Omega_{\rm m} = 0.304$, $\Omega_{\rm b} = 0.048$, $\Omega_\gamma \approx 0$, $\Omega_\Lambda = 0.696$, $H_0 = 68.1$ km/s/Mpc, which is consistent with the most recent *Planck* results (Planck Collaboration et al., 2016).

It can be noticed from the fiducial values of the cosmological parameters that the Universe is now dominated by a repulsive energy called dark energy, the main driver that keeps the Universe expanding in acceleration. For the rest matter component, the ordinary baryonic matter only contributes to $\sim 16\%$ of the total matter budget and most of the matter is "dark", which does not interact with photons and therefore

[3] A letter by Karl Popper to Helge Kragh at 10 June 1994, probably one of Popper's last, to reply some questions in a then-preparing book about the history of modern cosmology (Kragh 1996)

cannot be detected by photon-collected telescopes. Rather, dark matter acts as a dominant gravitational source that strongly influences the dynamics of the galaxies, such as the rotation curve of the HI disks, the velocity dispersion of galaxies in galaxy clusters, and the dynamical transition from disk to bulge (Zwicky, 1933; Roberts & Rots, 1973; Ostriker & Peebles, 1973; Rubin et al., 1980; Begeman, 1989)

As I described above, small perturbations embedded in CMB evolve in response to gravity. Overdense regions become denser by accreting materials from underdense regions, while underdense regions gradually become voids. Here, I define the overdensity at a given position r as:

$$\delta(r) = \frac{\rho(r)}{\bar{\rho}} - 1, \tag{1.2}$$

where $\bar{\rho}$ is the average energy density of the Universe. Because it is not possible to map any individual density fluctuation from the CMB to a single galaxy at present, the main idea is to study the statistical properties of the hierarchical structure and compare the ensemble average of the model predictions to observations. The best way to quantify the large scale spatial distribution of the density fluctuation is the two-point correlation function $\xi(r)$, or its Fourier pair, the power spectrum $P(k)$, which expresses the possibility of a given overdensity surrounded by other overdensities within a specific length scale r or wave number k. Inflation theory predicts that the quantum fluctuation follows a Gaussian random field with a simple primordial power-law power spectrum $P(k) \propto k$. This power spectrum is usually called "scale-invariant" because fluctuations of all wavelengths have the same amplitude when they enter the event horizon of the Universe. As soon as fluctuations reach a critical density $\delta_{\rm crit}$ where gravity wins over the cosmic expansion, these structures start to experience run-away collapse and create dark matter halos that potentially host individual galaxies. Given the shape of the primordial power spectrum, linear growth

of the fluctuation, and spherical collapse, Press & Schechter (1974) first estimated the halo mass function at different redshifts as a power-law with an exponential cutoff, so called "Schechter function":

$$\frac{dN}{dM} = \sqrt{\frac{2}{\pi}} \frac{\rho_m}{M} \frac{-d\ln\sigma(M)}{dM} \frac{\delta_{\rm crit}(z)}{\sigma(M)} e^{-(\delta_{\rm crit}(z)/\sigma(M))^2/2}, \tag{1.3}$$

where $\sigma(M)$ is the standard deviation of the Gaussian random field.

As soon as the gravitational bound halos form, the density in the central region becomes so large that the cooling timescale of the baryonic gas becomes much smaller than the Hubble time. Eventually the gas is compressed sufficiently to form stars (Gunn & Gott, 1972; Rees & Ostriker, 1977; White & Rees, 1978). Small galaxies continue to accrete materials from the ambient medium, merge with one another, and, gradually, build up into larger and larger structures. This "bottom-up" structure formation scenario is one of the main predictions of the Λ-CDM paradigm.

1.2 Current status of cosmological simulations

The purpose of computing is insight, not numbers.

Richard Wesley Hamming

The above story line sets the conceptual framework of the hierarchical structure formation of the Universe, but it is important to turn stories into scientific predictions and compare them with observations. Although some analytical models of structure formation I mentioned above can shed light on a small number of statistical properties of the galaxy distribution, their limitations are obvious. First, all of the above predictions are based on linear perturbation theory that assumes the growth of the fluctuation is smaller than unity. As the fluctuation grows into non-linear regime, $\delta > 1$, the linear theory is no longer valid and future evolution is generally too complex to solve analytically. Second, the hierarchical structure growth and the internal

structure of dark matter halos are strongly influenced by halo mergers that associate with phase mixing and violent relaxation. This many-body problem is very complicated and depends strongly on the initial conditions of halo mergers, which can only be calculated via N-Body simulations (e.g. White, 1978; Aarseth & Binney, 1978; May & van Albada, 1984). Finally, and most importantly, the gas and stellar components of the galaxies evolve under huge degrees of freedoms that are governed by baryonic physics of gas accretion, cooling, and heating. This is why numerical simulations are essential to tackle these complex non-linear, non-equilibrium processe (Bertschinger, 1998).

Thanks to the development of numerical algorithms and ever-increasing computing power (Moore's Law), today numerical simulations have become the main, if not the only, tool to study the detail process of galaxy formation and evolution. If the COBE observations mark the golden age of precise cosmology, the uses of supercomputers to solve galaxy formation problems mark the golden age of modeling galaxy formation. Current cosmological simulations that incorporate many relevant physical processes enable us overcome a huge dynamical range from the large scale structures of the cosmic web all the way down to individual star-forming regions. One striking example of the improvements in modeling galaxy formation during the last several decades is a comparison between simulations of galaxy merger in the early 1970s and in the 2010s shown in Figure 1.2. In Toomre & Toomre (1972), the simulated galaxy was represented by a handful of mass points that interact with each other via only gravity, while in Renaud et al. (2015), the merging galaxy was resolved by millons to billions of gas, star, and dark matter elements with sub-pc spatial resolution.

1.2.1 Numerical techniques

As we saw in Section 1.1, on very large scales, the space-time evolution is governed by the Friedman equation; on smaller scales, the dynamics and structures of the dark

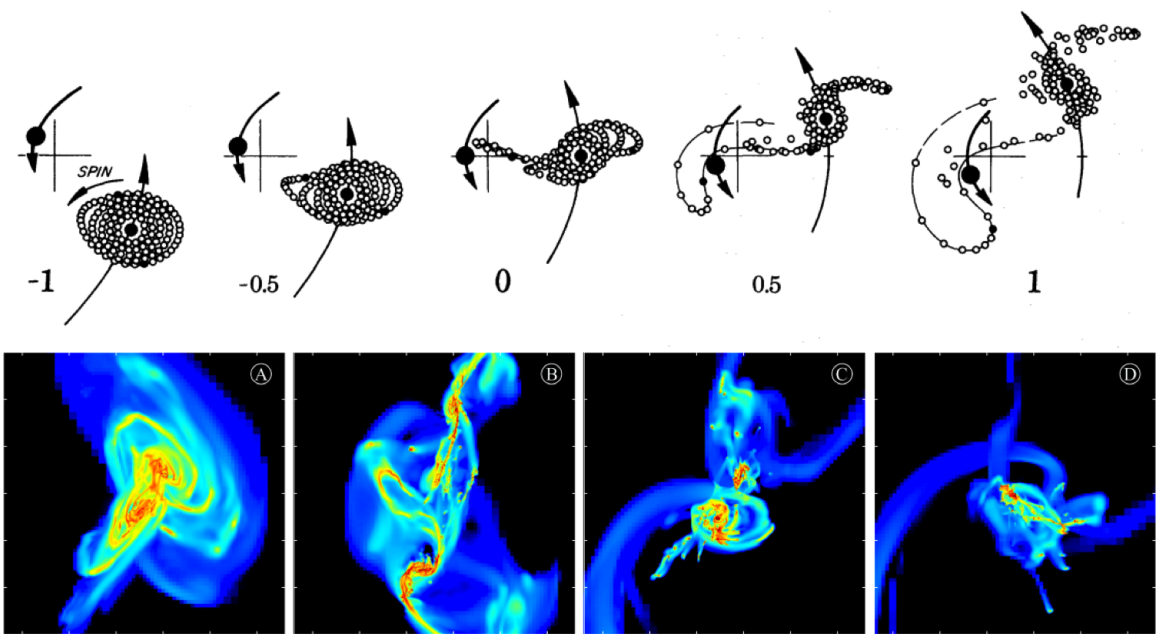

Figure 1.2: Numerical simulations of galaxy merger for Toomre & Toomre (1972) (upper panels) and Renaud et al. (2015) (lower panels). For each series of simulation snapshots, the time evolution of the merging system is shown from left to right.

matter halos are controlled by gravity; within each halo, the properties of the ISM are determined by the gas dynamics of the baryonic matter. Therefore, the whole process can actually be calculated by a combination of the Friedman, Poisson and Eulerian equations.

The Eulerian equations are a set of hyperbolic equations that describe the conservation of mass, momentum and energy:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0, \tag{1.4}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\nabla \Phi - \frac{\nabla P}{\rho}, \tag{1.5}$$

$$\frac{\partial E}{\partial t} + \nabla \cdot [(E + P)\boldsymbol{u}] = -\rho \boldsymbol{u} \cdot \nabla \Phi + (\Gamma - L), \tag{1.6}$$

where ρ is gas density, $\boldsymbol{u}$ is velocity, P is gas pressure, $E = \rho(\epsilon + u^2/2)$ is the total gas energy density, ϵ is the specific internal energy of the gas, Γ and L are the heating and

cooling rates per unit volume, Φ is the gravitational potential. The three equations above are closed by the equation of state of the gas and the Poisson equation:

$$\epsilon = \frac{1}{\gamma - 1}\frac{P}{\rho}, \tag{1.7}$$

$$\nabla^2\Phi = 4\pi G\rho_{\rm tot} - \Lambda, \tag{1.8}$$

where γ is the gas polytropic index and $\rho_{\rm tot}$ is the total density including both baryonic and dark matter components.

The dark matter component is usually represented by collisionless particles in the cosmological simulations that evolve under the influence of gravity only. The kinematics of the i-th dark matter particle located at $\boldsymbol{x}_i$ with velocity $\boldsymbol{u}_I$ can be described by the weak-field approximation of general relativity:

$$\frac{d\boldsymbol{x}_i}{dt} = \boldsymbol{u}_i;\ \frac{d\boldsymbol{u}_i}{dt} = \nabla\Phi. \tag{1.9}$$

Although seemingly simple, the above set of equations are computationally expensive due to the huge dynamical range (from the Mpc large scale structure to the pc scale star-forming regions), large number of resolution elements (from millions to several billions), and complex physical processes (gravity, hydrodynamics, radiation, star formation and feedback). Even pure N-body simulations that treat all matter as collisionless particles require sophisticate algorithms to efficiently and accurately estimate the gravitational interactions of all particles.

The first generation of N^2 particle-particle codes were developed in the 60s to 70s (Aarseth, 1963; Peebles, 1970; Groth et al., 1977). Since then the algorithms have been advanced dramatically and are typically classified into different categories: tree-based solvers (Barnes & Hut, 1986), mesh-based solvers (Klypin & Shandarin, 1983), and some adaptive refinement methods that extend the dynamical range without

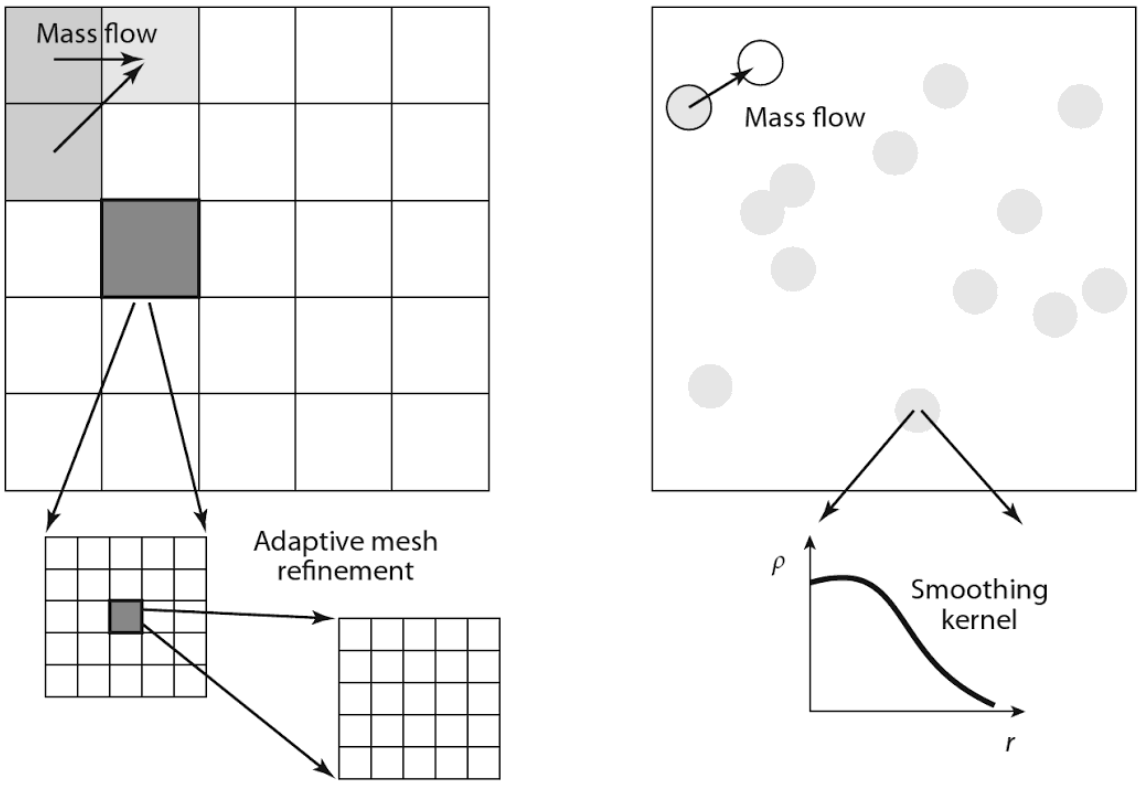

Figure 1.3: Numerical techniques for solving hydrodynamics: mesh-based with adaptive mesh refinement (left) vs. smoothed particle hydrodynamics (right) (Loeb & Furlanetto, 2013).

significantly increasing computational costs (Kravtsov et al., 1997; Bryan & Norman, 1998; Knebe et al., 2001). Since then, great advances in N-body simulations have been made to study the internal structure of the dark matter halos (or the NFW profile; Navarro et al., 1996) and the spatial distribution of galaxy clustering (Conroy et al., 2006; Springel et al., 2006)

Solving hydrodynamic equations are also categorized into different groups: e.g. particle- vs. mesh-based methods. A sketch of the two techniques is shown in Figure 1.3. In particle-based methods, or smoothed particle hydrodynamics (SPH), as implemented in the pioneering work done by Evrard (1988), the gas elements are treated as gas particles whose motion follows the gas flow. The density, momentum, and energy field are reconstructed via kernel density estimation by convoluting particle attributes with specific kernels. Alternatively, for all simulations present in this

thesis, I use the mesh-based Adaptive Refinement Tree (ART) code (Kravtsov et al., 1997; Kravtsov, 1999, 2003; Rudd et al., 2008) to solve the hydrodynamics of the galaxies. In ART, gas elements are treated as gas cells and mass flows are described by exchanging material through cell interfaces. To achieve high spatial resolutions in dense region, cells whose density becomes sufficiently high are divided into 8 children. This refinement process is done iteratively until cells at all levels have similar masses. Both SPH and mesh-basd codes have their own advantages and limitations. Comparing SPH and mesh-based methods is beyond the scope of this thesis, and I would like to refer the reader to some excellent review articles that address this issue (e.g. Springel, 2010; Teyssier, 2015). Following the philosophy of distinguishing the numerical evil from the physical reality, several efforts have been made to compare simulations performed with different codes from the same initial conditions, such as the Santa Barbara Cluster Comparison Project (Frenk et al., 1999) and the most recent AGORA project that I'm involved in (Kim et al., 2014, 2016). It is important to keep in mind that knowing exactly which results are physical and which are merely numerical artifacts are crucial to the interpretation of simulation outcomes.

1.2.2 Uncertainties of sub-grid models

The most important goal of cosmological simulations is to understand the evolution of galaxies, which are most commonly observed from their starlight. Therefore, simulations must take into account the process of star formation in which cold and dense gas in galaxies becomes stars. Following their formation, stars, especially massive ones, quickly deposit a large amount of mass, momentum, energy, radiation, and metals into the ISM during their lifetime. This so-called "stellar feedback" as well as star formation processes happens on sub-pc scales that are far from being spatially resolved by current cosmological simulations with box sizes larger than several Mpc. Therefore, these processes have to be taken into account by alternative numerical

implementations called "subgrid models".

Stars are formed in cold and dense GMCs. In cosmological simulations, the basic sub-grid model for star formation converts a fraction of star-forming gas into star particles at a rate that is calibrated by either observed KS relations or motivated by theoretical estimation of the gas free fall:

$$\dot{\rho}_* = \frac{\epsilon_{\rm ff}\rho_{\rm gas}}{\tau_{\rm ff}}, \tag{1.10}$$

where $\tau_{\rm ff} \propto \rho_{\rm gas}^{-0.5}$ is the local free-fall time, and $\epsilon_{\rm ff}$ is a free parameter that controls the star formation efficiency per free-fall time. This type of sub-grid model was first implemented by Katz (1992) and is little changed since then, with only some small variations taking into account different criteria for creating star particles or functional forms of Eq. (1.10).

Stellar feedback follows the formation of stars. Sub-grid models that mimic the stellar feedback process is another key ingredient in cosmological simulations. Katz et al. (1996) first attempted to model the feedback from supernovae (SNe) by depositing 10^{51} erg thermal energy per SN to the surrounding medium. They soon realized that this has almost no effect on either the properties of the ISM or the suppression of star formation, since the thermal energy from SNe is quickly radiated away by the rapid cooling of the dense gas cells around star particles. Because the inefficient feedback cannot counteract rapid cooling, gas that falls into the dark matter halo is quickly converted into stars. Therefore, given a constant baryonic fraction of the Universe, the stellar mass function of the simulated galaxies has the same shape of the halo mass function I described in Eq. (1.3), which is fundamentally different from observations in terms of both shape and normalization, see Figure 1.4. Besides the inconsistent stellar mass functions, other problems of the previous generations of simulations with inefficient feedback include: the absence of extended stellar disks but

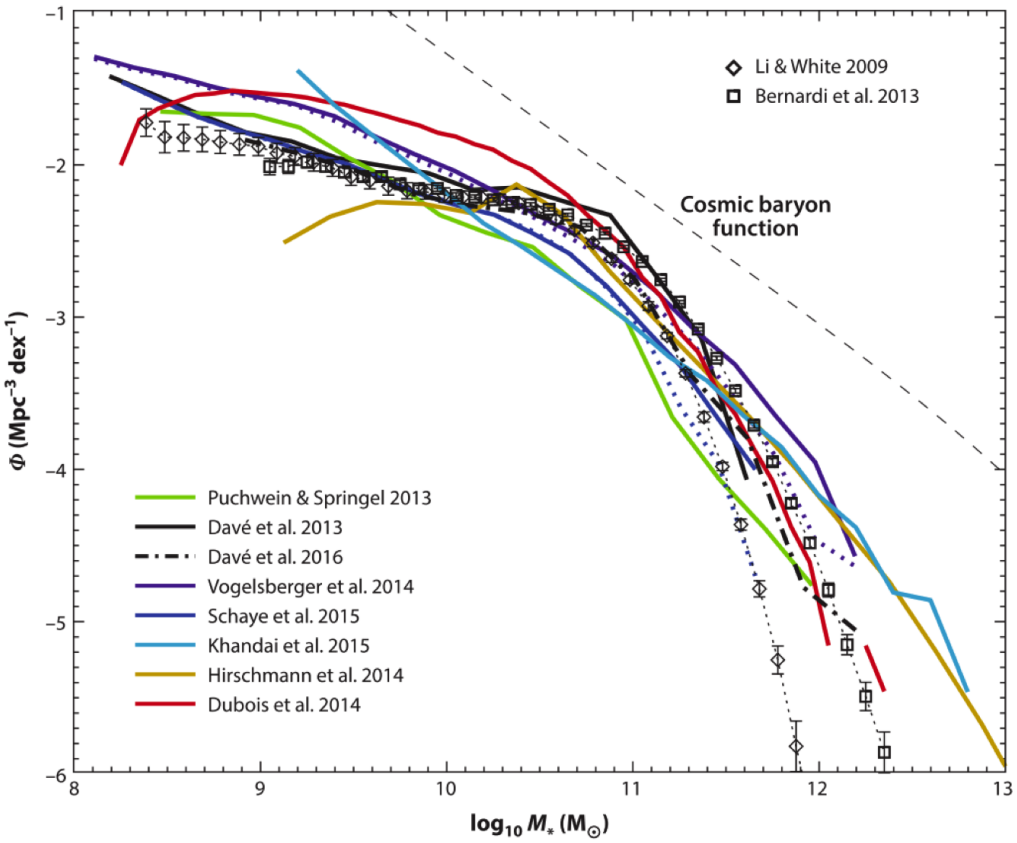

Figure 1.4: Galaxy stellar mass functions for a compilation of recent large-scale cosmological hydrodynamic simulations using different feedback schemes described in Section 1.2. The observed stellar mass functions from Li & White (2009) and Bernardi et al. (2013) are overplotted for reference. The dashed line shows the derived galaxy mass function from halo mass function in Eq. (1.3) with a constant baryonic fraction.

a large bulge component at the center of the z=0 galaxies, a centrally peaked rotational velocity profile, and overproduction of stars at high-z that leads to incorrect characterizations of cosmic star formation history (Katz et al., 1996; Somerville & Primack, 1999; Springel & Hernquist, 2003b; Kereš et al., 2009).

In response to these problems, many groups have developed various ways of effectively injecting feedback energy in cosmological simulations. Stinson et al. (2006) developed a blastwave feedback scheme, where radiative cooling is turned off for gas that is presumably swiped by SNe during the Sedov phase. With this delayed cooling model, star formation activities are significantly suppressed and disk-dominated systems emerge (Governato et al., 2007; Guedes et al., 2011), although this method is

often criticized for being nonphysical. Recently, a more physical-oriented superbubble feedback scheme has been introduced to partially justify the delayed cooling process (Keller et al., 2015).

Another widely used scheme proposed by Dalla Vecchia & Schaye (2012) and adopted by the EAGLE simulations (Schaye et al., 2015) is stochastic energy deposition. This method is used in many SPH simulations, where the gas particles receiving feedback from adjacent star particles have a small probability to be heated to a much higher temperatures than the mean value, which dramatically increases the cooling timescale. As an alternative to thermal energy deposition, Navarro & White (1993) introduced a momentum feedback scheme that add momentum to gas elements. This idea was later implemented in GADGET-2 by Springel & Hernquist (2003a) and was used in the Illustris simulations (Vogelsberger et al., 2014). Specifically, in the Illustris simulations, a new particle type called wind particles is introduced to mimic SNe shocks kicking the ambient medium. Assuming momentum-driven winds with a mass loading factor that scales with the velocity dispersion of the galaxies, wind particles can travel freely across the galactic disks and deposit mass and energy at larger radii, which mimics the behavior of the large scale galactic winds/outflows.

Besides SNe, there are other types of feedback from massive stars, such as stellar winds and radiative pressure. Indeed, the lifetime of GMCs in observations is typically smaller than a couple of free-fall times, even before the first SN explosions. This critical information inspired an exploration of early feedback from other sources prior to SNe. Agertz et al. (2013) estimated the energy budgets from stellar luminosity, winds, and SNe from a single stellar population with a Kroupa (2001) initial mass function, and found that the momentum injection rate from all three sources are comparable. They found that early feedback from radiative pressure and stellar wind reduces the gas density around star particles and enhances the effects of subsequent SN explosions. Similar results were also found in the FIRE simulations (Hopkins

et al., 2014), where the effects of feedback from different sources add non-linearly so that a full accounting of feedback sources is required to reproduce the various global properties of galaxies. Moreover, feedback schemes from non-thermal components, such as cosmic ray, have also been actively explored (Uhlig et al., 2012; Booth et al., 2013; Salem & Bryan, 2014).

As seen in the above description, promising progress has been made through explorations of various sources and numerical implementations of stellar feedback. It is now widely accepted that stellar feedback is indeed a remedy to excessive star formation and bulge-dominated galaxies. But a remedy is not a panacea. The idea that "feedback controls everything" is overly optimistic and misleading. It should be noted that the general agreement with the global properties of the observed galaxies from different simulations is attributed to the fine-tuning of the normalizations and scalings of the feedback models. It is actually troublesome that a surprisingly broad range of feedback models claim to match the same galactic properties, which limits the predictive power of the current cosmological simulations in different perspectives (Naab & Ostriker, 2016), see Figure 1.4.

It seems there are two main issues. First, typically subgrid models are calibrated by comparing the global properties of galaxies on scales larger than kpc scales, such as stellar mass functions, star formation history, stellar mass-halo mass relations, and KS relations. Even though some global relations can be reproduced by these simulations, it is still unknown whether the subgrid models they used are appropriate for capturing the physical properties of the ISM on smaller scales. As the spatial resolution of current simulations is approaching the pc scales (Hopkins et al., 2014; Read et al., 2015; Wetzel et al., 2016), it is critical to develop systematic methods to calibrate the sub-grid models on a similar scale.

Second, in contrast to the active exploration of stellar feedback schemes, the star formation prescriptions have not changed markedly for more than two decades since

the pioneering work of Katz (1992). I emphasis here that stellar feedback processes largely rely on star formation prescriptions. Inappropriate star formation prescriptions can lead to incorrect results, even when equipped with the most sophisticated feedback models. For example, Agertz & Kravtsov (2015) have already shown that only by changing the value of $\epsilon_{\rm ff}$, the star formation history as well as the morphology of the galaxies changes drastically.

Fortunately, this problem has recently been realized in the galaxy formation community and progress is begining to be made to achieve realistic modeling of the star formation process. A series of works by Gnedin et al. (2009); Gnedin & Kravtsov (2010, 2011) developed a phenomenological model of the formation and disruption of H_2 and showed that an H_2-dependent star formation recipe can explain the steepening of the Kennicutt-Schmidt relation in the low surface density regime ($\Sigma_{\rm gas} < 100 M_\odot {\rm pc}^{-2}$). Governato et al. (2007, 2010); Guedes et al. (2011) found that using higher density thresholds for star formation with higher spatial resolutions leads to more concentrated star formation and more effective feedback that significantly improve the ability to produce realistic galaxy disks. In other work, Semenov et al. (2015) implemented a subgrid scale turbulence model in ART and implemented a turbulence-based $\epsilon_{\rm ff}$ model. This model predicted a wide range of $\epsilon_{\rm ff} \sim 0.1 - 10\%$, which is consistent with observed variations of $\epsilon_{\rm ff}$ in local star-forming regions in the Milky Way (Heiderman et al., 2010; Lada et al., 2010; Murray, 2011; Evans et al., 2014).

It can be seen from the above narrative that all significant improvements in numerical modeling of galaxy formation are inspired by observational advances: the inefficient star formation in high-z galaxies leads to the exploration of strong feedback schemes; the observed short lifetime of GMCs implies the importance of early feedback prior to SNe. To deal with current issues in modeling star formation in cosmological simulations, we should step outside of the ivory tower of theoretical

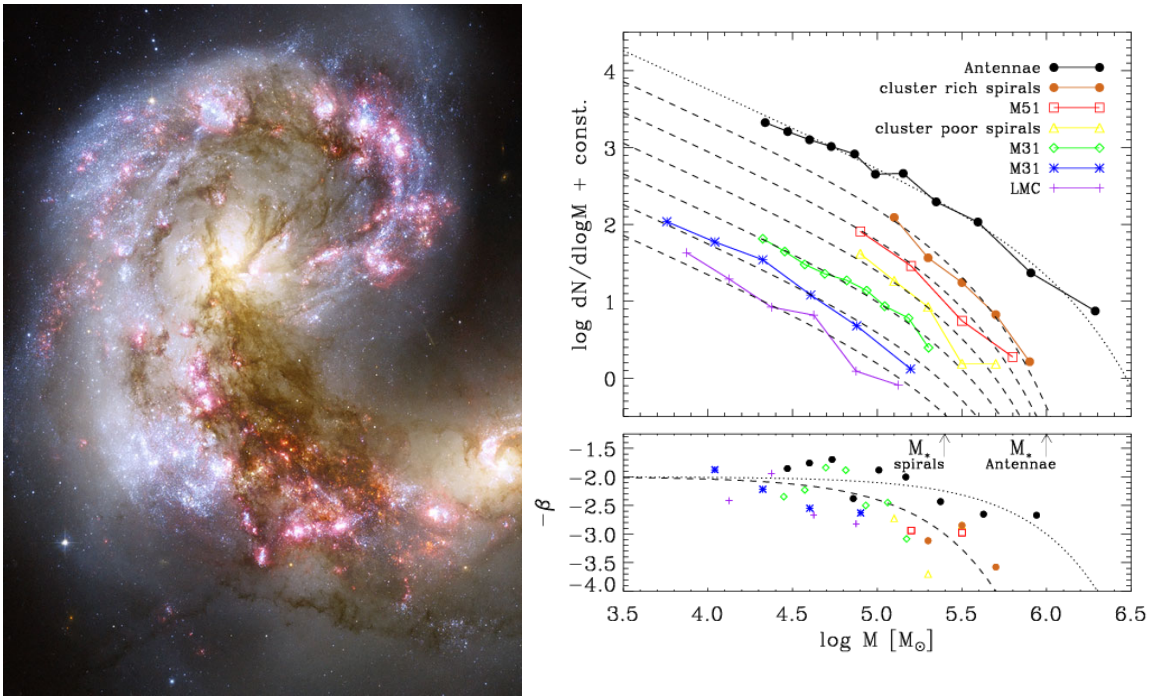

Figure 1.5: Left: The Antennae galaxies in collision by the *Hubble Space Telescope*. Right: Initial mass function of young clusters in different galaxies (Portegies Zwart et al., 2010).

speculations again and survey current observational advances in star formation.

1.3 Properties of young massive star cluster

Observation and theory get on best when they are mixed together, both helping one another in the pursuit of truth. It is a good rule not to put overmuch confidence in a theory until it has been confirmed by observation.

Sir Arthur Stanley Eddington

Look at the optical image of the Antennae galaxies taken by the *Hubble Space Telescope* in the left panel of Figure 1.5. It can be seen that the blue starlight coming from the youngest stars is far from evenly distributed. Instead, these young stars formed in clusters, each of which contains hundreds to even millions of stars. Why do stars form in this way? What are the properties of these clusters? How do these clusters with the size of several pc relate to cosmological simulations of Mpc scales?

These questions will be answered in this section.

In Sec. 1.1, I sketched a brief history of the Universe from the Big Bang to the formation of dark matter halos. Compared to the structure formation in the early Universe, the star formation process occurred later is far more complicated due to the complex interactions among dark matter, radiation, and gas. Pulled by the gravitational potential of the dark matter halo, gas is accreted along filaments. Gas that flows from different directions clashes and converts its kinetic energy into thermal energy so that the halo gas asymptotically approaches a quasi-virialized state. The halo gas dissipates its thermal energy via radiative cooling, loses its pressure support, and gradually falls into the center of the dark matter halos. Its angular momentum, however, cannot be dissipated, thus inducing the formation of the gas disk with a size that is determined by the intrinsic halo spin (Mo et al., 1998). Under the Toomre instability (Toomre, 1964) and spiral density waves (Lin & Shu, 1964), some densest parts of the disk fragment into GMCs. Each GMC is threaded by complex webs of dense filaments, whose intersections produce individual molecular cores. The cold and dense core starts run-away collapsing once its mass reaches the Jeans mass (Jeans, 1902), and individual stars or star multiples emerge. The Jeans mass is defined as:

$$M_J \approx 700 M_\odot \left(\frac{T}{200\mathrm{K}}\right)^{3/2} \left(\frac{n_\mathrm{H}}{10^4 \mathrm{cm}^{-3}}\right)^{-1/2}, \tag{1.11}$$

where T is the gas temperature and n_H is the number density of the hydrogen atoms. The formation of stars requires that gravity overcome the resistant forces of gas pressure, turbulent motions, and magnetic fields. It is not surprising that star formation is one of the most complicated and intriguing problems. Putting aside many details that are not relevant to this thesis, recent observations and theoretical investigations of nearby star-forming regions have led to some general consensus that can potentially be used as a guide to future exploration of sub-grid models and/or as a test

of these models. Indeed, recent observations of star-forming regions in the Milky Way and other nearby galaxies reveal a large number of cluster samples within which the information of their formation environment is embedded (Portegies Zwart et al., 2010).

Most stars are not formed in isolation, but in clustered environments (Lada & Lada, 2003). Quantitatively, the rate of star formation that occurs in clusters for the Milky Way is $\sim 3 \times 10^3 M_\odot \mathrm{Myr}^{-2}\mathrm{kpc}^{-2}$, comparable to the star formation rate surface density of the Galaxy (Miller & Scalo, 1979). This suggests that cluster formation is the dominant mode of star formation in the galaxy. On large scales, the efficiency of star formation in our galaxy is only a few percent (Zuckerman & Evans, 1974; Kennicutt, 1998; Evans, 1999; Evans et al., 2009). For a long time, it was believed that this low star formation efficiency is the result of a collective of low efficient star-forming regions, which is due to slow gravitational collapse hindered by magnetic field or turbulence (Krumholz & McKee, 2005; Krumholz & Tan, 2007). However, this quasi-equilibrium scenario of star formation is challenged by recent observations of star-forming regions.

Herbig et al. (1986) performed a systematic search for T Tauri stars that are presumably at least 1-3 Myr old in the GMCs and found none. Briceno et al. (1997) used X-ray surveys and found no "post-T Tauri stars" in their samples, suggesting that the GMCs they examined produce stars in less than 10 Myr. Hartmann (2009) compiled recent observations of cloud lifetime in his *Table 2.1.* It is striking that the typical lifetime of star-forming clouds in the solar neighborhood is in general shorter than 10 Myr. Ballesteros-Paredes & Hartmann (2007) found that young star clusters of ages ~ 10 Myr are commonly free of dust extinction, suggesting that their natal molecular clouds are dispersed very rapidly. Further, recent observations of nearby star-forming regions reveal a very short age spread of stars in many young star clusters, only a few Myr or a couple of free-fall timescale (Mac Low & Klessen,

2004; Hartmann et al., 2012; Hollyhead et al., 2015), suggesting that cluster formation is rapid, too.

This then raises the question of how such a rapid star formation process can lead to the low star formation efficiency observed in many nearby galaxies. It has recently been recognized that, while the star formation efficiency is low on global scales, many localized regions with high virial parameter tend to have much higher efficiencies, sometimes even reach more than 50% to produce bound star clusters (Murray, 2011). The current understanding of the origin of the low global star formation efficiency and long gas depletion time is caused by a combination of multiple SF-FB cycles and a relative large and wide range of efficiency within individual clouds (Semenov et al., 2017).

Similar to star formation efficiency, the mass of star clusters also presents a wide range spanning from $\sim 100 M_{\odot}$ for typical open clusters to larger than $10^6 M_{\odot}$ for young massive clusters (Portegies Zwart et al., 2010). The initial mass function of young clusters formed in the galaxies can be described by a Schechter function with a power-law slope and an exponential cutoff at high mass. The right panel of Figure 1.5 shows the CIMF of clusters in different galaxies. The power-law slopes of these mass functions are close to -2 and the cut-off mass correlates well with the star formation intensities of the host galaxies.

Here, I summarize some of the key observational indications that can shed light on improving the star formation prescription in cosmological simulations.

- Stars are formed in cluster environments.
- Star formation efficiency of individual GMCs varies dramatically from $\epsilon_{\rm ff} = 0.1\%$ to $\sim 50\%$.
- Cluster formation is a rapid and dynamical process with $\sim$ Myr cluster formation timescale.

- CIMF shows a universal Schechter-like shape with a -2 power-law slope.

The above observations highlight the need for a new star formation prescription that considers the mode of cluster formation. Instead of a fixed $\epsilon_{\rm ff}$ that is used in current cosmological simulations, the new prescription needs to consider an efficiency that varies based on the cloud properties. The rationale of the model and choice of model parameters should be indicated by whether it can reproduce the properties of clusters in observations, such as a short cluster formation timescale and Schechter-like CIMF. In Chapter II, I introduce my new implementation of a continuous cluster formation (CCF) prescription. In this prescription, a cluster particle grows continuously through gas accretion from its natal GMCs, at a rate determined by the cell gas properties. This work is published in Astrophysical Journal (Li et al., 2017). In Chapter III, I compare the model clusters with observations to constrain different choices of star formation and feedback parameters. This work is to be submitted. I also calculate the dynamical evolution of the model clusters and explore the connection between young massive clusters and globular clusters in Chapter IV.

1.4 From young massive clusters to globular clusters

The same as using the properties of young star clusters (YMCs) as testbeds to the modeling of galaxy formation, globular clusters (GCs) are thought for a long time to preserve fossil information on the assembly history of their host galaxies, due to their old age and compact structure (Harris, 2001). For example, the observed bimodal metallicity distribution of GCs in the Milky Way and nearby galaxies provide hints on the chemical composition of their host galaxies in the early Universe (Zinn, 1985; Harris, 1996; Larsen et al., 2001; Peng et al., 2006). The mass-to-light ratio of GCs indicates their dynamical evolution along their orbits around the galaxies (Vesperini & Heggie, 1997; Baumgardt & Makino, 2003; De Marchi et al., 2007; Kruijssen &

Mieske, 2009). However, compared to the concrete observational properties of YMCs, the use of GCs as a tracer of galaxy formation is not straightforward because of both the observational challenges and theoretical uncertainties. Observationally, for example, because of the age-metallicity degeneracy of the isochrone fitting (Worthey, 1994), determining an accurate age or metallicity of the GCs is challenging (e.g. Leaman et al., 2013; Forbes et al., 2015). Theoretically, revealing the mysteries of GC formation requires the understanding of stellar astrophysics, ISM physics, hierarchical structure formation, and cosmology. Therefore, although GC systems have been known for almost four centuries [4], their formation and evolution is still one of the most challenging unsolved problems in the interface among all major fields of astrophysics (e.g. Brodie & Strader, 2006).

Many theories have attempted to explain the physical origin of GC in the early time of the Universe. Peebles & Dicke (1968) modeled GC formation at the epoch after recombination when the cosmological Jeans mass of the largest structure is similar to the mass of GCs $\sim 10^6 M_\odot$. Later, Fall & Rees (1985) connected the formation of GCs with the thermal instability of the hot gas in the mini-halos during the very early stages of galaxy formation. The dark matter halos of these GC progenitors might be stripped and the central clusters survive because of their compact structure (Peebles, 1984). Unfortunately, there are two problems. First, GCs formed in this scenario from individual halos will have rotations much faster than the observed ones in the Galaxy. Second, models of this type cannot explain the existence of younger and metal-rich sub-population of GCs.

As many theoretical speculations continued in the 1990s, one of the key observations is the discovery of a large number of YMCs using the *Hubble Space Telescope* in the Antennae galaxies (Whitmore et al., 1999), see Figure 1.5, and other galaxies (e.g. Portegies Zwart et al., 2010). The CIMF in the Antennae galaxies shows a high-mass

[4] A German astronomer, Abraham Ihle, discovered a galactic globular cluster, M22, while observing Saturn in 1665.

tail that extends to a cluster mass as high as $\sim 3 \times 10^6 M_\odot$, which is massive enough to survive across cosmic time. The existence of such intense star formation activities in the current Universe offers the possibility that YMCs can potentially act as the progenitors of GCs, suggesting that the formation of GC systems may not require unique conditions that only exist in the early Universe. Recent observations in the Hubble Frontier Field reveals a lensed galaxy that contains many stellar clumps of a size < 100 pc with very high star formation rates at $z > 2$ (Johnson et al., 2017b). These sites can potentially host massive star clusters that become the progenitors of GCs observed today. Moreover, the presence of a large number of YMC in Antennae galaxies, an actively merging galaxy system, implies that YMC formation is likely triggered by strong shocks and/or high pressure environments in the ISM that usually happens during major mergers of two galaxies (Kruijssen, 2014), thus linking the origin of GCs to the hierarchical structure formation of the Universe.

The pioneering work of Ashman & Zepf (1992) first proposed that GCs are formed during the gas-rich major-merger events and predicted the bimodality of the GC metallicity distribution even before observations. To explore GC formation in the cosmological framework, Kravtsov & Gnedin (2005) post-processed snapshots of cosmological simulations and assigned GCs at the density peaks of the resolved GMCs. They found a roughly linear correlation between the total mass of GC and the mass of their host galaxies, which is later confirmed by observation of GC systems in different galaxies (e.g. Hudson et al., 2014). Since these simulations can only be carried out at high redshifts, the formation of GCs at later epochs and their evolution cannot be captured. Thus, based on several empirical observed scaling relations, Muratov & Gnedin (2010) built a semi-analytical model upon N-body simulations that reach $z = 0$ and reproduced the bimodality of GC metallicity distribution in the Milky Way.

GC systems have been found not only in our Galaxy but also various types of

galaxies from dwarf spheroids to giant ellipticals. Recent observations of Virgo galaxies reveal a commonly-existed multimodal GC color distributions over a wide range of galaxy masses (Peng et al., 2006). A natural extension to the work Muratov & Gnedin (2010) is to apply this semi-analytical framework to more halo merger trees and explore systematically the origin of the multimodality of GC colors in different galactic environments. This work is presented in Chapter V, which is a recompilation of the published version in Astrophysical Journal (Li & Gnedin, 2014).

CHAPTER II

Star Cluster Formation in Cosmological Simulations.

We present a new implementation of star formation in cosmological simulations, by considering star clusters as a unit of star formation. Cluster particles grow in mass over several million years at the rate determined by local gas properties, with high time resolution. The particle growth is terminated by its own energy and momentum feedback on the interstellar medium. We test this implementation for Milky Way-sized galaxies at high redshift, by comparing the properties of model clusters with observations of young star clusters. We find that the cluster initial mass function is best described by a Schechter function rather than a single power law. In agreement with observations, at low masses the logarithmic slope is $\alpha \approx 1.8 - 2$, while the cutoff at high mass scales with the star formation rate. A related trend is a positive correlation between the surface density of star formation rate and fraction of stars contained in massive clusters. Both trends indicate that the formation of massive star clusters is preferred during bursts of star formation. These bursts are often associated with major merger events. We also find that the median timescale for cluster formation ranges from 0.5 to 4 Myr and decreases systematically with increasing star formation efficiency. Local variations in the gas density and cluster accretion rate naturally lead to the scatter of the overall formation efficiency by an order of magnitude, even when

the instantaneous efficiency is kept constant. Comparison of the formation timescale with the observed age spread of young star clusters provides an additional important constraint on the modeling of star formation and feedback schemes.

2.1 Introduction

One of the most important problems in astrophysics is understanding the formation and evolution of galaxies. Although the Λ cold dark matter paradigm shows great success in reproducing observations on large scales and builds a solid theoretical framework (Planck Collaboration et al., 2014; Vikhlinin et al., 2009; Springel et al., 2006), many galaxy formation questions, especially those that involve baryonic physics, are not yet answered.

A self-consistent way of modeling the baryonic component of galaxies is to run simulations that include all relevant physics such as gravity, hydrodynamics, star formation, radiation transport, etc. This approach has already proved to be a very important tool. Current simulations are making rapid progress in reproducing different types of galaxies and global scaling relations, such as the Kennicutt-Schmidt relation and star formation history, by realizing the importance of stellar feedback in shaping the properties of interstellar medium (ISM; for a recent review, see Somerville & Davé, 2015). Various stellar feedback mechanisms suppress star formation activity at high redshifts and retain a gas reservoir for star formation at low redshifts. This feedback loop allows the self-regulation of star formation in galaxies.

Although many processes of stellar feedback have been explored (e.g. Katz, 1992; Navarro & White, 1993; Stinson et al., 2006; Governato et al., 2007, 2012; Scannapieco et al., 2008; Hopkins et al., 2011; Hummels & Bryan, 2012; Booth et al., 2013; Agertz et al., 2013; Stinson et al., 2013; Ceverino et al., 2014; Salem & Bryan, 2014; Hopkins et al., 2014; Keller et al., 2015), the star formation prescription has not changed for over two decades since Katz (1992) and Cen & Ostriker (1992). In most of the

current cosmological simulations, star particles are formed in cold and dense gas, with the rate that is calculated by assuming a fixed efficiency per free-fall time, $\epsilon_{\rm ff}$. One exception is the recent implementation of the turbulence-based $\epsilon_{\rm ff}$ by Semenov et al. (2016). Star particles are created by a Poisson process, with their masses set beforehand and therefore, unrelated to the feedback they produce. Yet, the efficiencies of star formation and feedback are closely interrelated, as both affect key properties of galaxies (Agertz & Kravtsov, 2015, 2016), and thus must be treated self-consistently.

One possible solution is to grow stellar particles over time and let their own feedback terminate star formation locally. This causes a number of issues that need to be resolved: How to choose the size of a star-forming region? How to treat interactions of neighboring regions? How to tell if the continuous model is more realistic than the instantaneous particle creation? Important constraints on these come from observations of clustering of young stars.

Most stars form in clusters and associations, which can be considered as building blocks of the stellar component of galaxies (e.g. Lada & Lada, 2003). Star clusters follow a well-defined initial mass function (CIMF), analogously to the mass function of individual stars. It follows an approximate power law at low masses but has a steeper falloff at high mass, which can be described by the Schechter function (e.g. Portegies Zwart et al., 2010). The maximum cluster mass scales with the star formation rate of its host galaxy. Clusters form within the dense parts (cores and clumps) of giant molecular clouds (GMCs). Stars in the youngest clusters, still partially embedded in the molecular gas, show a spread of ages of only a few Myr (Mac Low & Klessen, 2004; Hartmann et al., 2012; Hollyhead et al., 2015). Even though resolving the sizes of young clusters ($\sim$ 1 pc) is still beyond capabilities of current cosmological simulations, resolving GMC structure is within reach (e.g. Hopkins et al., 2014; Read et al., 2015; Wetzel et al., 2016). Therefore, the implementation of star formation can be made more realistic if stellar particles corresponded directly to individual clusters

forming within GMCs. The observed properties of young star clusters, such as the CIMF and the cluster formation timescale, can serve as tests of the star formation and feedback prescriptions on a scale much smaller than what is typically used in galaxy formation simulations (~ 1 kpc).

In this paper, we develop a new model for implementing star formation by considering star cluster as a unit of star formation. In contrast to all other cosmological simulations where stellar particles are formed instantaneously, the formation of a stellar particle in our model is resolved in time over a period of a few Myr and terminated by its own feedback. The final cluster masses are set self-consistently, given the implemented physics of stellar feedback. We show that the shape of the model CIMF is consistent with observations of young star clusters, but the cluster formation timescale decreases systematically with increasing local efficiency of star formation and strength of stellar feedback.

We also show that this model leads to the scatter of the cluster formation efficiency by an order of magnitude, even when the instantaneous efficiency per free-fall time is kept constant. This scatter is caused by local variations in the gas density and cluster accretion rate, which are affected by the cluster feedback. The variation of the efficiency is predicted by analytical models of star formation in supersonic turbulent GMCs (e.g. Krumholz & McKee, 2005; Padoan & Nordlund, 2011; Federrath & Klessen, 2012; Hennebelle & Chabrier, 2013), although detailed comparison with our results is beyond the scope of this paper. Small-scale simulations of turbulent cascade also predict systematic variation of $\epsilon_{\rm ff}$ with the virial parameter (e.g., Padoan et al., 2012; Kritsuk et al., 2013; Federrath, 2015). Observational constraints on $\epsilon_{\rm ff}$ on the GMC scale vary from $\sim 0.01\%$ to $\sim 10\%$ (e.g., Murray, 2011; Evans et al., 2014; Vutisalchavakul et al., 2016). Although estimates for some clouds reach values as high as $\epsilon_{\rm ff} \sim 30\%$, these values may be artificially inflated by different evolutionary phases of the GMC (Feldmann & Gnedin, 2011). The fact that our implementation

naturally leads to scatter of cluster formation efficiency paves the way towards more realistic modeling of star formation in the turbulent galactic ISM.

In Section 2.2, we describe the simulation setup and detailed implementation of this sub-grid model. In Section 2.3, we examine the global star formation history of the simulated galaxies. In Section 2.4, we investigate the shape of the cluster initial mass function and its environmental dependence, and compare them with observations of young massive star clusters in nearby galaxies. In Section 2.5, we demonstrate an important effect of the star formation efficiency on the cluster formation timescale. In Section 2.6, we discuss the origin of CIMF in the turbulent ISM and compare our implementation with previous models. Finally, we summarize our conclusions in Section 2.7.

2.2 Simulation Setup

The simulations were run with the Eulerian gasdynamics and N-body Adaptive Refinement Tree (ART) code (Kravtsov et al., 1997; Kravtsov, 1999, 2003; Rudd et al., 2008). The latest version of ART includes several new physical ingredients that make it suitable for investigating the detailed star formation processes in a cosmological context. It includes an updated version of three-dimensional radiative transfer of ionizing and ultraviolet radiation using the Optically Thin Variable Eddington Tensor approximation (Gnedin & Abel, 2001). Both the local ionizing radiation from star particles (Gnedin, 2014) and the extra-galactic background (Haardt & Madau, 2001) are considered as the ionization sources that feed into the radiative transfer solver. The ionization states of various species of hydrogen (HI, HII, H_2) and helium (HeI, HeII, HeIII), as well as the cooling and heating rates, are calculated based on the non-equilibrium chemical network. Finally, a phenomenological model of H_2 formation and self-shielding on dust grains (Gnedin & Kravtsov, 2011) allows us perform a more realistic modeling of star formation based on local molecular component. Re-

cently a subgrid-scale (SGS) model for the numerically unresolved turbulence was implemented in the ART code. Turbulence, generated by gravitational instabilities as well as kinetic feedback from stars and AGN, is an important ingredient for star formation (McKee & Ostriker, 2007). This SGS model has been tested in isolated disk simulations (Semenov et al., 2016) and in this paper it is applied to cosmological simulations.

We run cosmological simulations in a periodic box of size $L_{\rm box} = 4$ Mpc comoving. The initial condition is selected and tested by collisionless runs so that the central galaxy has total mass $M_{200} \approx 10^{12} M_\odot$ at $z = 0$. There are also a few satellite galaxies with halo masses in the range $10^{10} - 10^{11} M_\odot$.

This initial condition has a non-zero "DC mode" that corrects the deviation of the cosmological evolution due to the difference between the average matter density in the box and the average matter density in the whole universe (Gnedin et al., 2011). A constant parameter $\Delta_{\rm DC}$, which represents the density fluctuation level of the current box, determines the relationship between the expansion rate of the box and the expansion rate of the universe:

$$a_{\rm box} = \frac{a_{\rm uni}}{[1 + \Delta_{\rm DC} D_+(a_{\rm uni})]^{1/3}}, \tag{2.1}$$

where $a_{\rm box}$ and $a_{\rm uni}$ are the scale factors of the simulation box and the global universe, respectively. $D_+(a)$ is the linear growth factor of density perturbation at scale factor a. Our initial condition has $\Delta_{\rm DC} = -1.02$, that is, a slightly underdense region.

The ART code uses adaptive mesh refinement, which increases spatial resolution in high density regions during simulation runtime. All of our simulations start with a 128^3 root grid, which gives the mass of dark matter particle $m_{\rm DM} = 1.05 \times 10^6 M_\odot$ and size of the root cell 31.25 kpc comoving. We allow a maximum of ten additional refinement levels, which gives us spatial resolution of $L_{10} = 4 \times 10^6/128/2^{10} \approx 30$ pc

Table 2.1: Model Parameters

Models	SFE10 (fiducial)	SFE20	SFE100	LOWRHO	TURB50	TURBSF	TURBSF2
$\rho_{\rm crit}(\rm cm^{-3})$	1000	1000	1000	10	1000	-	-
$T_{\rm crit}(\rm K)$	20000	20000	20000	20000	20000	-	-
$f_{\rm H2,crit}$	0.5	0.5	0.5	0.5	0.5	-	0.5
$\epsilon_{\rm ff}$	0.1	0.2	1.0	0.1	0.1	-	-
$f_{\rm turb}$	0.1	0.1	0.1	0.1	0.5	0.5	0.5

Note. Other fixed parameters for all models: $\tau_{\rm max} = 15$ Myr; $M_{\rm th} = 10^3 M_\odot$; $D_{\rm GMC} = 10$ pc; feedback speed ceiling $v_{\rm fb}^{\rm ceiling} = 5000$ km/s; feedback temperature ceiling $T_{\rm fb}^{\rm ceiling} = 10^8$ K.

comoving. At $z \approx 4$, the physical size of a $10^{\rm th}$ level cell is about 6 pc, smaller than a typical size of a giant molecular cloud. We employ a combination of the Lagrangian refinement criteria (both DM and gas mass) and the Jeans refinement criteria in the simulation. The cell will be refined if either criteria is fulfilled. For the Lagrangian refinement criteria, the cell will be refined if either the DM mass of the cell exceeds $3 f_{\rm tol} m_{\rm DM} \Omega_{\rm m}/\Omega_{\rm DM}$, or the gas mass exceeds $3 f_{\rm tol} m_{\rm DM} \Omega_{\rm b}/\Omega_{\rm DM}$, where the refinement split tolerance $f_{\rm tol} = 0.6$. For the Jeans refinement criteria, cells larger than twice the local Jeans length $\lambda_{\rm Jeans}$ will be refined. The local Jeans length $\lambda_{\rm Jeans}$ is defined as

$$\lambda_{\rm Jeans} = v_{\rm gas} \tau_{\rm ff} \approx \sqrt{\frac{\pi (c_s^2 + \sigma_v^2)}{G \rho_{\rm gas}}}, \tag{2.2}$$

where c_s is the sound speed of the cell, and σ_v is the local gas velocity dispersion computed by the root mean square value of the velocity differences between a given cell and its six immediate neighbors.

We adopt a ΛCDM cosmology with $\Omega_{\rm m} = 0.304$, $\Omega_{\rm b} = 0.048$, $h = 0.681$, and $\sigma_8 = 0.829$ that is consistent with the most recent *Planck* result (Planck Collaboration et al., 2016).

2.2.1 Continuous Cluster Formation

We introduce a new model for the formation of stellar particles in cosmological simulations: continuous cluster formation. In this prescription, cluster particles are formed within a spherical region of fixed physical size, $D_{\rm GMC} \sim 10$ pc. This setup can avoid sudden changes of cell size by mesh refinement and gradual changes due to the expansion of the Universe.

In most of our runs, cluster particles are created only in cold ($T < T_{\rm crit}$), dense ($\rho > \rho_{\rm crit}$) cells with high molecular fraction ($f_{\rm H2} > f_{\rm H2,crit}$) [1]. Also, we allow particle creation only when the spherical cluster formation region is located at the local density maximum by comparing the density of a given cell with its 6 immediate neighbors. Once created, cluster particles are labeled "active" and will grow during the cluster formation timescale $\tau_{\rm max} = 15$ Myr. Clusters with age older than $\tau_{\rm max}$ will be labeled "inactive" and stop growing, because the observed age spread in young star clusters is typically below 15 Myr. Another constraint on cluster formation is that new cluster particle can be created only in a cell that does not already contain another active particle. Otherwise, we will grow the existing cluster instead of creating a new one.

All cluster particles in our simulations are considered to be collisionless elements, which go into the gravity solver. The Poisson equation is solved by FFT at the root level, and by the relaxation method at all refinement levels. These particles are mapped to Eulerian grids; therefore, they experience the same potential irrespective of their masses.

Since the resolution of the simulations can reach 30 comoving pc, a sphere of physical diameter 10 pc can be larger than a single cell at high redshifts. In this case, the neighboring cells that are covered by the sphere should also participate in cluster formation. We developed an algorithm that returns gas properties of central, as well

[1] In contrast to the fixed $\epsilon_{\rm ff}$ runs, the turbulence-based $\epsilon_{\rm ff}$ runs does not require an ad hoc density, temperature, or molecular fraction threshold for star formation, as $\epsilon_{\rm ff}$ is naturally suppressed in warm diffuse gas (see Semenov et al., 2016) .

as 26 neighboring cells, and calculates the total star formation rate (SFR) within the sphere. But the volume participating in cluster formation cannot be larger than the 27 cells.

The particle growth rate is calculated as follows. The SFR density of the central cell is:

$$\dot{\rho}_* = \epsilon_{\rm ff} \frac{f_{\rm H_2} \rho}{\tau_{\rm ff}}, \tag{2.3}$$

where $\tau_{\rm ff} = \sqrt{3\pi/32G\rho_c}$ is the free-fall time of the central cell, $\epsilon_{\rm ff}$ is the star formation efficiency per free-fall time, and $f_{\rm H_2}$ is calculated based on the non-equilibrium chemical network that is described in Gnedin & Kravtsov (2011). The total mass accumulation rate within the sphere is the sum of the contributions from the central as well as 26 neighboring cells:

$$\dot{M} = \sum_i f_{{\rm sp},i} V_i \dot{\rho}_{*,i} = \frac{\epsilon_{\rm ff}}{\tau_{\rm ff}} \sum_i f_{{\rm sp},i} V_i f_{{\rm H_2},i} \rho_{{\rm gas},i}, \tag{2.4}$$

where V_i is the volume of cell i, and $f_{{\rm sp},i}$ is the fraction of V_i that is located within sphere. Here the SFR density of the neighboring cells is estimated by the free-fall time of the central cell, since the collapse of the whole spherical cluster formation region is dominated by the central cell. The mass increment during simulation timestep Δt is then $\dot{M}\Delta t$.

The star formation efficiency $\epsilon_{\rm ff}$ in model SFE10 (fiducial) is set to be ten percent. Model SFE20 ($\epsilon_{\rm ff} = 20\%$) and SFE100 ($\epsilon_{\rm ff} = 100\%$) test the influence of the star formation efficiency on the global SFR as well as cluster particle properties. Recently, Padoan et al. (2012) found a relationship between $\epsilon_{\rm ff}$ and cloud virial parameter ($\alpha_{\rm vir}$) by analyzing MHD simulations of turbulent GMCs. Semenov et al. (2016) implemented a SGS turbulence model in the ART code to study the influence of turbulence on gas dynamics and explore the turbulence-based star formation efficiency in isolated disk simulations. In this paper, we setup a simulation TURBSF to test

this model in cosmological runs. The SFR density is evaluated as $\dot{\rho}_* = \epsilon_{\rm ff}(\alpha_{\rm vir})\rho/\tau_{\rm ff}$. We also apply this turbulence-based $\epsilon_{\rm ff}$ on molecular gas (TURBSF2) by adding the molecular fraction $f_{\rm H_2}$ in the above rate: $\dot{\rho}_* = \epsilon_{\rm ff}(\alpha_{\rm vir})f_{\rm H_2}\rho/\tau_{\rm ff}$. The parameters of the runs are listed in Table 2.1.

For all simulations discussed in this paper, we take 10 pc, a typical size of dense clumps in GMCs (e.g., Murray, 2011), as the diameter of the cluster formation sphere, $D_{\rm GMC}$. This value is consistent with the size of massive cluster-forming clouds that are observed in the recent sub-mm ATLASGAL survey (Urquhart et al., 2014). The range of the effective radius for massive proto-cluster candidates in that survey is about 5-10 pc. The value of $D_{\rm GMC}$ is tightly constrained. Any small sphere size would not contain enough material to form massive clusters, while a significantly larger sphere size could include several regions that should be forming separate clusters.

Because $D_{\rm GMC}$ is an important parameter in our model, we test its influence on the formation of cluster particles by varying its value around 10 pc. In the case of $D_{\rm GMC} = 5$ pc, the sphere is small enough so that it completely resides within the central cell and cannot reach any neighboring cells, for all redshifts of interest $z < 11$. We find that the masses of the model clusters are smaller than those in the case of $D_{\rm GMC} = 10$ pc by a factor of $1.8 - 2.0$. The CIMF exhibits a systematic shift to lower masses so that clusters with mass larger than several million solar mass are seldom formed. We have done additional test runs with $D_{\rm GMC}$ increasing from 5 to 15 pc, and found that the CIMF first extends to higher masses, and then gradually saturates at $D_{\rm GMC} > 10$ pc. This suggests that a sphere with $D_{\rm GMC} = 10$ pc contains majority of the molecular gas that will collapse into a single cluster-forming region.

However, when the size of the sphere is much larger than the gas cell, $D_{\rm GMC} > 3L_{\rm cell}$, the sphere will cover more than one layer of neighboring cells. Extracting gas properties of all cells within the sphere makes the book-keeping complicated and significantly reduces the computation efficiency. Therefore, we only allow cluster

particles to grow their mass from the gas within the closest 27 cells. With the choice of $D_{\rm GMC} = 10$ pc, the cell at the $10^{\rm th}$ level has size smaller than $D_{\rm GMC}/3$ at $z > 8$. In this case, some fraction of the material in the sphere that lies beyond the 27 cells cannot participate cluster formation. For this reason, we control the refinement criteria so that the smallest cell size is always larger than $\rm D_{GMC}/3$. This suggests that gas cells are not allowed to reach the $10^{\rm th}$ level of refinement until $z \lesssim 8$.

We also explored the possibility of varying sphere size based on the observed mass-size relation of local GMC (e.g. Larson, 1981). However, it requires several iterations to obtain the appropriate cloud size and sometimes fails to converge. Another uncertainty comes from the normalization of the mass-size relation for GMCs in high-z environment. Therefore, we did not apply it in our current simulations.

As we mentioned above, if there exists an active cluster particle in a cell, the cluster mass will grow with the rate given by Eq. (2.4). The momentum and metallicity increment will also be added to the active particle accordingly. If there are more than one active cluster particle in a cell, we only grow the one with the smallest velocity relative to the host cell. This method prevents adding mass to unrelated cluster particles that occasionally fly through the cell. We also explored an alternative scenario of treating multiple active particles in a single cell by merging them into one active cluster. We found that the CIMF is no longer a power-law shape, but highly inclined toward high masses. This is possibly because that active clusters that fly from other cluster formation regions are absorbed by the main cluster, and clusters in different cluster formation clouds are merged.

Before creating a new cluster particle, we predict its mass accumulated over the timescale ($\tau_{\rm max}$) by assuming a constant SFR: $M_{\rm est} = \tau_{\rm max}\dot{M}$. If this mass is smaller than the threshold mass $M_{\rm th}$, the cluster particle will not be formed in order to avoid small-mass particles. After each global timestep, we remove all inactive cluster particles if their masses are smaller than $M_{\rm th}$, and recycle their mass and momentum

back to the gas cells. To justify this process, we compare the scientific results and the computational efficiency with and without particle removal. We found that these inactive low-mass clusters contain only 3-5% of the stellar mass and have a negligible effect on the overall stellar component of the simulated galaxies, but they contribute ∼40% in paricle number and slow down the code. Therefore, it is appropriate to recycle such particles to improve the computational efficiency.

We would like to understand the accretion history of cluster particles. The typical timestep in our simulations is about 1000 years [2]. The active cluster formation period can be resolved by several thousand steps. Since it is not practical to store each local timestep, we record two integral timescales that characterize the mass accretion history. The first is the total duration of a star-forming episode, $\tau_{\rm dur}$, from the time of creation of the cluster particle to the last time before it becomes inactive. The second is the mass-weighted cluster formation timescale:

$$\tau_{\rm ave} \equiv \frac{\int_0^{\tau_{\rm dur}} t\,\dot{M}(t)\,dt}{\int_0^{\tau_{\rm dur}} \dot{M}(t)\,dt}, \tag{2.5}$$

where $\dot{M}(t)$ is the mass accumulation rate of a given cluster at time t, measured from the moment of particle creation. By construction, the duration $\tau_{\rm dur}$ cannot exceed the maximum formation time $\tau_{\rm max}$. If the mass accumulation history is a power-law, $\dot{M}(t) \propto t^n$, the relationship between $\tau_{\rm ave}$ and $\tau_{\rm max}$ can be estimated as $\tau_{\rm ave} = \tau_{\rm max}\frac{n+1}{n+2}$. For example, a time-independent $\dot{M}$ gives $\tau_{\rm ave} = \tau_{\rm max}/2$. If $\dot{M}$ decreases over time, as a result of self-feedback of newly formed stars, we expect $\tau_{\rm ave} < \tau_{\rm max}/2$. We discuss the accumulation history of star clusters in Section 2.5.

[2] For a gas cell with size $L_{\rm cell} \approx 10$ pc, Courant-Friedrichs-Lewy condition requires a timestep smaller than $\Delta t < L_{\rm cell}/v_{\rm gas} = 10^4{\rm yr}\ (L_{\rm cell}/10\,{\rm pc})\ (v_{\rm gas}/1000\,{\rm km/s})^{-1}$. Due to the discreteness of the timestep refinement among various levels, the timestep at the highest level is even shorter if it is passed from lower level cells with high velocity and high temperature. We find that the typical timestep for our fiducial run is about 1000 yr.

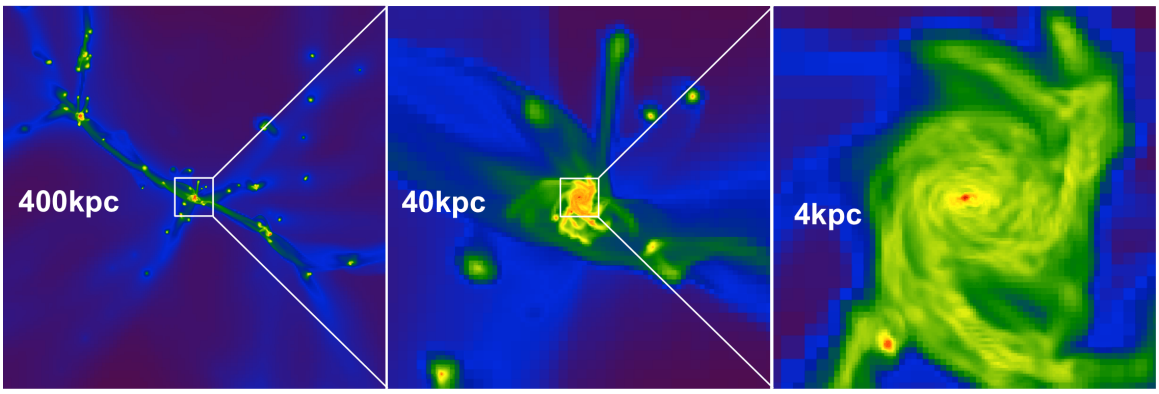

Figure 2.1: Gas density projection plots of the fiducial run centered on the main galaxy at $z \approx 3.3$. The physical sizes of box for the three panels, from left to right, are 400 kpc, 40 kpc, and 4 kpc, respectively.

2.2.2 Dynamical Disruption of Star Clusters

Above we described the detailed implementation of the creation and growth of cluster particles. Another key ingredient, especially for the study of globular clusters, is their subsequent dynamical evolution. Star clusters suffer tidal dissolution during the early stage when they are located within the gaseous disk, and experience gradual evaporation of stars after they escape to the galaxy halo. We create a new variable, the bound fraction $f_{\rm bound}$, to represent the fraction of mass that is still bound to the cluster particle at a given time, and update $f_{\rm bound}$ based on the local tidal field as well as the rate of internal evaporation (Gnedin et al., 2014). We assume that the unbound part remains near the cluster, so that the overall stellar mass distribution is unaffected by cluster disruption. Since this paper focuses mainly on the formation process of young star clusters, we will not discuss here the details of calculating cluster evolution, and leave this topic to a follow-up paper.

2.2.3 Stellar Feedback

The creation of cluster particles is followed by stellar feedback that returns mass, momentum, and energy into the surrounding medium. The feedback process has been

demonstrated to be crucial in maintaining steady star formation activity.

The sub-grid stellar feedback model used in our simulations is similar to that described in Agertz et al. (2013). It includes the injection of thermal energy, momentum, metals, and ionization radiation from supernova (SN) explosion, stellar winds, stellar luminosity, and radiative pressure onto surrounding gas and dust. The SN rate of a given cluster particle is calculated by assuming the Kroupa initial mass function (Kroupa, 2001), and the injected energy and momentum are calibrated by Padova stellar evolution models (see details in Agertz et al., 2013). An analytical fit by Gnedin (2014) is used to account for the time evolution of the ionization radiation from young stars. During the growth of active cluster particles, we use the mass-weighted age as the age of the stellar population, which determines the amount of feedback. We also include the SGS turbulence model that treats the unresolved turbulent energy as a separate hydrodynamic field. This field is similar to the "feedback energy variable" in Agertz et al. (2013), but in addition to isotropic non-thermal pressure support it includes anisotropic production of the SGS turbulence by cascade from the resolved scales, turbulent diffusion and dissipation of turbulence over local crossing time rather than fixed rate of decay assumed in Agertz et al. (2013) (see details in Semenov et al., 2016; Schmidt et al., 2014).

The momentum exerted from stellar particles is distributed evenly to 26 nearest neighbors surrounding the parent cell of the stellar particles. The feedback momentum is added directly into the cell if it has the same direction with the momentum of that cell. Otherwise, the feedback momentum is subtracted from the cell momentum and the kinetic energy associated with the canceled momentum is converted to thermal energy and added to the cell internal energy. When adding feedback energy or momentum onto a given cell, we also check whether the temperature (or velocity) of the cell is larger than $T_{\rm fb}^{\rm ceiling} = 10^8$ K (or $v_{\rm fb}^{\rm ceiling} = 5000$ km s^{-1}). To avoid creating unrealistic hot and fast flow, feedback does not add to cells whose temperature or

velocity has already reached the ceiling value.

Although the feedback model contains many physical ingredients from various sources, it should be noted that the only adjustable parameter is the fraction of SN energy that is converted into turbulent energy, $f_{\rm turb}$. In the fiducial run, we set $f_{\rm turb} = 10\%$ and we vary this factor to 50% in TURB50 model to test the effect of turbulence on our results.

2.3 Global properties of simulated galaxies

The simulations stall at low redshift because of the very short timestep. Our fiducial run has reached $z = 3.3$ and most of the results about it are from that snapshot. We use the ROCKSTAR halo finder (Behroozi et al., 2013c) to identify halos and substructures in the whole simulation box, and construct halo merger trees by the consistent-tree algorithm (Behroozi et al., 2013d).

Figure 2.1 shows the gas density projection plots of the main galaxy for the fiducial run at $z \approx 3.3$ in three different scales from 400 kpc to 4 kpc. The filamentary structure that connects galaxies is clearly seen in the left panel. In the central panel, the gaseous disk of the main galaxy is shown in a face-on view, with obvious spiral patterns. There are also substructures orbiting the main galaxy, including significant tidal streams. In the right panel, the simulation is zoomed in to 4 kpc, and the discreteness of outer gas cells becomes visible. In the outer edge of the galaxy, the density is low and the mesh is coarse, while at the center, the highest refinement level is reached due to the high density.

We also find that gas fragmentation due to gravitational instability generates a large number of dense clouds along the spiral arms. Massive star clusters are formed in these clouds. Their radiation ionizes the gas nearby, and some fraction of the radiation escapes the host galaxy and contributes to the extragalactic UV flux that reionized the universe at high redshift (e.g., Gnedin, 2016). The high spatial resolution and

three-dimensional radiative transfer implemented in our simulations make it possible for us to study the escape of radiation from young clusters, and in particular, estimate the escape fraction in the inner few hundred parsecs of the galaxy. Since the main goal of this paper is to study the properties of star clusters, we present a preliminary analysis of the escape fraction in Appendix.

Figure 2.2 shows the star formation histories (SFH) of the main progenitor for different models, compared to the SFH of an average $M_{\rm vir}(z = 0) = 10^{12}\,M_{\odot}$ halo derived from the abundance matching technique (Behroozi et al., 2013b). The star formation rate is calculated from the cluster particles within the main galaxy at each snapshot and smoothed over 50 Myr.

A similar feature of all SFHs is the general rise towards lower redshift, with periodic bursts of star formation. Two noticeable bursts at $z \approx 6 - 7$ and $z \approx 4 - 5$ coincide with two major-merger events that are identified from the merger trees. The star formation rates reflect a particular mass assembly history encoded in the initial conditions of our simulations, but also depend on the model parameters such as $\epsilon_{\rm ff}$ and $f_{\rm turb}$. For example, the peaks of the bursts shift to later times with decreasing $\epsilon_{\rm ff}$ and increasing $f_{\rm turb}$, and the fiducial run shows significant reduction of SFR after the second burst.

It also appears that the feedback in our simulations is not strong enough to match the average SFR predicted by Behroozi et al. (2013b). It could be due to the implementation of feedback processes, or the specific initial condition we chose for these runs. As we described in Section 2.2.3, subgrid turbulence in our simulations decays over the local crossing time. For strongly turbulent cells, this timescale can be much shorter than 10 Myr, a constant dissipation time that was assumed in Agertz & Kravtsov (2015). As a result, the sub-grid turbulence decays faster in our model and the turbulent pressure becomes insufficient to push the gas away from the cluster-forming regions. Therefore, the dynamical effect of the sub-grid turbulence in our

simulations is not as significant as that in Agertz & Kravtsov (2015). In addition, even simulations with stronger feedback implementation, such as FIRE (Hopkins et al., 2014), show a larger range of SFH variation than predicted by the abundance matching. For example, their run m12q has higher SFR at high redshift due to a particular choice of "quiescent" initial conditions.

2.4 Cluster initial mass function

2.4.1 Power law vs. Schechter function

One of the most fundamental properties of young star clusters is the CIMF. Observations of nearby galaxies have found that CIMF follows the Schechter function form: a power-law distribution with an exponential cutoff at $M_{\rm cut}$ (see Portegies Zwart et al., 2010, and references therein):

$$\frac{dN}{dM} \propto M^{-\alpha} \exp\left(-M/M_{\rm cut}\right). \tag{2.6}$$

The power-law index lies in a fairly narrow range $\alpha \approx 1.8 - 2.2$, gradually getting steeper for less massive galaxies. The cutoff mass varies significantly more from galaxy to galaxy and scales most strongly with the star formation rate of the host galaxy. The typical value of $M_{\rm cut}$ for Milky Way-sized spiral galaxies is about $2 \times 10^5\, M_\odot$ (e.g. Gieles et al., 2006), increasing to $M_{\rm cut} > 10^6\, M_\odot$ for luminous interacting galaxies (Bastian, 2008).

Figure 2.3 shows the model CIMF in the main galaxy for the fiducial run. The CIMF exhibits the Schechter function-like shape with a power-law slope similar to the observed. The normalization and cutoff mass vary with cosmic time, increasing at least until $z = 4$. Both variations are correlated with the galaxy SFR. All of these trends are consistent with the observations of young star clusters, which means our implementation of cluster formation, at least in its general features, is realistic.

To investigate the relation between SFR and the high mass end of the CIMF at different epochs, we make bins of clusters formed within 50 Myr of each other. As this interval is more than three times the longest cluster formation time, we can treat each bin as a roughly independent measurement. For each bin, we estimate the average SFR, maximum cluster mass $M_{\rm max}$, the best-fit power-law slope α, and cutoff mass $M_{\rm cut}$. The average value of the best-fit slope for all bins is $\alpha \approx 1.8$, similar to the observed.

Figure 2.4 shows a strong correlation between the SFR and $M_{\rm max}$, as well as $M_{\rm cut}$, for bins with SFR $> 1\,M_\odot$/yr. The best-fit power law relationship between SFR and $M_{\rm cut}$ is:

$$M_{\rm cut} \approx 1.4 \times 10^4\,M_\odot \left(\frac{\rm SFR}{1 M_\odot/{\rm yr}}\right)^{1.6}, \tag{2.7}$$

and the relationship between SFR and $M_{\rm max}$ is:

$$M_{\rm max} \approx 8.8 \times 10^4\,M_\odot \left(\frac{\rm SFR}{1 M_\odot/{\rm yr}}\right)^{1.4}. \tag{2.8}$$

At SFR $< 1\,M_\odot$/yr, the small samples of clusters may hinder accurate determination of the mass function shape.

Theoretically, for samples drawn from a given mass function, the maximum mass can be calculated by the integral equation: $N(\geq M_{\rm max}) = 1$. Comparing the estimated $M_{\rm max}$ with the actual value can help us validate or rule out specific functional forms of dN/dM. The normalization of CIMF comes from the total mass of clusters formed within our chosen time interval, which is simply $M_{\rm tot} = {\rm SFR} \times 50$ Myr. Then it is only the shape of CIMF that determines the maximum cluster mass.

For the case of a pure power-law, $dN/dM \propto M^{-\alpha}$, with $1 < \alpha < 2$ and $M_{\rm min} \ll M_{\rm max}$, simple integration gives $M_{\rm max} \approx \frac{2-\alpha}{\alpha-1}\,M_{\rm tot}$. From Figure 2.4, we can see that this theoretical $M_{\rm max}$ overestimates the actual maximum cluster mass in the simulations by more than one order of magnitude.

Alternatively, for the case of the Schechter function CIMF as described by Eq. (2.6), the relationship between $M_{\rm tot}$ and $M_{\rm max}$ is

$$M_{\rm tot} = \frac{\Gamma(2-\alpha, M_{\rm min}/M_{\rm cut}) - \Gamma(2-\alpha, M_{\rm max}/M_{\rm cut})}{\Gamma(1-\alpha, M_{\rm max}/M_{\rm cut})} M_{\rm cut}, \tag{2.9}$$

where $\Gamma(s, x)$ is the upper incomplete Gamma function. Taking $\alpha = 1.8$, $M_{\rm min} = 10^3\,M_\odot$, and the mean relation between SFR and $M_{\rm cut}$ from Eq. (2.7), the expected $M_{\rm max}$ can be obtained by solving Eq. (2.9) numerically. The result is shown by solid black line in Figure 2.4. An approximate solution is $M_{\rm max} \approx M_{\rm cut}\ \ln\left(M_{\rm tot}/M_{\rm cut}/\Gamma(2-\alpha)\right)$. This value is much less sensitive to the total mass than the power-law CIMF, and is closer to the cutoff mass.

It should be emphasized that the black line in Figure 2.4 is not the best-fit relation between SFR and $M_{\rm max}$, but the expected $M_{\rm max}$ calculated from the above procedure. The similarity between this expected $M_{\rm max}$ and the one obtained from the simulations suggests that the Schechter function is an excellent representation of the CIMF in our simulation, while a pure power-law distribution is ruled out.

2.4.2 Spatial variation

Another important aspect of CIMF is its spatial variation. Recently, Adamo et al. (2015) examined the mass function of young star clusters in M83 using multiband *Hubble Space Telescope* imaging data. They split the whole cluster sample into four radial bins and found significant steepening of the CIMF in the outer bins, as well as decrease in the maximum cluster mass. Both trends were related to the steady decrease of the surface density of SFR with radius.

Following a similar procedure, we divide our main galaxy at $z \approx 3.3$ by four concentric circles so that each annulus contains the same number of star clusters with age younger than 100 Myr. The CIMF of each bin is shown in Figure 2.5. The

mass function no longer has a universal shape, but presents a spectrum of power-laws with various slope: the inner bins ($R < 100\,\mathrm{pc}$) have $\alpha < 2$, while the outer bins ($R > 100\,\mathrm{pc}$) have $\alpha > 2$, in the mass range $10^4 - 10^6\,M_\odot$. The trend of the CIMF steepening with galactocentric distance is consistent with the spatial variation of CIMF in M83. Moreover, the maximum mass of clusters in the inner bins is much higher than that in the outer ones. The SFR density for the four bins from inner to outer ones is 460, 14, 0.06, and 0.02 $M_\odot\,\mathrm{yr}^{-1}\,\mathrm{kpc}^{-2}$, respectively. It is clear that the gradient of CIMF correlates with the local SFR density. It is a local manifestation of the same global trend shown in Figure 2.4.

2.4.3 Dependence on models of star formation and feedback

We now explore the variation of CIMF under different model parameters. In Figure 2.6, we show the CIMF of all models in the main galaxy at the same epoch. It is a higher redshift ($z \approx 5.3$) than previous plots for the fiducial model ($z \approx 3.3$) because the other runs did not advance as far in time, for reasons of computational efficiency.

Models SFE20 and TURBSF present similar CIMF to the fiducial run, indicating that our cluster formation prescription is not sensitive to model parameters when they are in a reasonable range. In detail, the slope at $M < 10^5\,M_\odot$ steepens somewhat with increasing ϵ_{ff} and, for the one model we explored (TURB50), with increasing f_{turb}. In all these cases the range of α is consistent with observations of different nearby galaxies. At high cluster masses the models converge even closer.

The only exception is Model SFE100, with 100% star formation efficiency per free-fall time, which has a dramatically different CIMF. It deviates from a power-law distribution at low mass, and has the truncation mass that is much lower than the other runs. Although the inability to create massive clusters with a very high star formation efficiency seems counter-intuitive, it reflects the complex nature of the

interplay between star formation and feedback: very high star formation efficiency leads to an early starburst; stellar feedback from this burst is strong enough to destroy molecular gas and expel material from the star-forming region; in a short time, the cluster growth is terminated and the final cluster mass is determined only by the first starburst episode. This scenario is confirmed by the analysis of cluster formation timescales in Section 2.5.

The differences between the low-SFE (TURBSF, TURBSF2) and high-SFE runs (SFE10, SFE20, TURB50) are reinforced at later times. These five runs have the last common output epoch at $z \approx 4.3$. The high-SFE runs have a slightly steeper slope $\alpha > 2$ at $M > 10^4\,M_\odot$, which gradually turns over at lower cluster mass and reaches $\alpha = 2$ at $M \approx 10^3\,M_\odot$. The low-SFE runs show the turnover of the slope to $\alpha < 2$ already at $M \sim 10^5\,M_\odot$. Lower SFE and weaker feedback runs also reach larger maximum cluster mass, up to a factor of two. These differences are consistent with the overall trend that more extended star formation events lead to larger eventual cluster mass.

2.4.4 Dependence on galaxy mergers

To investigate the environmental dependence of CIMF, we plot in Figure 2.7 the relationship between the star formation history of the main galaxy and the major merger events of its host halo. The SFR is split in two parts, contributed by low-mass and high-mass clusters. The merger ratio is defined as the differential increase of the main halo mass between adjacent snapshots: $R_m \equiv (M_{h,i} - M_{h,i-1})/M_{h,i-1}$, where $M_{h,i}$ is the mass of the main halo in the $i^{\rm th}$ snapshot. The progenitor and descendant information of the main halo is extracted from the merger tree calculated by ROCKSTAR. Only mergers with $R_m > 0.3$ are shown in the Figure, and the durations of the two merger events are determined by visual inspection of the dark matter density distribution across several snapshots. Quantitatively, we set the

starting point of a merger as the time when the host-satellite separation becomes less than ~ 30 kpc, following Behroozi et al. (2015) who found systematic enhancement of SFR for galaxy pairs with projected separation smaller than about 60 kpc, but not at larger separations.

We can verify indeed that the last merger with mass ratio $R_m = 0.6$ at $z \approx 4 - 5$ triggers starbursts with SFR peaks around $30\,M_\odot\,\mathrm{yr}^{-1}$, and creates a large fraction of massive clusters. There are several distinct bursts, at least 5, in the time interval of about 400 Myr that it takes to complete the merger. They are likely associated with two orbital passages of the merging galaxies and gravitational instabilities in the new combined gaseous disk. The first merger at early times ($z \approx 7$) does not produce many massive clusters, as the galaxies are still too small to contain sufficient amount of cold gas. Some massive clusters also form in the quiescent period following the last merger, when the galactic disk still holds large gas reservoir, and so the merger activity and associated potential perturbations are not unique requirements to create massive clusters. Still, our results show that the rate of cluster formation is enhanced during the periods of gas-rich major mergers.

To investigate possible differences in the shape of CIMF during major mergers versus quiescent periods, we split the model clusters into merger-generated ($t_{\mathrm{creation}} = 1250 - 1400$ Myr) vs. non-merger-generated ($t_{\mathrm{creation}} = 900 - 1200$ Myr) groups. Figure 2.8 shows that the cluster mass distribution has higher cutoff mass during the major merger event. This can be understood using the empirical SFR $- M_{\mathrm{cut}}$ relation described by Eq. (2.7): $M_{\mathrm{cut}} \propto \mathrm{SFR}^{1.6}$. In the non-merger case, the typical SFR is about $5\,M_\odot\,\mathrm{yr}^{-1}$ and the corresponding cutoff mass is $1.8 \times 10^5\,M_\odot$, whereas in the merger case, $M_{\mathrm{cut}} = 1.7 \times 10^6\,M_\odot$ for $\mathrm{SFR} = 20\,M_\odot\,\mathrm{yr}^{-1}$. The ratio of the two cutoff masses is almost exactly $(20/5)^{1.6}$. Physically, this trend is consistent with the picture of the formation of massive star clusters in high-pressure environment that is produced by major mergers (e.g., Muratov & Gnedin, 2010; Kruijssen, 2014; Renaud

et al., 2015).

2.4.5 Dependence on density of star formation rate

Analytical models of cluster formation predict that the fraction of clustered star formation increases with the intensity of star formation; specifically, with the SFR per unit area, $\Sigma_{\rm SFR}$, not just the total SFR in a given galaxy. In order to investigate the relationship between SFR density and the cluster fraction, Γ, we need to define Γ in the simulation analogously to how it is measured in observations. This is hard to do, because we represent all stellar distribution as clusters of various mass, whereas in observations clusters are typically determined using group-finding algorithms applied to continuous stellar maps. As a proxy for Γ we choose the fraction of all star formation contained in clusters more massive than a given mass $M_{\rm cl}$:

$$\Gamma_{\rm sim} \equiv \frac{\rm SFR(M > M_{cl})}{\rm SFR}. \tag{2.10}$$

It is not the same quantity as the observed Γ, and therefore, comparison between the two should be treated with caution. Our goal is to study the trend of $\Gamma_{\rm sim}$ with $\Sigma_{\rm SFR}$, and only loosely related it to the observed range of Γ.

We split the main and satellite galaxies into concentric circular bins, as we did to study the spatial variation of CIMF in Section 2.4.2. We calculate $\Sigma_{\rm SFR}$ by dividing the SFR by the area of the annulus. We calculate the SFR by counting all cluster particles in a given annulus with ages between 15 to 50 Myr and use $M_{\rm cl} = 10^4\,M_\odot$. This is a compromise value for the different choices used in the studies of star clusters in M83 (roughly $M > 10^3\,M_\odot$) by Adamo et al. (2015) and seven other nearby galaxies (roughly $M > 10^{4.7}\,M_\odot$) by Goddard et al. (2010). Including the cluster disruption would not affect our calculation, as the mass loss over only 50 Myr is expected to be minimal for all clusters more massive than $10^3\,M_\odot$.

Figure 2.9 shows the relationship between the SFR surface density and $\Gamma_{\rm sim}$ for the fiducial run at $z \approx 3.3$. The trend that higher $\Sigma_{\rm SFR}$ leads to higher $\Gamma_{\rm sim}$ is reproduced in our model over a large range of $\Sigma_{\rm SFR}$ from 10^{-3} to $10^{3}\,M_{\odot}\,{\rm yr}^{-1}\,{\rm kpc}^{-2}$. To test the robustness of this result, we vary $M_{\rm cl}$ from $5\times10^{3}\,M_{\odot}$ to $10^{5}\,M_{\odot}$ and find that, although the absolute value of $\Gamma_{\rm sim}$ changes with $M_{\rm cl}$, the positive correlation remains. We also find that, for the same galaxy, both $\Sigma_{\rm SFR}$ and $\Gamma_{\rm sim}$ are higher in the inner annuli than in the outer ones. This is consistent with the spatial variation of CIMF, as we discussed in Section 2.4.2. The shallower slope and higher cutoff mass of CIMF in the inner annuli naturally lead to higher cluster fraction.

2.5 Cluster formation timescale

Observations of star-forming regions in the Galaxy suggest that cluster formation process is quick (Lada & Lada, 2003; Mac Low & Klessen, 2004). Hartmann et al. (2012) compile age information of young stars in the Orion Nebula Cluster and find the age spread as short as only a few Myr. Such a narrow spread presents an additional test of the implementation of cluster formation and feedback.

We investigate the formation timescale and mass accretion history of model clusters by analyzing the distribution of two characteristic timescales, $\tau_{\rm dur}$ and $\tau_{\rm ave}$, defined by Eq. (2.5) in Section 2.2.1. The first is the full duration of the cluster formation episode, from the first star to the last. The second timescale is weighted by star formation rate and more closely corresponds to the observed age spread.

Figure 2.10 shows the cumulative probability distribution of $\tau_{\rm dur}$ and $\tau_{\rm ave}$ of all cluster particles in the main galaxy for all models at the last common epoch, $z \approx 5.3$. As can be seen from the left panel, more than half of the cluster particles stop accretion completely within 10 Myr for models SFE10, SFE20, SFE100, and TURB50. For model SFE100, majority of the cluster particles become inactive within only 2 Myr. This is because feedback from active cluster particles heats up and removes the gas

from their natal GMCs and changes the cell properties so that one or more star formation criteria are violated. For two turbulence-based star formation efficiency models, TURBSF and TURBSF2, star clusters are not able to completely shut down accretion within the maximum allowed time $\tau_{\rm max} = 15$ Myr, because there is no density threshold that is assigned to turn off star formation. So, even if the feedback from a cluster particle has already blown out majority of the nearby molecular gas, there is still a trickle of mass that can be added to the cluster particle.

A more useful estimate of the formation timescale is the mass-weighted value, $\tau_{\rm ave}$, which contains information on the mass accretion history of an individual cluster. The right panel of Figure 2.10 shows that the timescales of TURBSF and TURBSF2 models are much shorter, and more than 50% of the cluster particles assemble their masses within 5 Myr. For the fiducial run, most of the cluster particles have $\tau_{\rm ave}$ smaller than 4 Myr, consistent with observations.

An important feature, illustrated by the plot, is a systematic dependence of the cluster formation timescale on the star formation efficiency and strength of stellar feedback. It is clear in both panels that, the higher $\epsilon_{\rm ff}$ the faster the clusters are formed. Therefore, comparison of the formation timescale of model clusters with the observed age spread of young star clusters would provide additional constraints on the local efficiency of star formation and feedback schemes.

Although it is impractical to store mass increments of all cluster particles at every local timestep, $\sim 10^3$ yr, we randomly select 10% of active clusters at the last epoch and output their mass growth history $M(t)$. The goal is to check whether the conclusions drawn from the distributions of $\tau_{\rm dur}$ and $\tau_{\rm ave}$ are consistent with this detailed output. We normalize the mass evolution of each cluster by its final mass $M_{\rm final}$, split $\tau_{\rm max}$ into 100 intervals of equal duration, and calculate the $25^{\rm th}$ and $75^{\rm th}$ percentiles of $M(t)/M_{\rm final}$ at each interval. These quartile ranges for models SFE10, SFE100, and TURBSF, are shown in Figure 2.11. We find that most of the clusters

reach their maximum mass within 15 Myr. Especially, in SFE100 runs, clusters finish accretion in only 3 Myr. This is consistent with the distribution of $\tau_{\rm dur}$ in the left panel of Figure 2.10, and suggests that feedback from young stars extinguishes star formation effectively. We also find decreasing cluster formation timescale with increasing star formation efficiency, which is again consistent with the conclusion drawn from the distribution of $\tau_{\rm ave}$.

For all three models, the mass histories correspond to either constant or decreasing growth rate. None of them shows a linearly increasing rate, $\dot{M} \propto t$, as suggested by Murray & Chang (2015) for the collapse of a self-gravitating cloud. This is likely because stellar feedback slows down gas accretion as the mass of the cluster increases.

The overall shape of the mass history distribution appears to be linear for the most part, with a slower-increasing tail over the last few Myr. We can quantitative assess how close $\dot{M}$ is to a constant. If the growth rate of each cluster in a given run was truly linear, $\dot{M} = {\rm const}$, then it would reach its final mass in a time $t_{\rm final} = M_{\rm final}/\dot{M}$. The median time for the whole sample, $t_{\rm final}^{\rm med}$, would then be equivalent to the median duration timescale $\tau_{\rm dur}$ (because of linearity of $M(t)$), which in turn would be twice the mass-weighted timescale: $\tau_{\rm dur} = 2\tau_{\rm ave}$. The median mass accumulated within $\tau_{\rm ave}$ would be half of the final mass. The values of $t_{\rm final}^{\rm med}$ or $\tau_{\rm ave}$ would differ from model to model, according to the trend with star formation efficiency and feedback. Figure 2.11 demonstrates that the actual median mass $M(\tau_{\rm ave})$ is not exactly half of $M_{\rm final}$, but is not too different from it, for the three models shown. Thus, clusters spend most of their formation period accreting at a steady rate.

From the large variation of the cluster formation history shown in Figure 2.11, we conclude that different clusters have dramatically different growth history during their active period. This reflects the intrinsic variation of the physical conditions in different star-forming regions, as well as the complex interplay between gas accretion and stellar feedback. This variation is also reflected in the large scatter of the resulting

star formation efficiency, as we will discuss in Section 2.6.5.

2.6 Discussion

2.6.1 Comparison with other implementations of star formation

Most cosmological simulations model star formation in a way that is similar to Katz (1992): a fraction of cold and dense gas is converted to star particles with a constant efficiency. Some simulations adopt a fixed star particle mass and create particles with a Poisson random process at each timestep, while others use a relatively long sampling timescale (typically several Myr) and create star particles with varying masses. One of the few alternatives, Cen & Ostriker (1992) grow star particles over the galaxy dynamical time, according to a priori chosen star formation law. Moreover, in all of the above methods, star particle masses are determined before their feedback processes begin to influence the ambient environment. In contrast, in our model, cluster particles grow their mass continuously at each local timestep until their own feedback terminates the star formation episode. The mass accumulation history is resolved by a large number of timesteps, and the final particle mass is obtained self-consistently so that it can be considered a good proxy to the mass of a given cluster formation region.

In contrast to full galactic simulations, small-scale simulations of star and planet formation create collisionless sink particles that accrete matter over time. This approach has been used both in SPH (e.g. Bate et al., 1995) and grid codes (Krumholz et al., 2004; Federrath et al., 2010). Sink particles are created in local density peaks if the gas is Jeans-unstable and gravitationally bound. The particles continue to accrete gas as long as these criteria are satisfied, which leads to a significant fraction of all gas in the simulation box to be converted into stars. Such simulations model small gas clouds, typically under $10^3\,M_\odot$, which do not produce many massive stars.

Therefore, even when feedback is included, it does not drive strong outflows that would terminate a star formation episode. In this regard, our simulations differ as they model continuous inflow of gas onto large clouds, which generate many massive stars with stronger feedback.

Also, the sink particle approach usually requires that the accretion region be resolved by many cells, in order to determine the flow convergence and mass increment at each timestep. In a cosmological simulation, however, resolving cluster-forming regions requires a sub-parsec cell size, which is still computationally challenging. Instead, stellar particles in our model grow over a period of time, parametrized by $\epsilon_{\rm ff}$ and based on observations of individual GMCs on scales comparable to our spatial resolution (Krumholz et al., 2012).

2.6.2 On the shape of CIMF

In Section 2.4, we examined the model CIMF in different galaxies at various epochs and found that it can be well described by the Schechter function. The cutoff at high mass is related to the overall amount of cold dense gas available for star formation and therefore depends on the environment and evolutionary stage of the host galaxy. The power-law slope at low mass is much more stable, with only a limited range of variation between $\alpha \approx 1.8 - 2.0$. It is encouraging that this range is consistent with observations of young star clusters. We can now investigate the physical origin of this CIMF slope within our model, and the connection between the cluster mass and physical properties of the ISM.

We calculate the molecular gas mass within each cluster-forming sphere by summing over the contributions of the central cell as well as the 26 neighbors: $M_{\rm sp} = \Sigma f_{{\rm sp},i} V_i f_{{\rm H}_2,i} \rho_{{\rm gas},i}$. We also estimate the mass function of individual gas cells to compare it with the distribution of $M_{\rm sp}$. Since all clusters in our model accrete gas from spheres of fixed volume $V_{\rm sp} = \pi D_{\rm GMC}^3/6$, to make a fair comparison, we rescale the

total (or molecular) cell mass to match the same volume: $\rho_{\rm gas} V_{\rm sp}$ (or $f_{\rm H_2} \rho_{\rm gas} V_{\rm sp}$).

Figure 2.12 shows that the distribution of the rescaled total and molecular cell mass exhibits a steep slope at low masses, $\alpha > 3$. In contrast, the mass function of $M_{\rm sp}$ has a slope closer to that of the CIMF. This implies that our method in which the sphere encloses gas cells with various density is a more realistic way of modeling star cluster formation, than using simply the properties of the central cell.

However, the power-law slope of the mass function of $M_{\rm sp}$ is still steeper than $\alpha = 2$ and the maximum sphere mass is only $\sim 2 \times 10^6\, M_\odot$, several times smaller than the most massive model clusters. This discrepancy may stem from the fact that the instantaneous mass of the cluster-forming sphere does not contain the information of the evolving accretion rate due to feedback.

In order to connect the mass function of the spheres and CIMF, we present a simple cluster formation model by assuming a power-law mass accumulation history: $\dot{M}(t) = \dot{M}_0\, (t/t_0)^m$, where t_0 is the time over which a cluster maintains the initial growth rate $\dot{M}_0 = \epsilon_{\rm ff} M_{\rm sp}/\tau_{\rm ff}$. The expected cluster mass in this model is:

$$M = \int_0^{\tau_{\rm max}} \dot{M}(t)\, dt = \frac{8 \epsilon_{\rm ff}\, G^{1/2}}{\pi\, D_{\rm GMC}^{3/2}} \frac{\tau_{\rm max}^{m+1}}{(m+1)\, t_0^m}\, M_{\rm sp}^{3/2}. \tag{2.11}$$

The value of the index m does not affect the scaling with the initial sphere mass or density. Instead, the scaling $M \propto M_{\rm sp}^{3/2}$ reflects the dependence of the star formation rate on the free-fall time.

In Figure 2.11, we find that the median cluster formation history in the SFE10 run is generally described by a decreasing growth rate, roughly as $\dot{M} \propto t^{-0.5}$. We estimate the distribution of the expected cluster mass from the mass function of $M_{\rm sp}$ using Eq. (2.11), with $m = -0.5$, $\epsilon_{\rm ff} = 0.1$, $D_{\rm GMC} = 10$ pc, $t_0 = 1$ Myr, and $\tau_{\rm max} = 15$ Myr. Other choices of the index m would only slightly affect the normalization of the mass function, but not its slope. We find that the slope of the expected cluster

mass function ($\alpha \approx 1.9$) is very similar to that of the CIMF. Also the maximum expected mass can reach as high as $10^7\,M_\odot$, similar to the maximum cluster mass in the simulations. This consistency indicates that the origin of CIMF is a combination of the mass function of GMCs and the accretion history within the cluster-forming regions.

In the expected cluster mass function, the power-law shape extends to very high mass and there exists no clear cutoff as seen in the CIMF. The high mass cutoff in the simulation may be caused by strong feedback from the most massive cluster particles. This feedback may result in more variation of the growth history than assumed in our simple estimate.

2.6.3 On the similarities and differences with the constant-density-threshold model

Kravtsov & Gnedin (2005) proposed a different scheme for forming massive star clusters, using post-processing of cosmological simulations. Those simulations were run with an earlier version of the ART code, with a traditional star formation recipe and a relative low threshold density for creating stellar particles. Even though the spatial resolution was 9 times lower than in our present simulations, it was sufficient to resolve the structure of largest gas clouds within a galactic disk at the same epoch, $z \approx 3.3$. The pressure within the clouds was dominated by turbulent motions, with an approximately flat velocity dispersion profile. This means that although the clouds were not strictly isothermal because the thermal pressure was negligible compared to the turbulence, they still had roughly $\rho(r) \propto r^{-2}$ density profiles. Adopting these profiles for the unresolved inner structure of the clouds, Kravtsov & Gnedin (2005) suggested that the central regions above a particular very high density could be forming massive star clusters. In order to produce a gravitationally bound cluster, the fraction of the dense gas turning into stars had to be above 50% (and was taken

to be 60%); the rest of the gas is likely to be quickly blown out of the region by the winds and radiation of the young stars. The remaining stars would expand to a new hydrostatic equilibrium, and the final half-mass radius of the cluster R_h could be estimated as the size of the region above the threshold density, corrected for the expansion. The value for the high density threshold ($\rho_{\rm CSF} = 10^4\,M_\odot\,{\rm pc}^{-3}$) was taken such that after the expansion, the final half-mass density would match the median observed density of Galactic globular clusters ($\approx 3\times10^3\,M_\odot\,{\rm pc}^{-3}$). In that sense, the model had no free adjustable parameters.

Analysis of several outputs of that simulation showed a consistent shape of CIMF, similar to a power law with $\alpha \approx 2$, and also a sharper cutoff at high mass. An equally good fit was provided by a log-normal function.

What is similar and what is different in that model and our new model? In terms of technical execution, the old model was simpler, implemented on a few discrete outputs rather than in run-time, and based on a less sophisticated simulation. These differences aside, below we focus on the analysis of the results of the two models, as they relate to the origin of CIMF.

In the old model, the constant density threshold implies that the average density, M/R_h^3, is the same for all model clusters. The half-mass radius is set by the condition $\rho(R_h) = \rho_{\rm CSF} = {\rm const}$, or $\rho_{\rm cell}(L_{\rm cell}/R_h)^2 = {\rm const}$, which means it scales as $R_h \propto \rho_{\rm cell}^{1/2}$, and the cluster mass scales as $M \propto \rho_{\rm cell}^{3/2}$. In the new model, the cluster mass scales with cell density in exactly the same way, $M \propto \rho^{3/2}$, but for a different reason – because of the dependence of the SFR on the free-fall time. Of course, the time evolution of the cluster accretion rate in the continuous model can modify this scaling, but the overall dependence on cell density should still be similar in the two models. Thus, if the distribution of cloud density was the same in both simulations, the two models would result in the same CIMF.

At the highest densities in the old simulation, the probability distribution function

could be fit by a single power law, $dN/d\log\rho \propto \rho^{-n}$, with $n \approx 1.4$. Given the scaling of cluster mass on density described above, this results in the CIMF obeying a power law $dN/d\log M \propto M^{-1-2n/3}$. The value of the slope $\alpha = 1 + 2n/3 \approx 2$.

In the new simulations the slope of the density PDF is steeper and changes, as shown by Figure 2.12, from $n \approx 3.3$ for $\rho \sim 10^3\, m_p$ cm^{-3} to $n \approx 2$ for $\rho > 10^4\, m_p$ cm^{-3}. The ART code now incorporates substantially revised physics of gas heating and cooling and includes transfer of ionizing radiation through the ISM. Thus the differences in the density PDF are not surprising. Applying the old model to the new simulations would result in a steeper CIMF ($\alpha \approx 2.3 - 3.2$). This again emphasizes the need for a continuous model of cluster accretion that we developed in this paper.

The advantage of the Kravtsov & Gnedin (2005) model was the ability to estimate cluster size, in addition to mass. In our new simulations we can only set an upper limit on the cluster size being smaller than the radius of the sphere, $D_{\rm GMC}/2$. With this limit, we can estimate the lower limit on the average cluster density in our continuous model,

$$\rho_{\rm av} = \frac{M}{4\pi/3\,(D_{\rm GMC}/2)^3} \propto \rho^{3/2},$$

and compare it with the central cell density. Calculating all the coefficients results in

$$\frac{\rho_{\rm av}}{\rho} \approx 0.48 \left(\frac{\epsilon_{\rm ff}}{0.1}\right) \left(\frac{\tau_{\rm max}\, t_0}{15\,{\rm Myr}^2}\right) \left(\frac{\rho}{10^3\, m_p\,{\rm cm}^{-3}}\right)^{1/2}.$$

At low densities, this estimated $\rho_{\rm av}$ can be lower than the cell density, because the local efficiency of star formation $\epsilon_{\rm ff}$ is low and only fraction of the gas is being converted into stars. Conversely, at densities above $\rho \sim 4 \times 10^3\, m_p\,{\rm cm}^{-3}$, the cluster accretes enough material to compensate for low $\epsilon_{\rm ff}$ and becomes dense than the initial cloud. Note that the actual half-mass density may be a factor of few higher than our estimate, because of additional dissipation on sub-grid scales during cluster formation.

2.6.4 On the high mass end of CIMF

Although various observations of young massive clusters in nearby galaxies suggest a roughly power-law mass function, it is still debated whether there exists a high mass cutoff in the CIMF and whether the cutoff mass is environment-dependent.Larsen (2002) found a correlation between SFR and the V-band magnitude of the brightest clusters and suggested that cluster formation is a stochastic process in which the maximum cluster luminosity can be explained by statistical sampling from a universal CIMF, the so-called "size-of-the-sample" effect. Other observations in various environments suggest that the variation of the high mass end of the CIMF may reflect different physical condition of their host galaxies (e.g. Bastian, 2008; Goddard et al., 2010; Adamo et al., 2015). More recently, Sun et al. (2016) analyzed the cluster mass-galactocentric distance relation in four galaxies and concluded that, at least statistically, it is hard to rule out the environment-independent cluster formation scenario.

In our simulations, we have a large number of cluster samples formed at different epochs in different galaxies. Therefore, we can study the influence of environment on the model CIMF. First, we find a strong correlation between the fraction of massive clusters $\Gamma_{\rm sim}$ and the SFR surface density, $\Sigma_{\rm SFR}$, shown in Figure 2.9. If the shape of the CIMF does not change with environment and the most massive clusters are formed merely due to the "size-of-the-sample" effect, average $\Gamma_{\rm sim}$ will be constant for different SFR and have larger scatter for lower SFR cases. However, the apparent positive correlation between $\Gamma_{\rm sim}$ and $\Sigma_{\rm SFR}$ suggests that massive clusters are preferentially formed in starbursts. This in turn implies an environment-dependent cluster formation process and disfavors the stochastic scenario. Second, there exists a positive correlation between the cutoff mass $M_{\rm cut}$ and SFR of the host galaxy, which directly reveals environmental dependency of CIMF. The maximum cluster mass would be largely overestimated by assuming a pure power-law CIMF, but it is in agreement

with the predicted value when assuming a Schechter function with $M_{\rm cut}$ that is inferred from the empirical $M_{\rm cut}$-SFR relation. All these results validate a Schechter function fit with the environment-dependent cutoff mass as best description of CIMF.

2.6.5 On the variation of star formation efficiency

Recent observations of star-forming complexes suggest that the star formation efficiency on GMC scales varies significantly over 2-3 orders of magnitude (e.g. Lada et al., 2010; Murray, 2011; Evans et al., 2014; Vutisalchavakul et al., 2016). Such a large scatter may come from either the variation in cloud properties (e.g. Krumholz & McKee, 2005; Padoan & Nordlund, 2011; Federrath & Klessen, 2012; Hennebelle & Chabrier, 2013), or the time-evolution of self-gravitating clouds (Feldmann & Gnedin, 2011; Lee et al., 2015), or both. Realistic modeling of star formation in galaxy formation simulations should be able to reproduce this range of variation. Here we examine how the differences in the formation history of each model cluster affect the value of the efficiency that would be inferred for it at the time it begins star formation.

We define the integral efficiency $\epsilon_{\rm int}$ such that the final cluster mass, M, can be obtained by a constant growth rate $\epsilon_{\rm int} M_{\rm sp,0}/\tau_{\rm ff,0}$ over the mass-weighted cluster formation timescale $\tau_{\rm ave}$:

$$\epsilon_{\rm int} \equiv \frac{M/\tau_{\rm ave}}{M_{\rm sp,0}/\tau_{\rm ff,0}}. \tag{2.12}$$

Here $M_{\rm sp,0}$ is the initial mass of the cluster-forming sphere, and $\tau_{\rm ff,0}$ is the free-fall time when the cluster particle is created.

Figure 2.13 shows the distribution of $\epsilon_{\rm int}$, normalized by the intrinsic star formation efficiency $\epsilon_{\rm ff}$, for both fixed $\epsilon_{\rm ff}$ and turbulence-based $\epsilon_{\rm ff}$ runs. A mass-weighted average $\epsilon_{\rm ff} \approx 1.4\%$ is used for TURBSF run. The median values of $\epsilon_{\rm int}/\epsilon_{\rm ff}$ for all models are smaller than unity. This is possibly due to the outflows from the cluster-forming region caused by stellar feedback, which reduce the amount of gas available for star formation. More importantly, we find a large variation of $\epsilon_{\rm int}$ around the median.

For the fixed $\epsilon_{\rm ff}$ runs (SFE10, SFE20, SFE100), $\epsilon_{\rm int}$ varies over more than 2 orders of magnitude. This variation is a natural consequence of the evolution of properties of cluster-forming spheres caused by the continuous gas accretion, cluster formation, and stellar feedback. The variation is even larger in TURBSF run, where $\epsilon_{\rm ff}$ is re-evaluated at each timestep during the cluster formation period. Although we cannot compare distributions of efficiencies in our simulations and observations directly, the large variation shown by the models in Figure 2.13 is in good qualitative agreement with observational estimates. This demonstrates that modeling continuous formation of star cluster particles in galaxy formation simulations results in a more realistic distribution of cluster masses.

2.6.6 On the origin of globular clusters

Very massive star clusters, with $M \gtrsim 2 \times 10^5\,M_\odot$, have lifetime comparable to the age of the Universe (e.g., Muratov & Gnedin, 2010). Therefore, they can be considered as progenitors of present-day globular clusters (GCs). Studying the formation of these clusters at high redshift gives us an opportunity to investigate the origin of GC populations. For example, Ashman & Zepf (1992) proposed a model of GC formation in the gas-rich major mergers. Muratov & Gnedin (2010) and Li & Gnedin (2014) incorporated this scenario in the semi-analytical model to study GC properties in the Milky Way and the Virgo Cluster galaxies, and successfully reproduced the multi-modal metallicity distributions of their GC systems.

Following the idea that cluster formation is a strong environment-dependent process, there are two requirements for the formation of massive clusters. First, the CIMF needs to have a high cutoff mass so that the probability of sampling massive clusters is not too low. For instance, in order to obtain $M_{\rm cut} = 10^6\,M_\odot$, the corresponding SFR needs to reach as high as $14\,M_\odot\,{\rm yr}^{-1}$, based on Eq. (2.7). The SFR of this level is hard to achieve during the quiescent evolutionary stage of a Milky Way-sized

galaxy, but it can be reached when the galaxy experiences major mergers. Second, to form such massive clusters, the host galaxy needs to have a large reservoir of cold gas. Indeed, as we can see from Section 2.4.4, massive clusters in our simulations are preferentially formed during the gas-rich major merger events.

Another possible dynamical effect of major mergers on GC formation is that the merger kicks massive clusters out of their birthplace in the galactic disk with high gas and stellar density, and places them into the halo where the tidal field is weak. This process helps to protect the newly formed clusters from being quickly destroyed by the intense tidal field of the disk. Such a two-stage formation and evolution scenario (Kravtsov & Gnedin, 2005; Kruijssen, 2015) can be tested in our model by analyzing the evolution of cluster bound fraction, as discussed in Section 2.2.2. In a follow-up paper we will investigate the transition from young massive clusters at high redshift to the GCs that we observe in the local Universe.

2.7 Summary

We introduced a new star formation implementation, in which star cluster is considered a unit of star formation and cluster particles accumulate mass from a fixed sphere similar in size to GMC clumps. The mass growth of a given particle is terminated by its own feedback, so that its final mass is obtained self-consistently and represents the actual mass of a newly formed star cluster within the GMC. We implemented this cluster formation model in the ART code, and performed several high-resolution cosmological simulations of a Milky Way-sized galaxy under different model parameters. We analyzed the properties of young massive clusters in the simulated galaxies at high redshift ($z > 3$) and the main results are summarized below.

- The CIMF is best described by a Schechter function, with a power-law slope

$\alpha \approx 1.8$, in agreement with observations of young massive clusters in nearby galaxies. The shape of the CIMF is not sensitive to the model parameters, except for model SFE100. The CIMF of SFE100, with 100% star formation efficiency per free-fall time, has a cutoff mass that is much smaller than the other models and is inconsistent with observations.

- We find a positive correlation between the SFR and the cutoff mass: $M_{\rm cut} \propto {\rm SFR}^{1.6}$. The maximum cluster mass also increases with the SFR, $M_{\rm max} \propto {\rm SFR}^{1.4}$, consistent with the value expected for a Schechter-like CIMF. We also find a tight correlation between the SFR surface density and the fraction of star formation contained in massive star clusters. The maximum mass decreases with distance from the galaxy center, as the SFR density decreases. All these trends suggest that cluster formation is a local process that strongly depends on the galactic environment. The scenario that clusters are formed solely by stochastic sampling from a universal CIMF is ruled out.

- Maximum cluster mass in a given galaxy reaches $10^7\,M_\odot$ when ${\rm SFR} > 10\,M_\odot\,{\rm yr}^{-1}$, but falls below $10^6\,M_\odot$ when ${\rm SFR} < 3\ M_\odot\,{\rm yr}^{-1}$.

- Feedback from young stars extinguishes star formation in a GMC within 4 Myr, consistent with the observed age spread of young star clusters. The cluster formation timescale decreases systematically with increasing local efficiency of star formation, $\epsilon_{\rm ff}$, to a minimum of 0.5 Myr. This systematic trend indicates that cluster formation is strongly influenced by its own feedback. The power-law slope of the CIMF arises from a combination of the mass function of GMC clumps and the feedback-regulated mass accretion history.

- We show that our cluster formation model leads to a large variation of the integral star formation efficiency (Eq. 2.12) even when $\epsilon_{\rm ff}$ is kept constant. The range of this variation spans about two orders of magnitude, similar to

recent observations of star formation efficiency within GMCs. It is a natural consequence of the evolution of gas density in star-forming regions, caused by continuous gas infall, cluster formation, and stellar feedback.

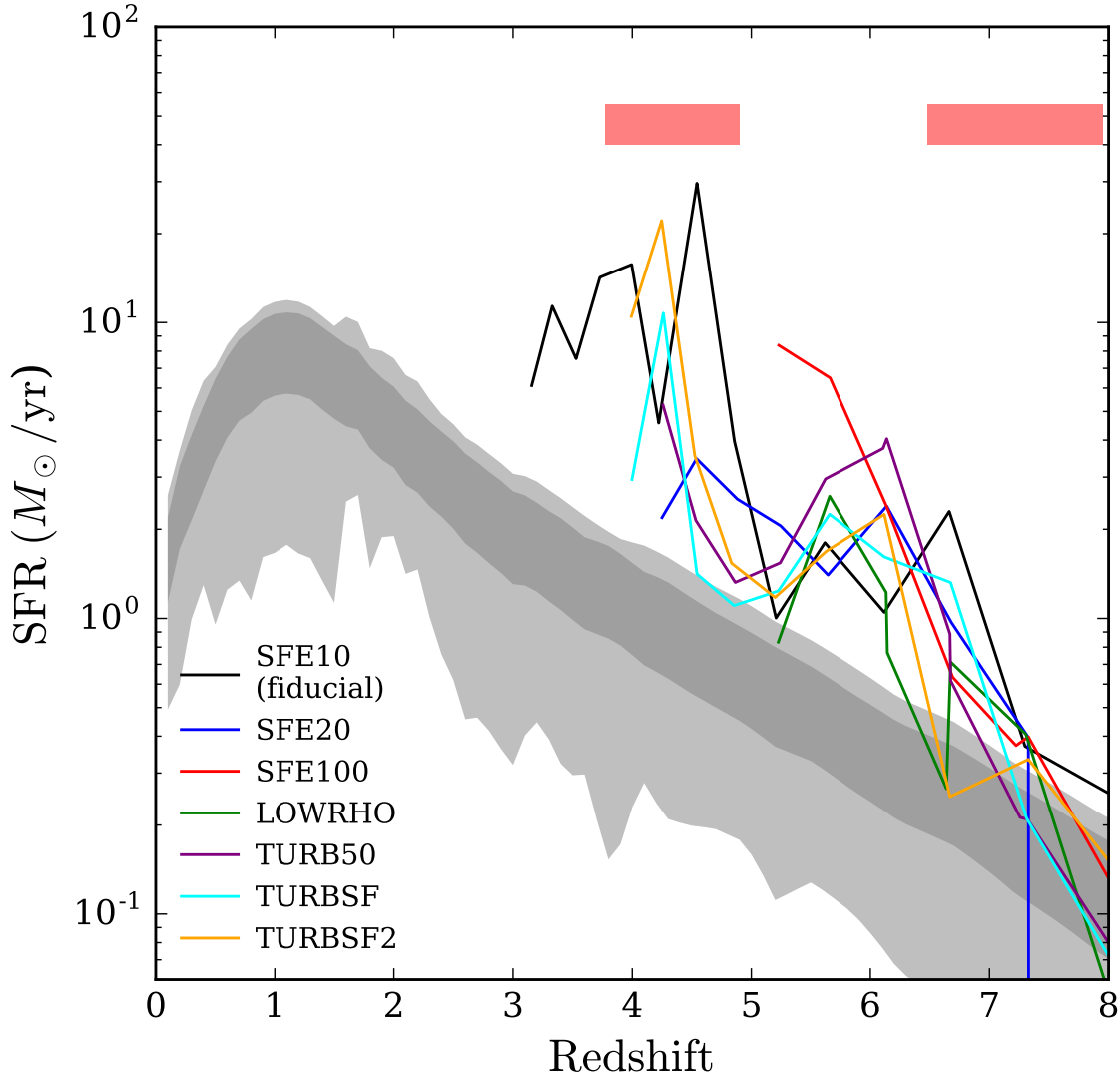

Figure 2.2: Simulated star formation history of the main galaxy, smoothed over 100 Myr bins, for different models (see legend). The star formation history for a $M_{\rm vir}(z = 0) = 10^{12} M_\odot$ halo from the abundance matching technique (Behroozi et al., 2013b) is overplotted by shaded regions. The dark and light regions show the one- and two-sigma confidence intervals, respectively. Red horizontal bars show the epochs of two major-merger events (with mass ratio larger than 0.3) that are identified from the merger trees of the main halo. The span of the bar represents the duration of each merger, see Sec. 2.4.4 for detailed description.

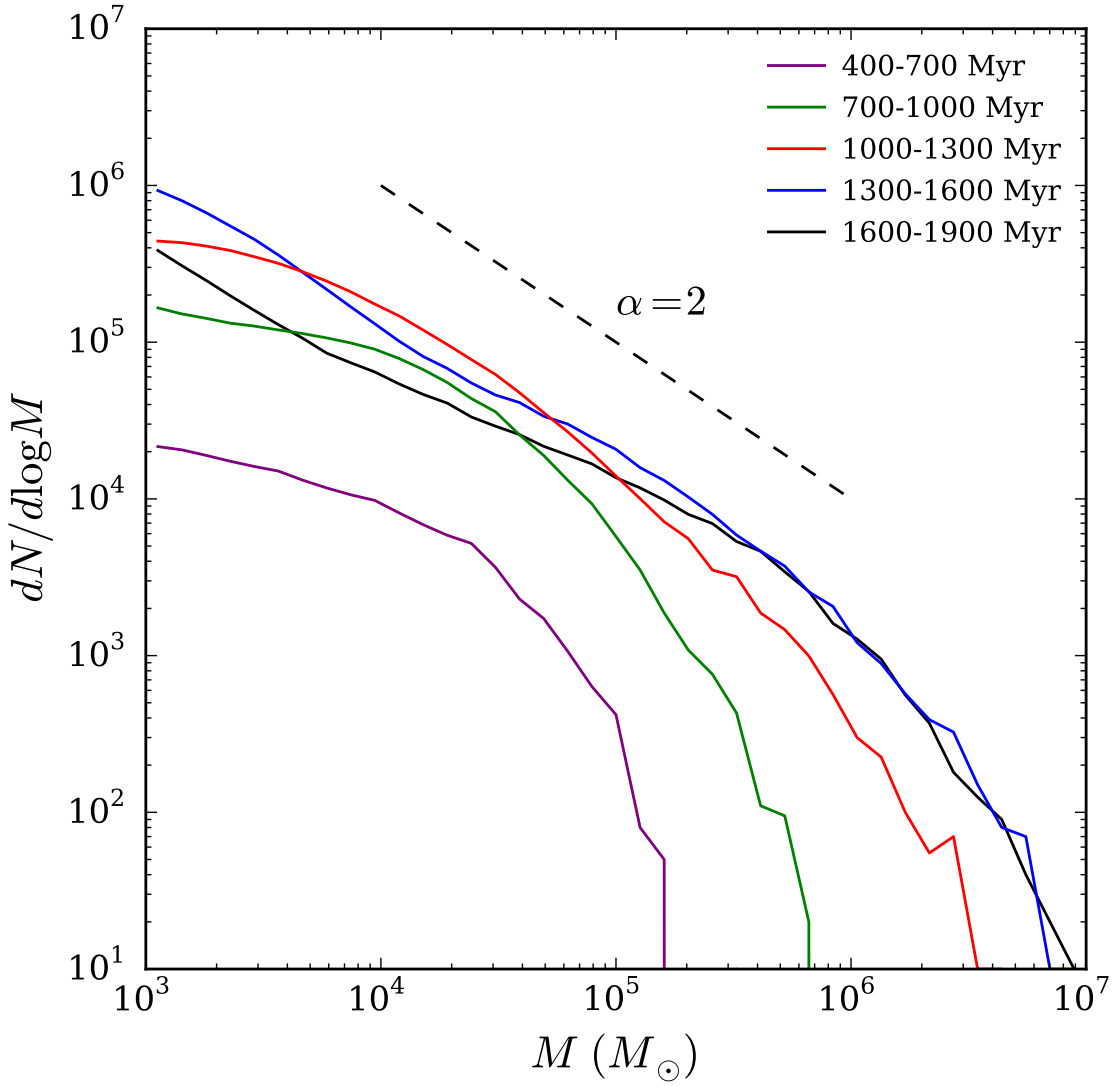

Figure 2.3: Cluster initial mass function for different cluster age bins in the main galaxy for the fiducial run. Cluster disruption is not included. A power-law distribution with slope of 2 is plotted by a dashed line. The mass functions exhibit a stable shape across a large range of cosmic time, and both the power-law slope and the high mass cutoff are consistent with observed cluster samples in nearby galaxies.

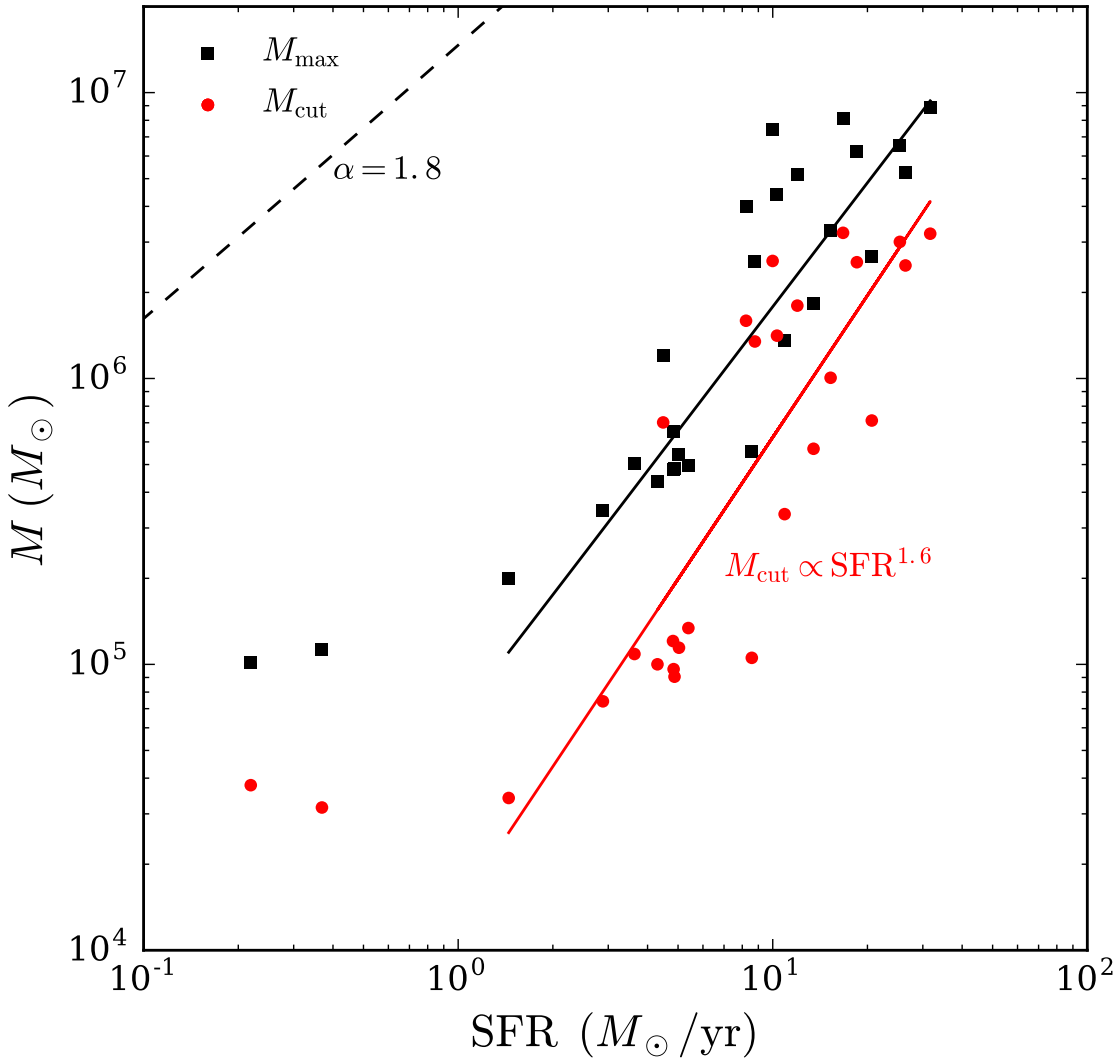

Figure 2.4: SFR vs. maximum cluster mass ($M_{\rm max}$) and cutoff mass ($M_{\rm cut}$) for the fiducial run. The SFR is averaged over 50 Myr and $M_{\rm max}$ is chosen from the clusters found in the same time interval. The initial mass function of clusters in each interval is fitted by Eq. (2.6), and both the power-law slope and cutoff mass are obtained. The red solid line shows the best-fit relation between $M_{\rm cut}$ and SFR for samples with SFR $> 1 M_\odot\,{\rm yr}^{-1}$. The theoretical maximum masses for given SFRs for both pure power-law and Schechter mass function are shown by dashed and solid black lines, respectively (See Section 2.4.1 for detailed calculations). Note that the black solid line is not the best-fit between SFR and $M_{\rm max}$, but the expected $M_{\rm max}$ by assuming the Schechter CIMF with $\alpha = 1.8$ and $M_{\rm cut}$ that is estimated by the empirical $M_{\rm cut}$-SFR relation (red line).

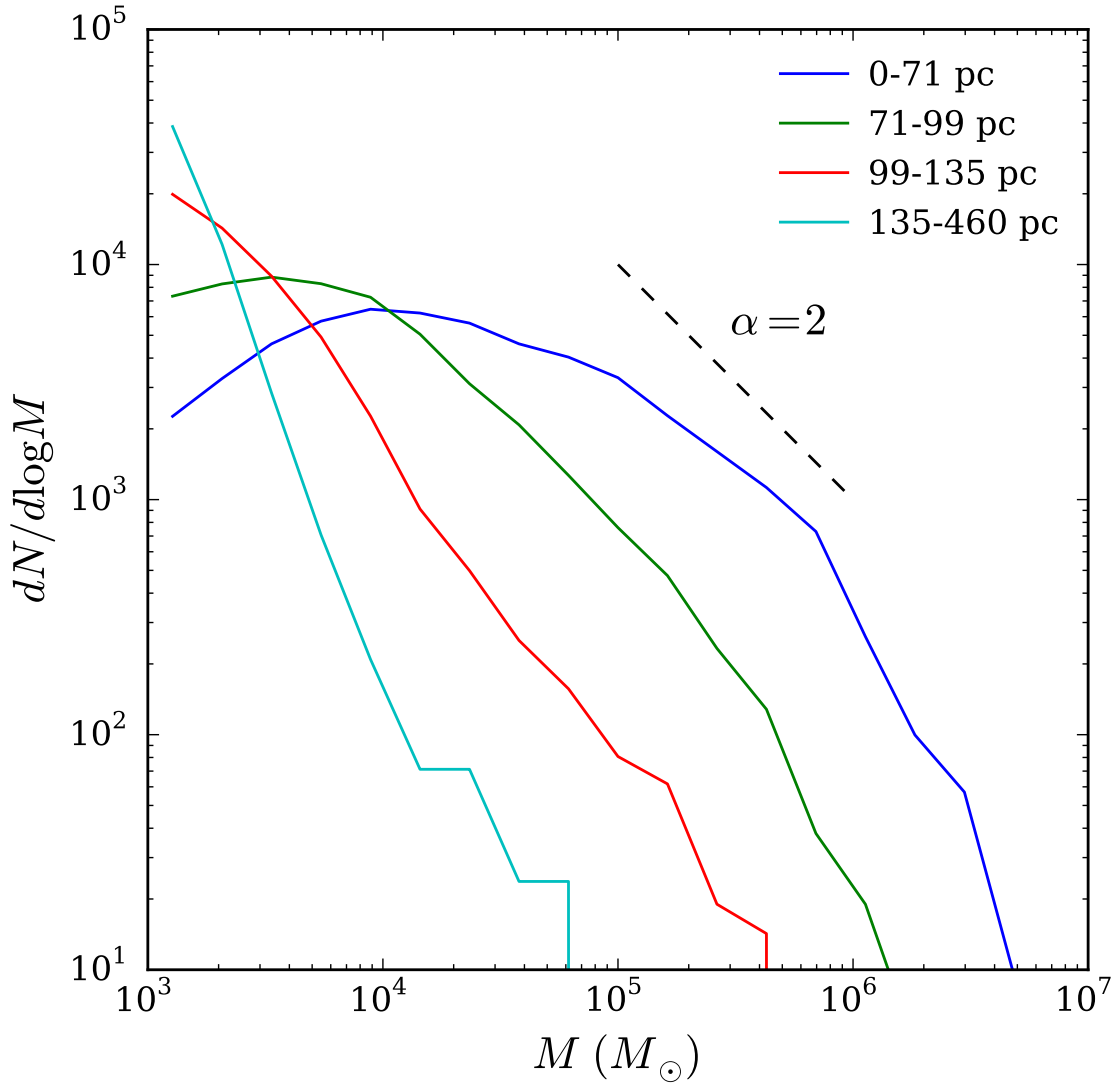

Figure 2.5: Initial mass function of young clusters (¡100 Myr) in four radial bins of equal cluster number, for the main galaxy in the fiducial run. The mass functions show clear steepening from the inner annulus (blue line) to the outer ones as well as the decreasing maximum cluster mass with radius.

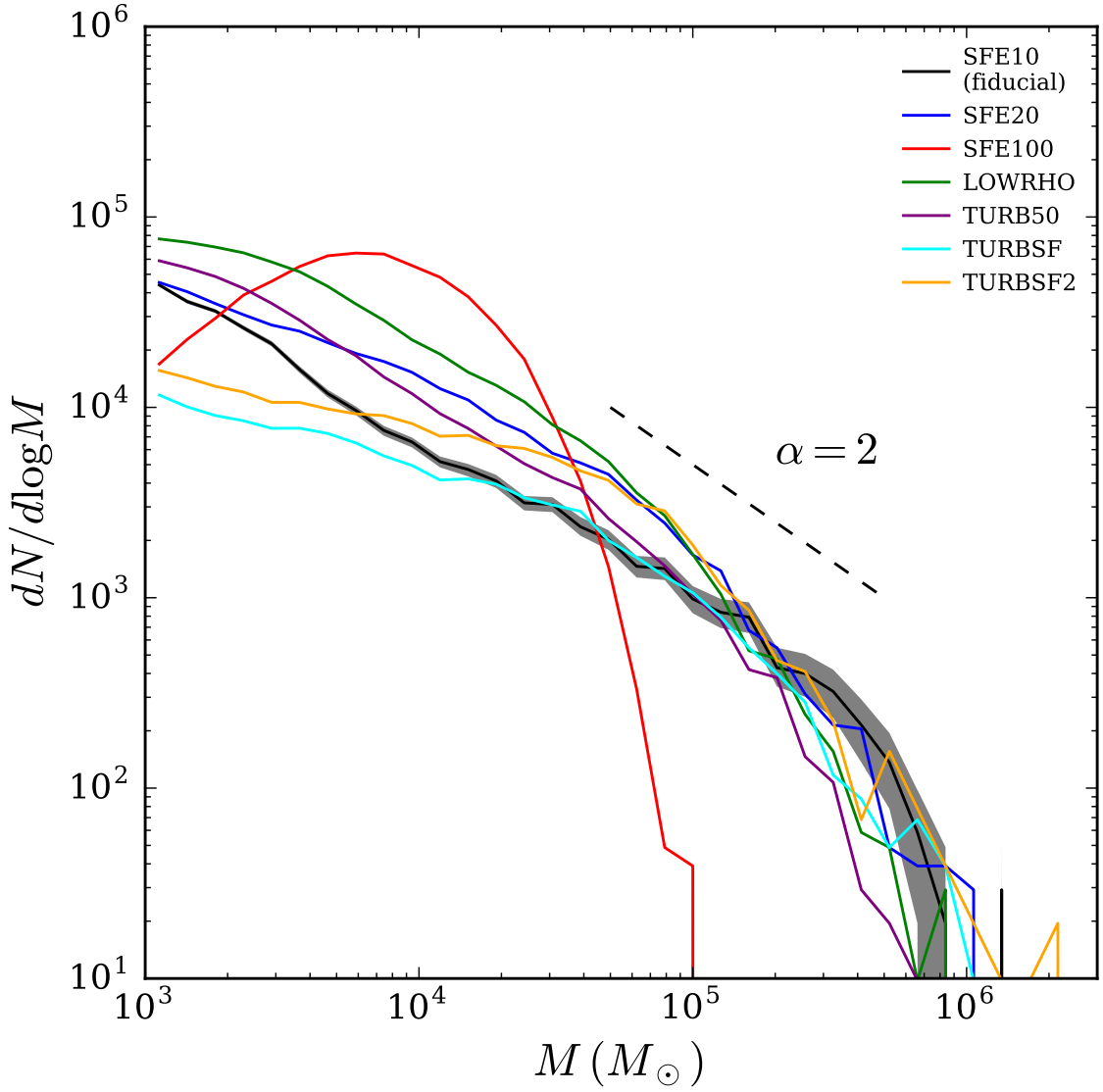

Figure 2.6: Cluster initial mass function of all seven models (see legend for color codes) in the main galaxy at the same epoch ($z \approx 5.3$). The grey shaded region around the black line is the 95% confidence interval of the CIMF constructed by bootstrap resampling for the SFE10 model (fiducial).

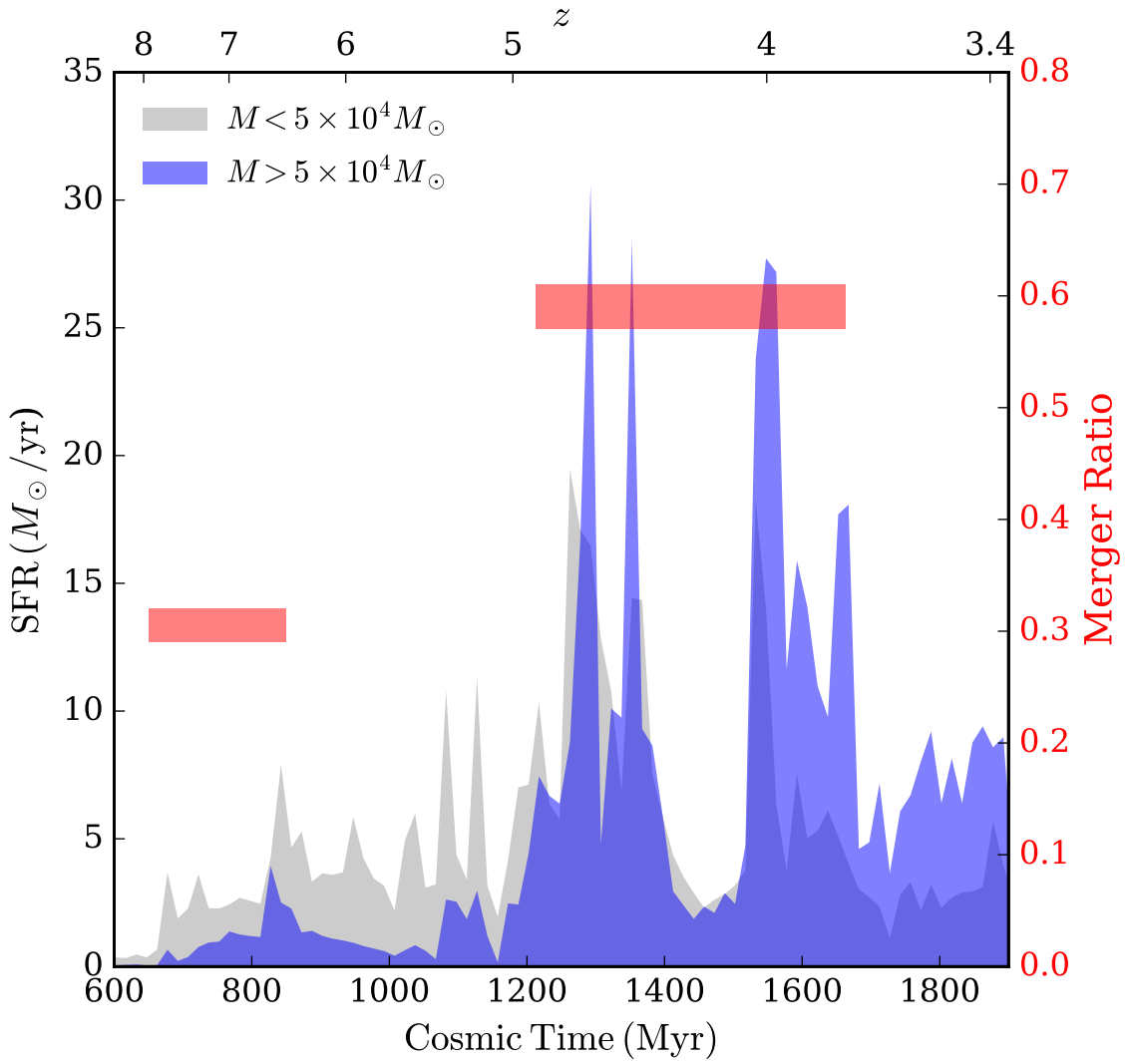

Figure 2.7: Star formation history of the main galaxy split into massive (blue shaded; $M > 5 \times 10^4 M_\odot$) and less massive clusters (gray shaded; $M < 5 \times 10^4 M_\odot$) for the fiducial run. Instead of calculating SFR of the main galaxy at each snapshot as in Section 2.3, here we calculate the formation history of all cluster particles located within the main galaxy at the last snapshot. The SFR is smoothed over 15 Myr. Two major merger events with mass ratio larger than 0.3 are labeled by red bars. The duration of merger is represented by the horizontal length of the bar, while the merger mass ratio that is indicated by the vertical position of the bar according to the scale on the right y-axis.

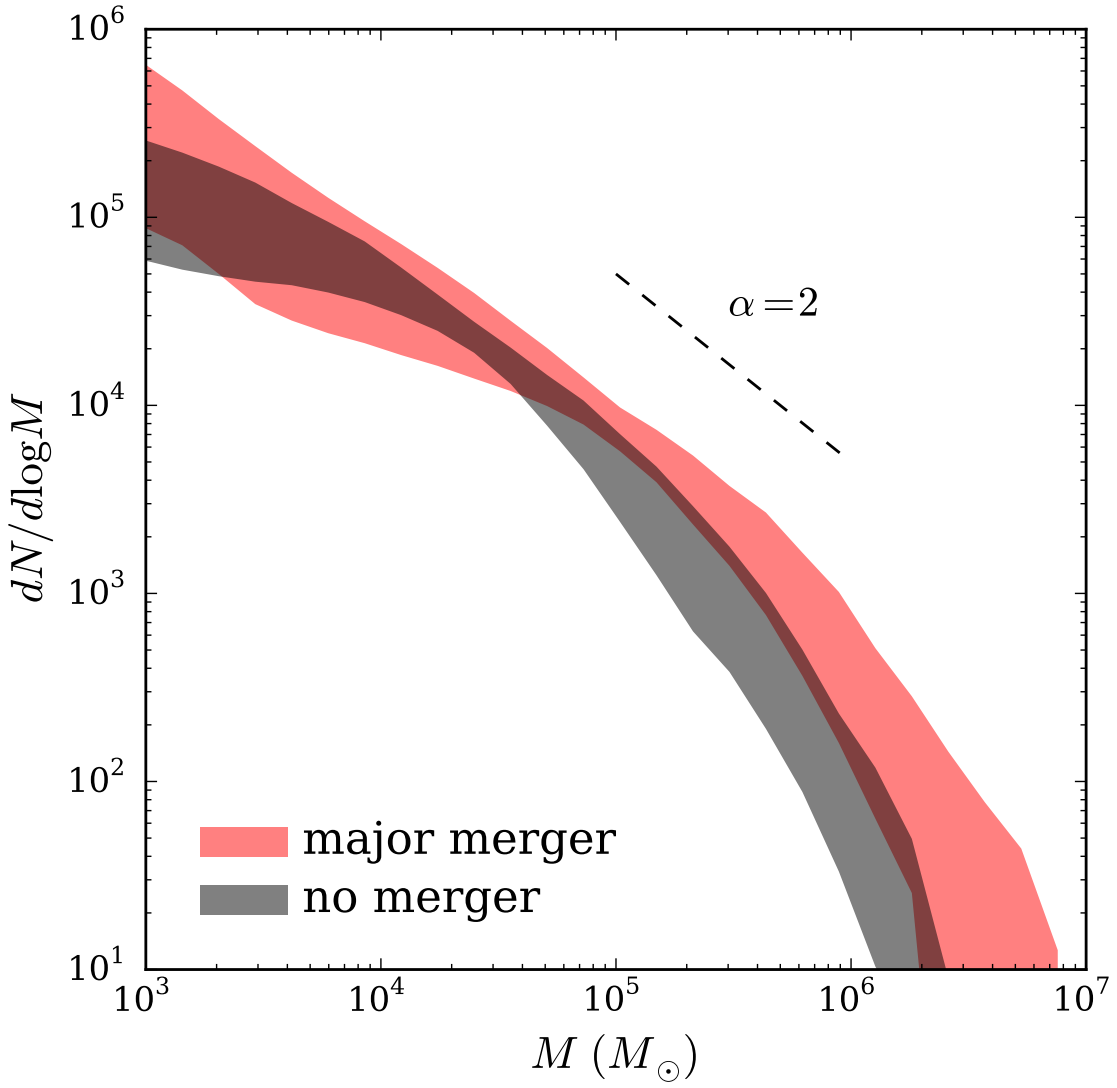

Figure 2.8: Cluster initial mass function split by merger activity: the mass function during major merger (black) and between mergers (red). Each band shows the standard deviation of the mass functions around the mean value of models SFE10 SFE20, TURB50, TURBSF, and TURBSF2.

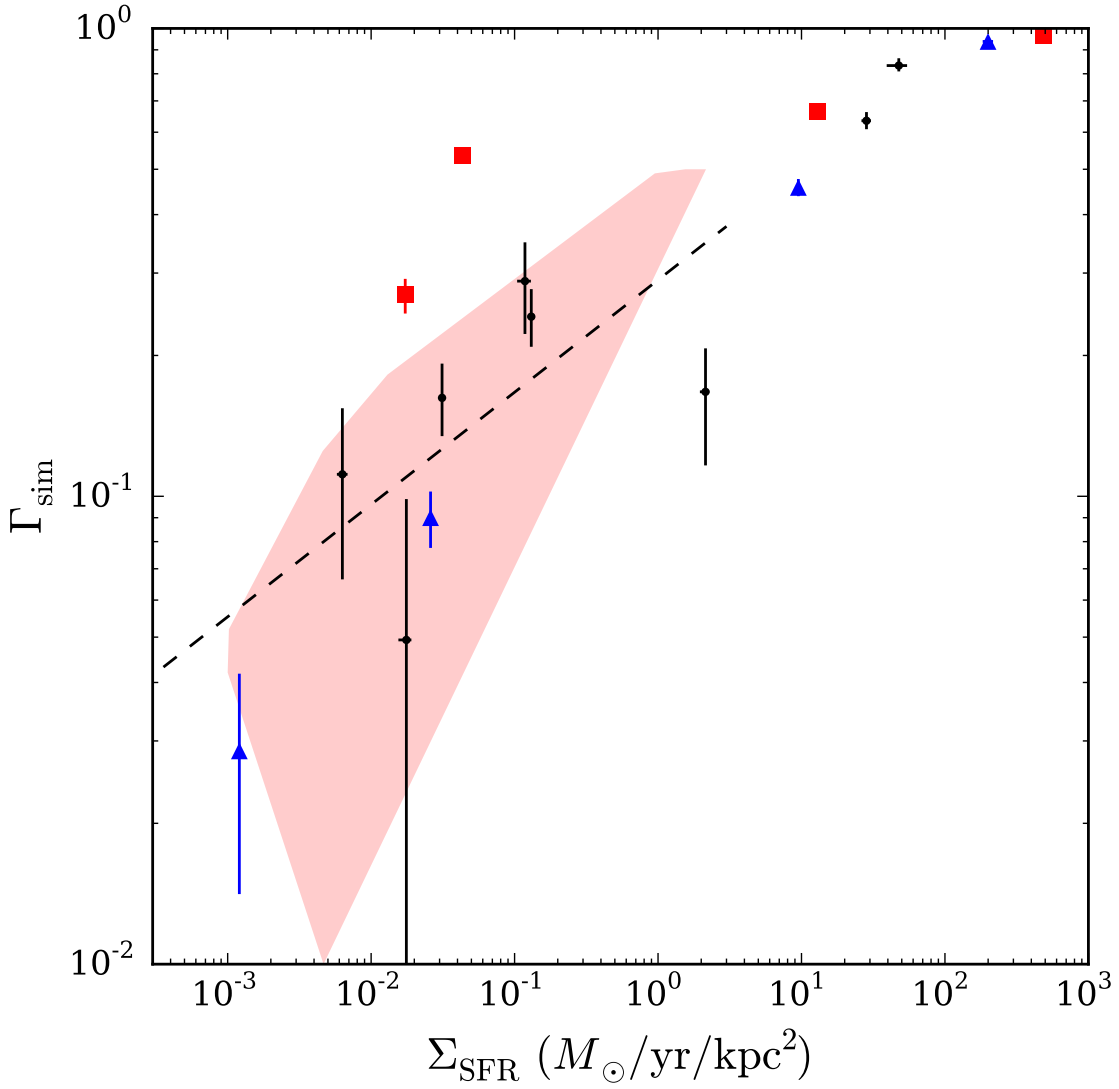

Figure 2.9: SFR density vs. fraction of young massive star clusters ($M > 10^4\,M_\odot$) in both main and satellite galaxies for the fiducial run at $z \approx 3.3$. The red and blue points represent the clusters in the main and the second largest galaxies, respectively. Black points show the other satellite galaxies. Dashed line shows the empirical relation for the observed star cluster populations in 7 nearby galaxies by Goddard et al. (2010), while the pink shaded region shows the envelope that covers the data points compiled in Adamo et al. (2015).

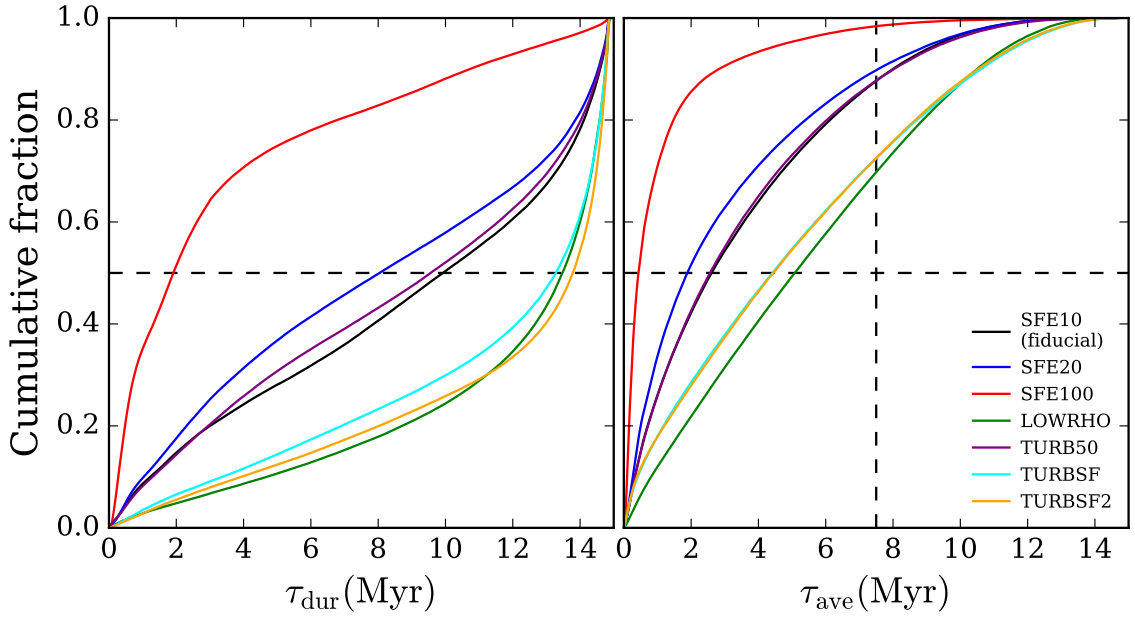

Figure 2.10: Cumulative distribution function of the cluster formation duration $\tau_{\rm dur}$ (left panel) and the mass-averaged cluster formation timescale $\tau_{\rm ave}$ (right panel). Different colors correspond to the different models described in legend. The cluster samples are selected within the main galaxy at $z \approx 5.3$ for all models.

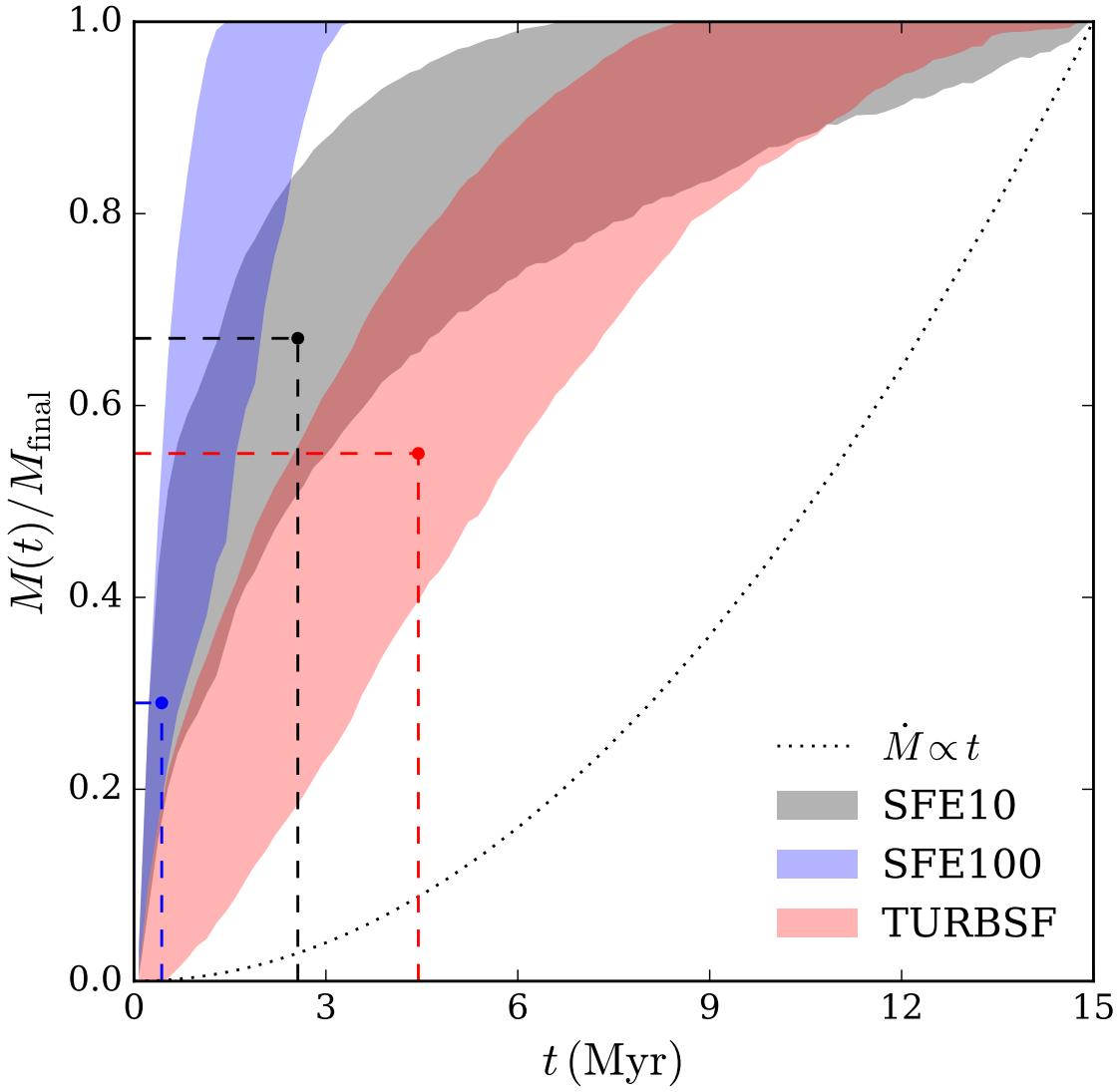

Figure 2.11: The quartile (25-75 percentile) ranges of mass growth history for active clusters in SFE10 (black), SFE100 (blue), and TURBSF (red) runs, respectively. Median mass-averaged cluster formation timescales, $\tau_{\rm ave}$, for each model are shown as vertical dashed lines. If the mass growth of all clusters was exactly linear, $\dot{M}$ = const, these lines would intersect the median mass track at $M(\tau_{\rm ave})/M_{\rm final} = 1/2$. The actual points at which they intersect vary from 0.29 to 0.67. A linearly increasing mass growth history, $\dot{M} \propto t$, predicted by Murray & Chang (2015) for the collapse of self-gravitating turbulent cloud is overplotted for comparison.

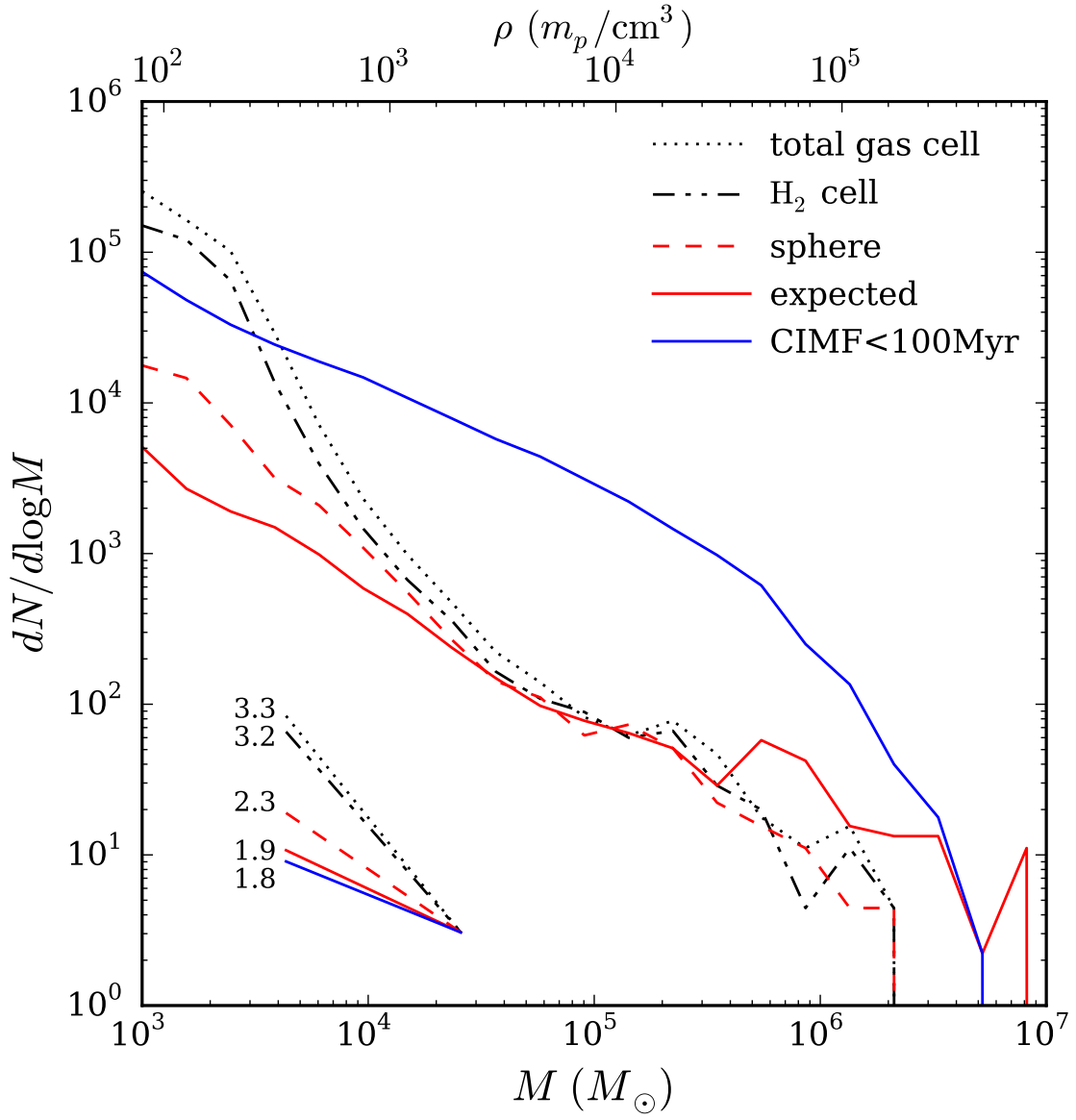

Figure 2.12: Mass functions of gas cells (black), cluster-forming spheres (red), and star clusters (blue) within the virial radius of the main galaxy at $z \approx 3.3$. Dotted and dashed-dotted black lines show the distribution of total and molecular gas mass in simulation cells, rescaled to the volume of the cluster-forming sphere. Red dashed line is the mass function of molecular gas within the spheres. Red solid line shows the expected distribution of cluster mass from a simple growth model discussed in Section 2.6.2. Blue solid line shows CIMF for all clusters younger than 100 Myr. We choose this longer time interval to accumulate sufficient cluster number to characterize CIMF, but this means the normalization of CIMF differs from the other displayed mass functions. The plot illustrates only the differences in the shape of the mass functions. Dynamical disruption of clusters would also reduce the normalization of CIMF. The best-fit slopes of all mass functions in the range $10^3 - 10^5\,M_\odot$ are shown in the lower left corner with the corresponding line styles.

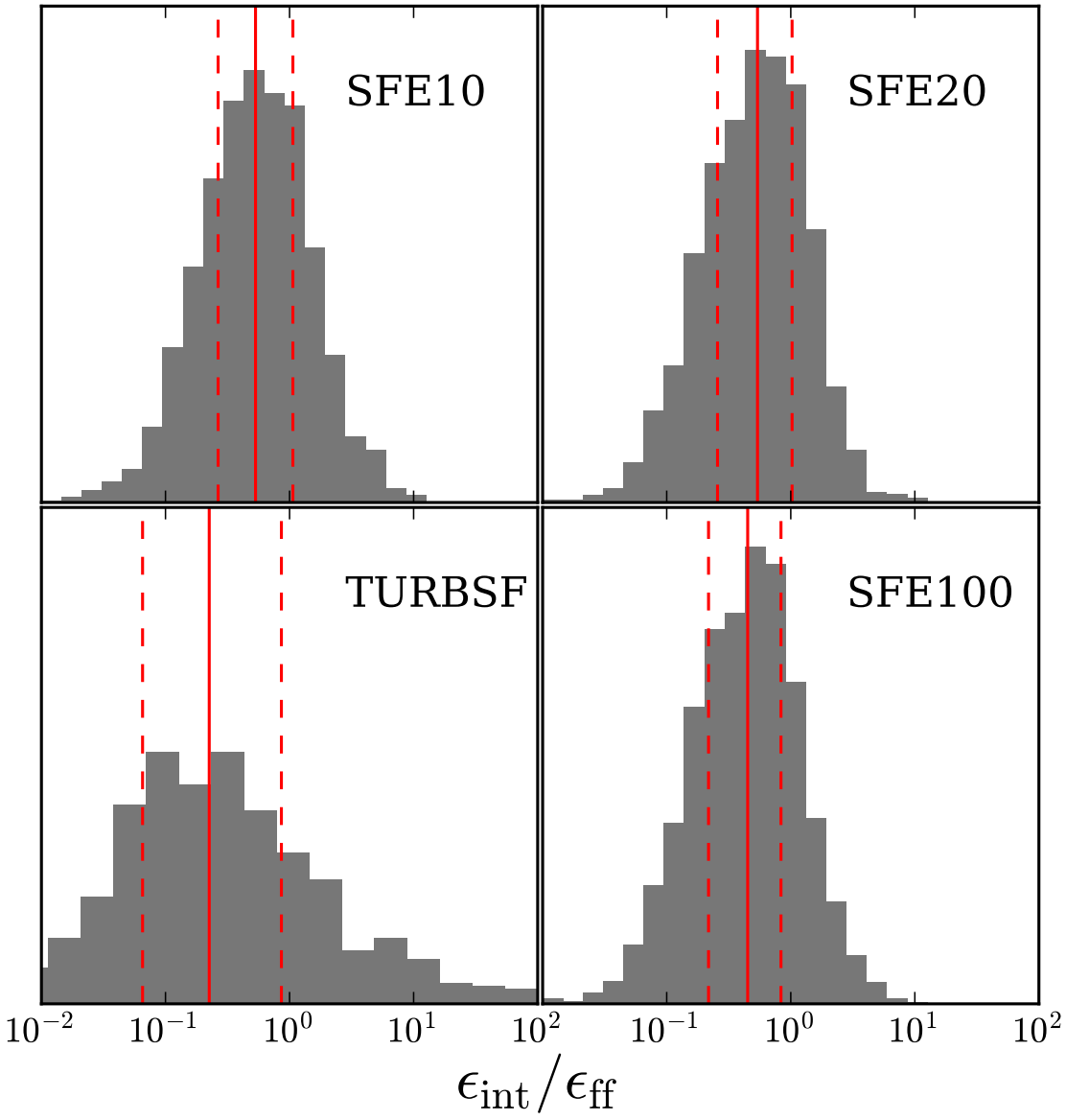

Figure 2.13: Distribution of the integral star formation efficiency $\epsilon_{\rm int}$ for four models with different local efficiency $\epsilon_{\rm ff}$. We show the ratio $\epsilon_{\rm int}/\epsilon_{\rm ff}$ to emphasize the spread of values resulting from different accretion histories. Vertical lines represent the median and the 25-75 percentile range for each distribution.

CHAPTER III

Effects of Star Formation Efficiency and Stellar Feedback

3.1 Introduction

With the advance of observational discoveries, such as the anisotropy of the cosmic microwave background (e.g. Komatsu et al., 2011; Planck Collaboration et al., 2016) and the large-scale galaxy clustering (e.g. Geller & Huchra, 1989; Bond et al., 1996; Gott et al., 2005), the Λ-Cold Dark Matter cosmology and hierarchical structure formation framework have been widely accepted and served as the starting point of theoretical investigations of galaxy formation. Among all theoretical methodologies, numerical simulations have become the main tool to study the formation and evolution of galaxies (see Somerville & Davé, 2015).

Built upon the first generation of N-body only simulations that explored the growth of large-scale structure under gravity (Springel et al., 2005; Boylan-Kolchin et al., 2009a; Stadel et al., 2009; Klypin et al., 2011), recent cosmological hydrodynamical simulations with state-of-the-art suites of physical ingredients and numerical techniques started to reproduce various stellar and gaseous properties of the observed galaxies, such as the stellar mass-halo mass relation, the average star formation histories (SFHs), and Kennicutt-Schmidt relations (KSRs), not only at $z \approx 0$ but also at

high redshifts (e.g. Agertz et al., 2013; Vogelsberger et al., 2014; Schaye et al., 2015; Hopkins et al., 2014). This success is achieved partly with the accurate numerical treatments of complex baryonic physics, such as gravity, gasdynamics, and radiative heating and cooling, but largely due to novel implementations of sub-grid models that describe the star formation and stellar feedback processes that cannot be spatially or temporally resolved in these simulations (e.g. Katz, 1992; Cen & Ostriker, 1992; Navarro & White, 1993; Katz et al., 1996; Springel & Hernquist, 2003a).

During the last two decades, great efforts have been made to explore the sources and implementations of stellar feedback processes in cosmological simulations (Stinson et al., 2006; Governato et al., 2007; Scannapieco et al., 2008; Agertz et al., 2011, 2013; Guedes et al., 2011; Aumer et al., 2013; Booth et al., 2013; Ceverino et al., 2014; Keller et al., 2014; Roškar et al., 2014). These works have demonstrated the important role played by the energetic feedback to suppress star formation at high redshift and launch galactic winds (e.g. Muratov et al., 2015). However, it is troublesome that a surprisingly broad range of feedback models claims to match the same global properties of galaxies by fine-tuning parameters of feedback prescriptions, reducing the predictive power of galaxy formation modeling (Naab & Ostriker, 2016). Yet, it is still unknown whether the implementations used in these simulations are appropriate for capturing the physical properties of the interstellar medium (ISM) on smaller scales. As the spatial resolution of current simulations is approaching the scales of individual star-forming regions (e.g. Hopkins et al., 2014; Read et al., 2015; Wetzel et al., 2016), it is critical to develop systematic methods to calibrate the sub-grid models on a similar scale.

In Li et al. (2017, hereafter, Paper I) we have introduced a new prescription for modeling star formation by considering star clusters as a unit of star formation, following the general consensus that most stars form in cluster environments (Lada & Lada, 2003). In this prescription, a cluster particle grows continuously through

gas accretion from its natal giant molecular cloud (GMC). The growth of a cluster particle is resolved in time and is terminated by its own energy and momentum feedback. Thus, the final particle mass is set self-consistently and can be considered as the mass of a single star cluster formed within the GMC. Since the cluster growth is determined by both the efficiency of star formation and the strength of stellar feedback, comparing key properties of model clusters with observations provides us with a unique opportunity to constrain these sub-grid models on scales of cluster-forming regions, instead of the kpc scales.

Recent observations of star-forming regions in the Milky Way and other nearby galaxies reveal a large number of cluster samples that contain the information about their formation environment (Portegies Zwart et al., 2010). Star clusters follow a well-defined initial mass function (CIMF) that can be described by the Schechter function with a power-law slope of ~ -2 and an exponential cutoff at high mass end. Indeed, as we have shown in Paper I, the slope of the CIMF reflects the slope of the density PDF of the star-forming gas, modulated by the feedback effects. Moreover, the high mass cutoff is also related to the intensity of star formation activity of the host galaxies (e.g. Larsen, 2002; Adamo et al., 2015; Johnson et al., 2016).

For a long time, it has been believed that the galaxy-wide low star formation efficiency was caused by the delay of gravitational collapse by magnetic or turbulent support (Krumholz & McKee, 2005; Krumholz & Tan, 2007). However, recent observations reveal a very short age spread of stars, a few Myr or a couple of free-fall timescale, in many young star clusters (Mac Low & Klessen, 2004; Hartmann et al., 2012; Hollyhead et al., 2015), suggesting that cluster formation is a rapid and dynamical process. The cluster formation timescale is determined by both the speed of gas accretion and the intensity of stellar feedback, thus providing an additional test of the star formation and feedback implementation in the simulations.

In this paper, we further revise and improve the cluster formation model. We

describe the major updates to the model in Section 3.2, and present results from new set of simulations in Section 3.3. We perform simulations with a wide range of the sub-grid model parameters: local star formation efficiency and boost of supernova momentum feedback. The CIMF, the fraction of star formation in bound clusters, and the age spreads present new constraints on the sub-grid parameters, which we discuss in Section 3.4. We summarize our results and conclusions in Section 3.5.

3.2 Simulations

3.2.1 Overview of cluster formation

The full description of the simulation setup is presented in Paper I. Here we first briefly summarize the common aspects and then describe the improvements to the model. Most of the simulations in this paper are new and have distinct features from those discussed in Paper I.

The simulations are run with the Eulerian gasdynamics and N-body Adaptive Refinement Tree (ART) code, which includes several key physical ingredients, such as three-dimensional radiative transfer, non-equilibrium chemical network, phenomenological molecular hydrogen formation and destruction, and a subgrid-scale turbulence model. We run the cosmological simulations from the initial condition that contains a main halo with the total mass $M_{200} \approx 10^{12}\,M_{\odot}$ at $z = 0$, in a periodic box of 4 comoving Mpc in size. All simulations start on a 128^3 root grid, which sets the dark matter particle mass $m_{\rm DM} = 1.05 \times 10^6\,M_{\odot}$. High dynamic range of spatial resolution is achieved by adaptive mesh refinement, where additional cell levels are added to the root grid. We apply a Lagrangian refinement criteria for both dark matter and gas components so that the mass of all gas cells varies only within a narrow range at all times. In addition, we also adopt a Jeans refinement criterion with which cells larger than twice the local Jeans length will be refined.

In Paper I, we have developed a novel algorithm for the formation of star clusters in cosmological simulations: continuous cluster formation (CCF). In CCF, each star particle represents a single star cluster, formed at a local density peak of the molecular gas. In order to avoid dependence on the time-variable physical size of a cell, $L_{\rm cell}$, the cluster particle is allowed to grow its mass via gas accretion within a spherical region of fixed physical size. We interpret this sphere as the dense part of a GMC that would form a bound stellar system in free-fall collapse. The optimal value of the sphere size was investigated in Paper I, $R_{\rm GMC} = 5\,{\rm pc}$, which is similar to the sizes of observed massive cluster-forming clouds (e.g. Urquhart et al., 2014). We will henceforth refer to this star-forming sphere as the "GMC".

At the highest refinement level, the GMC sphere fully includes the peak-density ("central") cell and partially overlaps with its 26 neighbor cells from a $3 \times 3 \times 3$ cube configuration. Accessing more than these immediate neighbors is computationally prohibitive in the ART code. The growth rate of a given cluster depends on the H_2 density of the overlapping cells:

$$\dot{M} = \sum_{\rm cell} f_{\rm GMC}\, V_{\rm cell}\, \dot{\rho}_* = \frac{\epsilon_{\rm ff}}{\tau_{\rm ff}} \sum_{\rm cell} f_{\rm GMC}\, V_{\rm cell}\, f_{\rm H_2}\, \rho_{\rm gas}, \tag{3.1}$$

where $V_{\rm cell}$ is the volume of each neighbor cell, $f_{\rm GMC}$ is the fraction of $V_{\rm cell}$ that overlaps with the GMC, $f_{\rm H_2}$ is the mass fraction of hydrogen in molecular phase, $\rho_{\rm gas}$ is the total gas density, and $\epsilon_{\rm ff}$ is the star formation efficiency per free-fall time. The free-fall time is defined as

$$\tau_{\rm ff} \equiv \left(\frac{32G\rho}{3\pi}\right)^{-1/2} \approx 1.6\,{\rm Myr}\, \left(\frac{n_{\rm H}}{10^3\,{\rm cm^{-3}}}\right)^{-1/2}, \tag{3.2}$$

where $n_{\rm H}$ is the total number density of hydrogen. The mass growth of the cluster particle is calculated and added to the particle mass at each local timestep, typically $\Delta t \sim 100$ yr. The mass accumulation history of each cluster is thus temporally

resolved with many thousands of steps.

Cluster formation is allowed only in cells above the number density $n_{\rm H,th} = 10^3\,{\rm cm}^{-3}$. This is close to the observational estimate by Kainulainen et al. (2014) of the H_2 density threshold for star formation in nearby GMCs of $\sim 5 \times 10^3\,{\rm cm}^{-3}$. In addition, to avoid the creation of very small and numerous clusters, new particles are created only if their expected final mass is above threshold $M_{\rm th}$. The expected mass is calculated as the initial rate $\dot{M}$ times the maximum allowed formation time, $\tau_{\rm max} = 15\,{\rm Myr}$. Since the actual duration of cluster formation is significantly shorter in the new runs, we discuss and revise $M_{\rm th}$ in Section 3.2.2.4.

3.2.2 Improvements to Paper I methodology

3.2.2.1 Gas cell refinement

In Paper I, we employed quasi-Lagrangian refinement criteria, which keep the mass of all cells (except for cells at the finest level) within a similar range. We examined the influence of $R_{\rm GMC}$ on CIMF by varying $R_{\rm GMC}$ from 2.5 to 7.5 pc and found that the physical size of the gas cells ($L_{\rm cell}$) involved in star formation should be comparable to the size of the GMC sphere. Therefore, we tune the refinement strategy such that the physical size of gas cells at the finest refinement level remains around $R_{\rm GMC} = 5\,{\rm pc}$. Initially, we allow 9 refinement levels on the 128^3 root grid at high redshift until $z \approx 9$. As simulations run toward lower redshift, the $10^{\rm th}$, $11^{\rm th}$, and $12^{\rm th}$ refinement levels are added at $a_{\rm uni} \approx 0.1$, 0.2, and 0.4, respectively. This refinement method keeps $L_{\rm cell}$ of the finest level in the range between 3 and 6 pc at all cosmic times of interest in this paper.

3.2.2.2 Redshift-independent star formation efficiency

The mass accretion rate of a cluster is given by Equation (3.1), which contains parameter $\epsilon_{\rm ff}$. However, the meaning of $\epsilon_{\rm ff}$ is quantitatively different from what

is commonly used in galaxy formation simulations. In the traditional prescription, stellar particles are given the mass calculated as the star formation rate density $\dot{\rho}_*$ times the volume of the cell containing the particle. The cell size is fixed in comoving coordinates but expands in proper physical coordinates, so that for the same gas density (typically, near the threshold for star formation) the particle mass increases with time. In contrast, in our model, each star cluster grows its mass over several million years by accreting material within a cloud of fixed physical radius, $R_{\rm GMC}$. Our $\epsilon_{\rm ff}$ is applied to a fixed volume and does not vary with cosmic time. This is an important improvement.

We can connect our value of $\epsilon_{\rm ff}$ with that on the scale of a cell via the differences in volume. Consider the case when the diameter of a GMC sphere is smaller than a star-forming cell, and therefore the GMC is completely embedded in the cell. The smaller volume allowed to participate in star formation translates into the smaller efficiency for the whole cell:

$$\epsilon_{\rm ff,cell} = \frac{V_{\rm GMC}}{V_{\rm cell}}\,\epsilon_{\rm ff} = \frac{4\pi}{3}\left(\frac{R_{\rm GMC}}{L_{\rm cell}}\right)^3 \epsilon_{\rm ff}. \tag{3.3}$$

For example, the seemingly high value $\epsilon_{\rm ff} = 50\%$ in our CCF model is equivalent to $\epsilon_{\rm ff,cell} \approx 1\%$ for a cell of 30 pc, a typical size for current highest-resolution cosmological simulations.

We allow cluster formation at the three finest refinement levels, which is consistent with our adopted star formation density threshold: $n_{\rm crit} = 10^3\,{\rm cm}^{-3}$. The average number density of hydrogen plus helium atoms for cells at level l and redshift z in our simulation box is:

$$n(z) \approx 900\,{\rm cm}^{-3} \times 8^{l-9}\left(\frac{1+z}{3}\right)^3. \tag{3.4}$$

We can see that at $z = 2$, cells at level 9 (the third finest level at that time) have

the average density close to $n_{\rm crit}$. Cells create new clusters as soon as they exceed the density threshold, which usually happens already at the third finest level. The physical cell size at that level is in the range $L_{\rm cell} = 12 - 24$ pc, which is larger than $2R_{\rm GMC}$. Thus, the corresponding cell efficiency is $\epsilon_{\rm ff,cell} = (0.038 - 0.30)\,\epsilon_{\rm ff}$.

This is an important point for the comparison with the low inferred efficiency per free-fall time in Galactic star forming regions, $\epsilon_{\rm ff} \approx 1\%$ (Kennicutt, 1998; Krumholz et al., 2012). In our model, clusters begin forming with a similar efficiency. As the gas continues to collapse in free fall, the density increases and cells are refined to the highest level. Then the effective $\epsilon_{\rm ff,cell}$ increases as well, in agreement with results of Murray (2011). As in every numerical simulation, at the highest refinement level we are not properly resolving gas collapse, and therefore underestimate the density and overestimate the free-fall time. This numerical effect also requires adopting larger value of $\epsilon_{\rm ff}$ to maintain the same ratio $\epsilon_{\rm ff}/\tau_{\rm ff}$.

3.2.2.3 Restriction on creation of new clusters near existing active clusters

In Paper I, cluster particles were seeded in cold dense cells that contain local density peaks. We follow the same procedure here, but with addition algorithmic improvements.

As we discussed in Section 3.2.2.1, the physical size of gas cells at the finest refinement level is always kept in the range of $3 - 6\,$pc. This means that clusters in the smallest cells can accrete gas from all 27 cells that overlap with the GMC sphere (see Figure 3.1). To be consistent with our interpretation that the peak-density cell represents the star-forming part of a single GMC, the neighboring cells should not produce new clusters that would compete for gas supply with an existing cluster. Therefore, we prohibit a cell from creating a new particle if it already has an actively growing cluster in a neighbor cell.

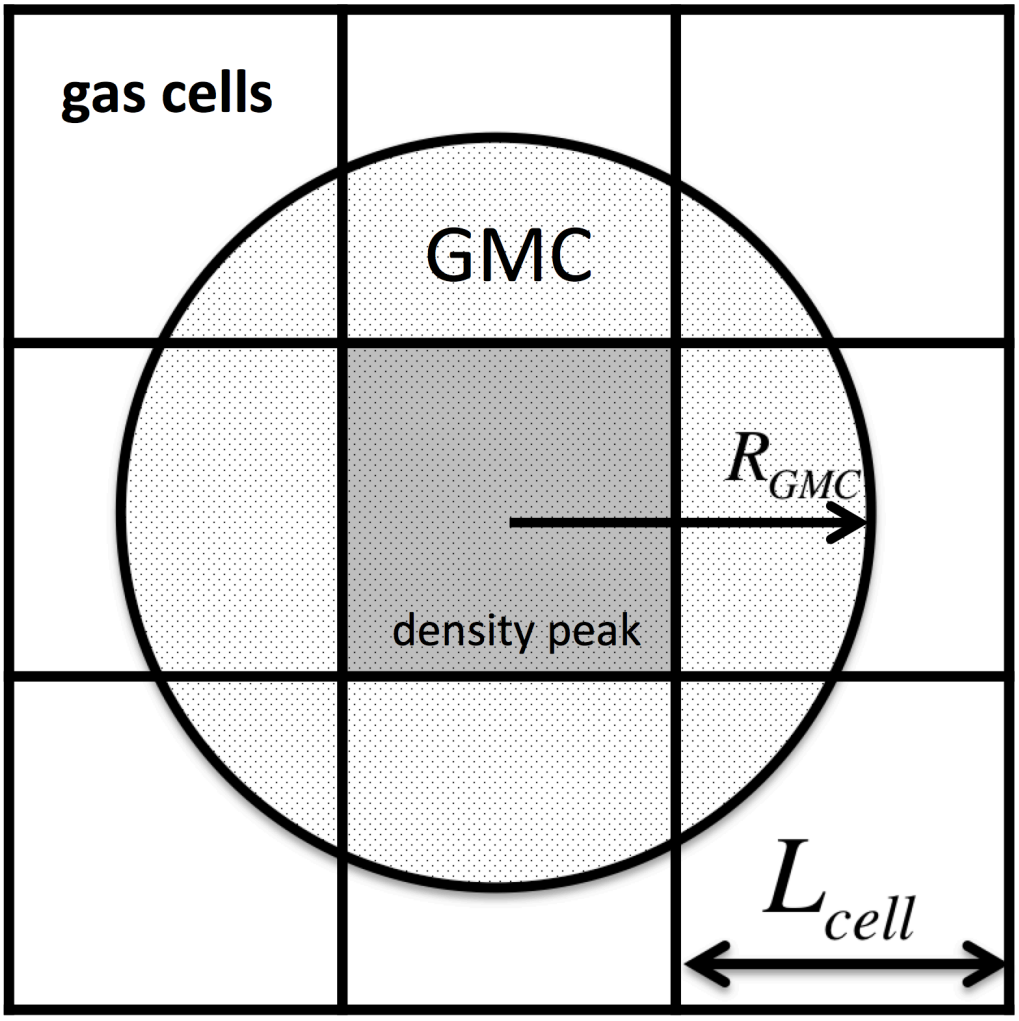

Figure 3.1: Sketch of the star-forming GMC sphere laid out on the gas cell structure in the simulations.

This restriction makes sense only for the finest refinement level cells. If cluster particles are formed in coarser cells, which enclose a large fraction of volume of the GMC sphere, then there is no competition with neighboring cells. In this case, we allow new clusters to be created.

We also do not apply this restriction to the finest-level cells that are separated by more than one cell size from the central cell. These are the diagonal cells in the cube. Thus, the only cells that are prohibited from creating new clusters are the six ones, "face-touching" the central cell.

To describe the above criterion for cluster seeding more quantitatively, we define an overlap fraction, $f_{\rm over}$, of the volume of a given "face-touching" neighbor that is occupied by the central GMC sphere. If a given cell satisfies the star formation criterion but has any "face-touching" neighbor with a significant overlap, $f_{\rm over} > 20\%$, that already contains an actively growing cluster, the above cell is not allowed to create a new particle. If the overlap fraction is small, $f_{\rm over} < 20\%$, we do not apply this restriction.

All these checks are meant to minimize the impact of the neighbor restriction as much as is reasonable. However, we show in Section 3.3 that it still has a significant impact on the shape of the CIMF, relative to the results in Paper I.

3.2.2.4 No removal of low-mass clusters

In the Paper I algorithm, any inactive cluster particles less massive than threshold mass $M_{\rm th} = 10^3\,M_\odot$ were recycled to speed up the simulations. Recycling meant that the material converted into stars was returned to the ISM and available for future star formation. However, some amount of energy and momentum from the stars of these failed clusters was deposited into the surrounding gas while the clusters were forming and their final mass was unknown, and this feedback could not be undone. That is, failed clusters produced some feedback that was not real. The contribution of these low-mass clusters to the stellar mass budget and overall galaxy dynamics was negligible, and therefore we can expect the impact of their feedback to be small.

With a stronger stellar feedback implementation described below in Section 3.2.2.8, both the galaxy stellar mass and the number of cluster particles are significantly reduced. Therefore, the recycling of low-mass clusters is not needed. In the new runs in this paper we have eliminated it.

We also boosted $M_{\rm th}$ from $10^3\,M_\odot$ to $6 \times 10^3\,M_\odot$, because the effective cluster formation duration is much shorter than the maximum time $\tau_{\rm max}$ that is used to

predict the mass of a cluster before its formation. Although the effective timescale for individual clusters varies, we show in Section 3.3.7 that $\tau_{\rm ave} = 2.5\,{\rm Myr}$ is a good upper limit for clusters less than $\sim 10^5\,M_\odot$. Therefore, the minimum formation rate $\dot{M}_{\rm min} = 10^3\,M_\odot/15\,{\rm Myr}$ in Paper I produced clusters with typical mass $\sim \dot{M}_{\rm min} \times 2.5\,{\rm Myr}$, a factor of 6 below out intended threshold. With the new value $M_{\rm th} = 6 \times 10^3\,M_\odot$ we eliminate most of small clusters with final mass below $10^3\,M_\odot$.

3.2.2.5 No molecular fraction threshold for continuing cluster growth

In Paper I, we employed a threshold on molecular fraction of hydrogen so that clusters could only form and grow in cells with high molecular fraction $f_{\rm H_2} > 50\%$. This meant that if the molecular fraction of a given cell fell below 50%, not only new clusters were not allowed to form, but also an existing active cluster in the cell was not allowed to continue to grow.

After detailed analysis of the growth history of individual cluster particles, we found that this molecular fraction threshold leads to intermittent gas accretion for some clusters, when $f_{\rm H_2}$ oscillates around the threshold value. It also generates many gap periods without any mass growth, especially for the most massive clusters. These gaps affect the calculation and interpretation of the cluster formation timescales. Indeed, we found that the long formation timescales of some massive clusters in Paper I were not due to slow star formation but due to the cessation of star formation by this molecular threshold.

Here we revise the implementation of molecular threshold by adopting it only as the criterion of initial particle creation. Once a cluster is formed, its subsequent accretion is no longer affected by the molecular fraction. Thus, active clusters can continue to grow their mass at the rate that is proportional to the molecular fraction, via Equation (3.1), until the gas density drops below the threshold $n_{\rm crit} = 10^3 {\rm cm}^{-3}$.

3.2.2.6 Gap elimination

As we described above, imposing molecular fraction threshold to active clusters can lead to gaps in their mass growth history. This is due to an artificial implementation of the threshold. Another reason for the existence of long accretion gaps is due to the motion of clusters.

In high-density regions, such as the inner part of the galaxy, gravitational forces accelerate particles to high velocities. For a velocity dispersion on the order of 20 km s^{-1}, during the active formation period, $\tau_{\max} = 15\,\mathrm{Myr}$, clusters can travel up to 300 pc, which is much larger than the cell size or the GMC size. This means that active clusters can travel through and accrete gas from multiple GMCs, different from the one in which they had originally formed. This is not physically correct because a cluster should accrete gas only from one GMC, and each GMC should host a distinct cluster. Moreover, when a cluster travels through low-density regions between dense clouds, the accretion stops because the gas density falls below the threshold density for star formation, thus creating artificial gaps in the mass growth history.

To eliminate accretion of a given cluster particle from multiple GMCs, in the simulations presented in this paper we monitor the accretion process for each active cluster. We deactivate a cluster even before it reaches the maximum time $\tau_{\max}$ if we find that it completely stopped growing mass for more than 1 Myr. By design, this procedure eliminates all gaps longer than 1 Myr, and systematically reduces the cluster formation times even for most massive clusters, as we show in Section 3.3.7.

3.2.2.7 Mass loss rate due to stellar evolution

In Paper I, stellar particle with initial mass M_i lost mass due to stellar winds and stellar evolution at the rate given by:

$$\frac{dM}{dt} = \frac{\eta}{t + \tau_{\mathrm{loss}}} M_i, \tag{3.5}$$

where η is the typical fraction of mass loss and $\tau_{\rm loss}$ is the characteristic timescale, both of which depend on the initial mass function of stars. For a Kroupa (2001) IMF, we used $\eta = 0.046$ and $\tau_{\rm loss} = 2.76 \times 10^5\,{\rm yr}$. The time-evolution of the cumulative fraction of stellar mass loss is then obtained by integrating Equation (3.5) over time:

$$f_{\rm loss}(t) = \frac{M_{\rm loss}}{M_i} = \int_0^t \frac{\eta}{t + \tau_{\rm loss}} dt = \eta \ln\left(\frac{t}{\tau_{\rm loss}} + 1\right). \tag{3.6}$$

Figure 3.2 shows this mass loss history, together with the results from a detailed stellar population synthesis model FSPS[1] for the same IMF (Conroy et al., 2009; Conroy & Gunn, 2010). We find that the previous implementation overestimates the mass loss rate over the whole stellar lifetime. The overestimation is more prominent for younger stellar populations. For example, after 3 Myr, $\sim 10\%$ of the stellar mass is lost in Paper I, but less than 1% in FSPS modeling.

Here we introduce a new mass loss model that fits the time-evolution of the cumulative fraction of mass loss in FSPS with a second-order polynomial. We write $f_{\rm loss}(t) = ax^2 + bx + c$ where $x = \log_{10}(t/{\rm yr})$, and obtain the best-fit parameters: $a = -0.010$, $b = 0.288$, and $c = -1.42$. The mass loss rate at a given age t is obtained by differentiating $f_{\rm loss}(t)$ with respect to t:

$$\frac{dM}{dt} = \left(\frac{b}{\ln 10} + \frac{2a}{\ln^2(10)} \ln(t)\right) \frac{M_i}{t}. \tag{3.7}$$

Note that this expression works for stars in the age range from 2.75 Myr to 13.7 Gyr. For ages younger than 2.75 Myr, we assume no appreciable mass loss.

3.2.2.8 Changes to stellar feedback

As in Paper I, the feedback sub-grid model consists of mass, momentum, and energy injections from stellar winds, radiation pressure, and supernova explosions.

[1]Flexible Stellar Population Synthesis, `https://github.com/cconroy20/fsps`

The main update in this paper is in the implementation of supernova remnant (SNR) feedback. In the new SNR model, we estimate the partition of the thermal, kinetic, and turbulence energies of SNR using a parameterization model calibrated by high-resolution hydrodynamic simulations in inhomogeneous turbulent medium by Martizzi et al. (2015). The energy and momentum input from SNRs depend on the ambient density and spatial resolution of the simulations, as described in detail in Semenov et al. (2016).

One caveat of the SNR model is that it is calibrated by simulations of SN explosion in isolation, rather than in a more complex star formation region. In reality, stars are formed mainly in clusters and massive cluster-forming regions usually produce a large number of massive stars that undergo supernova explosion over several Myr. Gentry et al. (2017) found that such clustering of supernovae can enhance momentum feedback by an order of magnitude relative to that delivered by an isolated supernova. Therefore, in some of our simulations, we boost the strength of momentum feedback of the SNR model in Martizzi et al. (2015) by a factor $f_{\rm boost}$ =3, 5, or 10, as listed in Table 3.1.

3.2.2.9 Initial bound fraction

In Paper I, we assumed that the whole accreted mass of cluster particles during their active growth is gravitationally self-bound when the particles emerge from their natal clouds. That is, the particle mass is the bound cluster mass. This assumption does not take into account the complex dynamical evolution of star clusters in the early phase, when the boundedness is affected by the hierarchical structure of the ISM and gas expulsion due to stellar feedback. Recent observations and numerical simulations of turbulent clouds suggest that the fraction of mass in a given star-forming complex that remains bound after a few Myr depends strongly on the integral star formation efficiency on the scale of GMCs (Goodwin, 1997; Geyer & Burkert,

2001; Goodwin & Bastian, 2006; Smith et al., 2011; Kruijssen et al., 2012).

In this paper we introduce the *initial bound fraction*, f_i, and assign it to each stellar particle. We adopt a linear dependence of f_i on the local star formation efficiency:

$$f_i \equiv \min\left(\frac{\epsilon_{\rm int}}{\epsilon_{\rm core}}, 1\right) \tag{3.8}$$

where $\epsilon_{\rm core} = 0.5$ is the correction factor for mass loss due to protostellar outflows, suggested by Kruijssen (2012). The integral star formation efficiency, $\epsilon_{\rm int}$, is defined as the ratio between the final mass of the stellar particle ($M_{\rm f}$) and the maximum baryon mass of the GMC throughout the whole course of cluster accretion:

$$\epsilon_{\rm int} \equiv \frac{M_{\rm f}}{\max_t(M_* + M_{\rm gas})}. \tag{3.9}$$

It is important to emphasize that the GMC mass in our simulations varies on the timescale of less than a Myr, as the central part is converted into stars and outside gas flows in, in essentially free fall. The sum of stellar and gas mass usually increases over several Myr, before declining as the feedback of young stars disperses the remaining gas. This situation is very different from simulations of star formation in isolated clouds, where the total baryon mass is fixed as the initial cloud mass. The initial cluster mass in this paper then refers to the bound cluster mass that takes into account the initial bound fraction: $f_i M_{\rm f}$.

The fraction of stars remaining bound to the cluster continues to evolve due to the dynamical evaporation and tidal stripping. We calculate this process with a subgrid model and will discuss the evolution of the cluster mass function in a follow-up paper.

3.2.2.10 Alternative definition of cluster formation timescale

In Paper I, the average duration of cluster formation was calculated as the mass-weighted star formation timescale:

$$\tau_{\rm ave} \equiv \frac{\int_0^{\tau_{\rm max}} t\,\dot{M}(t)\,dt}{\int_0^{\tau_{\rm max}} \dot{M}(t)\,dt}, \tag{3.10}$$

where $\dot{M}(t)$ is the cluster SFR at time t. This definition best describes steady mass accretion. For example, in the case of constant SFR, $\tau_{\rm ave} = \tau_{\rm max}/2$.

Although the actual star formation history of individual star clusters is difficult to determine in observations, recent theoretical models and hydrodynamic simulations suggest a dynamic, time-variable process. Self-gravitating collapse of the dense parts of GMCs is thought to accelerate star formation until stellar feedback, such as stellar winds and radiative pressure/heating, pushes the accreting gas out. The whole process happens quickly, as indicated by the observed age spread of stars in embedded clusters within 3-4 Myr, only a couple free-fall times of their natal clouds; see Section 3.2.4.1.

One caveat of the original definition of cluster formation time described in Equation (3.10) is that, in the case of increasing and then decreasing SFR, $\tau_{\rm ave}$ actually records the epoch when $\dot{M}$ reaches its peak, rather than the duration of the process. Here we define a new quantity that better describes the width of the star formation history. The age spread is defined as the ratio between the final particle mass $M_{\rm f}$ and the mass-weighted SFR over the whole growth history $\langle\dot{M}\rangle$:

$$\tau_{\rm spread} \equiv \frac{M_{\rm f}}{\langle\dot{M}\rangle} = \frac{M_{\rm f}}{\int_0^{\tau_{\rm max}} \dot{M}^2 dt / M_{\rm f}}. \tag{3.11}$$

For a power-law mass accretion history, $\dot{M} \propto t^{\alpha}$, both $\tau_{\rm ave}$ and $\tau_{\rm spread}$ can be explicitly evaluated as $\tau_{\rm ave} = \frac{\alpha+1}{\alpha+2}\tau_{\rm max}$ and $\tau_{\rm spread} = \frac{2\alpha+1}{(\alpha+1)^2}\tau_{\rm max}$, respectively, for $\alpha \neq -1$. However, in the case when $\dot{M}$ exhibits a peak, e.g. a Gaussian function $\dot{M} \propto$

$\exp(-(t-t_0)/2\sigma^2)$, $\tau_{\rm ave}$ is

$$\tau_{\rm ave} = t_0 + \sqrt{\frac{2}{\pi}}\sigma \frac{\exp(-t_0^2/2\sigma^2) - \exp(-(t_0-\tau_{\rm max})^2/2\sigma^2)}{\mathrm{erf}(t_0/\sqrt{2}\sigma) + \mathrm{erf}((\tau_{\rm max}-t_0)/\sqrt{2}\sigma)}, \tag{3.12}$$

while the new definition of $\tau_{\rm spread}$ is

$$\tau_{\rm spread} = \sqrt{\pi}\sigma \frac{[\mathrm{erf}(t_0/\sqrt{2}\sigma) + \mathrm{erf}((\tau_{\rm max}-t_0)/\sqrt{2}\sigma)]^2}{\mathrm{erf}(t_0/\sigma) + \mathrm{erf}((\tau_{\rm max}-t_0)/\sigma)}. \tag{3.13}$$

The relationship between the intrinsic width of the mass accretion history σ and the derived average ages $\tau_{\rm ave}$ and $\tau_{\rm spread}$ is shown in Figure 3.3.

When $\sigma \to 0$, the SFH reduces to a δ-function located at $t = t_0$. In this case, $\tau_{\rm ave} \to t_0$, while $\tau_{\rm spread} \to 2\sqrt{\pi}\sigma$. This extreme case illustrates the problem with our previous definition of the formation timescale. Rather than recording the age spread of stars, $\tau_{\rm ave}$ actually reflects the epoch of the peak of star formation. Instead, the new definition follows the true width of the age distribution.

As σ increases, the SFH becomes flatter. At the other extreme, if $\sigma > \tau_{\rm max}$, the SFR can be considered constant. The above equations then reduce to $\tau_{\rm ave} \to t_0/2$ and $\tau_{\rm spread} \to t_0$.

3.2.3 New runs

In this paper we present new runs with the above improvements. All start with the same initial conditions. Their key physical parameters are listed in Table 3.1. The number after "SFE" in their name corresponds to the local $\epsilon_{\rm ff}$ in percent. The standard value of SNR momentum boost factor is $f_{\rm boost} = 5$. In run SFE50-SNR3, we adopt $f_{\rm boost} = 3$ to test the dependence of global SFR on this factor. In addition to these new runs with full updates, we include also for comparison the fiducial run "old-fid" from Paper I, and "SNR10-old" run with $f_{\rm boost} = 10$ and some intermediate degree of update. These latter runs illustrate progression in the development of our

Table 3.1: Model runs

Name	z_f	$\epsilon_{\rm ff}$	Feedback	$L_{\rm cell}^{\rm com}$ (pc) at $z = 2$
SFE10	1.7	0.1	early[a]+5*SNR[b]	15
SFE50	1.5	0.5	early+5*SNR	15
SFE50-SNR3	2.0	0.5	early+3*SNR	15
SFE100	1.5	1.0	early+5*SNR	15
SFE200	1.5	2.0	early+5*SNR	15
SFEturb	2.7	variable	early+5*SNR	15
fid-PaperI	3.2	0.1	early+old SN[c]	60
SNR10-old	2.0	0.1	early+10*SNR	15

Notes: a. "early": early feedback schemes including stellar wind and radiative pressure. b. "SNR": resolution-dependent SNR feedback scheme in Martizzi et al. 2015; the number above "SNR" is the boosting factor. c. "old SN": previous SN thermal and momentum feedback with the amount that is calibrated by Agertz & Kravtsov (2015).

algorithm.

- fid-PaperI: The fiducial run in Paper I with $\epsilon_{\rm ff} = 10\%$. As shown in Figure 3.5, this run overestimates the expected SFR of the main galaxy by more than one order of magnitude at $z = 4 - 8$. Due to the high SFR, a large number of clusters more massive than $10^5\,M_\odot$ are formed at high z. The inability to suppress SFR at high-z is caused by the weak feedback that is used in this run. Another sign of the ineffectiveness of the feedback comes from the mass-weighted cluster formation timescale $\tau_{\rm ave}$. In Paper I, we showed that the median value of $\tau_{\rm ave}$ for all clusters in this run is about 3 Myr, consistent with observational constraint that is described in Section 3.2.4.1. Detailed analysis suggests that this median value is dominated by a large number of less massive cluster particles. For clusters more massive than $10^5\,M_\odot$, however, $\tau_{\rm ave}$ is much longer, sometimes even longer than $\tau_{\rm max}/2 = 7.5\,$Myr, suggesting that the stellar feedback cannot terminate the gas accretion process for massive clusters within the allowed accretion timescale that is assigned in the simulation.

- highr: The same as "old-fid", except (1) a refinement strategy that keeps the ratio of $D_{\rm GMC}/L_{\rm cell}$ roughly constant, but one level coarser than that described in Section 3.2.2.1; and (2) a new SNR feedback prescription described in Section 3.2.2.8 with momentum boosting factor $f_{\rm boost} = 5$. We found that the SFH of this run follows the abundance matching results until $z \approx 2.5$, when a major merger happened in the main galaxy. This merger brings so much cold gas into the inner regions of the galaxy that stellar feedback is unable to disperse. An intense starburst occurs and a prominent stellar spheroid forms at the galactic center. As we discuss in the next section, adding one additional refinement level prevents this star burst and leads to a reasonable SFH. Another result is that, although the SFR of this run is much smaller than "fid-PaperI" due to the adoption of the new feedback model, there are still many clusters, especially massive ones, that have very long cluster formation timescale. This finding inspires the detail investigation of the origin of the long timescale and the implementation of the gap elimination algorithms as we described in Section 3.2.2.5 and 3.2.2.6.

- SNR10-old: The same as "highr", but with a larger boosting factor $f_{\rm boost} = 10$. The goal of this run is to probe the upper limit of $f_{\rm boost}$. As shown in Figure 3.5, the SFH is always under the abundance matching result. This underestimate of SFR sets an upper limit on the reasonable choice of $f_{\rm boost}$.

  For comparison, we have also run a simulations with the standard $f_{\rm boost} = 5$ but boosted the momentum feedback from radiative pressure by a factor of 10. This boost produced no apparent effect on SFH, unlike the SN boost.

Table 3.2: Properties of nearby young star clusters

Name	Age (Myr)	Age Spread (Myr)	Mass ($M_\odot$)	Reference
Orion Nebula Cluster	4-5	1-3	2000	Jeffries et al. 2011
R136	4-5	3	4.5e5	Massey & Hunter 1998
LH95	4	2.8-4.4	?	Da Rio et al. 2010
NGC 1569A	< 5	free of dust extinction	1e6	Maoz et al. 2001
Westerlund 1	5	0.4 or < 1	6.3e4	Negueruela et al. 2010
NGC 4103	?	2-4	?	Forbes 1996
NGC 3603 YC	2	0.1	?	Kudryavtseva et al. 2012
NGC 3603 HII	1	< 1	1.9e4	Pang et al. 2013
W3 Main	?	2-3	?	Bik et al. 2012
Antennae clusters	?	$< 6, A_V = 1$ mag	?	Whitmore & Zhang 2002
NGC 4449 clusters	?	$< 5, A_V = 0.5 - 1.5$ mag	0.5-5e4	Reines et al. 2008
M83 clusters	?	< 4	?	Hollyhead et al. 2015

3.2.4 New observational constraints

3.2.4.1 Observed age spread of young clusters

To use the predicted formation duration as an additional test of the models, in Table 3.2 we compile recent measurements of the spread of relative ages of stars in several young star clusters. Although obtaining accurate age measurements is still challenging, current observations suggest that the age spread should be less than 6 Myr for various star-forming regions in different galaxies. This upper limit also agrees with small-scale hydrodynamic simulations of cluster formation, which suggest star formation process should proceed on the timescale comparable to the free-fall time (e.g. Hartmann et al., 2012; Grudić et al., 2017). This relatively short timescale provides a strong constraint on the implementation of star formation and stellar feedback.

3.2.4.2 SFR vs. maximum cluster mass in nearby galaxies

The observed data of star formation rate surface density and V-band absolute magnitudes of the brightest young clusters in different galaxies is compiled in Adamo

et al. (2015). Note that this compilation contains observations of galaxy-wide measurements as well as spatially resolved samples. To convert the magnitudes of the brightest clusters to the corresponding stellar masses, we used the V-band mass-to-light ratio for young star clusters from Lieberz & Kroupa (2017).

3.3 Results

Most of the simulations presented here were performed at high-performance computing center Flux at the University of Michigan. We highlight here that, although our simulations incorporate many state-of-the-art physical processes, such as non-equilibrium chemical networks and radiative transfer, with very high spatial resolution, the total computing time is not huge. For reference, runs with $\epsilon_{\rm ff} \geqslant 0.5$ take about $30,000 - 50,000$ CPU hours to reach $z \approx 1.5$.

Figure 3.4 shows the gas surface density of the inner 4 kpc of the main galaxy at $z \approx 2$ for six different runs. The projection is taken along the X-axis of the simulation box, which is close to the intrinsic rotation axis of the disk. Compared to the fiducial run in Paper I, most of the density maps here do not exhibit well-defined gaseous disks. Instead, due to the stronger feedback implementation, the most prominent structures are the kpc-scale low-density cavities surrounded by rings of higher-density beads and filaments. These cavities are created by multiple SNe from young star clusters. Shock waves of the "superbubbles" can travel several kpc through these low-density regions, compress the gas located around the edge of the bubble, and possibly trigger subsequent star formation. They also generate large-scale outflows from the inner regions. The densest regions of the galaxies, therefore, are not centrally concentrated, but are distributed as many kpc-scale clumps, which is consistent with recent Hubble Ultra Deep Field observations of $z \approx 2$ star-forming galaxies (Guo et al., 2012).

On the other hand, the SFE50-SNR3 run with weaker feedback (upper right panel)

shows a more regular disk with centrally-peaked gas distribution and prominent spiral arms. This contrast illustrates that the large-scale gas distribution is very sensitive to variation of the momentum feedback parameters, even by a factor of two.

3.3.1 Star formation history of the main galaxy

In Figure 3.5, we show the star formation history of the main halo for several runs with different values of the star formation and feedback parameters. The SFH is calculated from all cluster particles within $0.5\,R_{\rm vir}$ of the main galaxy from the last available snapshot of each run, where $R_{\rm vir}$ is the virial radius of the dark matter halo. The rate of star formation is averaged over 100 Myr around a given epoch in order to smooth out stochasticity.

In general, simulations with $f_{\rm boost} = 5$ agree well with the abundance matching results, in the sense of both SFR and the cumulative stellar mass growth. This suggests that the strength of stellar feedback used in these runs is appropriate to reproduce the inefficient star formation at high redshift. Changing the intensity of feedback has a measurable effect on the SFH. The run with $f_{\rm boost} = 3$ systematically overestimates SFR at all times, while the run with $f_{\rm boost} = 10$ underestimates SFR by an order of magnitude. The resulting SFHs of the two runs are well beyond the 1σ confidence intervals of the abundance matching result. Therefore, the choice of $f_{\rm boost}$ in our simulations is tightly bracketed.

On the other hand, varying $\epsilon_{\rm ff}$ from ~ 0.01 to 2.0 has almost no systematic effect on SFH. This insensitivity to the value of $\epsilon_{\rm ff}$ has already been found in several recent cosmological simulations with the implementation of strong stellar feedback and high spatial resolution (Agertz et al., 2013; Agertz & Kravtsov, 2015; Hopkins et al., 2013). The feedback-controlled low-efficiency star formation activity can be interpreted as the short lifetime of star-forming regions, which then requires a large number of star-forming cycles to convert all the gas into stars (Semenov et al., 2017).

However, the lack of sensitivity of SFH to the local star formation efficiency does not mean that one can assign an arbitrary value to $\epsilon_{\rm ff}$ in galaxy formation simulations. As we show in the rest of this section, $\epsilon_{\rm ff}$ has dramatic effects on the properties of individual star-forming regions, and in turn, the properties of young star clusters formed within.

3.3.2 Kennicutt-Schmidt relation

Above we showed that with $f_{\rm boost} = 5$ our simulations can reproduce the SFH of Milky Way-sized galaxies, suggesting that the galaxy-wide star formation activity is well suppressed with our feedback implementation. Another commonly-used representation of inefficient star formation is the "Kennicutt-Schmidt relation" (KSR), which is a relationship between the gas surface density and the surface density of star formation rate.

To derive the KSR for molecular gas in our simulations, we split the main galaxy disk into several 1×1 kpc square areas and calculate $\Sigma_{\rm H_2}$ and $\Sigma_{\rm SFR}$ within each area. $\Sigma_{\rm SFR}$ is estimated by using clusters younger than 20 Myr. We test the effect of different spatial smoothing scales on the KSR. We find that using smaller spatial smoothing scales between 100 to 500 pc gives larger scatters on the KSR, but with similar median value as the 1 kpc case. We collect measurements from $z \approx 9$ to the last available snapshot in each run and calculate the median value of $\Sigma_{\rm SFR}$ for a given range of $\Sigma_{\rm H_2}$. Figure 3.6 shows the KSR in six simulations with different $\epsilon_{\rm ff}$.

The observed KSR does not vary significantly with metallicity or galaxy type, when expressed through molecular H_2 gas (Bigiel et al., 2008, 2011). The observed relation in nearby spiral galaxies at $z = 0$ is consistent with linear, and we can expect it to hold for higher redshift and smaller galaxies. Our simulations produce an approximately linear relation but with systematically higher $\Sigma_{\rm SFR}$ than observed, by about a factor of two to one order of magnitude. We also find that the simulations

with higher $\epsilon_{\rm ff}$ tend to have somewhat higher $\Sigma_{\rm SFR}$ for a given $\Sigma_{\rm H_2}$.

3.3.3 Initial bound fraction

Figure 3.7 shows the initial bound fraction predicted for our model clusters. We find a strong positive correlation between f_i and particle mass for all runs with a wide range of model parameters. This trend is not sensitive to either the formation epoch or host galaxy mass, suggesting that f_i is mostly independent of the global galactic environment and reflects only local properties of individual star-forming regions.

Instead, we find that the initial bound fraction varies strongly with $\epsilon_{\rm ff}$. At $M = 10^5\,M_\odot$, f_i can reach above 50% for $\epsilon_{\rm ff} \geq 0.5$, but is limited to only $1-10\%$ for runs with $\epsilon_{\rm ff} < 0.5$. This difference in normalization is due to the corresponding scaling of the integral star formation efficiency, because of our assumption $f_i \propto \epsilon_{\rm int}$ (Equation 3.8).

The dependence of $\epsilon_{\rm int}$ on particle mass has also been noticed in recent GMC-scale simulations. For example, Grudić et al. (2017, hereafter, G17) ran a series of MHD simulations of isolated turbulent clouds with different initial gas surface density. They found a tight relation between $\epsilon_{\rm int}$ and $\Sigma_{\rm gas}$ and parametrized it by the following relation:

$$\epsilon_{\rm int} = \left(\frac{1}{\epsilon_{\rm max}} + \frac{\Sigma_{\rm crit}}{\Sigma_{\rm gas}}\right)^{-1}, \tag{3.14}$$

with best-fit parameters $\epsilon_{\rm max} = 0.77$ and $\Sigma_{\rm crit} = 2800\,M_\odot\,{\rm pc}^{-2}$. The critical surface density $\Sigma_{\rm crit}$ above which the efficiency rises significantly corresponds to the cloud mass $M_{\rm crit} \approx 2.2\times10^5\,M_\odot\,(R/5\,{\rm pc})^2$. Using our definition of the initial bound fraction, the above equation can be re-written as a relationship between f_i and scaled particle mass $y \equiv M/M_{\rm crit}$:

$$f_i = \min\left(\frac{\sqrt{y^2 + 4y\epsilon_{\rm max}} - y}{\epsilon_{\rm max}}, 1\right). \tag{3.15}$$

We emphasize that the radius R used in G17 is not the same as the radius of our

GMC sphere, $R_{\rm GMC}$. In their simulations of isolated clouds, both the mass and size of the cloud are fixed by the initial condition, while our star-forming GMC has an open boundary through which gas flows in and out, controlled by gravity and stellar feedback. Our $R_{\rm GMC}$ is a lower limit to R since the cluster can accrete more distant gas, e.g. from all 27 neighbor cells. We can estimate a radius of the corresponding accretion region by taking it to be a sphere of the same volume as the 27 gas cells. Since the physical size of our cells at the finest refinement level varies between 3 and 6 pc (see Section 3.2.2.1), we take the average length to be with 4.5 pc. This gives the effective radius $R \approx 8.4$ pc. We use this value to compare the G17 scaling with ours.

The right panel of Figure 3.7 shows a good agreement between the relation given by Equation (3.15) and our runs with $\epsilon_{\rm ff} \geq 1$. The lower efficiency runs cannot reach the relatively high values of f_i predicted by G17. We fit the relation between particle mass and initial bound fraction as $f_i \propto M^a$, and find the power-law slope for our clusters in the range $a = 0.43 - 0.51$. This is also consistent with the G17 result $a \approx 0.5$ for clusters in the mass range $M < 10^5\,M_\odot$.

The initial bound fraction in our models depends somewhat on the SNR boosting factor $f_{\rm boost}$. The run with $f_{\rm boost} = 3$ shows systematically higher f_i than the corresponding $f_{\rm boost} = 5$ run, although the difference in the bound fraction is only $\sim 20\%$. The dependence of f_i on $f_{\rm boost}$ can be understood as the balance between gravity and feedback: with lower $f_{\rm boost}$, a cluster in one GMC of a given mass requires more stellar mass to reach the same amount of momentum of SNe to fully disperse the gas. Therefore a cluster in SFE50-SNR3 run reaches higher final mass for the same GMC mass, which implies higher values of $\epsilon_{\rm int}$ and f_i.

There are several consequences of introducing the initial bound fraction to determine the initial cluster mass. First, due to the mass-dependence of f_i, the shape and normalization of the CIMF is different from the particle mass function. Second, the

sensitivity of f_i to the choice of $\epsilon_{\rm ff}$ leads to a different integrated cluster formation efficiency and maximum cluster mass for runs with different $\epsilon_{\rm ff}$. These observables can be used to constrain $\epsilon_{\rm ff}$, as we discuss below.

3.3.4 Cluster initial mass function

CIMF is one of the key properties of young star clusters. Here we examine the CIMF of model clusters in the main halo, summed over the central galaxy and satellite galaxies. In Paper I we showed that the Schechter function provides a good description of the shape of the cluster mass function. Here we also fit all CIMFs with a Schechter function using the maximum likelihood method. Because the shape of the CIMF for small clusters now deviates more from a single power law, we restrict the fit only to clusters more massive than a certain minimum mass, above which the CIMF is best described by the Schechter form. The best-fit slopes, the cutoff masses, as well as the choice of the minimum mass at different epochs for all runs are listed in Table 3.3.

In Figure 3.8, we show the combined CIMF of all clusters formed at all times up to the last available snapshot of each run. Here we distinguish between the stellar particle mass (M) and the cluster mass ($f_i\,M$) that takes into account the initial bound fraction. Because f_i itself depends positively on particle mass, the power-law slope of the CIMF is in general shallower than that of the particle mass function. Among different runs, we find that the high-mass end of the CIMF is also strongly affected by the initial bound fraction. Simulations with lower $\epsilon_{\rm ff}$ tend to have lower f_i for a given cluster mass, thus modifying the mass function more strongly at all masses. We notice that the CIMFs here extend only to $\sim 10^6\,M_\odot$ for $\epsilon_{\rm ff} \geq 0.5$ runs rather than $\sim 10^7\,M_\odot$ in Paper I. This is mainly due to the lower star formation rates caused by stronger feedback implementation in the current simulations. The maximum mass is even as small as $\sim 10^5\,M_\odot$ for SFEturb and SFE10 runs because of the small value of f_i.

Table 3.3: CIMF best-fit parameters

Runs	$M_{\rm min}/\,M_\odot$	α	$M_{\rm cut}/\,M_\odot$
	full-particle		
SFE10	6×10^4	3.44	1.28×10^5
SFE50	6×10^4	3.27	1.63×10^5
SFE50-SNR3	4×10^4	3.67	4.50×10^5
SFE100	4×10^4	3.06	4.52×10^5
SFE200	4×10^4	3.55	NA[a]
	full-cluster		
SFE10	5×10^3	2.58	3.61×10^4
SFE50	2×10^4	2.47	3.08×10^5
SFE50-SNR3	1×10^4	2.86	1.31×10^5
SFE100	2×10^4	2.54	2.25×10^5
SFE200	4×10^4	3.27	NA[a]
	$z\approx2.0$-cluster		
SFE10	3×10^3	2.64	2.44×10^4
SFE50	8×10^3	2.18	3.99×10^4
SFE50-SNR3	1×10^4	1.99	1.25×10^4
SFE100	1×10^4	1.68	3.52×10^4
SFE200	6×10^3	1.15	2.54×10^4
	$z\approx5.3$-cluster		
SFE10	2.5×10^3	1.13	1.61×10^4
SFE50	6×10^3	1.28	1.30×10^5
SFE50-SNR3	6×10^3	1.81	1.73×10^5
SFE100	5×10^3	1.44	2.49×10^5
SFE200	3×10^4	1.51	8.45×10^5

a. the cutoff mass of the mass function of all clusters in SFE200 run cannot be obtained since it can be described by a pure power-law with no apparent exponential cutoff at high mass end.

To compare more directly with observations of young star clusters formed in the same star formation episode, in Figures 3.9 and 3.10 we show the CIMF of clusters younger than 100 Myr at $z \approx 2$ and $z \approx 5.3$. We choose an epoch around $z \approx 2$, when the main galaxy has not experienced any major mergers for more than 500 Myr. The CIMF of young clusters reflects the properties of the ISM that gradually grows its mass. We find a wide range of the power-law slopes for different runs, from $\alpha \approx 2.6$ for SFE10 run to $\alpha \approx 1.2$ for SFE200 run. There exists a systematic trend that higher $\epsilon_{\rm ff}$ leads to a shallower CIMF slope. Shallower slope means that, for a given galactic SFR, clusters with higher mass are more likely to be created. That is the reason why the runs with higher $\epsilon_{\rm ff}$ tend to have higher mass tails in the CIMF.

Figure 3.10 shows the CIMF of young clusters around $z \approx 5.3$, when the main galaxy experiences a major merger. The overall normalization of the mass function is lower than that at $z \approx 2$, because the galaxy at $z \approx 5.3$ is much smaller and contains less gas. However, the maximum cluster mass at this epoch is roughly one-order-of-magnitude higher than in the corresponding runs at $z \approx 2$. This is because the power-law slopes of the CIMF at this epoch, 1.1-1.8, is much shallower than the non-merger case at $z \approx 2$. The cutoff masses of these CIMF are all above $10^5\,M_\odot$, which is several times larger than these at $z \approx 2$. We already saw such enhancement of the formation of massive clusters in gas-rich galaxy mergers in the previous runs presented in Paper I. This result is even more pronounced in the new runs with stronger feedback.

3.3.5 Fraction of clustered star formation

In Paper I, we concluded that cluster formation is an environmentally-dependent process where the high-mass end of the CIMF depends strongly on the star formation activity of the host galaxy. Both recent observations (Adamo et al., 2015; Johnson et al., 2016, 2017a) and theoretical work (Kruijssen, 2015) suggest that the SFR

surface density ($\Sigma_{\rm SFR}$), rather than SFR itself, better represents the intensity of star formation and physical conditions of the galactic ISM. Therefore, here we explore the effects of $\Sigma_{\rm SFR}$ on such cluster properties as the fraction of clustered star formation, Γ, and maximum cluster mass, $M_{\rm max}$.

Analogously to constructing the KSR in Section 3.3.2, we split the disk of the main galaxy into the 1×1 kpc square grid. For each square area, we calculate the surface density of star formation rate averaged over 20 Myr. The fraction of clustered star formation is defined as the ratio between the mass in bound clusters and total stellar mass formed within the same time interval, 20 Myr:

$$\Gamma \equiv \frac{\sum f_i M}{\sum M}. \tag{3.16}$$

We also test larger averaging timescale from 20 Myr to 50 Myr and find that the results are not sensitive to the choice of the timescale.

This Γ is different from the quantity we used in Paper I in two ways. First, to estimate Γ we split the galactic disk into concentric circular bins in Paper I. However, due to the stronger feedback implemented in this paper, no well-defined gas disk presents in the main galaxy. Instead, the disks are clumpy and asymmetric, often showing features of strong outflows. Therefore, it is not possible to find a definitive center and cylindrical-symmetric axis to construct circular bins. Second, in Paper I, we defined Γ simply as the fraction of clusters more massive than $10^4\,M_\odot$, not all of which are necessarily gravitationally bound. In this paper, with the introduction of the initial bound fraction, we have a more physical way of estimating the clustered fraction as it is defined in observations.

Figure 3.11 shows a positive correlation between $\Sigma_{\rm SFR}$ and Γ for all the runs. We find that Γ changes by about one order of magnitude over the range $\Sigma_{\rm SFR} = 10^{-3} - 1\,M_\odot\,{\rm yr}^{-1}\,{\rm kpc}^{-2}$. At $\Sigma_{\rm SFR} > 1\,M_\odot\,{\rm yr}^{-1}$, the values of Γ stop increasing and

saturate to the value that is set by the maximum value of initial bound fraction and the mass fraction of the most massive clusters. This saturation of Γ at high $\Sigma_{\rm SFR}$ is also found in the analytical model of cluster formation efficiency in Kruijssen et al. (2012). It should be noted that, the physical origin of this saturation for our and their models is different. In Kruijssen et al. (2012), the saturation is caused by the "cruel cradle effect", in which young clusters formed within less dense regions are destroyed by the tidal interaction with other star forming regions during their embedded phase, while, in our model, the saturation is caused by the mass-dependent initial bound fraction.

It is clear that Γ also depends strongly on $\epsilon_{\rm ff}$. Run SFE200 with $\epsilon_{\rm ff} = 2.0$ shows a very high Γ from 0.05 to 0.6, while runs SFE10 and SFEturb do not have any star-forming regions with $\Gamma > 0.1$. This sensitivity to the choice of $\epsilon_{\rm ff}$ makes Γ an excellent observable to constrain this parameter.

3.3.6 Maximum cluster mass

Figure 3.12 shows the relationship between $\Sigma_{\rm SFR}$ and the maximum mass of clusters formed within a 100 Myr interval. We find a clear positive correlation, going roughly as

$$M_{\rm max} \propto \Sigma_{\rm SFR}^{2/3}.$$

The power-law slope is similar to the best-fit value, ≈ 0.7, for the observations of maximum cluster mass described in Section 3.2.4.2. The similarity of the slope among all runs reveals a robust result that the high-mass end of the mass function of model clusters depends uniquely on the intensity of the star formation activity. The normalization of this relation also scales monotonically with $\epsilon_{\rm ff}$, because of the $\epsilon_{\rm ff}-$dependence of the initial bound fraction.

Note that the runs with $\epsilon_{\rm ff} \leq 0.1$ cannot produce clusters more massive than $10^5\,M_\odot$ at all considered epochs for $z > 1.5$. This indicates that no clusters in these

Table 3.4: Cluster formation timescale, $\tau_{\rm ave}$, in Myr

Runs	low mass 10%-50%-90%	high mass 10%-50%-90%
SFEturb	0.25-1.07-3.13	0.43-2.07-8.62
SFE10	0.14-0.57-1.83	0.96-2.12-3,85
SFE50	0.07-0.26-0.84	0.52-1.44-2.53
SFE50-SFR3	0.09-0.26-0.80	0.45-1.26-2.35
SFE100	0.06-0.19-0.60	0.32-1.43-2.42
SFE200	0.04-0.15-0.46	0.55-1.59-2.29

runs would survive to the present time to become globular clusters. It is another evidence against such low values of $\epsilon_{\rm ff}$.

3.3.7 Cluster formation timescale

The upper panel of Figure 3.13 shows the cumulative distributions of cluster formation timescales of $\tau_{\rm ave}$. we find a clear trend that the higher $\epsilon_{\rm ff}$, the shorter the timescale, especially for low mass clusters ($< 10^5\,M_\odot$). This means that the growth history of these clusters is dominated by the gas accretion, which is in turn controlled by $\epsilon_{\rm ff}$. This result, obtained after many algorithmic updates described in Section 3.2, is still consistent with that in Paper I.

More quantitatively, we split the whole cluster sample into low ($< 10^5\,M_\odot$) and high mass clusters ($> 10^5\,M_\odot$) and calculate the 10%, 50%, and 90% percentiles of the distribution of $\tau_{\rm ave}$. The results for both low and high mass clusters are listed in Table 3.4.

In general, the timescales for low mass clusters is very short. The median values of $\tau_{\rm ave}$ are all smaller than ~ 1 Myr. To better understand the changes across different runs, instead of just comparing the median values we quantitatively evaluate the differences of the probability density distributions. For instance, for samples s_1 and s_2, we find the factor $f_{\rm best}$ iteratively to maximize the p-value of the Kolmogorov-Smirnov

test between s_1 and $f_{\rm best}\, s_2$. This method can obtain $f_{\rm best}$ so that the cumulative distribution of s_1 and $f_{\rm best}\, s_2$ are most similar to each other. We find that $\tau_{\rm ave}$ in SFE200, SFE100, and SFE50 runs are shifted a factor of 0.27, 0.34, and 0.45 times relative to the values in SFE10 run. This suggests a strong anti-correlation between $\epsilon_{\rm ff}$ and $\tau_{\rm ave}$, with a scaling relation that can be best described as $\tau_{\rm ave} \propto \epsilon_{\rm ff}^{-0.45}$. This is interesting because, although the instantaneous gas accretion rate of clusters is linearly correlated with $\epsilon_{\rm ff}$ in Equation (3.1), the correlation between $\epsilon_{\rm ff}$ and $\tau_{\rm ave}$ is nonlinear.

On the other hand, the comparison of the distributions of $\tau_{\rm ave}$ for SFE50 and SFE50-SNR3 runs shows that the boosting factor of SNe feedback $f_{\rm boost}$ has a negligible effect. This is expected since $f_{\rm boost}$ only controls the intensity of the momentum feedback from SNe that explode after 3 Myr, which is longer than the typical formation timescale of clusters in this mass range.

For massive clusters ($M > 10^5\, M_\odot$), however, the distributions of cluster formation timescale are very similar for runs with $\epsilon_{\rm ff} \geq 0.5$. They all have a median timescale around 1.5 Myr and truncate at $\tau_{\rm ave} \approx 3$ Myr, the time when the first SNe explode. This suggests that the growth history of massive clusters is determined by both gas accretion and stellar feedback. However, an immediate termination of gas accretion by SNe only happens with high $\epsilon_{\rm ff}$. Slower star formation cannot accumulate enough massive stars, and therefore enough SNe within the first 3 Myr to disperse the natal GMCs. This can be seen from the long tail of $\tau_{\rm ave}$ in SFE10 and SFEturb runs. Although the median value is only $\approx 2-3$ Myr, there are many clusters with $\tau_{\rm ave} >$ 3 Myr in SFE10 run and $\tau_{\rm ave} > 5$ Myr in SFEturb run. Such long formation timescales are inconsistent with the observed age spread of young star clusters in nearby star-forming regions, as discussed in Section 3.2.4.1.

The same as $\tau_{\rm ave}$, another definition of the formation timescale, $\tau_{\rm spread}$, shows a similar positive correlation to $\epsilon_{\rm ff}$. However, the median values of $\tau_{\rm spread}$ for different

runs are even smaller than that of $\tau_{\rm ave}$. The median values range from 0.1 to 0.3 Myr for low mass clusters and from 0.5 to 2.5 Myr for high mass ones.

3.4 Discussion

Now we discuss the constraints on $\epsilon_{\rm ff}$ and $f_{\rm boost}$ resulting from the tests described above.

3.4.1 Global properties: Star Formation Rate and Kennicutt-Schmidt relation

In Section 3.3.1 we found that the SFH of the main galaxy in runs with $f_{\rm boost} = 5$ is consistent with the abundance matching results. This SFH is very sensitive to the choice of $f_{\rm boost}$, which gives a tight constraint on its value: $3 < f_{\rm boost} < 10$. On the other hand, the SFH depends little on the value of $\epsilon_{\rm ff}$ over the whole two orders of magnitude range.

As long as the feedback energy is strong enough to trigger these star formation–feedback cycles, the global properties of galaxies and the inefficient star formation activity can be reproduced, and the result is not sensitive to the detailed implementation. This phenomenon was also found in recent high-resolution cosmological simulations by Hopkins et al. (2014).

In contrast to good match of the SFH, the normalization of the KS relation in all of our simulations is overestimated by a factor of 2 to 20. In the analysis, we used 20 Myr averaging timescale to calculate $\Sigma_{\rm SFR}$. We chose this short timescale because we found that the structure of the ISM in our simulations is vulnerable to the feedback from young clusters, such that even the kpc-scale gas distribution changes very rapidly. Also, since we used the high density threshold for cluster formation, $n_{\rm th} = 10^3\,{\rm cm}^{-3}$, the star formation regions are concentrated into the densest regions. Therefore, the 1×1 kpc grid used to derive the KS relation sometimes contains

only a couple of star-forming complexes. This situation is very different from the observations of low-redshift galaxies, where comparable areas contain many small star-forming regions and both $\Sigma_{\rm H_2}$ and $\Sigma_{\rm SFR}$ are averaged over different regions in different phases of star formation.

Analogously to the instantaneous star formation rate in Equation (3.1), higher $\epsilon_{\rm ff}$ runs tend to have higher $\Sigma_{\rm SFR}$ for a given $\Sigma_{\rm H_2}$. However, the slope of the molecular KS relation is nearly linear and consistent with the observed linear relation, despite the non-linear scaling $\dot{\rho}_* \propto \rho_{\rm gas}^{3/2}$ used in Equation (3.1).

3.4.2 Slope of the star cluster mass function

In Section 3.3.4, we showed that the power-law slope of the CIMF is sensitive to the choice of $\epsilon_{\rm ff}$. Observations of CIMF in nearby galaxies reveal a Schechter-like function with a power-law slope close to -2 (Portegies Zwart et al., 2010). SFE10 run shows a power-law slope as steep as ≈ 2.5, while SFE200 run has a very shallow slope smaller than 1.5. Therefore, under the current simulation setup, we find that the range $\epsilon_{\rm ff} = 0.5 - 1.0$ matches the observations best.

The normalization of the CIMF also depends strongly on $\epsilon_{\rm ff}$. Because of the positive correlation between $\epsilon_{\rm ff}$ and the initial bound fraction, runs with lower $\epsilon_{\rm ff}$ create clusters that have lower f_i for a given particle mass. Therefore, applying this initial bound fraction shifts the particle mass function to smaller masses and leads to the apparent lower normalization of the CIMF for lower $\epsilon_{\rm ff}$ runs.

3.4.3 Effects of major mergers

In Section 3.3.4, we find that there is a significant difference of CIMFs produced during major mergers and during quiescent periods. In a merger, the slope of the CIMF becomes shallower while the truncation mass rises much higher. This promotes the formation of most massive clusters, $M > 3 \times 10^5\,M_\odot$. This is similar to the result

we found in Paper I, despite the many updates to the algorithm in this paper.

The existence of more than 100 GCs at present gives additional constraint on the creation of most massive clusters at high-z. As can be seen in Figure 3.8 and 3.12, potential GC progenitors, clusters with mass larger than $\sim 3 \times 10^5\,M_\odot$, can only be created by runs with $\epsilon_{\rm ff} \geq 0.5$.

3.4.4 Fraction of clustered star formation

In Paper I, we presented a positive correlation between $\Sigma_{\rm SFR}$ and the fraction of young clusters with mass above $10^4\,M_\odot$. Roughly speaking, this fraction is a reasonable proxy to the fraction of clustered star formation that we estimated in Section 3.3.3. This is because massive clusters tend to have higher f_i, and therefore, most of the bound cluster mass is contributed by the most massive clusters. Since the shape of the cluster particle mass function is not affected strongly by the value of $\epsilon_{\rm ff}$ in Paper I, the fraction of massive clusters is not sensitive to the choice of $\epsilon_{\rm ff}$. However, in this paper, the strong correlation between $\epsilon_{\rm ff}$ and the newly-implemented initial bound fraction f_i leads to a strong positive correlation between $\epsilon_{\rm ff}$ and the integrated cluster fraction, Γ. Thus, Γ as a function of $\Sigma_{\rm SFR}$ provides us a new diagnostic to constrain $\epsilon_{\rm ff}$.

The positive correlation between $\Sigma_{\rm SFR}$ and Γ found in Section 3.3.5 is caused by two effects. First, in high $\Sigma_{\rm SFR}$ areas the CIMF is likely to extend to higher masses, as demonstrated in Figure 3.12. Second, high-mass clusters typically have larger initial bound fraction, which leads to higher efficiency Γ. Therefore, this correlation comes about from a combination of the variation of CIMF in different environments and the mass-dependent initial bound fraction.

We find that the simulations with $\epsilon_{\rm ff} = 0.5 - 1.0$ reproduce the observed values of Γ over a wide range of $\Sigma_{\rm SFR} = 10^{-3} - 1\,M_\odot\,{\rm yr}^{-1}\,{\rm kpc}^{-2}$. However, there are three data points from Adamo et al. (2015) that show very high $\Gamma \sim 0.5$ at $\Sigma_{\rm SFR} \sim$

$1\,M_\odot\,\mathrm{yr}^{-1}\,\mathrm{kpc}^{-2}$, which cannot be reached by either SFE50 or SFE100 run. It is important to keep in mind that these three data points are from galaxy-integrated rather than spatially-resolved measurements, as we did in this paper. Note also that the data compilation shown in Figure 3.11 combines measurements with different SFR tracers, different cluster identification criteria, and different averaging spatial and time scales. Future observations with consistent methodology and higher spatial resolution are required to place better constrains on the star formation models.

3.4.5 Maximum cluster mass

In Section 3.3.6, we find a positive correlation between the SFR surface density and maximum mass of young clusters. The normalization of this relation scales with $\epsilon_{\rm ff}$. We find that runs with $\epsilon_{\rm ff} \geq 0.5$ are consistent with the observations of clusters in nearby galaxies over a large range of $\Sigma_{\rm SFR} = 10^{-4} - 2\,M_\odot\,\mathrm{yr}^{-1}\,\mathrm{kpc}^{-2}$. Due to the small observation samples, it is hard to constrain $\epsilon_{\rm ff}$ into narrower range, but at least $\epsilon_{\rm ff} \leq 0.1$ is safely ruled out.

3.4.6 Mass accretion history inferred from different definition of cluster formation timescales

As we discussed in Section 3.2.2.10, different definitions of the cluster formation timescale are sensitive to different mass accretion behavior of individual clusters.

The first experiment is to check whether the mass accretion history can be described by a simple power-law, $\dot{M} \propto t^\alpha$, as suggested by Murray & Chang (2015). Such power-law mass accretion gives $\tau_{\rm ave} = \frac{\alpha+1}{\alpha+2}\tau_{\rm max}$ and $\tau_{\rm spread} = \frac{2\alpha+1}{(\alpha+1)^2}\tau_{\rm max}$. Therefore, if the mass accretion history of individual clusters can indeed be described by a power-law, the indexes α derived from $\tau_{\rm ave}/\tau_{\rm max}$ and $\tau_{\rm spread}/\tau_{\rm max}$ should be consistent with each other. However, we find that majority of the clusters do not show consistent α from these two definitions, suggesting that most clusters experience accretion

histories that cannot be described by a simple power-law.

The same as Paper I, we explore the detail output of the mass accretion rate at each local time steps for a fraction of massive clusters larger than $10^5\,M_\odot$. We find a general pattern that the accretion rate at the beginning of cluster formation is relatively small. As the density of the GMC becomes higher, the accretion rate increases until it reaches its peak by either exhausting the star-forming gas or stellar feedback. This pattern can definitely not be fitted by a single power-law, but it is more suitable to be described by a peaked distribution, such as a Gaussian shape. In this case, $\tau_{\rm ave}$ and $\tau_{\rm spread}$ actually reflect different characteristics of the mass accretion of individual clusters. $\tau_{\rm ave}$ indicates the timescale that a given cluster reaches the peak of its mass accretion, while $\tau_{\rm spread}$ quantifies the width of the peak.

3.4.7 Cluster formation timescale

The same as Paper I, the strong-dependence of cluster formation timescale on $\epsilon_{\rm ff}$ appears in the new runs. This robust relationship that can be used to constrain $\epsilon_{\rm ff}$. Unfortunately, since clusters in most runs have cluster formation timescale that is shorter than 3 Myr, they are all within the range of the observed age spread for young star cluster. One exception is the SFEturb run, in which $\sim 30\%$ of massive clusters have a timescale larger than 6 Myr, which is not consistent with observations.

3.4.8 Combination of all constraints

Here we summarize the constraints of star formation and feedback parameters, $f_{\rm boost}$ and $\epsilon_{\rm ff}$, from different observables:

- globular SFH: $f_{\rm boost} = 3 - 10$.
- slope of CIMF: $\epsilon_{\rm ff} = 0.5 - 1.0$.
- cluster formation efficiency: $\epsilon_{\rm ff} = 0.5 - 2.0$.

- maximum cluster mass: $\epsilon_{\rm ff} \geq 0.1$.

- cluster formation timescale: $\epsilon_{\rm ff} \geq 0.1$.

Based on the joint constraints, We find that $f_{\rm boost} = 3 - 10$ and $\epsilon_{\rm ff} = 0.5 - 1.0$ is favored in our model.

We emphasis again that $\epsilon_{\rm ff}$ used in this paper is different than that in other cosmological simulations: (1) $\epsilon_{\rm ff}$ only applies within the star-forming spheres of a fixed physical volume rather than time-varying cell volume. (2) although $\epsilon_{\rm ff}$ is a fixed value for most of the runs, the mass growth rate of a given clusters changes significantly during the active accretion period due to the changes of local gas density. This time-resolved accretion leads to a broad scatter of integrated star formation efficiency of individual clusters as is shown in Paper I.

3.5 Summary

We have described an improved implementation of star cluster formation in the cosmological code, ART. We update the cluster formation algorithm to eliminate the accretion gaps during the active period of clusters, which significantly reduce the long cluster formation timescales found in Paper I. We adopt a new SNR feedback model so that the SFH of the main galaxy follows the abundance matching result. We also introduce a new prescription that takes into account the initial bound fraction of clusters. We perform a series of cosmological simulations of a Milky Way-sized galaxy with different star formation and feedback parameters. Various properties of model clusters are used to constrain these parameters. The results from this new suite of simulations are summarized here:

- The global properties of the galaxies, such as the galaxy morphology and SFH, are strongly affected by the strength of feedback, but is not sensitive to star formation efficiency $\epsilon_{\rm ff}$.

- To match the SFH from the abundance matching result of a Milky Way-sized galaxy, the momentum boosting factor of SNR feedback is tightly constrained to be in the range $f_{\rm boost} = 3 - 10$.

- The initial bound fraction f_i increases with the cluster particle mass, irrespective to their formation epochs and host galaxy masses. f_i also positively correlates to the value of $\epsilon_{\rm ff}$.

- CIMF of model clusters presents a Schechter-like shape with slope that depends on $\epsilon_{\rm ff}$. We find that runs with $\epsilon_{\rm ff} = 0.5 - 1.0$ have a CIMF with a slope similar to -2, best matches the observed slopes.

- We find, during the major merger, the slope of the CIMF becomes shallower and the CIMF itself extends to higher masses, thus create more massive clusters that the epochs without merger.

- Integrated cluster formation efficiencies Γ are estimated based on the newly-implemented f_i within 1×1 kpc grids across the main galaxy disks. We find a positive correlation betewen $\Sigma_{\rm SFR}$ and Γ. The normalization of this correlation depends strongly on $\epsilon_{\rm ff}$, with $\epsilon_{\rm ff} = 0.5 - 1.0$ best matches the observational measurements.

- The same as Paper I, the cluster formation timescale depends strongly on $\epsilon_{\rm ff}$ for clusters less massive than $10^5\,M_\odot$. For massive clusters, the timescale truncates to 3 Myr, the time when first SNe explode in runs with $\epsilon_{\rm ff} \geq 0.5$. Runs with $\epsilon_{\rm ff} \leq 0.1$ show a long tail of longer timescales, therefore not favored by our model.

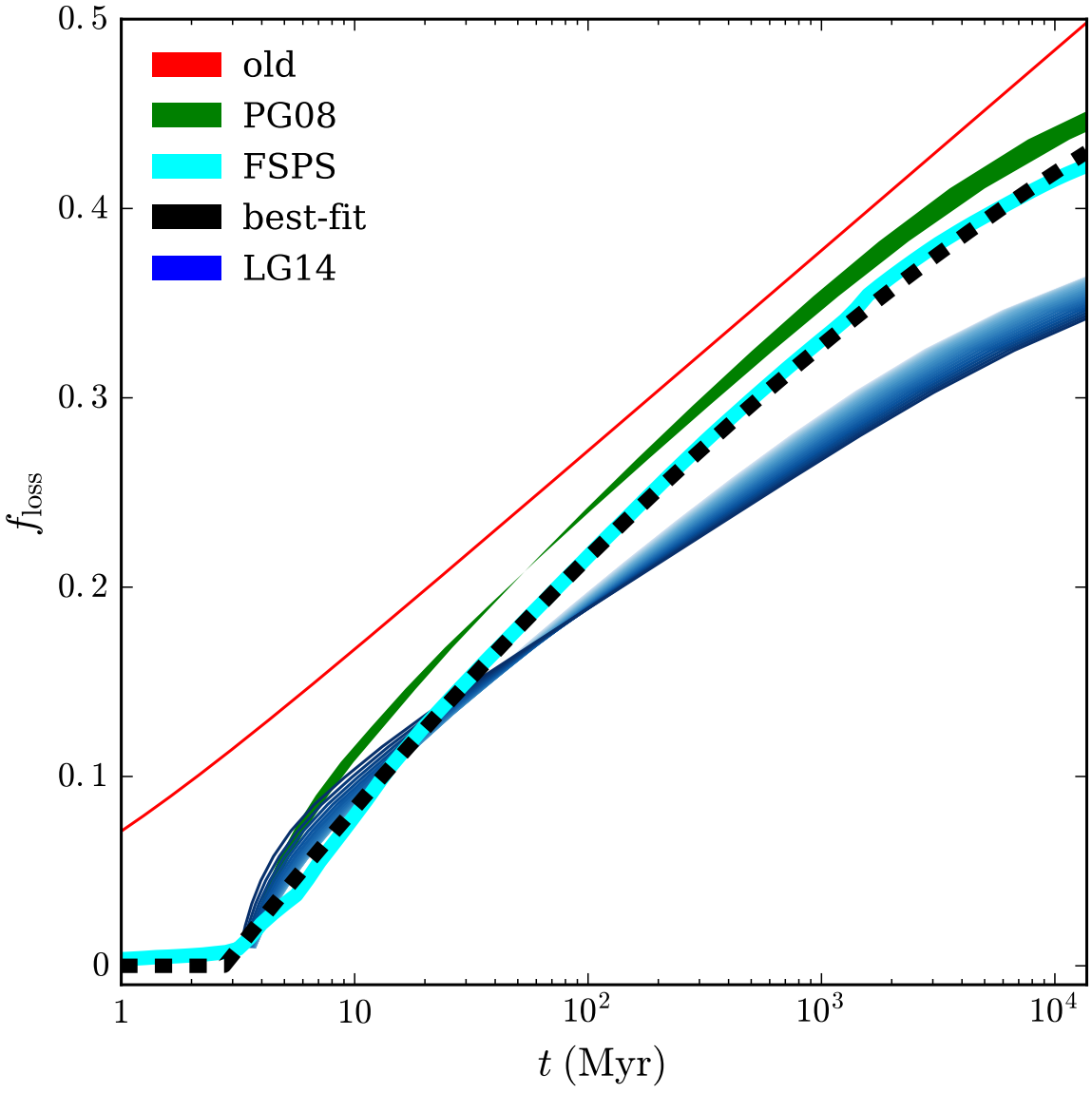

Figure 3.2: The cumulative fraction of mass loss due to stellar evolution for a single stellar population with Kroupa (2001) IMF: old prescription used in Paper I (red), Prieto & Gnedin (2008) (green), FSPS (cyan), and LG14 (blue). The line width for each source represents the mass loss from stars of different metallicity, from zero to solar. The best-fit expression to the average FSPS results (described by Equation 3.7) is overplotted by the thick dashed line.

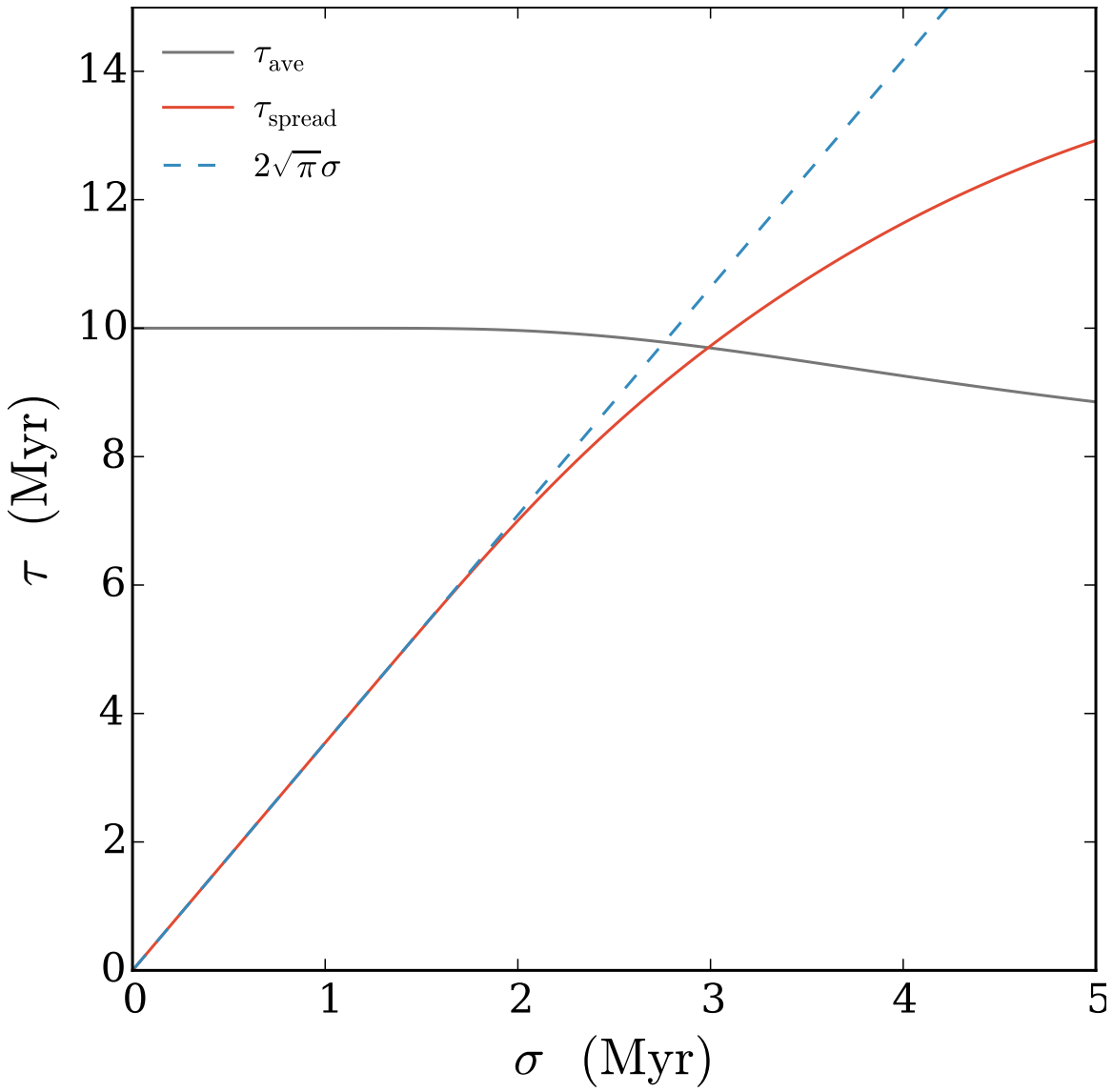

Figure 3.3: An example of the old ($\tau_{\rm ave}$, black) and new ($\tau_{\rm spread}$, red) definitions of cluster formation timescale as a function of Gaussian width of the SFR, peaked at a relatively late time $t_0 = 10$ Myr. The new definition closely follows the width of the accretion rate when σ is small compared to the total duration of the star formation episode.

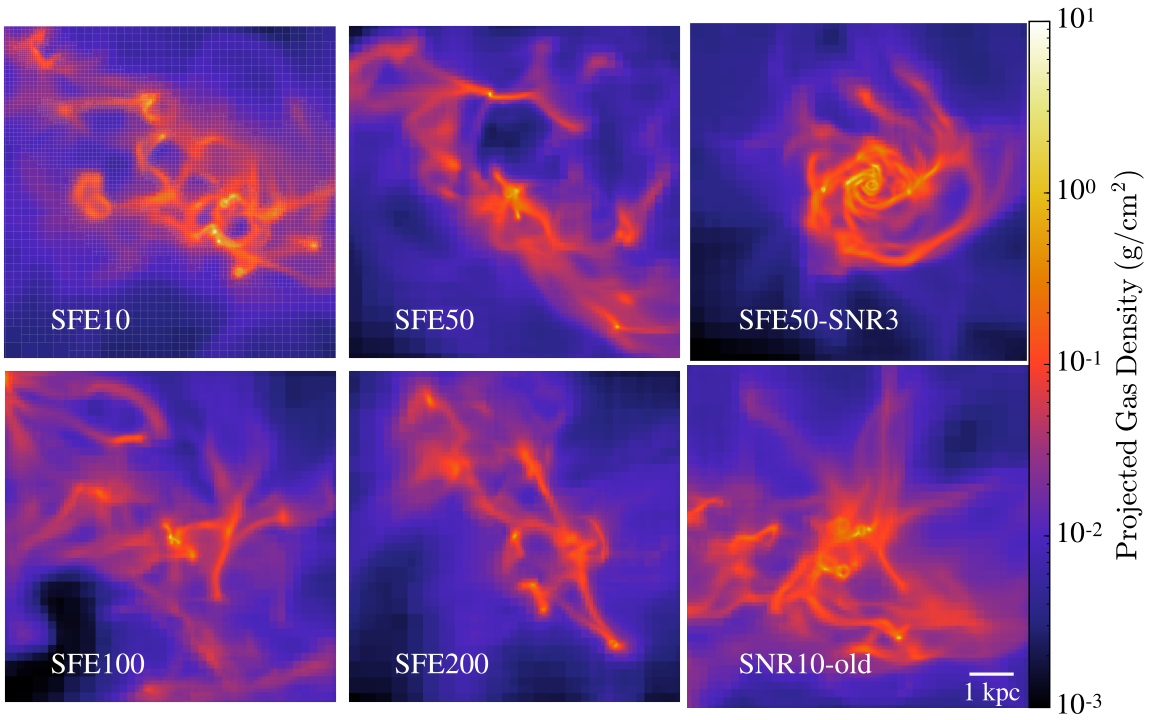

Figure 3.4: Gas density projection plots of the main galaxy at $z \approx 2$ for different runs. The adaptive refinement structure of the oct-tree code is shown in the upper left panel and the length scale of 1 kpc is shown in the lower right panel.

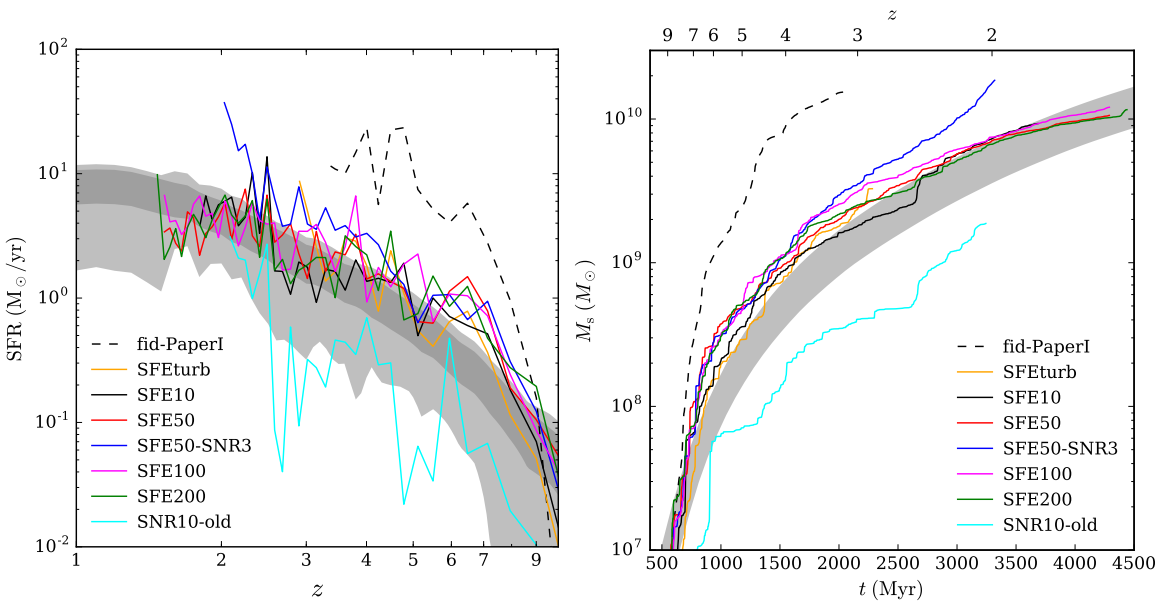

Figure 3.5: *Left:* Star formation history of the main galaxy for runs with different star formation and feedback parameters. The SFR is derived from all stellar particles within the main galaxy at the last snapshot, see Section 3.3.1 for details. Dark and light shaded areas are 1-σ and 2-σ confidence intervals of the expected SFR for an average $10^{12}\,M_\odot$ halo from abundance matching (Behroozi et al., 2013b). *Right:* Cumulative stellar mass history of the main galaxy. The shaded area is the 1-σ confidence interval from the abundance matching result.

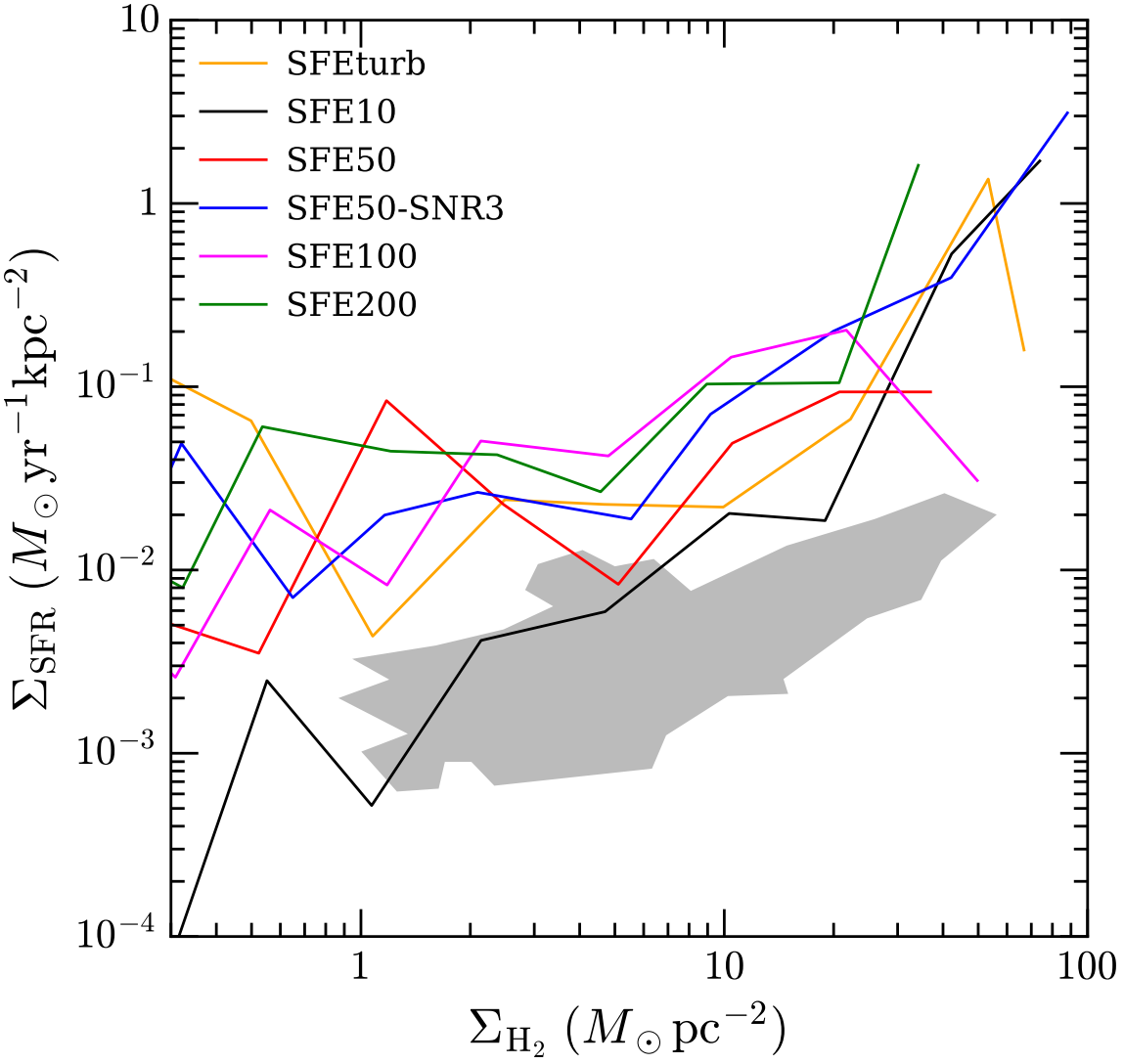

Figure 3.6: Kennicutt-Schmidt relation for the main galaxy. SFR is calculated over 20 Myr, spatial averaging is over 1 kpc squares. Gray regions shows the range of observed values in nearby spiral galaxies by (Bigiel et al., 2008, 2011).

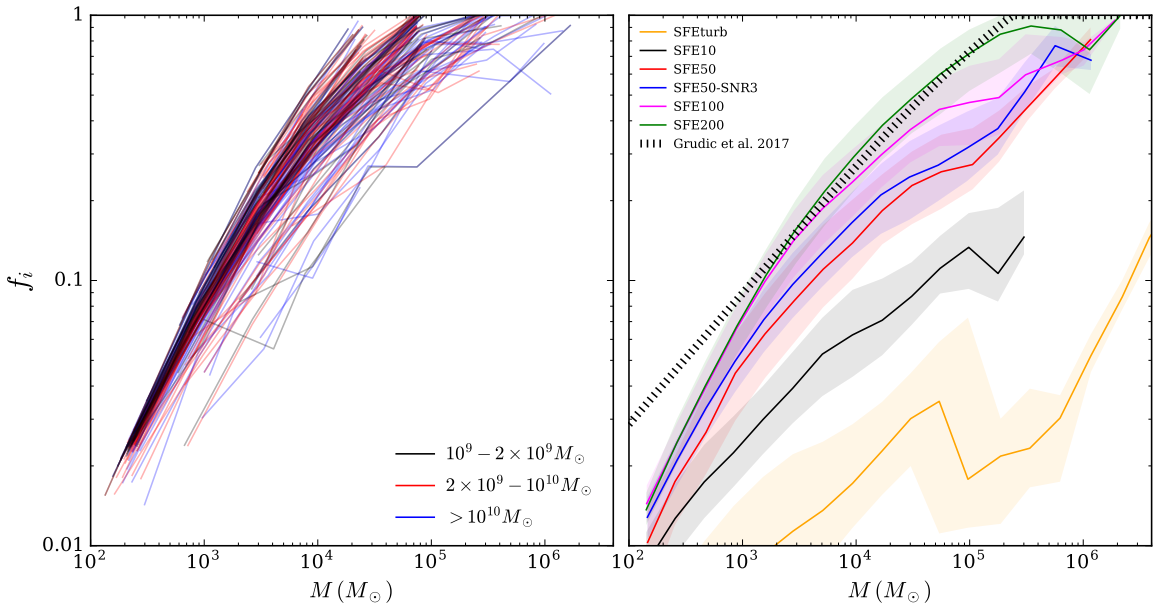

Figure 3.7: *Left:* Initial bound fraction vs. particle mass in SFE200 run. Clusters formed within galaxies of different halo mass are labeled by different color. *Right:* Same as left panel, but for runs with different star formation and feedback parameters. Striped line shows the relation calculated from MHD simulations of star formation in isolated GMCs by Grudić et al. (2017).

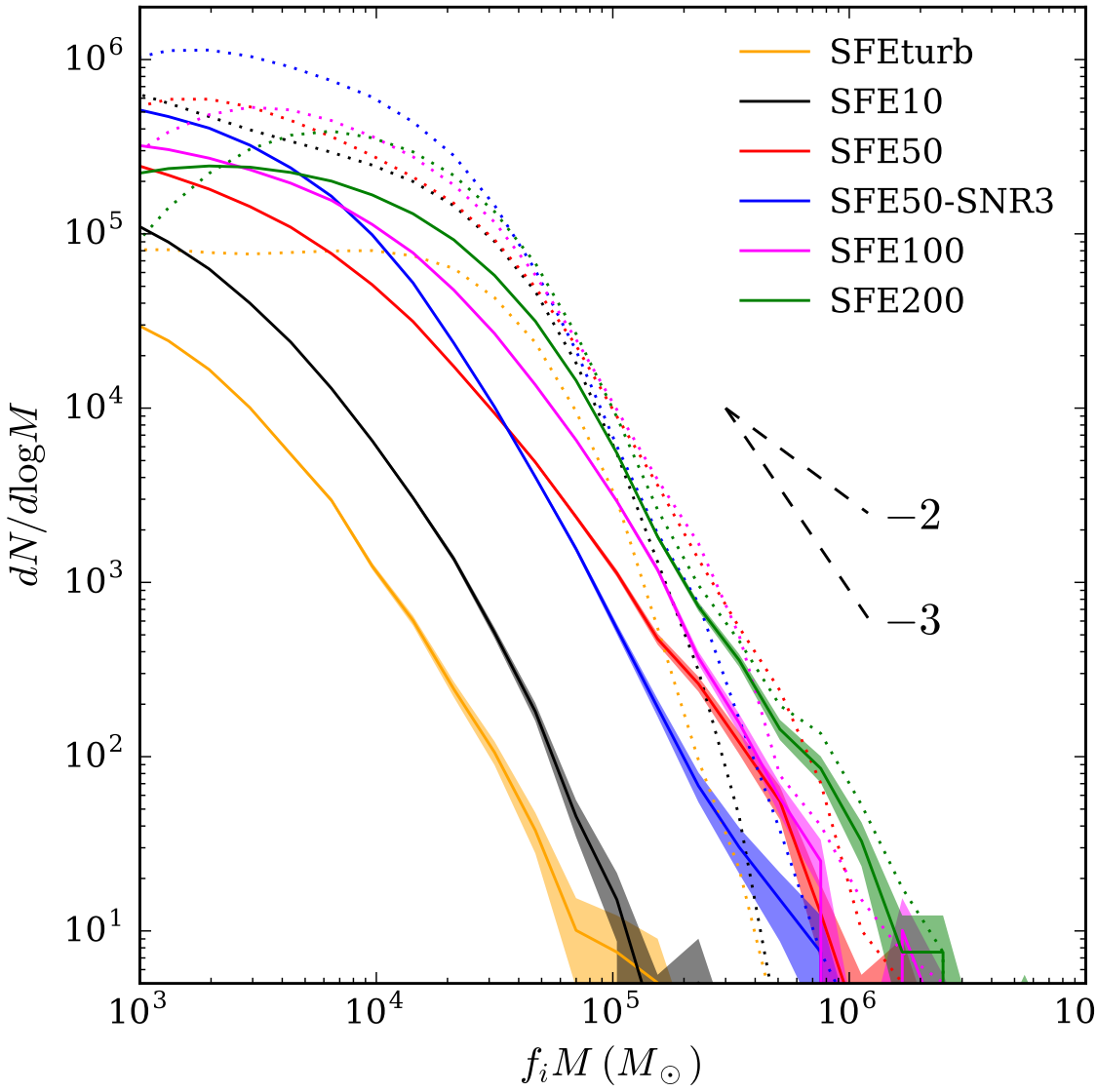

Figure 3.8: CIMF of all clusters ($f_i\,M$, solid lines) within the main galaxy from the last available snapshot of each run. Shaded areas show the binomial counting errors in mass bins of 0.16 dex. In contrast, dotted lines show the distribution of stellar particle mass (M), without considering the initial bound fraction. The power-law distributions with slope $\alpha = -2$ and -3 are overplotted as dashed lines for reference.

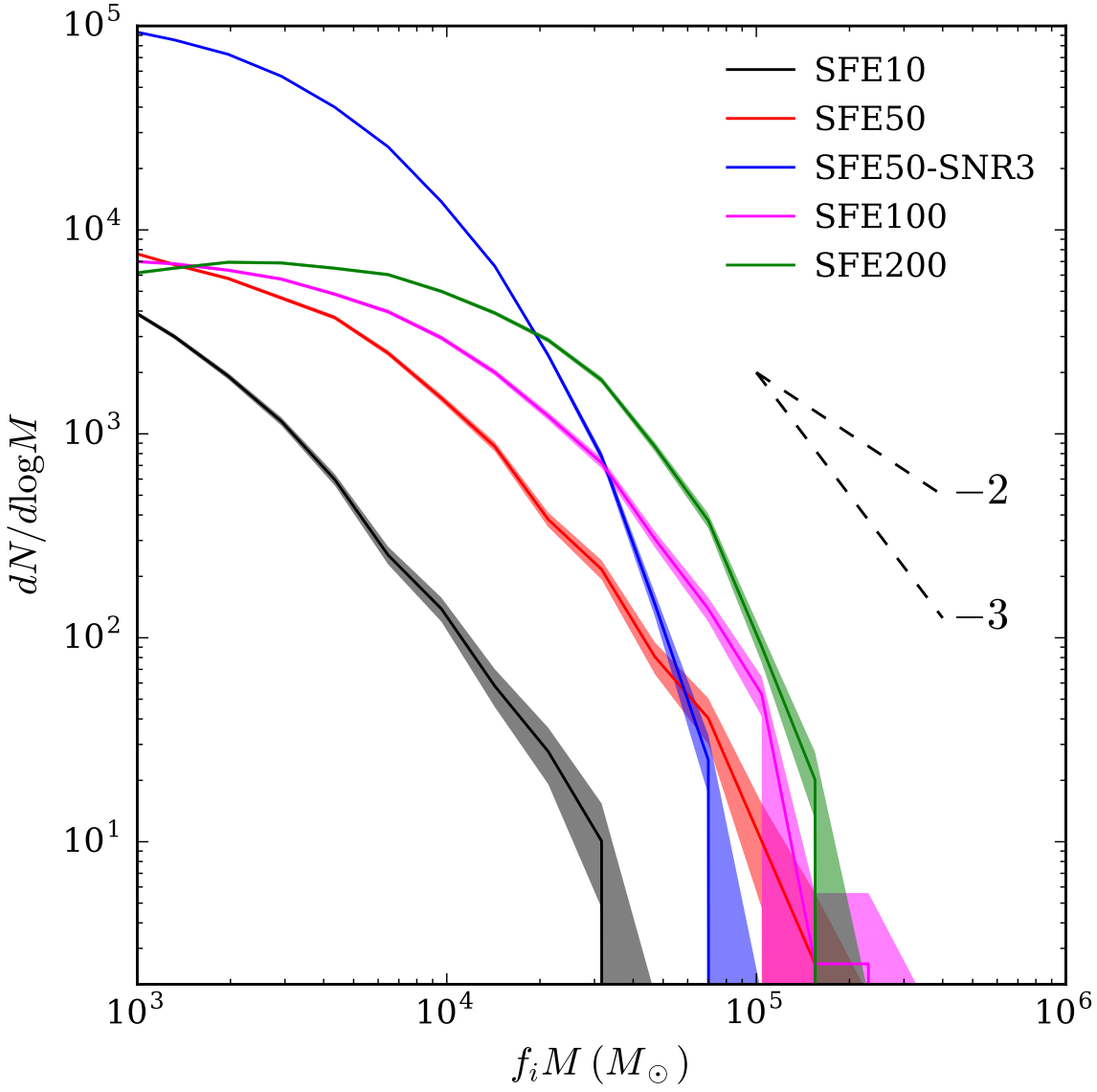

Figure 3.9: Same as Figure 3.8, but only for clusters younger than 100 Myr within the main galaxy at a quiescent epoch around $z \approx 2$.

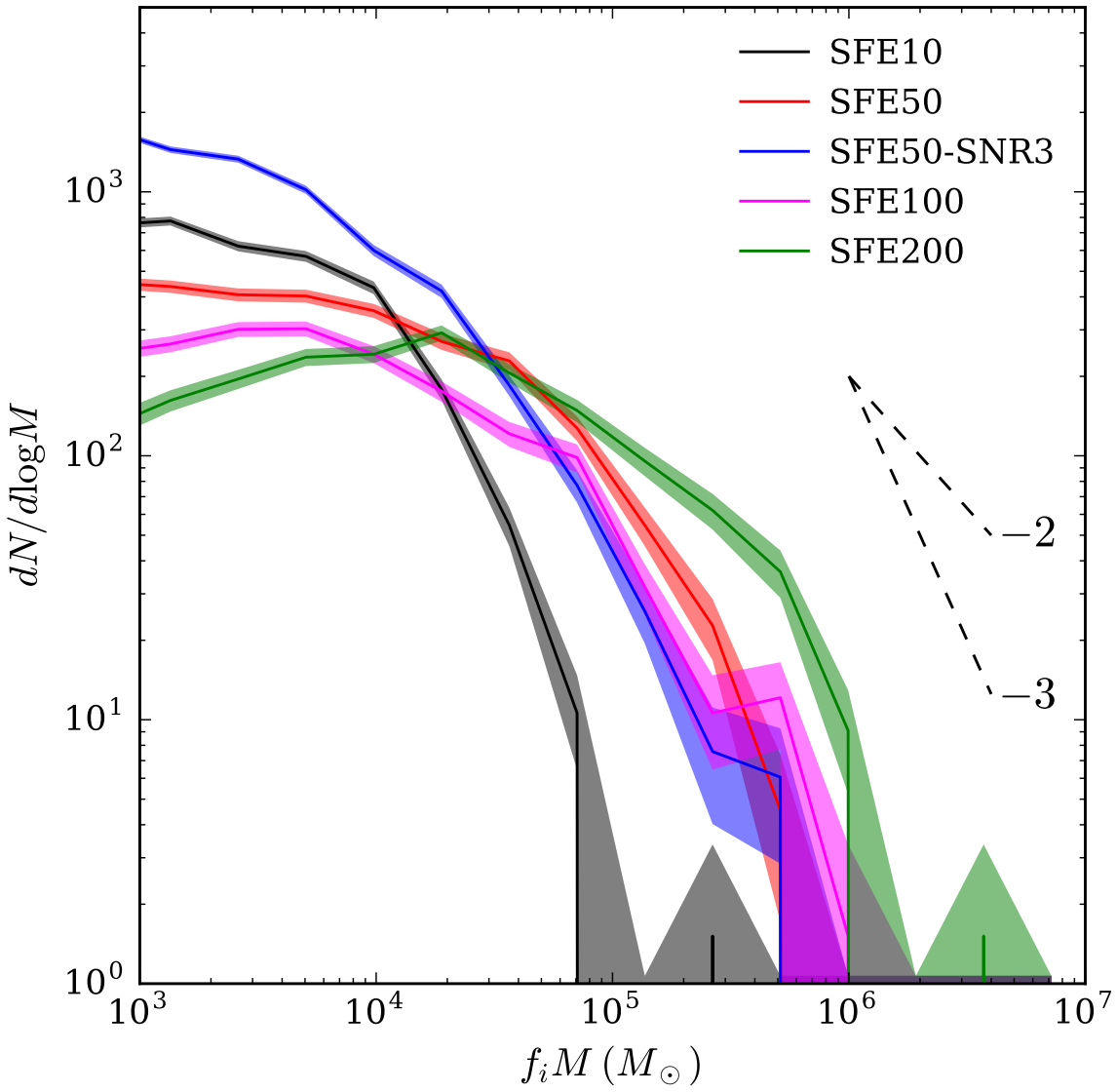

Figure 3.10: Same as Figure 3.9, but during a major merger epoch at $z \approx 5.3$.

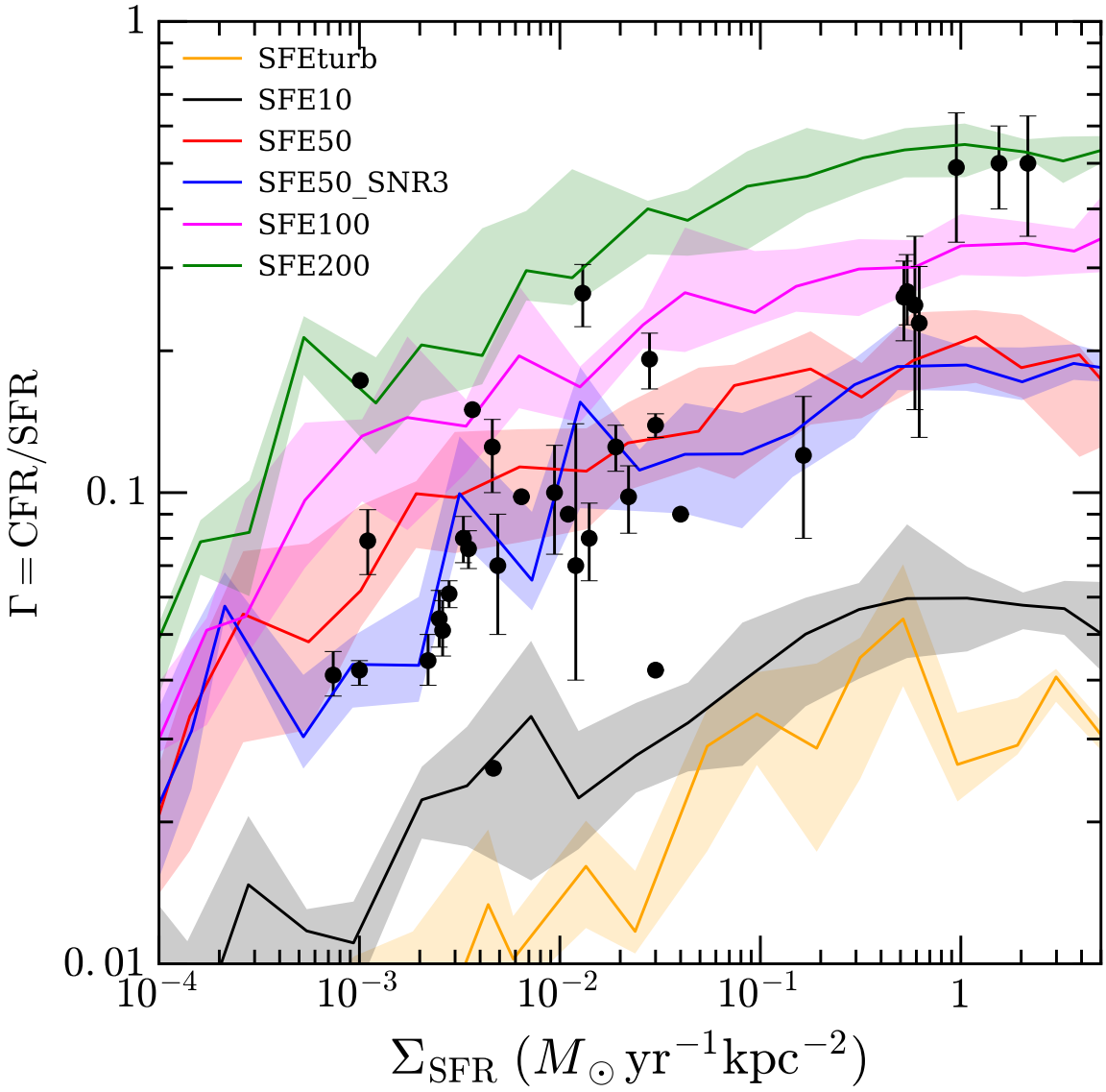

Figure 3.11: Fraction of clustered star formation as a function of star formation rate surface density. Σ_{SFR} is estimated on a spatial scale of 1 kpc for stars younger than 20 Myr. Solid lines and shaded areas show the median and 25-75% interquartile range of the distribution of Γ for a given Σ_{SFR} bin. The observed values (symbols with errorbars) are from a compilation of both galaxy-wide and spatially-resolved measurements of cluster samples in nearby galaxies (Goddard et al., 2010; Adamo et al., 2015; Silva-Villa & Larsen, 2011; Johnson et al., 2016).

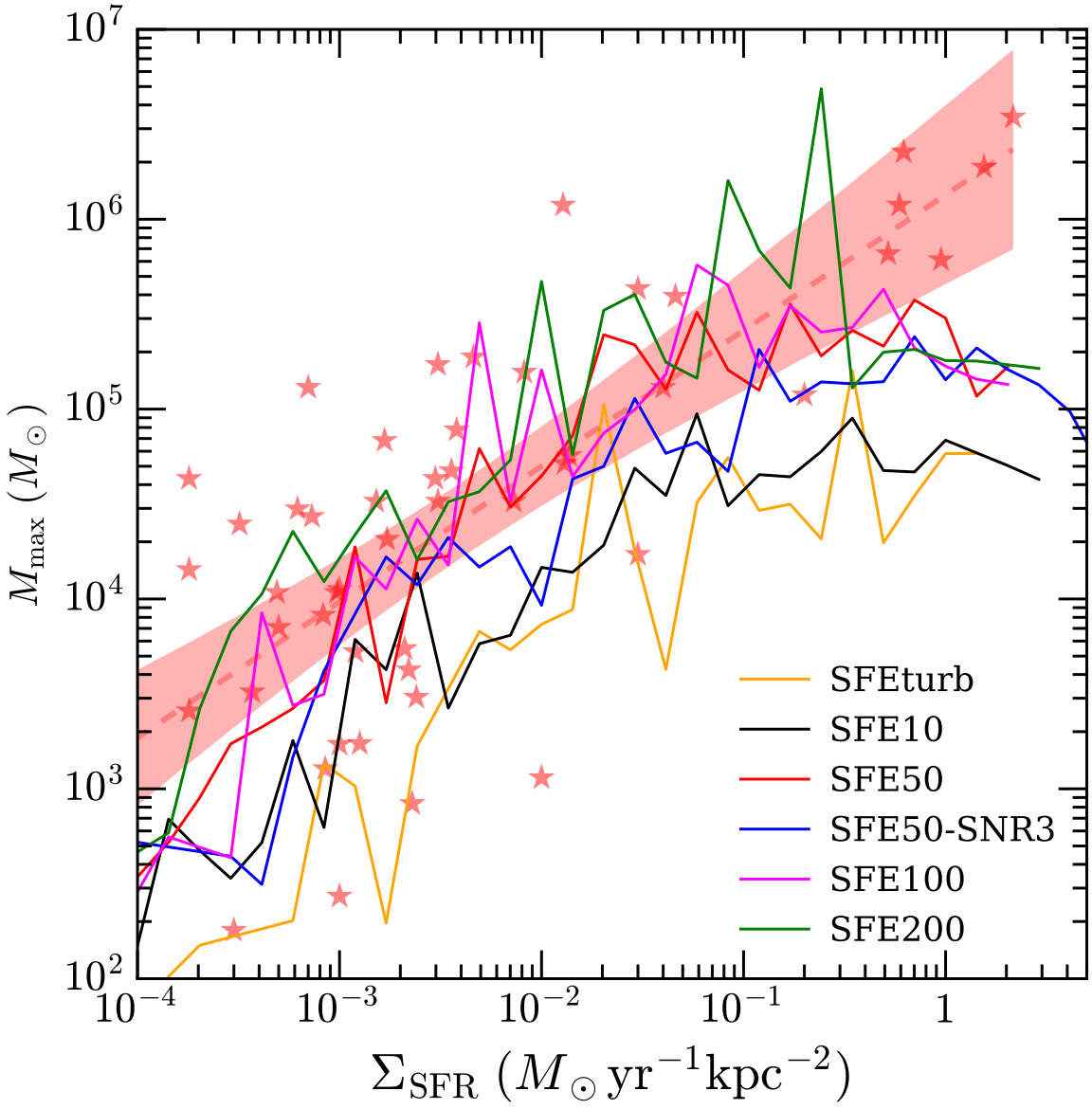

Figure 3.12: Maximum bound cluster mass vs. star formation rate surface density. $\Sigma_{\rm SFR}$ is calculated on the 1 kpc scale for clusters younger than 100 Myr across the disk of the main galaxy from z=10 to the last available snapshots for different runs. The compilation of observed maximum clusters mass in different galaxies from Appendix B in Adamo et al. (2015) is shown in red stars. The best linear fit to the data is overplotted as red dashed line, along with its $1\,\sigma$ confidence interval (red shaded area).

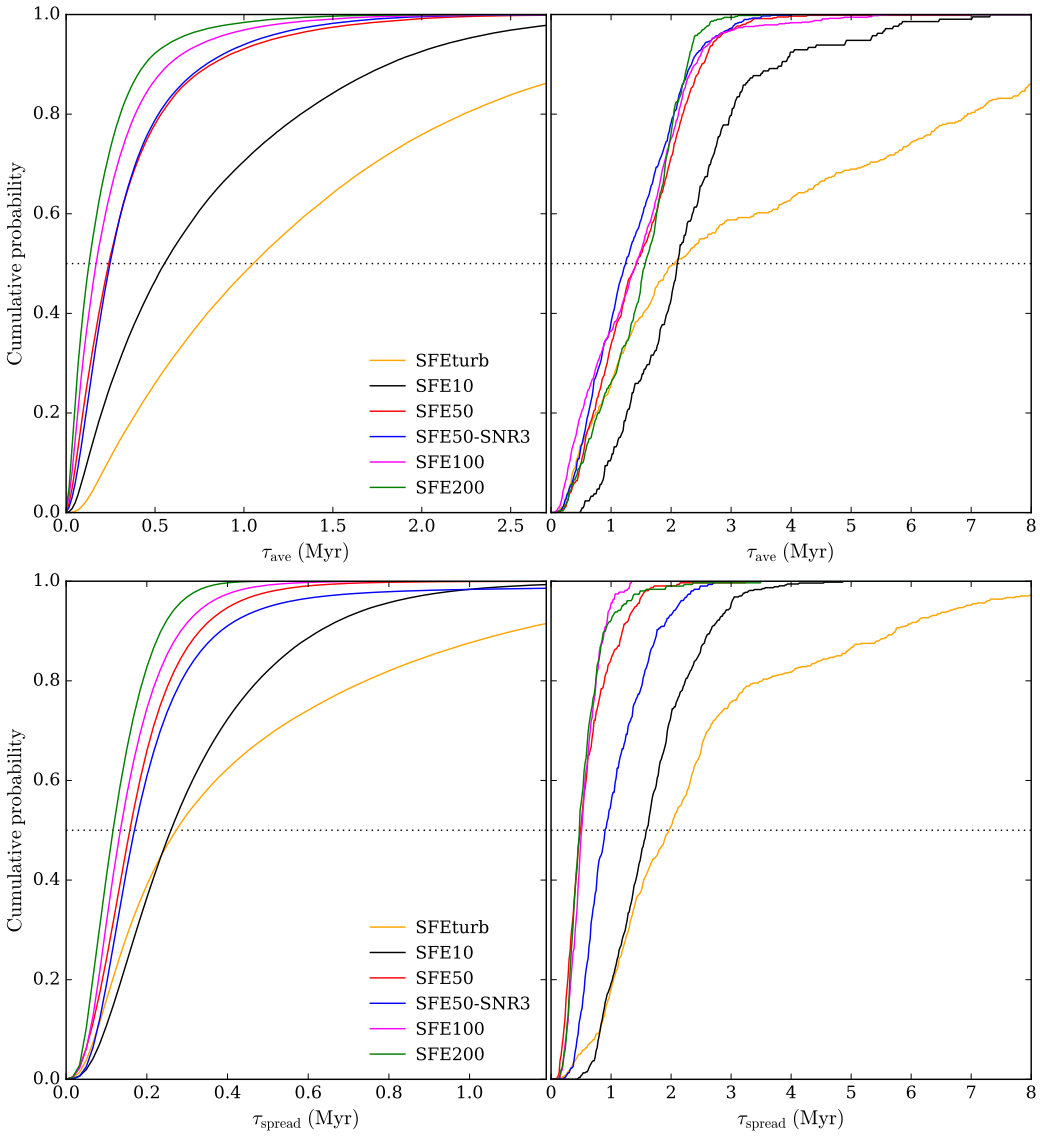

Figure 3.13: Cumulative distribution of mass-weighted cluster formation timescale $\tau_{\rm ave}$ (upper panels) and newly-defined age spread $\tau_{\rm spread}$ (lower panels) for clusters with mass smaller (left panels) or larger (right panels) than $10^5\,M_\odot$.

CHAPTER IV

Dynamical and Chemical Evolution of Massive Star Clusters

4.1 Introduction

For a long time, globular clusters (GCs) are thought to preserve fossil information on the assembly history of their host galaxies at early times (Harris, 2001). However, though GC systems have been known for almost four centuries, their formation and evolution is still one of the most challenging unsolved problems in the interface of stellar astrophysics, galaxy formation and cosmology (e.g. Brodie & Strader, 2006). One major uncertainty that hinders our understanding of the origin of GCs is their dynamical evolution across cosmic time. The dynamical disruption of globular clusters has been studied using various methods from orbit-averaged Fokker-Planck models to Monte Carlo models to direct $N-$body simulations (references). It is general agreed that the observed log-normal mass function of GCs at present is the relic of the power-law initial mass function of young massive clusters formed at high redshift (Prieto & Gnedin, 2008; Muratov & Gnedin, 2010; Li & Gnedin, 2014). The details of this process, however, is larger unknown.

Previous work on calculating tidal disruption of GCs usually assumes a analytical tidal field experienced by clusters along the orbit around the host galaxy. This method

cannot capture the grandness of the gravitational potential of the galaxy, therefore are not able to predict the tidal interaction between the dense structures on the disk, such as giant molecular clouds (GMCs), and star clusters, which is believed to be an important mechanism of disrupting star clusters. Recently, Renaud et al. (2017) post-processed the snapshots of cosmological hydrodynamical simulations and found that the average strength of tidal field increases as the simulations approach lower redshift, thus the dynamical disruption of GCs at $z > 1$ is negligible. The main issue of their approach is that the simulations with a spatial resolution ~ 200 pc, cannot resolve the densest structure in the galactic disk so that the strength of the tidal field is significantly underestimated. This issue is severe especially when clusters are still around their natal GMCs.

In Li et al. (2017) and Li et. al. in prep., we have developed and updated a new algorithm for modeling the formation of star clusters in cosmological simulations. The mass of individual cluster particles is determined by both the gas accretion and their own energy and momentum feedback. Thus, the final mass is set self-consistently and can be considered as the mass of a single star cluster formed within its natal GMC. We used a high spatial resolution, 3-6 pc, that are comparable or even smaller than the size of the GMCs. This resolution allows us to accurately calculate the tidal field around individual star clusters.

In this paper, we introduce a new algorithm for calculating the strength of the tidal field along the orbit of each cluster particle and estimate the mass loss rate at each global timestep of the simulations, see Section 4.2. This is the first time that the tidal disruption process has been calculated in the runtime in cosmological simulations. Some preliminary results are shown in Section 4.4.

4.2 Dynamical disruption of star clusters

The description of the simulation setup and detail implementation of the cluster formation algorithm is presented in Paper I and II. The simulations are performed with the Eulerian gasdynamics and N-body Adaptive Refinement Tree code. The cosmological simulations are run from the initial conditions that contains a main dark matter halo of total mass $M_{200} \approx 10^{12}\,M_{\odot}$ at $z = 0$ with a periodic box of 4 comoving Mpc in size. The simulations start from a 128^3 root grid with the mass of dark matter particles $m_{\rm DM} = 1.05 \times 10^6\,M_{\odot}$. We use adaptive mesh refinement technique to achieve high dynamical range of the spatial resolution. We add 9, 10, 11, and 12 additional refinement levels at $z = \infty$, 9, 4, and 1.5, so that the size of cells at the finest refinement level is in the range between 3 and 6 pc, comparable to the typical size of the star-forming regions in nearby galaxies. The descriptions of simulations with different star formation and feedback parameters can be found in Paper II.

We use CCF to create young star clusters in the simulations. After cluster particles are created in the simulations, they start to lose mass via various physical processes, such as evolution of individual stars and escape of stars due to the dynamical interaction to other stars within the clusters and the external tidal field in the galaxy. The fraction of the initial cluster mass that is gravitationally bound to the star clusters is defined as the bound fraction, $f_{\rm bound}(t) = f_{\rm initial} f_{\rm se}(t) f_{\rm dyn}(t)$, where $f_{\rm initial}$ is the initial bound fraction due to the gas expulsion during cluster formation (defined and calculated in Paper II), $f_{\rm se}$ accounts for the mass loss due to stellar evolution, and $f_{\rm dyn}$ accounts for tidal stripping and evaporation of stars (Li & Gnedin, 2014).

The stellar evolution included in the simulation assumes a time-dependent stellar wind and stellar explosion model that is calibrated by stellar population synthesis, e.g. FSPS. An integrated mass loss rate of a single stellar population with a Kroupa (2001) initial mass function used in our simulations is described in Paper II.

Independent of stellar evolution, the dynamical destruction of star clusters consists of two parts, internal evaporation and external tidal disruption. In this paper, these two processes are characterized by tidal disruption timescale ($t_{\rm tid}$) and evaporation timescale ($t_{\rm iso}$). Following the calculation in Gnedin et al. (2014), the rate of the dynamical destruction from the two processes can be expressed as:

$$\frac{dM}{dt} = -\frac{M}{\min(t_{\rm tid}, t_{\rm iso})} \tag{4.1}$$

.

In the rest of this section, we will describe

- how are the tidal tensor and strength of the tidal field estimated in the run time of the simulations?
- how are $t_{\rm tid}$ and $t_{\rm iso}$ calculated for each clusters?
- how is the bound fraction $f_{\rm bound}$ updated at each global timestep?

4.2.1 Tidal field around cluster particles in realistic galactic environment

The strength and orientation of the tidal field in the galactic environment can be fully characterized by tidal tensors. The general formalism of a tidal tensor at position $\mathbf{X}_0$ under a given gravitational potential $\Phi(\mathbf{X})$, is defined as

$$T_{ij}(\mathbf{X}_0, t) = -\frac{\partial^2 \Phi(\mathbf{X})}{\partial x_i \partial x_j}\Big|_{\mathbf{X}=\mathbf{X}_0}, \tag{4.2}$$

where i and j are two orthogonal directions in the Cartesian coordinates.

In the simulations, tidal tensors of all inactive cluster particles ($t_{\rm age} > t_{\rm max} = 15$ Myr) are estimated in the runtime of the simulations at each global time step [1].

[1]The duration of globular time steps varies during the course of the simulations. A good estimate about it would be around 1-3 Myr.

Algorithmically, the second-order differentiation of the gravitational potential are estimated across the $3 \times 3 \times 3$ cells centered on a given cluster particle. For the diagonal terms, e.g.:

$$\begin{aligned} T_{xx} &= -\frac{\partial^2 \Phi(x,y,z)}{\partial x^2} \\ &\approx -\frac{\Phi(x+L_{\rm cell},y,z)+\Phi(x-L_{\rm cell},y,z)-2\Phi(x,y,z)}{L_{\rm cell}^2}, \end{aligned}$$

where $L_{\rm cell}$ is the size of the central cell. For non-diagonal terms, e.g.:

$$\begin{aligned} T_{xy} &= -\frac{\partial^2 \Phi(x,y,z)}{\partial x \partial y} \\ &\approx -\frac{1}{4L_{\rm cell}^2}\left[\Phi(x+L_{\rm cell},y+L_{\rm cell},z)+\Phi(x-L_{\rm cell},y-L_{\rm cell},z)\right. \\ &\qquad \left. -\Phi(x+L_{\rm cell},y-L_{\rm cell},z)-\Phi(x-L_{\rm cell},y+L_{\rm cell},z)\right]. \end{aligned}$$

We then calculate all three eigenvalues of the tidal tensor, $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$, which represents the intensity of the tidal field in the direction of the corresponding eigenvectors. The tidal disruption timescale can then be referred from these eigenvalues.

4.2.2 Disruption timescales and cluster bound fraction

Following Eq. (4) in Gnedin et al. (2014), the tidal disruption timescale depends strongly on the angular frequency of the cluster orbit around a point mass tidal field, $\Omega_{\rm tid}$:

$$t_{\rm tid} \approx 10 \text{ Gyr} \left(\frac{M(t)}{2\times10^5 M_\odot}\right)^{2/3} \left(\frac{\Omega_{\rm tid}}{\Omega_\odot}\right)^{-1}, \tag{4.3}$$

where $\Omega_\odot \approx 41.4$ Gyr^{-1} is the angular frequency of the stellar orbits around the solar neighborhood. $\Omega_{\rm tid}$ solely depends on the strength of the tidal field around a given cluster. One option is to use the dynamical timescale within the Roche lobe of the

cluster: $\Omega_{\rm tid} \approx \sqrt{|\lambda_1|/3}$.

Eq. (4.3) is obtained from N-body simulations of star clusters in strong tidal fields. In the location where the tidal field is weak, e.g. when clusters are orbiting in the outer halo, the internal evaporation dominates the overall mass loss rate. Again, we follow Gnedin et al. (2014) to estimate the evaporation timescale:

$$t_{\rm iso} \approx 17 \text{ Gyr} \left(\frac{M(t)}{2 \times 10^5 M_\odot} \right). \tag{4.4}$$

Considering both the disruption in tidal fields and evaporation in isolation, the time evolution of the bound mass is determined by Eq. (4.1). In the simulation, the bound fraction is updated accordingly at each global timestep with timescale dt:

$$f_{\rm dyn}^{i+1} = \exp\left(-\frac{dt}{\min(t_{\rm tid}, t_{\rm iso})} \right) f_{\rm dyn}^{i}, \tag{4.5}$$

where i represent the i-th global timestep.

4.3 Metallicity of star clusters

In our simulations, cluster particles launch feedback after their formation in the forms of depositing mass, momentum, energy, and metals to their ambient medium. Instead of following the production of individual elements, to reduce the costs of computing memory, we recorded the metal enrichment using two separate variables that take into account the contributions from Type II and Type Ia supernovae, respectively. The metallicity of specific elements is calculated by post-processing using the preferred metal yield tables of Type II and Type Ia SNe (e.g. Woosley & Weaver, 1995; Nomoto et al., 1997, 2006). This simplified approach also gives us the flexibility to use different yield tables and compare the results without rerunning the whole simulations.

Comparing to Paper I and II, we change the maximum stellar mass that trigger Type II SNe from $100\,M_{\odot}$ to $40\,M_{\odot}$. This change has little effect on the strength of energy and momentum feedback since the number of SNe is dominated by low mass stars. However, because of the yield calculation in ART, this change leads to a 0.3-0.5 dex difference on the total SNII metallicity for a single stellar populations with a Kroupa IMF.

4.4 Preliminary results

4.4.1 Tidal field evolution over cosmic time

Figure 4.1 shows the time evolution of the tidal strength experienced by three star clusters reside in the different locations of the simulation. The intensity of the tidal field is represented by the absolute value of the maximum eigenvalue, $|\lambda_1|$, of the tidal tensor that is derived with the method described in Section 4.2.1.

We find that the strength of the tidal field is much higher than the typical value around the solar neighborhood. There is also a clear trend that the tidal field becomes weaker as the simulation runs toward lower redshift. The value of $|\lambda_1|$ decreases for more than two orders of magnitude for all three particles located at different places in the last snapshot. We emphasis that the weak tidal field at lower-z is not caused by the decrease of the intrinsic tidal field of the galaxy itself, but is caused by the evolution of galactocentric distance of these clusters. Due to the dynamical evolution, star clusters formed from the dense GMCs in the galactic disk are gradually scattered from the central part of the galaxy to its less dense outskirt. We also find a quasi-periodical oscillation of $|\lambda_1|$ with a period of ~ 100 Myr. This timescale is consistent with the dynamical time of the galaxies and is caused by the orbital motion of clusters around the galaxy.

Comparing the cluster in the outer part of the dark matter halo and on the disk,

we find that the tidal field on the gaseous disk is stronger than that in the halo. We also find that the tidal field in the satellite is smaller than that in the main disk. These two differences are consistent with the gas density distribution in different part of the simulations.

4.4.2 Evolution of the cluster mass function

The strong tidal field found in the above section, especially at early time after clusters formed in the dense environments, suggests that the mass loss via tidal disruption is critical to the mass evolution and the survival probability of the model clusters. Considering all processes of cluster mass losses mentioned in Section 4.2, we examine the changes of the shape of cluster mass function in our simulations.

Figure 4.2 shows the time evolution of the mass function of the model clusters in SFE200 run. We first clarify different cluster masses: cluster particle mass M, cluster initial mass $M_i = f_{\rm initial} M$, bound cluster mass at the last snapshot $M_{\rm bound} = f_{\rm bound} M = f_{\rm initial} f_{\rm se} f_{\rm dyn} M$, and the projected cluster mass at $z = 0$ ($M_{z=0}$) that is obtained from evolving $M_{\rm bound}$ from the last snapshot to $z = 0$ by considering only the internal evaporation in Eq. (4.4). We find that the shape of the mass function changes significantly from a power-law initial mass function to a log-normal mass function at $z = 0$. We find that tidal disruption reduces the number of clusters larger than $10^5\,M_\odot$ dramatically and the evaporation remove all clusters with mass smaller than $\sim 10^5\,M_\odot$. The log-normal shape of the projected mass function at $z = 0$ peaked at around $10^5\,M_\odot$ is very similar to the shape of the mass function of GCs. This similarity suggests that the young massive clusters formed at high redshift is a promising candidate of the progenitor of GCs and dynamical disruption is the main driver on transforming the shape of the mass function.

4.4.3 Stellar mass-stellar metallicity relation

Before examining the metallicity distribution of star clusters in the simulations, we first check the global metallicity evolution of the main galaxy across cosmic time. In Figure 4.3, we show the stellar mass-stellar metallicity relation in different runs. The stellar mass is obtained using all cluster particles within $0.5r_{\rm vir}$ of the main galaxy, while the metallicity, [Z/H], is derived from the total metal mass from both SNII and SNIa divided by the total stellar mass for clusters younger than 50 Myr. Each line shows the evolution track of the main galaxy from $z \sim 10 - 1.5$. We find a strong positive trend that the higher the stellar mass, the higher the metallicity, consistent with current observations of stellar metallicity in different types of galaxies (e.g. Gallazzi et al., 2005; Kirby et al., 2013).

It should be noted that the trend presented here is a combination of the stellar mass-stellar metallicity relation and its redshift evolution. However, due to the small number of galaxies in our simulations, it is very hard to separate these two effects. We also compare our results to that of the FIRE simulations. For example, the main galaxy with stellar mass $\sim 10^{10}\,M_{\odot}$ at $z \approx 1.5$ have total metallicity [Z/H] ~ -0.45 for runs with $\epsilon_{\rm ff} > 0.5$, similar to the value obtained in the FIRE simulations.

However, we find that run SFE50-SNR3 with weaker feedback shows a systematic 0.3 dex higher metallicity than runs with $\epsilon_{\rm ff} > 0.5$ for stellar mass $< 10^{10}\,M_{\odot}$. For stellar mass $> 10^{10}\,M_{\odot}$, its metallicity quickly rises to even super-solar. This is caused by a strong star burst concentrated at the very center of the galaxy so that the metallicity cannot be efficiently dispersed to larger radius due to the inefficient stellar feedback. SFE10 also shows a slight metal-rich than runs with $\epsilon_{\rm ff} > 0.5$, which do not exhibit clear different among each other.

4.4.4 Metallicity distribution of survival clusters

Figure 4.4 shows the metallicity distribution of all clusters in the simulation box at $z = 1.5$ for SFE200 run. The clusters more massive than $10^5\,M_\odot$ shows a broad range of metallicity from [Z/H] = -3.0 to -0.5. This range of metallicity is consistent with the range of GCs in the Milky Way, although with some missing of most metal-rich ones. Given that the simulations just reaches $z = 1.5$, we would expect more metal-rich clusters can be created during the later epochs.

Interestingly, we find that dynamical evolution remove majority of the clusters with metallicity between [Z/H] $= -1.5 - -1.0$ and generate a dip in the metallicity distribution, which leads to a bimodal shape metallicity distribution.

4.5 Summary

We have developed an algorithm to calculate the strength of the tidal field experienced by star cluster particles formed in cosmological simulations. Based on the tidal field of the realistic galactic environment in the simulations, the mass loss of the model clusters by stellar evolution, internal evaporation, and tidal disruption is estimated accurately in the simulation runtime. With high spatial resolution of the simulations, the tidal interaction between clusters and dense structures on the gaseous disk can be captured. We study the dynamical evolution of the massive star clusters across cosmic time and characterize the evolution of its metallicity distribution. Our main results are summarized here:

- The strength of the tidal field decreases as cluster particles orbiting away from the central, dense part of the galaxy. In contrast to the result in Renaud et al. (2017), tidal disruption in our simulations is crucial to the mass loss of clusters during the first Gyr.

- With the combination of tidal disruption and internal evaporation, the shape of

the bound cluster mass function changes from power-law to log-normal across cosmic time. The obtained log-normal mass function is similar to the observed mass function of GCs, suggesting that young massive clusters formed at high redshifts are promising candidates of the progenitor of GCs.

- We find a scaling relation between the stellar mass of the main galaxy and the average metallicity of young clusters at different epochs. This stellar mass-stellar metallicity relation is consistent with observations of high-z galaxies and other high resolution cosmological simulations.

- We find that the metallicity of massive clusters presents a wide range from $[\mathrm{Z/H}] = -3.0 - -0.5$. Interestingly, dynamical evolution preferentially removes clusters with median metallicity and generates a dip at $[\mathrm{Z/H}] = -1.5 - -1.0$. Thus a bimodal shape metallicity distribution is emerged for survival clusters, very similar to the metallicity distribution of the Milky Way GC populations.

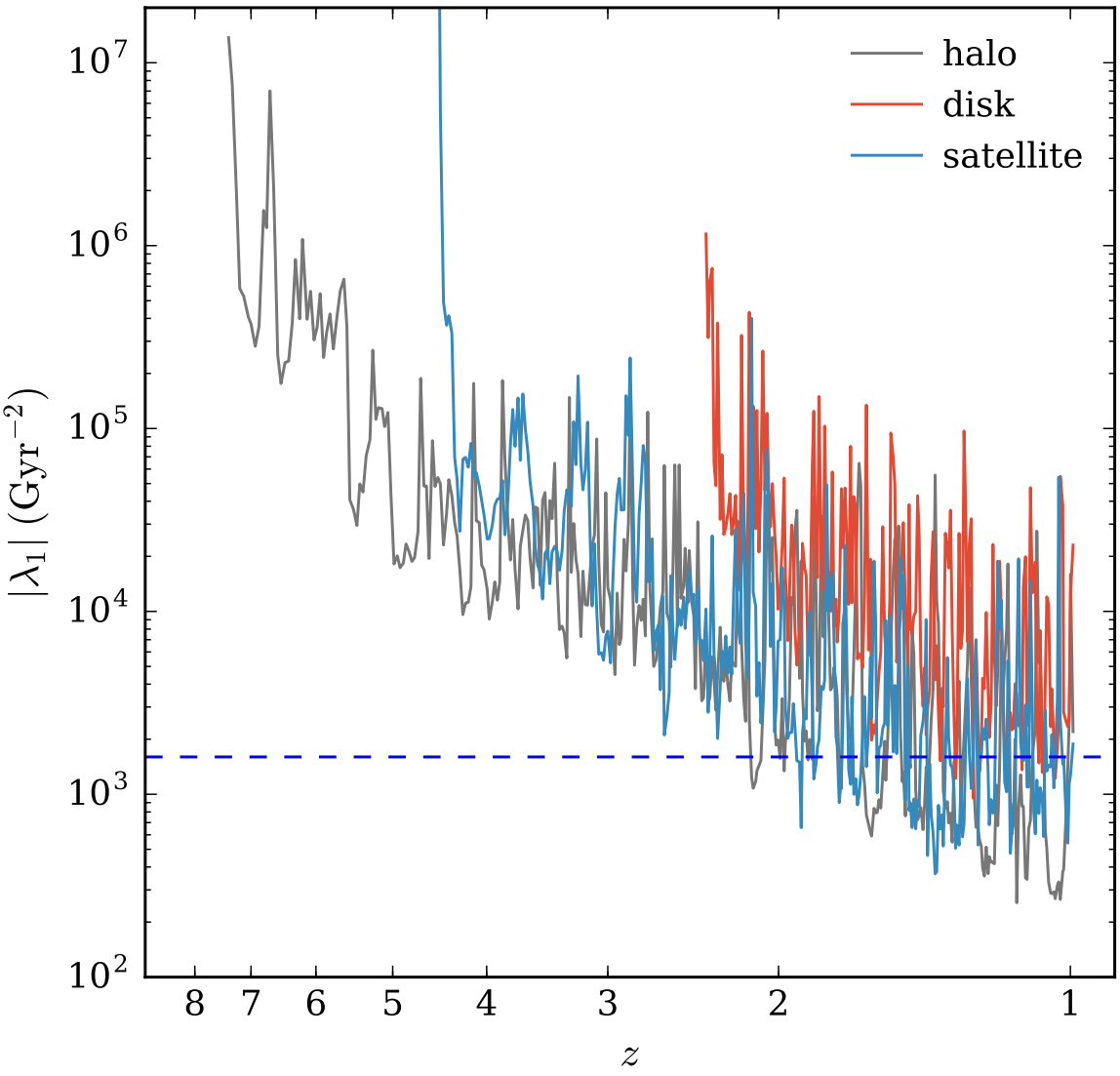

Figure 4.1: Evolution of the absolute value of the maximum of the three eigenvalues, $|\lambda_1|$, defined in Section 4.2.1 for three representative cluster particles from $z = 8 - 1$. The three clusters are selected so that they locate in different regions of simulations at $z \sim 1$: in the outer part of the main halo (black), in the disk of the main galaxy (red), in the satellite (blue). For reference, the intensity of the tidal field at the position of the Sun at present is shown as the horizontal dashed line.

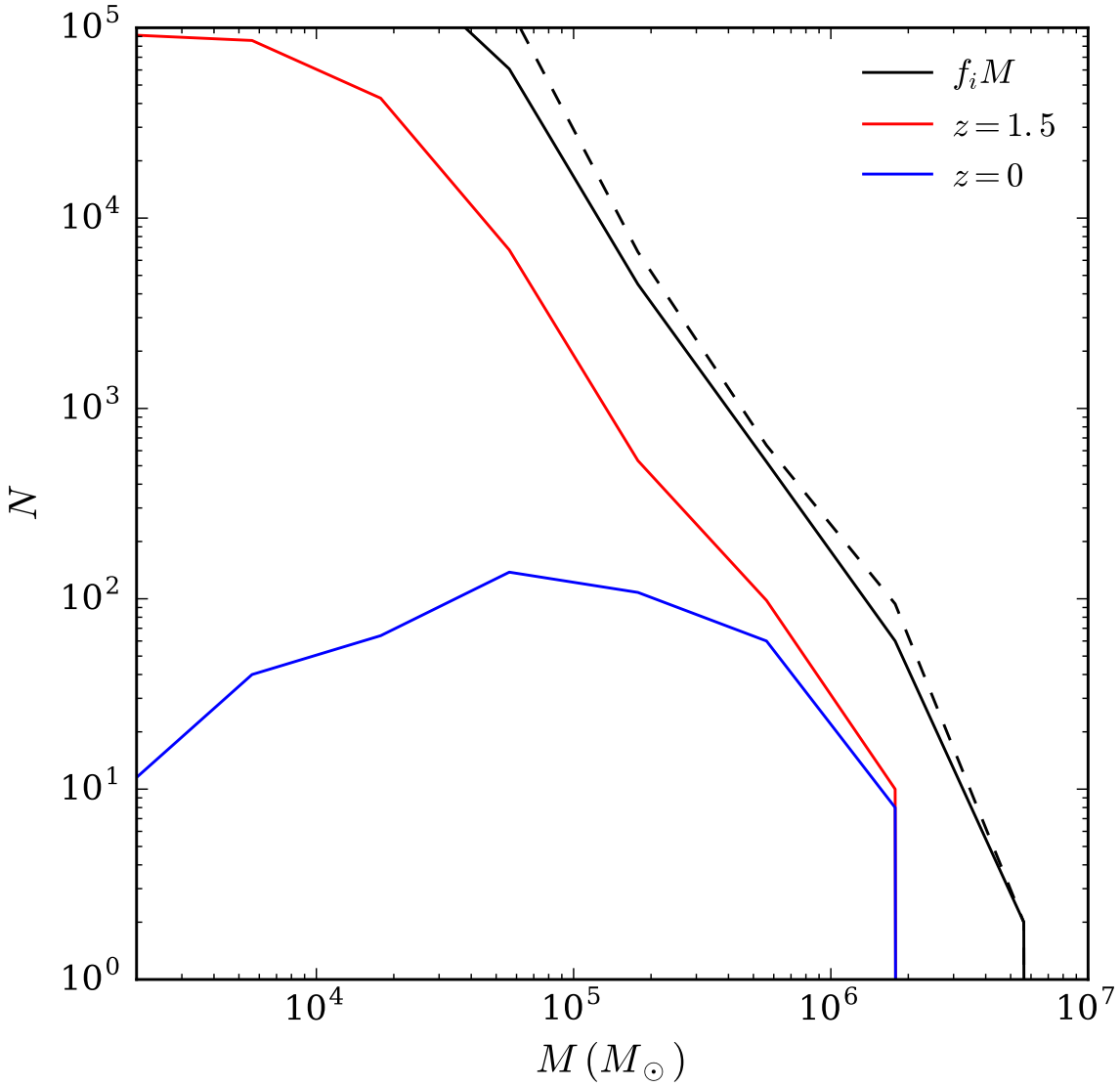

Figure 4.2: Evolution of mass functions of all clusters in the whole simulation box from the cluster initial mass function (M_i, black solid) to the bound mass function at $z \approx 1.5$ ($M_{\rm bound}$, red) and the mass function of clusters that evolve to $z = 0$ ($M_{z=0}$, blue). The initial mass function of cluster particle is shown in black dashed line. The definition of the different masses is described in Section 4.4.2.

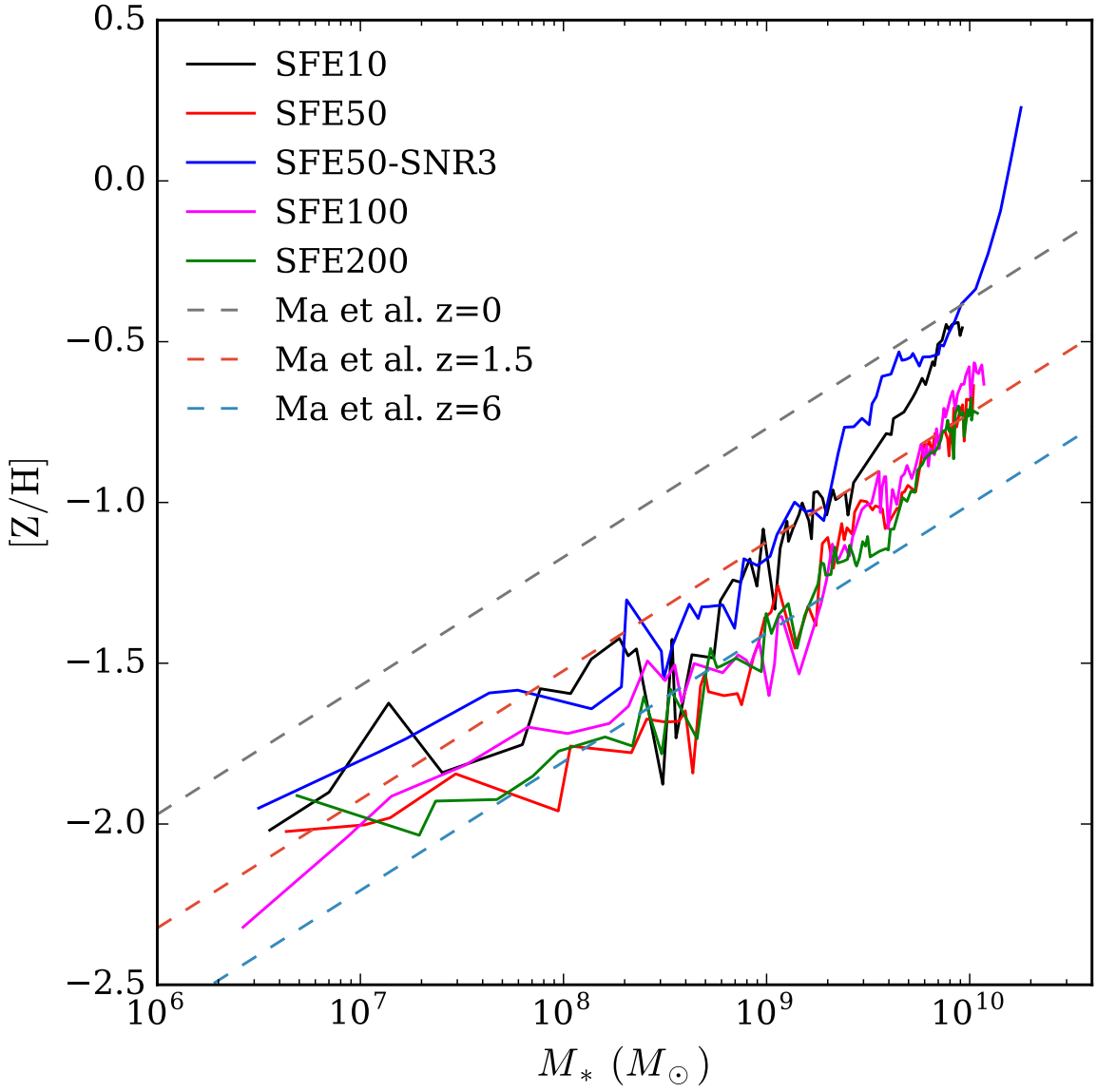

Figure 4.3: Stellar mass-metallicity relation of the main galaxy for runs with different star formation and feedback parameters. The stellar mass of the main galaxy is the total mass of cluster particles within $0.5r_{\rm vir}$ and the stellar metallicity is the mean metallicity of cluster particles younger than 50 Myr for a given epoch. For reference, the stellar mass-stellar metallicity relation at $z = 0, 1.5, 6$ from the FIRE simulations (Ma et al., 2015) is overplotted in the same figure.

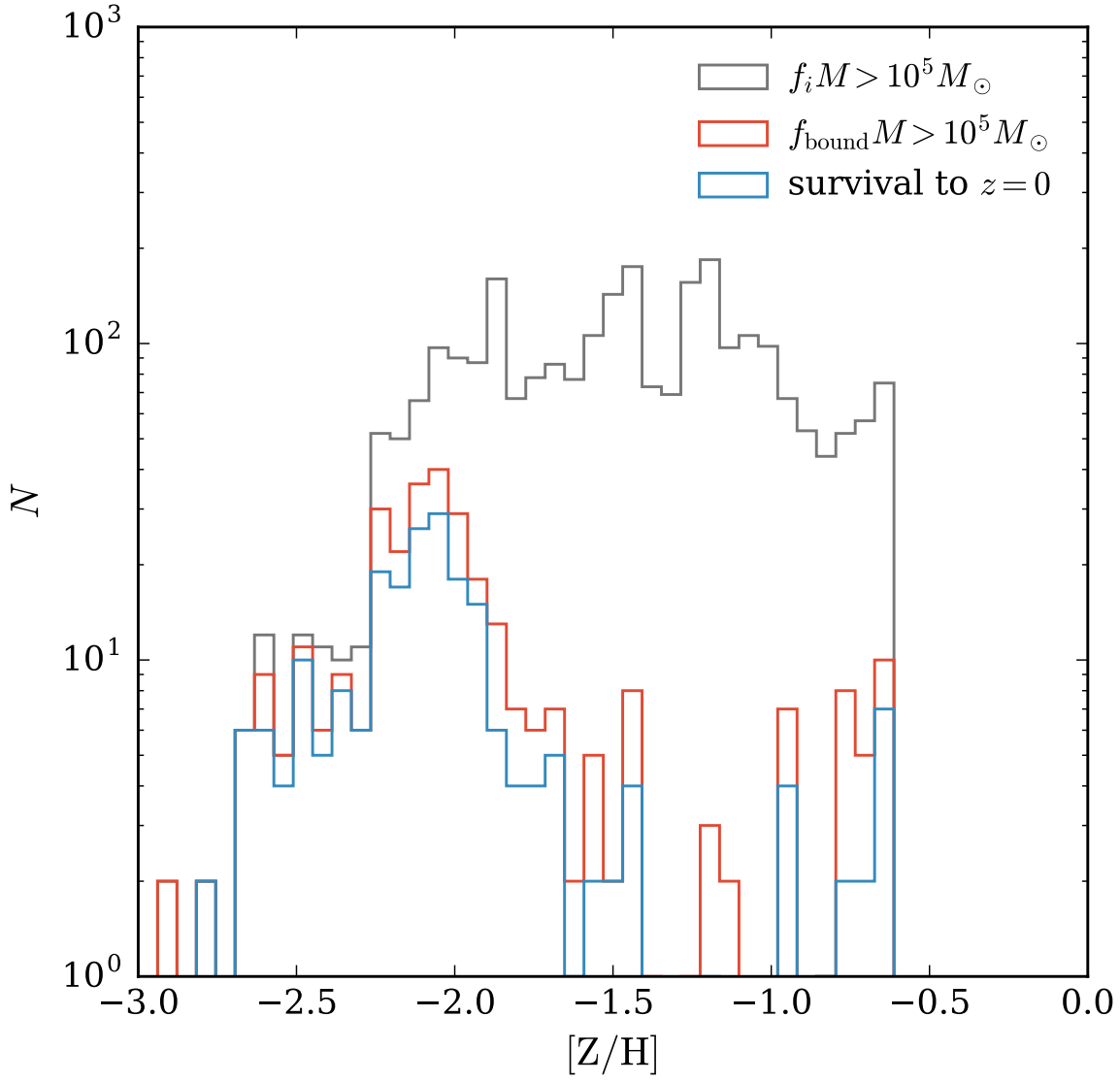

Figure 4.4: Metallicity distribution of star clusters formed in SFE200 run: clusters with initial mass larger than $10^5\,M_\odot$ (black), clusters with bound mass larger than $10^5\,M_\odot$ at $z \approx 1.5$ (red), clusters that survive at $z = 0$ considering only internal evaporation in Eq. (4.4) from $z = 1.5 - 0$.

CHAPTER V

Modeling the Formation of Globular Cluster Systems in the Virgo Cluster

The mass distribution and chemical composition of globular cluster (GC) systems preserve fossil record of the early stages of galaxy formation. The observed distribution of GC colors within massive early-type galaxies in the ACS Virgo Cluster Survey (ACSVCS) reveals a multi-modal shape, which likely corresponds to a multi-modal metallicity distribution. We present a simple model for the formation and disruption of GCs that aims to match the ACSVCS data. This model tests the hypothesis that GCs are formed during major mergers of gas-rich galaxies and inherit the metallicity of their hosts. To trace merger events, we use halo merger trees extracted from a large cosmological N-body simulation. We select 20 halos in the mass range of 2×10^{12} to $7 \times 10^{13}\,M_{\odot}$ and match them to 19 Virgo galaxies with K-band luminosity between 3×10^{10} and $3 \times 10^{11}\,L_{\odot}$. To set the [Fe/H] abundances, we use an empirical galaxy mass-metallicity relation. We find that a minimal merger ratio of 1:3 best matches the observed cluster metallicity distribution. A characteristic bimodal shape appears because metal-rich GCs are produced by late mergers between massive halos, while metal-poor GCs are produced by collective merger activities of less massive hosts at early times. The model outcome is robust to alternative prescriptions for cluster formation rate throughout cosmic time, but a gradual evolution of the mass-metallicity

relation with redshift appears to be necessary to match the observed cluster metallicities. We also affirm the age-metallicity relation, predicted by an earlier model, in which metal-rich clusters are systematically several billion years younger than their metal-poor counterparts.

5.1 Introduction

Globular cluster (GC) systems have been found in various types of galaxies. Because of their old age and compact structure, GCs are believed to carry information on galaxy assembly history at early times (e.g., Brodie & Strader, 2006). In particular, the colors and metallicities of GC systems provide a unique record of the early star formation and chemical enrichment in their host galaxies. One of the remaining puzzles is the origin of the commonly seen bimodal distribution of GC colors, within galaxies ranging from spirals to giant ellipticals. The bimodality in color is indicative of bimodality in metallicity, which has been used to separate GCs into two subpopulations: metal-poor and metal-rich (Harris, 2001; Peng et al., 2006; Harris et al., 2006).

In general, GCs have systematically lower metallicity than the field stars of their host galaxy. Therefore, they must have formed earlier than the bulk of stars, at least from the chemical evolution point of view. Early galaxies were smaller and less metal-enriched than those of today. Motivated by this fact, we test a hypothesis that major mergers of gas-rich galaxies (which happened more frequently at high redshift, in the hierarchical galaxy formation framework) are predominantly responsible for the formation of GCs. Ashman & Zepf (1992) predicted the metallicity bimodality resulting from galaxy mergers even before observations revealed it. In their model, the two subpopulations can be produced by distinct star-forming events, which could naturally occur in hierarchical structure formation. Based on this framework, Muratov & Gnedin (2010, hereafter MG10) modeled the metallicity distribution of Galactic GCs

using the mass assembly history from a cosmological N-body simulation, coupled with observational scaling relations for galaxy stellar mass and metallicity. This model incorporates both the formation and disruption of GCs in the progenitor galaxies of a host halo with the mass similar to the Milky Way. The model has successfully reproduced both the bimodal metallicity distribution and the log-normal distribution of cluster mass.

During the past decade, observations of GC systems outside the Local Group have advanced significantly with the Hubble Space Telescope (HST). A comprehensive study of galaxies in the Virgo cluster, the ACS Virgo Cluster Survey (ACSVCS), examined the photometric properties of 100 early-type galaxies, along with their GC systems. Multi-modal GC distributions are present in all target galaxies with absolute magnitude of $-22 < M_B < -15$. The peak metallicities of the two main modes follow a systematic (but weak) trend with galaxy luminosity, implying a possible common origin of these subpopulations. This data set gives us a good opportunity to investigate the formation of GC systems in massive elliptical galaxies.

The GC system in the Milky Way shows only weak bimodality: 30% of the clusters are in the metal-rich group. In contrast, giant Virgo ellipticals have comparable numbers of red and blue clusters, and therefore, they present better tests for the origin of the metallicity distribution.

In this paper, we extend the model of MG10 to more massive early-type galaxies and adopt the mass assembly history from a large cosmological Millennium-II (MM-II) simulation. The model is based on the calculation of the galaxy cold gas mass, the mass-metallicity relation (MMR), the cluster fraction, and the initial mass function (IMF), which we describe in Section 5.2. The final model is even simpler than MG10 and has only four adjustable parameters. In Section 5.3, we add the dynamical disruption of individual clusters, by considering two-body relaxation and stellar evolution. We apply this updated model to 20 halos selected from MM-II, with

total mass in the range of $10^{12} - 10^{14}\,M_{\odot}$, appropriate for massive elliptical galaxies. We also investigate variants of our model in Section 5.4. In Section 5.5 we compare the model cluster populations with the ACSVCS observations of 19 corresponding galaxies. We summarize and discuss the main results in Sections 5.6 and **??**.

5.2 Model for Globular Cluster Formation

We update the framework of the MG10 model using several independent realizations of halo assembly history and recent observational relations for galaxy stellar and gas masses. Halo merger trees are obtained from the Millennium Database[1]. For merger events that meet the required formation criteria, clusters are created by Monte Carlo sampling at the epoch of the merger and share the metallicity of their host galaxies, with an additional scatter. Both central and satellite halos are followed in the model, and clusters formed within both are collected into the final system. The cluster formation efficiency is linearly proportional to the mass of available cold gas, which in turn is set by the halo mass and redshift. Galaxy metallicity is set by the observed stellar MMR. All the details of the model are described below.

5.2.1 Mass Assembly History

We construct the mass assembly history of dark matter halos using the MM-II simulation (Boylan-Kolchin et al., 2009b). MM-II is a collisionless simulation within a $100h^{-1}\,$Mpc box, which contains halos up to $\sim 10^{15}\,M_{\odot}$. The particle mass, $6.9 \times 10^{6}\,M_{\odot}$, makes it possible to trace the formation of halos as small as $\sim 10^{9}\,M_{\odot}$. The cosmological parameters used in the simulation, and adopted in this paper, are $\Omega_{\Lambda} = 0.75$, $\Omega_{\rm m} = 0.25$, $h = 0.73$, and $\sigma_8 = 0.9$.

At first we tried to use the halo catalog of the original Millennium simulation (Springel et al., 2005) but found that its lower mass resolution ($8.6 \times 10^{8}\,M_{\odot}$) does

[1] `http://gavo.mpa-garching.mpg.de/Millennium`

not allow us to track satellite halos less massive than $10^{10}\,M_{\odot}$. Since $10^{9}\,M_{\odot}$ halos can still contribute to forming $10^{5}\,M_{\odot}$ star clusters that would likely survive to the present day (see Equation (5.5)), it is more accurate to use the MM-II catalogs to capture all merger events capable of producing massive star clusters.

The masses of central and satellite progenitors are collected at all 67 outputs from $z = 127$ to $z = 0$. Parent and child halos are connected with each other in the database by the identifiers `descendantId` and `lastProgenitorId`. We apply the tags `firstProgenitorId` and `nextProgenitorId` to find the most massive and second most massive progenitors of a given halo, and use their masses to calculate the merger ratio, R_m. This halo merger tree is the starting point of our model.

Although we do not require the halos specifically to be located in a Virgo-sized cluster, it should not bias our comparison with the ACSVCS galaxies. Cho et al. (2012) showed that the colors and luminosities of GC systems of early-type galaxies in low-density regions are similar to those in the Virgo cluster. While the environment has a small effect via the morphology-density relation, the properties of GC systems are primarily dominated by the host galaxy mass.

5.2.2 Stellar and Gas Masses

To all progenitors in a given merger tree, we assign the mass of stars and cold gas according to the following analytical prescriptions.

The stellar mass – halo mass relation, $M_*(M_{\rm h}, z)$, is based on the abundance-matching technique, using a parameterization of Behroozi et al. (2013b, their Equation (3)) for the Sloan Digital Sky Survey (SDSS) measurements of the galaxy luminosity function.

Note that Kravtsov et al. (2014) have recently found that the total luminosity, and stellar mass, of central galaxies in clusters (with $M_{\rm h} > 10^{14}\,M_{\odot}$) has been underestimated in the SDSS photometry pipeline, mainly due to over-subtraction of the

background light in extended galactic envelopes. The correction is substantial and can reach a factor of $2-4$ at $M_* \gtrsim 10^{12}\,M_\odot$. The magnitude of the corresponding correction at $z > 0$ is not yet known. We have decided not to include this correction, because our sample contains only one central cluster galaxy, and more importantly, we use the galaxy stellar mass only as a proxy for estimating the cold gas mass and metallicity from the observed scaling relations, as we describe below. These relations were derived for the stellar luminosities measured by the SDSS. In order to apply these relations consistently, we use the Behroozi et al. (2013b) equations as published.

To derive the mass of cold gas, $M_{\rm g}$, in a galaxy with stellar mass M_*, we combine recent results from the ALFALFA survey (Papastergis et al., 2012) with additional Arecibo observations of nearby starforming galaxies by J. Bradford & M. Geha (in preparation). These observations measure the mass of neutral HI gas, to which we add the corresponding HeI mass. We take the measured $M_{\rm g}$ as a proxy for the reservoir of gas available for star formation. The data at $z \approx 0$ show that the mean ratio $M_{\rm g}/M_*$ at a given stellar mass exhibits a bend at $M_* \approx 10^9\,M_\odot$, and cannot be described by a single power law. A satisfactory fit is provided by a double power law:

$$\eta \equiv \frac{M_{\rm g}}{M_*} \approx 1.8 \left(\frac{M_*}{10^9\,M_\odot} \right)^{-\alpha(M_*)}, \tag{5.1}$$

with a steeper slope $\alpha = 0.68$ for $M_* > 10^9\,M_\odot$, and a shallower slope $\alpha = 0.19$ for the less massive galaxies with $M_* < 10^9\,M_\odot$. The high-mass slope is consistent with the relation used by MG10 ($\alpha = 0.7$), but for dwarf galaxies the gas mass is reduced relative to the MG10 prescription.

The amount of cold gas in high-redshift galaxies is very uncertain. We can rewrite the gas-to-stellar fraction as $\eta(z) = {\rm sSFR(z)} \times {\rm t_{dep}(z)}$, where ${\rm sSFR} \equiv {\rm SFR/M_*}$ is the specific star formation rate (SFR), and $t_{\rm dep} \equiv M_{\rm g}/{\rm SFR}$ is the gas depletion timescale. The empirical evolution of the sSFR is consistent with ${\rm sSFR(z)} \propto (1+{\rm z})^{2.8}$ up to

$z \sim 2$ (Magdis et al., 2012; Tacconi et al., 2013), while $t_{\rm dep}(z)$ is consistent with being approximately constant for starforming galaxies at all redshifts (e.g., Bigiel et al., 2011; Feldmann, 2013). Thus we obtain

$$\eta(z) = \eta_0 \, (1+z)^n, \tag{5.2}$$

with $n = 2.8$.

An alternative derivation by Tacconi et al. (2013), from a CO survey of molecular gas of actively star-forming galaxies at $z \approx 1-3$, suggests a variable gas depletion timescale, decreasing roughly as $t_{\rm dep}(z) \approx 1.5\,{\rm Gyr}\,(1+{\rm z})^{-1}$, which in turn implies $\eta(z) = \eta_0 \, (1+z)^{1.8}$. Given the current uncertainty in the gas evolution, when constructing our model we consider both possibilities, $n = 2.8$ and $n = 1.8$.

At even higher redshift ($z \gtrsim 2-3$), there is evidence that the gas fraction saturates at a maximum value (e.g., Magdis et al., 2012). Accordingly, we limit η for very high redshifts: $\eta(z > 3) = \eta(z = 3)$.

In addition to the mean relations given by Equations ((5.1) and (5.2)), we include random scatter of 0.3 dex to account for the combined dispersion of the local MMR, specific SFR, and gas depletion time.

The stellar and gas mass fractions in a given halo are then defined as:

$$f_* \equiv \frac{M_*}{f_{\rm b}\, M_{\rm h}}, \quad f_g \equiv \frac{M_g}{f_{\rm b}\, M_{\rm h}} \tag{5.3}$$

where $f_{\rm b} \approx 0.16$ is the universal baryon fraction (e.g., Hinshaw et al., 2013).

A final constraint of the mass fractions is that the sum of the two cannot exceed the total accreted baryon fraction, $f_{\rm in}$:

$$f_*(z) + f_{\rm g}(z) \leq f_{\rm in}(z), \tag{5.4}$$

where $f_{\rm in} \leq 1$ is limited by photoheating by the extragalactic UV background, as described in MG10. In cases when the baryonic fraction ($f_* + f_g$) calculated from Equations (5.1) and –(5.2) exceeds $f_{\rm in}$, we revise the gas fraction to be $f_{\rm g,revised} \equiv f_{\rm in} - f_*$. This constraint only affects halos less massive than $M_{\rm h} \sim 10^{10}\,M_\odot$.

Figure 5.1 shows our derived gas and stellar fractions of halos ranging from $10^{9.5}\,M_\odot$ to $10^{14}\,M_\odot$ at redshifts $z = 0 - 5$. The new prescription is similar, but not identical, to that in MG10. The stellar fraction reach its maximum for the Milky Way-sized halos $M_{\rm h} \sim 10^{12}\,M_\odot$, and decreases at both higher and lower mass. The new $M_* - M_{\rm h}$ relation also depends much less strongly on redshift than that used in MG10.

The evolution of the gas fraction with redshift for the case $n = 2.8$ is faster than what is needed to account for the increase of stellar mass of galaxies at $z \lesssim 3$, resulting from the abundance matching method (Figure 4 of Behroozi et al., 2013b). We will test the sensitivity of our results to this prescription by considering an alternative calculation of the cold gas mass in Section 5.4.2.

5.2.3 Cluster Formation

In our model, clusters are formed during epochs of enhanced star formation following halo mergers. We trace both mergers between a satellite and a central halo, as well as mergers between two satellites. For a halo with mass $M_{{\rm h},i}$ at the ith simulation output, the mass of its main progenitor and (possible) second largest progenitor at output $i-1$ are $M_{{\rm h},i-1}$ and $M_{{\rm h2},i-1}$, respectively. The merger ratio is defined as $R_m = M_{{\rm h2},i-1}/M_{{\rm h},i-1}$, if the second progenitor is found. Otherwise, we use the differential increase in halo mass as a proxy: $R_m = (M_{{\rm h},i} - M_{{\rm h},i-1})/M_{{\rm h},i-1}$.

A cluster formation event is triggered by a gas-rich major merger, when the merger ratio exceeds a threshold value: $R_m > p_3$. We expect the threshold to be in the range of $p_3 = 0.1-0.5$ to have sufficient influence on the structure of the interstellar medium

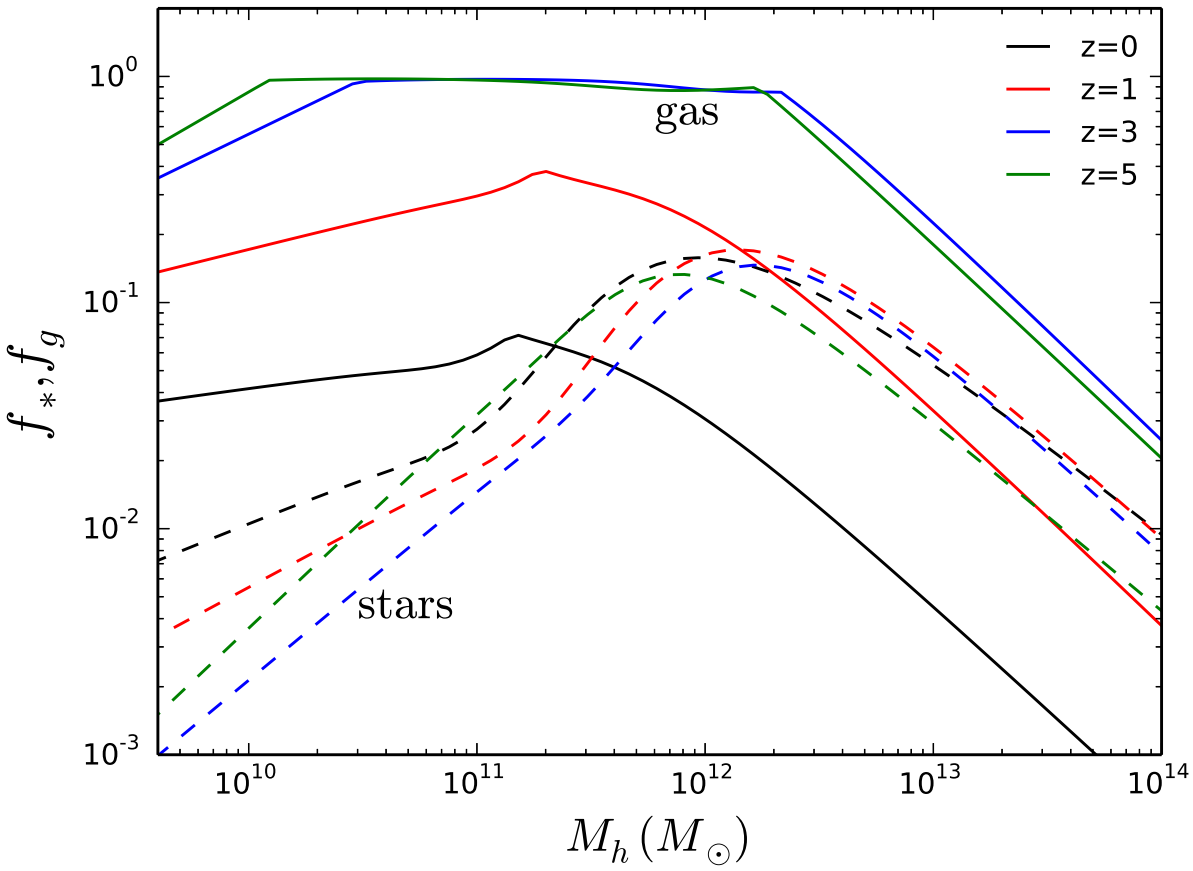

Figure 5.1: Adopted relation for the fraction of galaxy gas mass (solid lines) and stellar mass (dashed lines), in units of the universal baryon fraction (Equations (5.1)–(5.3), with $n = 2.8$) vs. halo mass, at several redshifts: $z = 0$ (black), $z = 1$ (red), $z = 3$ (blue), $z = 5$ (green).

that could trigger condensation of giant molecular clouds, but the exact value of p_3 is an adjustable parameter of the model.

The MG10 model used an additional parameter to set a minimum cold gas fraction of the merging halos (at the level of 4%). We have experimented with including this constraint, but found that it is automatically satisfied by the requirement to have enough gas mass to form a cluster with $M > 10^5\,M_\odot$, according to Equation (5.5). Any value of the gas fraction threshold below 10% gave similar results, and therefore, we set it to zero and eliminate it as a model parameter.

In another departure from the MG10 model, we do not include a "Case-2" for-

mation channel here, whereby clusters could form without a detected merger but in extremely gas-rich galaxies with a cold gas fraction above $\approx 98\%$. Instead, in Section 5.4.3 we will investigate an alternative scenario for continuous cluster formation.

The cluster formation rate scales approximately linearly with the mass of cold gas available for star formation, as indicated by detailed hydrodynamic simulations (Kravtsov & Gnedin, 2005):

$$M_{\rm GC} = 3 \times 10^{-5} p_2 \, f_{\rm b}^{-1} \, M_{\rm g}, \tag{5.5}$$

where $p_2 \sim 1$, the normalization factor, is another adjustable parameter in our model. This relation gives us the total mass of all GCs formed in a given galaxy at a given epoch. The normalization factor p_2 is necessary because the galaxy formation cannot be smoothly captured by processing of discrete outputs of the MM-II simulation. Note that the definition of p_2 differs from MG10, where it was written as $1 + p_2$. Here we explore a wider range of this parameter, allowing for $p_2 < 1$.

The total mass $M_{\rm GC}$ is then distributed into individual GCs by using a Monte Carlo method and adopting a power-law initial cluster mass function, $dN/dM = M_0 \, M^{-2}$. The minimum mass of individual clusters is set to $M_{\rm min} = 10^5 \, M_\odot$, below which a typical cluster is expected to be completely evaporated over the Hubble time. After fixing $M_{\rm GC}$ and $M_{\rm min}$, the maximum cluster mass $M_{\rm max}$ (also equal to the normalization M_0) is evaluated from the integral constraint $M_{\rm GC} = M_{\rm max} \, \ln(M_{\rm max}/M_{\rm min})$.

5.2.4 Mass-Metallicity Relation of Host Galaxies

The metallicity of individual model GC is determined by the metallicity of the host galaxy at the epoch when the GC is created. The stellar MMR of galaxies can be used to estimate the iron abundance of the host. Based on the observations in the Local Group, the MG10 model adopted a linear relation between the iron abundance

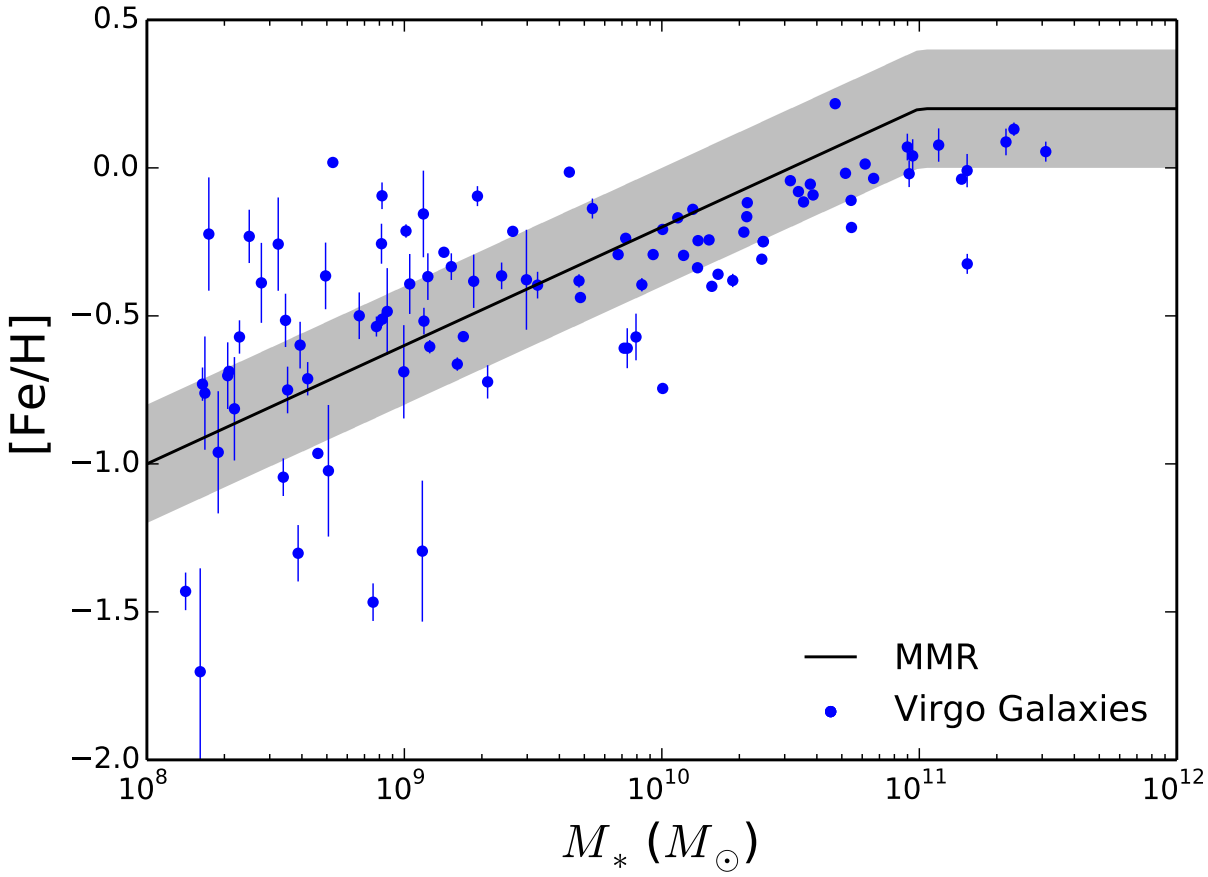

Figure 5.2: Adopted mass-metallicity relation at redshift zero (solid line) and scatter $\sigma_{\rm met} = 0.2$ dex (light shading). Points with errorbars represent the metallicity of the Virgo cluster galaxies, derived from their $(g-z)$ color (see details in Section 5.2.4).

and log stellar mass at zero redshift: $[\mathrm{Fe/H}] = 0.4\,\log(\mathrm{M_*}/10^{10.5}\,\mathrm{M_\odot})$, along with a gradual evolution of this relation with redshift.

Recent observational evidence suggests that the MMR is better described as a two-dimensional projection of a three-dimensional fundamental metallicity relation (FMR) of stellar mass, metallicity, and SFR (e.g. Lara-López et al., 2010, 2013), or alternatively, HI gas mass (Bothwell et al., 2013). No additional redshift evolution is needed. By combining this FMR with the evolution of the gas-to-stellar mass ratio, we can derive an explicit redshift-dependence of the projected MMR. As suggested by Bothwell et al. (2013), a new variable $(\log M_* - 0.35\log M_{\rm g})$ minimizes the scatter

in the FMR. In the notation of our Equation (5.2) this variable can be rewritten as $(\log M_* - 0.54\,n\log(1+z))$. Substituting it for $\log M_*$ in the local MMR, we obtain the evolving relation:

$$[\mathrm{Fe/H}] = 0.4\,\log\left(\frac{\mathrm{M_*}}{10^{10.5}\,\mathrm{M_\odot}}\right) - 0.216\,\mathrm{n}\log(1+\mathrm{z}). \tag{5.6}$$

The power-law slope, and amount of evolution to $z \approx 0.7$, is consistent with the recent measurements of Gallazzi et al. (2014). The redshift-dependence of our MMR is also consistent with the observed evolution of the gas-phase (O/H) abundance in the AGES survey (Moustakas et al., 2011) and 3D-HST survey (Cullen et al., 2014): about 0.3 dex lower metallicity at $z \approx 2$ relative to $z = 0$, at a fixed stellar mass. Zahid et al. (2014) derived somewhat different slopes of the mass and redshift dependence using the DEEP2 and COSMOS data, but the average amount of evolution to $z \approx 1.6$ is the same as in our relation, 0.25 dex.

It should be noted that the linear scaling with log stellar mass is not valid for very massive galaxies, whose metallicity tends to saturate at a supersolar value. Accordingly, in our model we limit the galaxy metallicity to the maximum value [Fe/H] = 0.2 for $M_* > 10^{11}\,M_\odot$.

Figure 5.2 shows our adopted MMR, along with derived metallicities of the Virgo cluster galaxies. The latter are calculated from the luminosity-weighted $(g-z)$ color using an empirical color-metallicity relation obtained by Peng et al. (2006), Equation (5.7) below, from the Galactic GC data. Applying this relation to the Virgo galaxies containing a mixture of stellar populations is justified because the resulting metallicities are clustered around the observed galactic MMR, without systematic bias. The stellar mass of these galaxies is converted from the K-band luminosity and the color-dependent mass-to-light ratio from Bell et al. (2003).

Kirby et al. (2013) showed that nearby dIrr galaxies follow the same mean MMR

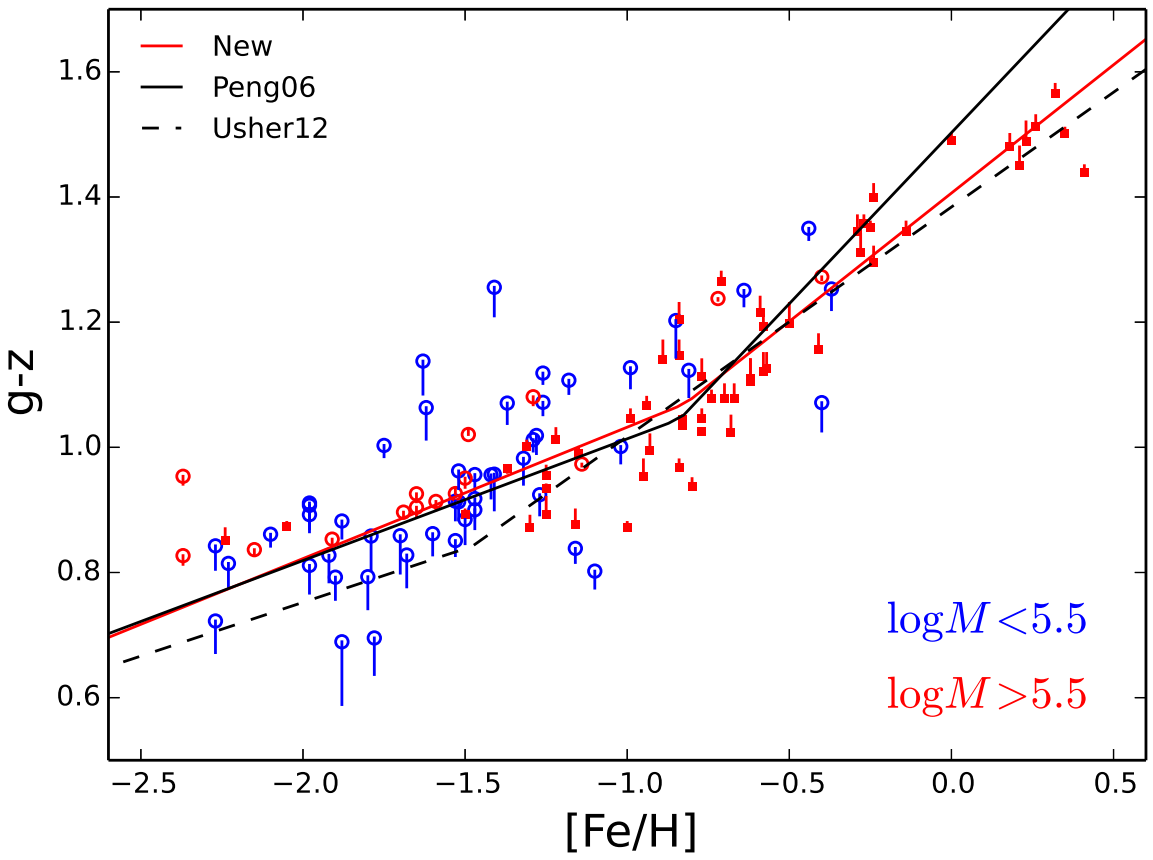

Figure 5.3: Color-metallicity relation of Galactic (open circles) and extragalactic (squares) GCs with spectroscopically measured [Fe/H]. Symbols show the $(g-z)$ color corrected for the evolutionary mass dependence, described by Equation (5.8) and corresponding text in Section 5.2.5. Vertical lines leading to the symbols show the amount and direction of this correction. Our new color-metallicity relation is overplotted (red line; Equation (5.9)) together with the other relations derived by Peng et al. (2006) (black solid line) and Usher et al. (2012) (black dashed line).

as dSph galaxies, with the power-law slope ≈ 0.3. The normalization of their MMR matches our MMR at $M_* \approx 10^8\,M_\odot$, but falls a little lower for higher mass galaxies. Given the considerable dispersion in the derived metallicities of the local and Virgo cluster galaxies, the two relations are not necessarily inconsistent.

5.2.5 Revisiting the GC Color-Metallicity Relation

The metallicity distribution of our model GCs should be compared with the observed populations. Unfortunately, spectroscopic measurements of GC metallicity outside the Local Group are rare. A common approach for Virgo cluster galaxies is to convert the observed GC colors via an empirical color-metallicity relation. Peng et al. (2006) derived such an empirical relation based on the calibration with the Galactic GCs and the additional clusters in Virgo galaxies M49 and M87 with available spectroscopy. The relation is nonlinear, with a steeper slope for metal-poor clusters ([Fe/H] < -0.8):

$$\begin{aligned} [\mathrm{Fe/H}] &= -6.21 + 5.14(g-z), \text{ if } 0.70 < (\mathrm{g}-\mathrm{z}) < 1.05 \\ [\mathrm{Fe/H}] &= -2.75 + 1.83(g-z), \text{ if } 1.05 < (\mathrm{g}-\mathrm{z}) < 1.45 \end{aligned} \tag{5.7}$$

Usher et al. (2012) used calcium triplet-based spectroscopy to determine the metallicity of 903 GCs in 11 early-type galaxies. They found a similarly nonlinear color-metallicity relation but with a different break point and slopes. Most recently, Vanderbeke et al. (2014) obtained updated SDSS photometry of 96 Galactic GCs and suggested a cubic polynomial fit for the relation.

However, based on the models for GC evolution, Goudfrooij & Kruijssen (2014) point out that as clusters lose preferentially low-mass (red) stars by evaporation, their color tends to get bluer. The process is driven by two-body relaxation, which is faster in lower mass clusters. This introduces a possible mass-dependent bias in the color-metallicity relation. To test for this bias, we took the data from Vanderbeke et al. (2014) for 96 Galactic GCs with the metallicity from the updated Harris (1996) catalog, and added the clusters in M49 and M87 from Peng et al. (2006). The mass of an individual GC can be estimated by using the color-dependent mass-to-light ratio from Bell et al. (2003): $\log(M/L_z) = 0.322\,(g-z) - 0.171$. We fit a linear model of two

variables, $(g-z) = \beta_0 + \beta_1\,[\mathrm{Fe/H}] + \beta_2\,\log \mathrm{M}$, and test whether the mass dependence is significant based on the ANOVA variance method. The p-value of the test statistic is $\lesssim 10^{-5}$, which means the null hypothesis that the color is independent of mass can be safely rejected. Thus the available data support the theoretical expectation of Goudfrooij & Kruijssen (2014) that low-mass GCs develop a "blue-shift" relative to high-mass GCs. This evolutionary change of color is in addition to the gradual reddening due to the passive stellar evolution, which is independent of cluster mass.

In order to use the color-metallicity relation to infer [Fe/H] of the observed clusters, we need to "undo" this evolutionary effect. We introduce a simple correction to the color

$$(g-z)_{\mathrm{cor}} = (g-z)_0 - 0.03\,\log\left(\frac{M}{10^6\,M_\odot}\right) \tag{5.8}$$

that minimizes the mass-dependence of the color-metallicity relation. Then we refit the nonlinear relation and obtain

$$\begin{aligned}(g-z)_{\mathrm{cor}} &= 0.21([\mathrm{Fe/H}]+0.82) - 1.07,\ \text{if } [\mathrm{Fe/H}] < -0.82\\ (g-z)_{\mathrm{cor}} &= 0.41([\mathrm{Fe/H}]+0.82) - 1.07,\ \text{if } [\mathrm{Fe/H}] \geq -0.82\end{aligned} \tag{5.9}$$

shown in Figure 5.3. We use this relation to determine GC metallicities in the Virgo galaxies and compare them to our model.

5.3 Dynamical Disruption

Although GCs are relatively stable and long-lived self-gravitating systems, they still gradually lose stars and dissolve into the field. In this paper, we include two sources of the mass loss: the evaporation of stars via two-body relaxation, which reduces the number of stars N_* within the cluster, and stellar winds and explosions, which reduce the average stellar mass $\bar{m}$. These two mechanisms can be quantified

by the mass continuity equation:

$$\frac{1}{M}\frac{dM}{dt} \equiv \frac{1}{N_*}\frac{dN_*}{dt} + \frac{1}{\bar{m}}\frac{d\bar{m}}{dt} = -\nu_{\rm ev}(M) - \nu_{\rm se}\frac{\bar{m}(0)}{\bar{m}}, \tag{5.10}$$

where $\nu_{\rm ev}$ and $\nu_{\rm se}$ are the evaporation rate and mass-loss rate due to stellar evolution, respectively. The time-dependent mass-loss rate $\nu_{\rm se}$ for a Kroupa (2001) IMF is derived in Prieto & Gnedin (2008). The cluster evaporation rate is estimated via the half-mass relaxation time:

$$\nu_{\rm ev} = \frac{\xi_e}{t_{\rm rh}} = \frac{7.25\,\xi_e\,\bar{m}\,G^{1/2}\,\ln\Lambda}{M^{1/2}\,R_{\rm h}^{3/2}}, \tag{5.11}$$

where $\xi_e = 0.033$ is the escape fraction of stars per relaxation time, $R_{\rm h}$ is the half-mass radius, $\bar{m} = 0.87\,M_\odot$ is the average stellar mass for a Kroupa IMF, and $\ln\Lambda = 12$ (e.g., Spitzer, 1987).

By assuming that the stellar evolution mass-loss timescale is much shorter than the evaporation timescale, the decrease of the initial cluster mass with time can be calculated as

$$M(t) = M(0)\left[1 - \int_0^t \nu_{\rm se}(t')dt'\right]\left[1 - \frac{1+3\delta}{2}\nu_{\rm ev,0}\,t\right]^{2/(1+3\delta)}, \tag{5.12}$$

where we take $\delta = 2/3$, as in MG10. The above parameterization provides a good fit to the results of direct N-body simulations of tidally limited clusters.

5.4 Alternative Models

Thus far, we have constructed a model for cluster formation during gas-rich merger events. We will refer to the above prescription as Model 1. There are three major uncertainties in this model: the evolution of the MMR with cosmic time (Equation (5.6)), the evolution of the cold gas fraction (Equation (5.2)), and the need for

major mergers to trigger cluster formation.

5.4.1 Model 2: No Metallicity Evolution

Although there is a consensus that more massive galaxies have higher metallicity, recent studies suggest that the MMR is more complex than was expected before. As summarized in Section 5.2.4, there exists a fundamental plane for star-forming galaxies in the three-dimensional parameter space: SFR, metallicity, and stellar mass. On the other hand, scarcity of spectroscopic observations of high-redshift galaxies leaves the evolution of the MMR with time very uncertain. In order to test the sensitivity of our model to this relation, we consider an extreme case of no-evolution and apply the local MMR, $[\mathrm{Fe/H}] = 0.4 \log(\mathrm{M_*}/10^{10.5}\,\mathrm{M_\odot})$, at all redshifts. We keep the rest of the prescriptions as in Model 1, and refer to this new case as Model 2.

5.4.2 Model 3: Gas Mass from SFR

In Model 1, the gas mass-stellar mass relation is used to derive the cold gas fraction of the halos. An alternative way of determining the gas mass is to use the gas depletion timescale, defined as the ratio between the cold gas mass and SFR, $t_{\rm dep} \equiv M_{\rm g}/\mathrm{SFR}$. Bigiel et al. (2011) found a constant timescale $\tau_{\rm Dep} \approx 2.35\,\mathrm{Gyr}$, with 1σ scatter of 0.24 dex. This depletion time, together with the SFR required to match the growth of the stellar mass of galaxies (Behroozi et al., 2013a), can be used to determine the amount of gas available for star formation. Note that such an empirical derivation of the SFR is independent of the direct measurements in Lyman-break galaxies, discussed in Section 5.2.2. In particular, it leads to much lower gas fraction in halos with $M_{\rm h} < 10^{11}\,M_\odot$ at redshift $z > 2$, relative to that in Model 1. These halos are crucial for cluster formation, as they are the typical hosts of the metal-poor GC population. Observational surveys of low-mass high-redshift galaxies are greatly incomplete and may be underestimating the stellar mass

growth. To correct for the possible incompleteness, we set the gas fraction in halos with $M_{\rm h} < 10^{11}\,M_{\odot}$ at $z > 2$ to be the same as in a $10^{11}\,M_{\odot}$ halo at $z = 2$. We keep the rest of the prescriptions as in Model 1, and refer to this new case as Model 3.

5.4.3 Model 4: No Mergers

Since all of the models above are based on the assumption that GCs are formed in gas-rich mergers, one may ask whether the merger scenario is a required channel for cluster formation or just one of several possible ways to reproduce the GC bimodality. Another plausible formation channel is during galactic starbursts, characterized by enhanced SFR, regardless of whether they are caused by mergers or continuous gas accretion. In this starburst case, the trigger for cluster formation can be either an SFR or specific SFR exceeding a critical threshold, $\rm SFR_{crit}$ or $\rm sSFR_{crit}$. We calculate the sSFR (SFR) for the halos in the whole merger tree from the differential stellar mass growth, as described in Behroozi et al. (2013a), and trigger cluster formation when the sSFR (SFR) exceeds the threshold. The latter is a new parameter of this alternative model, to which we refer as Model 4.

5.5 Results

5.5.1 Galaxy-Halo Matching

The above prescriptions allow us to create GCs within the MM-II halos with different mass assembly histories. We then compare the metallicity distribution of the model GC population to the observed GC systems in the Virgo cluster galaxies. In order to match the galaxies to the halos, we take the stellar mass of the Virgo galaxies, obtained from their color and K-band magnitude (Section 5.2.4), and calculate the expected mass of their halo using the stellar mass-halo mass relation (Behroozi et al., 2013b). As discussed in Section 5.2.2, this relation may be based on an underesti-

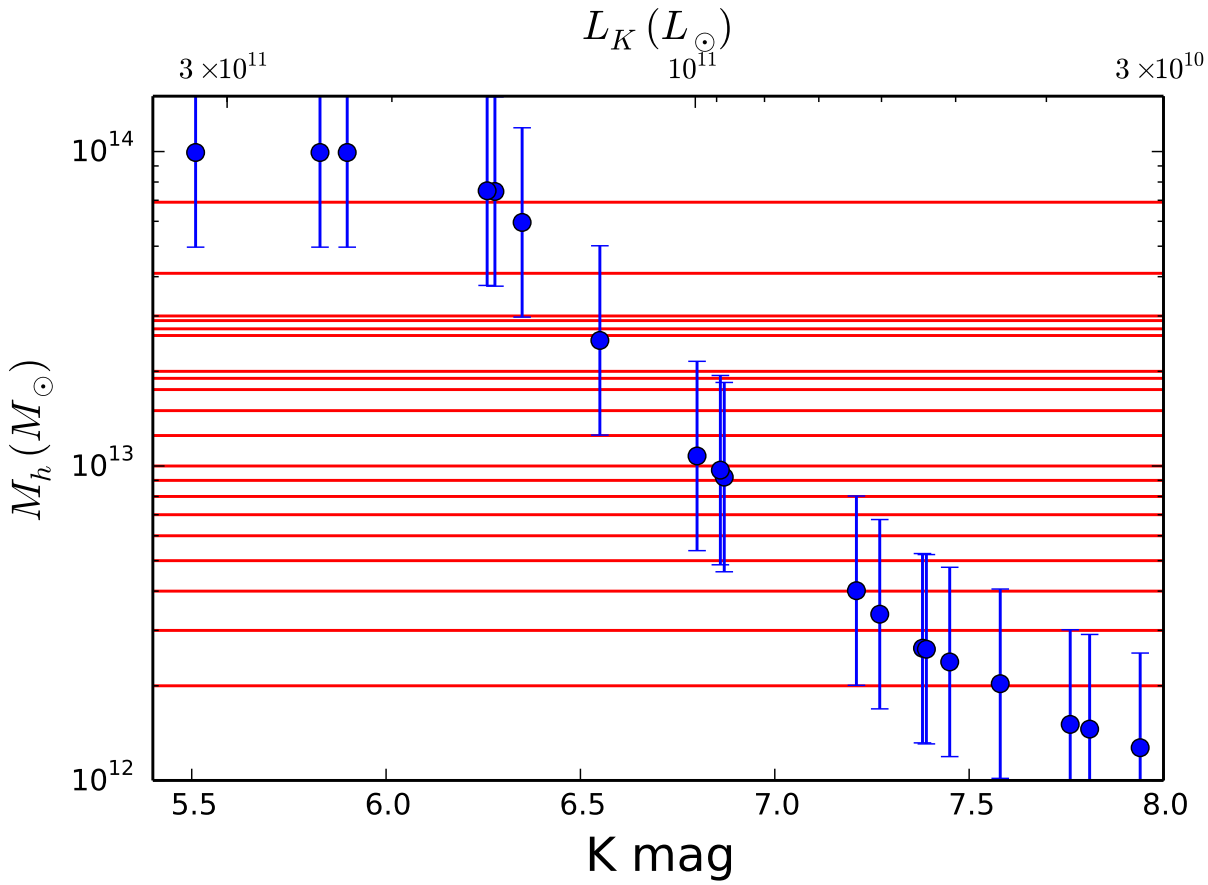

Figure 5.4: K-band apparent magnitude vs. derived halo mass for 19 Virgo galaxies, using the abundance matching technique (see Section 5.5.1). Red horizontal lines show the masses of the 20 halos selected from the MM-II simulation.

mated stellar mass of giant galaxies, which would lead to an overestimate of derived halo mass. To correct for this, we set the maximum halo mass at $\sim 10^{14}\,M_\odot$. The photometry of the Virgo galaxies, such as K-band magnitude and color, are from the 2MASS catalog provided on the ACSVCS Website[2]. The effective radii R_e are obtained by fitting the Sersic profile (Ferrarese et al., 2006). These data are reproduced in Table 5.1, along with the derived stellar and halo mass. The GC Virgo catalogs with the SDSS g and z-band magnitudes are from Jordán et al. (2009).

Because of the uncertainty in both the mass-to-light ratio and $M_* - M_{\rm h}$ relation,

[2] `https://www.astrosci.ca/users/VCSFCS/Home.html`

we add random scatter (0.1 dex or 0.2 dex) for each conversion step. The number of matching galaxy-halo pairs increases with increasing scatter, but the best-fit model parameters are not sensitive to the exact value. The relation between the observed K-band magnitude and calculated halo mass for the Virgo galaxies is plotted in Figure 5.4, together with the MM-II halos. We selected only halos with $M_{\rm h} > 10^{12}\,M_{\odot}$ to model massive elliptical galaxies that contain largest samples of clusters.

The galaxy-halo matching procedure is straightforward. We match each galaxy to all MM-II halos that fall within its calculated mass range, shown by the error bars in Figure 5.4 and listed in Table 5.1. For each pair, we compare the model and observed GC metallicity distributions using the Kolmogorov-Smirnov (KS) test. Then we combine the p-values of the KS probability for all pairs into a joint set and calculate the fraction of pairs with the p-value larger than 0.01. This fraction defines the "goodness" of our model, $G_{0.01}$. The 1% level of the KS probability is generally considered to indicate that the model is not inconsistent with the data. A "goodness" value of $G_{0.01} = 50\%$ means that half of the model realizations are consistent with the observed GC metallicities. The best-fit model parameters are determined by maximizing the "goodness" value.

5.5.2 Exploration of the Parameter Space

The adjustable model parameters are summarized in Table 5.2. In order to explore the space of parameters p_2 and p_3, we calculate $G_{0.01}$ on a two-dimensional grid spanning the range of $1 \leq p_2 \leq 5$, $0 < p_3 \leq 0.5$. Smaller values of the formation rate factor, $p_2 < 1$, lead to an insufficient amount of gas to form enough clusters to match the observations. At first, we fix the other two parameters, $\sigma_{\rm met} = 0.2$ and $n = 2.8$, and investigate them in detail later.

Figure 5.5 shows the contours of $G_{0.01}$, up to the maximum value of about 40%. This is a significant enough fraction of galaxy-halo pairs with the model GC metal-

Table 5.1: Host Galaxy Properties

VCC ID	K mag	$(g-z)$	R_e (kpc)	M_* ($10^{10}\,M_\odot$)	$M_{\rm h}$ ($10^{12}\,M_\odot$)
1226	5.51	1.60	17.0	31.0	$50-198$
1316	5.90	1.60	13.7	21.7	$50-198$
1978	5.83	1.62	8.0	23.3	$50-198$
881	6.28	1.57	35.3	15.3	$37-149$
798	6.26	1.38	12.9	15.4	$38-150$
763	6.35	1.56	12.7	14.6	$30-119$
731	6.80	1.53	9.9	9.4	$5.4-21$
1535	6.55	1.59	10.0	11.9	$12.6-50$
1903	6.87	1.53	10.1	9.0	$4.6-18$
1632	6.86	1.61	6.9	9.1	$4.9-19$
1231	7.27	1.53	1.5	6.2	$1.7-6.7$
2095	7.45	1.44	1.1	5.2	$1.2-4.8$
1154	7.21	1.54	2.4	6.6	$2.0-8.0$
1062	7.38	1.53	1.1	5.5	$1.3-5.3$
2092	7.58	1.50	2.5	4.7	$1.0-4.1$
369	7.94	1.57	0.6	3.4	$0.64-2.5$
759	7.81	1.54	2.2	3.8	$0.73-2.9$
1692	7.76	1.53	0.8	3.9	$0.76-3.0$
1030	7.39	1.49	0.8	5.4	$1.3-5.2$

Table 5.2: Fiducial Parameters of Model 1

Parameter	Best value	Range considered	Effect
$\sigma_{\rm met}$	0.2	$0.1-0.2$	Scatter of MMR
p_2	2.6	$1-5$	Normalization of cluster formation rate
p_3	0.33	$0-0.5$	Minimum merger ratio
n	2.8	$1.8-2.8$	Index of cold gas evolution

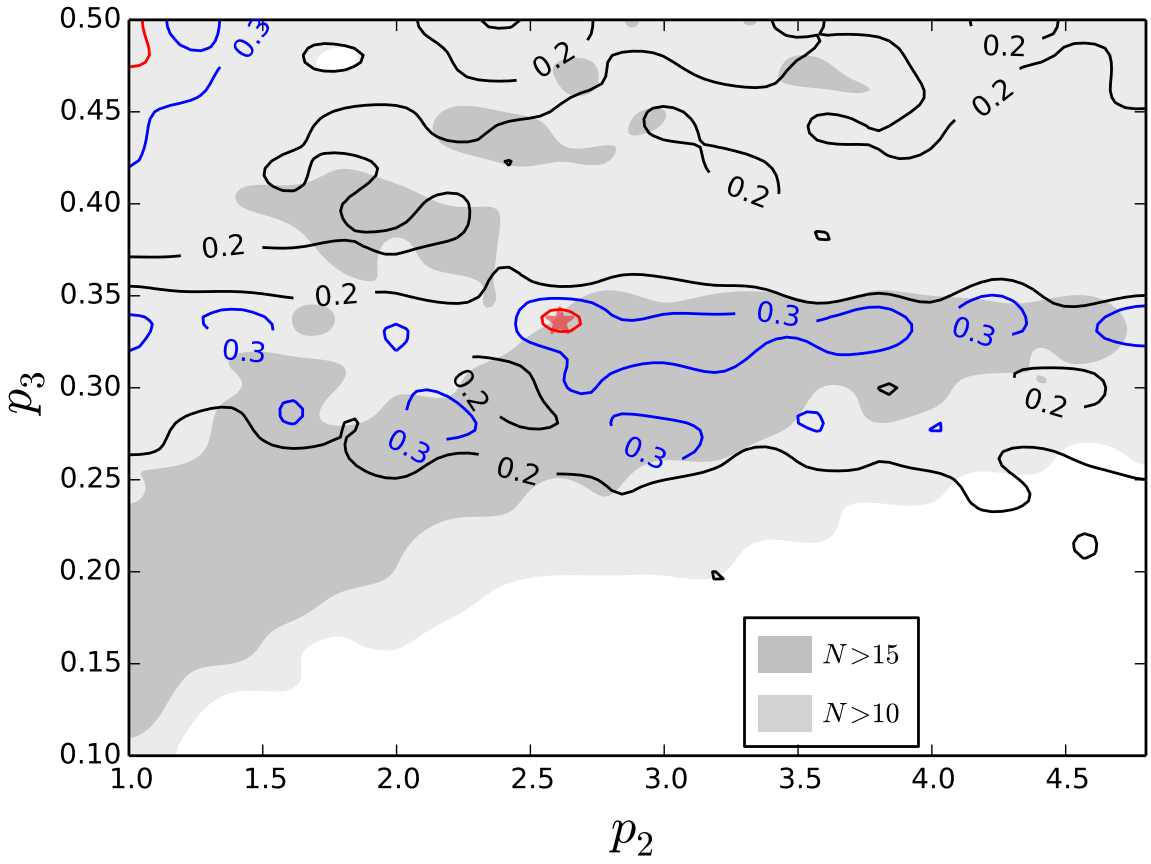

Figure 5.5: "Goodness" contours on the $p_2 - p_3$ parameter plane for Model 1, with fixed $\sigma_{\rm met} = 0.2$ and $n = 2.8$. For example, a contour marked with "0.3" encloses the range of parameters with $G_{0.01} > 30\%$. Shaded regions show the number of galaxies with the size of their GC system sufficiently similar to the observed (see Section 5.5.2 for details). The fiducial model with the best-fit parameters $p_2 = 2.6$ and $p_3 = 0.33$ is labeled by a red star. This model has both the highest $G_{0.01}$ value and the largest number of sufficient GC systems.

licities matching the observations.

In addition to the metallicity distribution, an important statistic is the total number of clusters surviving dynamical disruption to redshift zero, that is, the size of the current GC system. It would be very unlikely for any model to produce exactly the observed number of clusters in a given galaxy. Therefore, we introduce a "tolerance" of 0.2 dex on the logarithm of the ratio of the number of model clusters to observed clusters. If $|\log(N_{\rm GC,model}/N_{\rm GC,obs})| < 0.2$, we consider it a "sufficient" GC system.

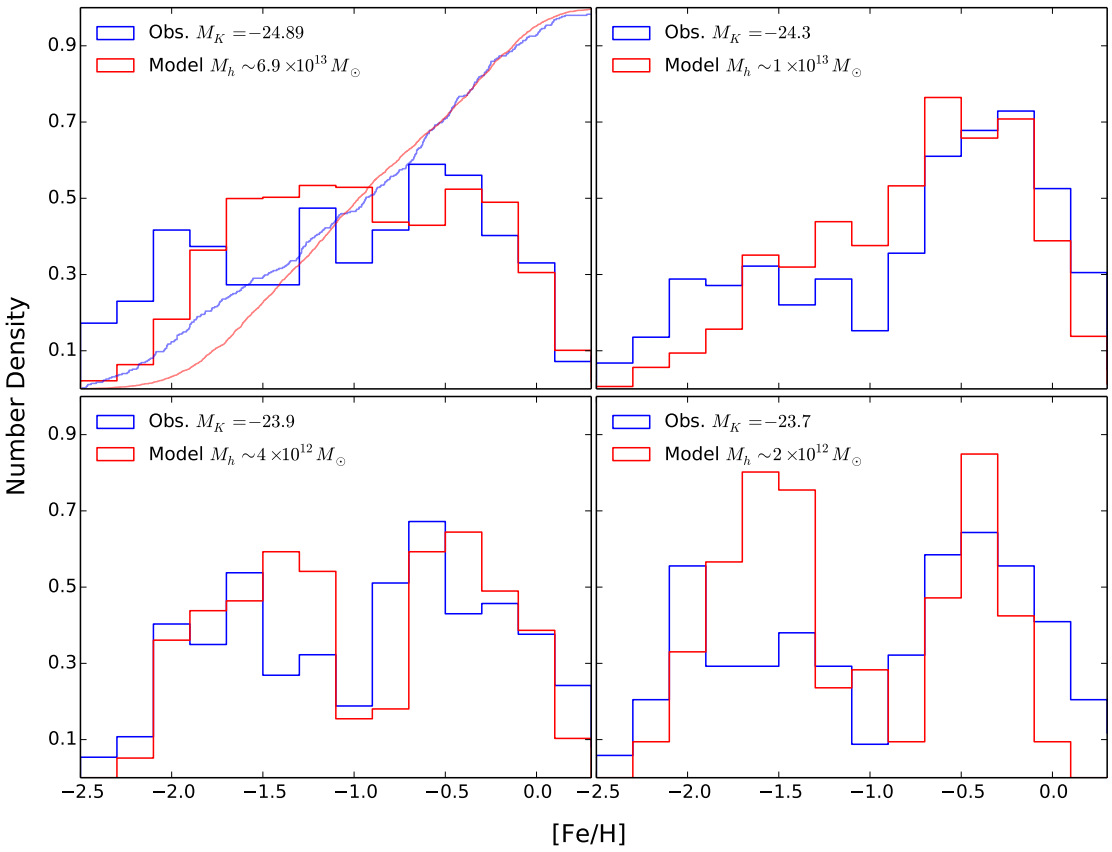

Figure 5.6: Comparison between the observed and modeled GC metallicity distributions. Different panels represent the halos of different mass, from largest to smallest, and their matched Virgo cluster galaxies. In the upper left panel, we overplot the cumulative distributions to show that the model is consistent with the data. The KS test probabilities for the four panels are $p_{KS} \approx 1\%, 4\%, 49\%, 7\%$, in order of decreasing mass.

Shaded regions in Figure 5.5 show the number of galaxies with the sufficient GC systems. The darkest shade indicates an almost complete match: 16, 17, 18, or 19 systems for our total sample of 19.

The region with the highest numbers of sufficient systems lies near the line pointing from $(p_2, p_3 = 1.0, 0.2)$ to $(4.5, 0.35)$. This trend can be easily understood: the larger the boosting factor p_2, the more GCs are created in the model, which need to be compensated by fewer merger events, and therefore, larger threshold ratio p_3.

The number of sufficient systems N_s helps us select the parameters of the best-

fit fiducial model. There are two sets of parameters with equally high goodness $G_{0.01} \approx 40\%$: $(p_2, p_3 = 2.6, 0.33)$ marked by a red star, and $(1.0, 0.48)$. However, the second set has significantly lower N_s, and consequently, we discard it.

We have also varied the scatter of the MMR $\sigma_{\rm met} = 0.1$ & 0.2, and found that $\sigma_{\rm met} = 0.2$ gives a higher goodness value, simply because it can spread the metallicity range to reach the low ([Fe/H] < -2) and high ([Fe/H] > 0) tails of the observed distribution.

The index of the cold gas fraction n (Equation (5.2)) is suggested to have two values: 1.8 and 2.8, as we discussed in Section 5.2.2. Intuitively, $n = 1.8$ will lead to a slower increase of $f_{\rm g}$ toward high redshift, which would suppress the formation of GCs. It will also slow down the metallicity evolution, and in turn bring closer the red and blue peaks of the [Fe/H] distribution. We have repeated the model-selection procedure for the $n = 1.8$ case and explored the goodness contours in the $p_2 - p_3$ plane, for different $\sigma_{\rm met}$. The largest goodness value is only $G_{0.01} = 0.16$, much smaller than that in the $n = 2.8$ case. Thus we conclude that $n = 2.8$ is favored in our model.

5.5.3 Removal of GCs of Satellite Galaxies

A caveat to our comparison is that the observed samples cover only inner parts of the Virgo galaxies. The ACSVCS data were obtained from single pointings of the HST/ACS camera with the field of view $202'' \times 202''$, which corresponds to $\sim 16 \times 16\,$kpc at the distance of the Virgo cluster. For 8 of our 19 galaxies, this scale lies within the effective diameter of the stellar distribution and therefore many GCs may be located outside. Peng et al. (2008) extrapolated the GC number density profiles to larger distances to estimate total counts and found that blue clusters extend further out than red clusters, as is typical of nearby GC systems. Unfortunately, we cannot extrapolate the missing cluster metallicities. Instead, we can reduce our model samples to match the observational setup as close as possible. Since we

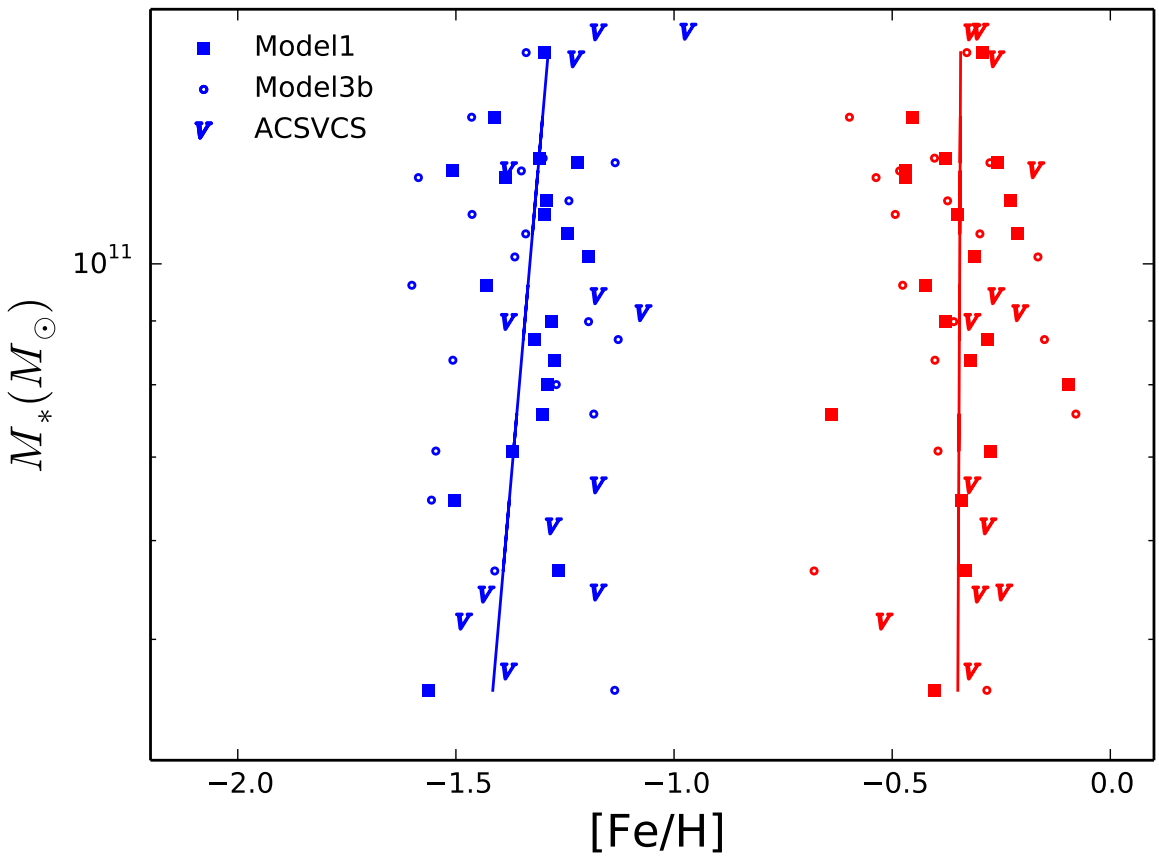

Figure 5.7: Peak metallicities of the blue and red cluster subpopulations within the Virgo galaxies (symbols V) and the fiducial Model 1 (squares) and Model 3 (circles) halos. Solid lines show linear fit to Model 1 points.

have no information on the spatial distribution of model clusters within individual halos, all we can do is remove clusters brought in by satellite halos and presumably deposited outside 8 kpc from the center. Most of them were already excluded by construction: when extracting merger trees from the MM-II database, we did not include any satellite halos within the virial radius of the central halo at $z = 0$.

In addition, there could be satellites in the central halo merger tree at high redshift that have not had sufficient time to migrate toward the center and deposit their GCs within the ACS field. To identify such satellites, we estimate the dynamical friction timescale of all halos in the tree based on their mass and position information in the MM-II database. The expression for the inspiral time is derived from

the Chandrasekhar (1943) formula, with numerical corrections based on cosmological simulations, e.g., Equation (5) of Boylan-Kolchin et al. (2008):

$$\frac{\tau_{\rm merge}}{\tau_{\rm dyn}} = 0.216 \frac{(M_{\rm host}/M_{sat})^{1.3}}{\ln(1+M_{\rm host}/M_{sat})} e^{1.9\eta} \left(\frac{r}{r_{\rm vir}}\right), \tag{5.13}$$

where $\eta = j/j_c$ is the orbital circularity and $r/r_{\rm vir}$ is the distance between host and satellite halo in the unit of host virial radius. The most likely value of the circularity is $\eta \approx 0.5$, based on the orbital analysis in the simulations of Boylan-Kolchin et al. (2008). We discard the GCs of each satellite halo at every epoch since $z = 2$ (at $z > 2$ all satellites have sufficiently short merging time) that had inspiral time longer than the available time until $z = 0$. An alternative expression for $\tau_{\rm merge}$ is given by Jiang et al. (2010), but the result of this satellite removal process is very similar in both cases.

This procedure affects only 5%-10% of GCs, mainly in the metallicity range of $-2.3 < {\rm [Fe/H]} < -1.0$, and does not remove clusters with higher metallicity. Such small changes can be understood intuitively. First, massive satellites have short merger timescales so that they are almost guaranteed to merge. Small satellites with longer inspiral times contain few GCs and cannot affect the overall metallicity distribution. Second, the metallicity of these discarded GCs is roughly in the middle of the distribution, not the poorest which come from the high redshift galaxies and not the richest which come from the central halos on the main branch of the merger tree. This dip in the middle of the metallicity distribution helps to sharpen the appearance of bimodality, although the effect is small. After the exclusion of the satellite GCs, the best fit parameters of Model 1 remain the same as those in Table 5.3, with a similar goodness value $G_{0.01} = 0.38$.

In the remainder of this paper we present the results for our full model samples, because the exclusion correction is small and involves additional steps (such as the

dynamical friction time estimate) that unnecessarily complicate the model.

5.5.4 Metallicity Distribution

Figure 5.6 shows the calculated GC metallicities within four representative MM-II halos for our fiducial Model 1, and the observed samples of the Virgo galaxies matched to these halos. This figure illustrates that the model produces realistic GC populations with the multi-modal metallicity distribution. Quantitatively, both the height and location of the blue and red peaks match well with the corresponding observations, in the full halo mass range from $M_{\rm h} \sim 10^{12}$ to $10^{14}\,M_{\odot}$.

Using the Gaussian Mixture Modeling code (Muratov & Gnedin, 2010), we determine the metallicities of the two modal peaks for all 20 halos, as well as for the Virgo galaxies. We also fit the relation between the stellar mass and the peak metallicity of the blue and red populations, $\mathrm{Z} \propto \mathrm{M}_{*}^{\gamma}$. Figure 5.7 shows that the model matches the observed locations of both peaks and follows the weak trend of increasing peak metallicity with galaxy stellar mass. The best-fit slopes are $\gamma = 0.24 \pm 0.17$ and $\gamma = 0.01 \pm 0.21$ for the metal-poor and metal-rich populations, respectively.

5.5.5 Mass Distribution

In this section we examine the mass distribution of model GCs. Figure 5.8 shows the initial and final mass functions of the largest halo with $M_{\rm h} = 6.9 \times 10^{13}\,M_{\odot}$. The dynamical erosion of the GC system turns the initial power-law shape to a peaked shape, which is consistent with the theoretical expectations and observations. We fit the final mass distribution by a conventional log-normal function:

$$\frac{dN}{d\log M} = \frac{1}{\sqrt{2\pi}\sigma_M} \exp\left[-\frac{(\log M - \overline{\log M})^2}{2\sigma_M^2}\right], \tag{5.14}$$

with best-fit parameters $\overline{\log M} = 5.10$ and $\sigma_M = 0.69$. The GC mass function within VCC 1226, obtained from the $(g-z)$ color and the color-dependent mass-to-light ratio, is overplotted in the same figure. The observed distribution is similar to the modeled one for GCs more massive than $10^5\,M_\odot$. For lower mass clusters, the observed distribution falls off sharply. It is most likely due to incompleteness of the flux-limited sample. Deeper observations are needed to investigate whether the true mass function in VCC 1226 is described by a similar log-normal.

5.5.6 Best-fit Parameter Sets for Other Models

In Section 5.4, we introduced alternative models by modifying particular assumptions of Model 1. We have repeated the model comparison procedure to explore the parameter space for each of these models. Their best-fit parameters are listed in Table 5.3.

In Model 2, the no-evolution of the MMR causes a systematic shift of model GCs to higher metallicity, which cannot match the observed values of the red and blue modal peaks. As we show in Section 5.6.1, the metal-poor population, in general, is formed at high redshift around $z \approx 4-8$, when the MMR evolution would lower [Fe/H] by about 0.5 dex (see Equation (5.6)). On the other hand, the most metal-poor GCs formed in the smallest halos with stellar mass $M_* \sim 2\times10^6\,M_\odot$ can only reach [Fe/H] ≈ -1.7 without MMR evolution. Thus, the observed GCs with [Fe/H] $= -2.5$ cannot be recovered even after adding the scatter $\sigma_{\rm met} = 0.2$. The "goodness" statistic for Model 2 is quite low (2%) even when we vary all the parameters in a wide range. Such a significant difference from the results of Model 1 may imply that the moderate evolution of the MMR with cosmic time is favored by the GC systems of massive early-type galaxies. However, at the current stage it is difficult to constrain the exact amount of the evolution because of other intrinsic uncertainties in the model.

Is the metallicity scatter necessary? An alternative way of populating the metal-

poor tail without the scatter is to apply stronger MMR evolution. To investigate this possibility, we set $\sigma_{\rm met} = 0$ and vary the coefficient in the second term of Equation (5.6) from 0.216 to 0.5. However, for all of these values, the goodness of fit is low ($G_{0.01} < 5\%$), which indicates that boosting the MMR evolution alone cannot substitute for the effect of scatter. Stronger MMR evolution creates several problems. First, both blue and red GC populations are shifted to lower metallicity, which leads to incorrect peak positions. At the same time, the metal-rich clusters with [Fe/H] > 0 are difficult to form without the scatter. Second, the scatter not only helps to fill both tails of the distribution ([Fe/H] < -2 and [Fe/H] > 0) but it also regulates the width of the two populations. Without the scatter, the metallicity distribution is more like a sum of delta functions rather than a Gaussian shape, especially for the metal-rich GCs formed by late discrete mergers.

For both Model 3 and Model 4, the contours of goodness $G_{0.01}$ are shown in Figure 5.9. Model 3 has two peaks with $G_{0.01} > 40\%$ (marked by red stars), which are as good as Model 1. However, the peak at $(p_2, p_3 = 2.0, 0.32)$ does not have as high a number of sufficient GC systems. We have determined the red and blue peak metallicities for this model and added them to Figure 5.7 for comparison with Model 1. The average peak metallicities are similar, but the dispersion in Model 3b is much larger. The combination of these comparisons makes us prefer Model 1 as the fiducial model.

The highest goodness of Model 4 is only $G_{0.01} = 18\%$, which means that fewer than one in five of the galaxy-halo pairs have acceptable metallicity distributions, significantly lower than in both Model 1 and Model 3. The best-fit critical sSFR to trigger cluster formation is $\mathrm{sSFR_{crit}} \approx 1.3\ \mathrm{Gyr}^{-1}$. We also tried using the critical SFR as the trigger parameter and found it to be even more difficult to reproduce the observations. The low goodness of Model 4 indicates that the major merger scenario may indeed be a dominate formation channel of GCs, at least in our model.

It should be mentioned that since our GC formation model is based on the merger tree extracted from the MM-II simulation with only 67 outputs along the whole cosmic time, the SFR we derive here is the average between two adjacent outputs. This averaging smoothes out short starburst events. Until we have simulations with high enough time resolution, the short-duration effects cannot be incorporated correctly. It is still interesting to investigate the differences between results of the merger (Model 1) and starburst (Model 4) scenarios. Figure 5.10 shows the formation redshifts of GCs in the two models for a $2 \times 10^{12}\, M_{\odot}$ halo. In Model 4, GC formation activity increases continuously toward relatively low redshift, $z \approx 1-2$. In contrast, Model 1 shows two formation epochs, one at low redshift when the last major merger happened between massive halos and another at higher redshift ($z \approx 5$) when mergers among small halos happened frequently. These differences in the formation history, together with the halo mass growth and MMR evolution, translate into the final GC metallicity distribution. Figure 5.11 illustrates how the metallicity bimodality is produced by discreteness of the late mergers.

In contrast, continuous field star formation, during and between mergers, does not lead to a bimodal metallicity distribution. In order to show it within our framework, we modeled the field metallicity distribution as a mass-weighted sum of stellar populations formed at each simulation output. We calculated the mass increments of all halos (central and satellite) in the merger tree between successive outputs, converted them to stellar mass using the stellar mass-halo mass relation (Behroozi et al., 2013a), and evaluated the metallicity of such a stellar population using the evolving MMR (Eq. (5.6)). The sum of these contributions roughly represents the metallicity distribution of the field stars. This distribution is clearly unimodal, in agreement with well-known observations (e.g. Harris & Harris, 2002). For example, the peak metallicity of a $2 \times 10^{12}\, M_{\odot}$ halo is at [Fe/H] ≈ -0.22 and the median is around [Fe/H] ≈ -0.28. In contrast, as we can see in Figure 5.7, the two peaks of

Table 5.3: Comparison of Best-fit Model Parameters

Model	p_2	p_3	sSFR (Gyr^{-1})	Goodness $G_{0.01}$
Model 1	2.6	0.33	–	0.40
Model 2	2.6	0.33	–	0.02
Model 3a	2.0	0.22	–	0.46
Model 3b	2.0	0.32	–	0.48
Model 4	1.6	–	1.3	0.18

the GC metallicity distribution are at [Fe/H] ≈ -1.54 & -0.4 for the metal-poor and metal-rich populations, respectively, with the median at [Fe/H] ≈ -1.30. This comparison shows that, although GCs and field stars are forming concurrently, the gas-rich merger-driven cluster formation filters a bimodal metallicity distribution from an extended unimodal one.

To further investigate the quality of fit of different models, in Figure 5.12 we present the full cumulative distribution of the KS test p-values for the best parameters of each model (Table 5.3). Most of the p-values of Model 4 are below 10^{-3}, so that the cumulative probability at $p_{KS} \gtrsim 1\%$ is already far above that of Model 1 and Model 3. The performance of Model 1 and Model 3a is fairly similar, which indicates that our major merger scenario is not too sensitive to the details of the cold gas modeling. Model 3b appears a little better than the other two, but as we discussed above, it cannot reproduce the number of GCs as well and has a larger scatter of the modal peak metallicities (Figure 5.7).

5.6 Discussion

Ashman & Zepf (1992) proposed the idea that GCs can be formed in mergers between gas-rich galaxies, since such mergers can perturb the gravitational poten-

tial, shock and compress the ISM within the two galaxies, and trigger large-scale starbursts. HST observations have already demonstrated many interacting galaxies with young massive star clusters, whose formation was likely triggered by merging (e.g., Holtzman et al., 1992; Whitmore, 2004; Larsen, 2009). Our best-fitting Model 1 suggests a minimum merger ratio of 1:3 for triggering cluster formation, consistent with this major merger scenario.

Tonini (2013) proposed a model for metallicity bimodality based on the observed number of clusters as a function of galaxy mass. She adopted the merger scenario and used Monte Carlo sampling to build merger trees for the progenitor galaxies. She suggested that the origin of GC bimodality is related mainly to the galactic MMR and hierarchical mass assembly history. Using our model, we reach a similar conclusion that the merger history plays a key role. However, to separate the red and blue peaks, the Tonini (2013) model requires a very strong evolution of MMR, such that [Fe/H] increases by 0.5 dex at high mass ($M_* \sim 10^{11.5}\,M_\odot$) and up to 1.5 dex at low mass ($M_* \sim 10^9\,M_\odot$), between $z = 3.5$ and $z = 0$. The available observations discussed in Section 5.2.4 support much smaller changes of [Fe/H] at a given stellar mass.

Our model also relies on the evolution of MMR, but the evolution we need is more moderate (≈ 0.3 dex). The key to separating the metal-poor and metal-rich subpopulations in our model is mainly due to the differentiation of cluster hosts. The metal-poor GCs come preferentially from the early mergers among small halos with lower metallicity, while the metal-rich GCs come from the late mergers between massive halos, which in turn have higher metallicity.

To demonstrate this effect, we select two halos with the highest and lowest mass ($6.9 \times 10^{13}\,M_\odot$ and $2 \times 10^{12}\,M_\odot$, respectively) and separate their GC systems by the merger epoch in which they were produced. Figure 5.13 shows the result for the best-fit parameters of Model 1. Although the dynamical destruction significantly reduces the number of surviving clusters, the shape of the metallicity distribution does not

change much from that imprinted at birth. On the other hand, the merger events that produced the clusters leave a clear mark. The metal-rich GCs are mainly produced by the most recent merger between massive halos, while the collection of early mergers among less massive halos contributes the bulk of the metal-poor clusters.

5.6.1 Age-metallicity Relation

Absolute ages of GCs can be determined using isochrone fitting of the H-R diagram, which requires resolved observations of individual stars. Until recently, the age measurements of Galactic GCs did not show a significant correlation between age and metallicity (e.g. Forbes & Bridges, 2010). New deep HST/ACS data reveal some intriguing trends of decreasing age with increasing metallicity (Dotter et al., 2011; VandenBerg et al., 2013; Leaman et al., 2013). In our model we have the full formation history of all GCs, which allows us to investigate any possible age-metallicity trends. Figure 5.14 shows a stack of all model GCs within the 20 halos in the fiducial Model 1. Although the majority of clusters are old, there is a significant tail of metal-rich clusters that are younger by up to 5 Gyr. The bulk of metal-poor clusters are formed as early as redshift $z = 4 - 6$, but the metal-rich clusters are formed over an extended epoch continuing to $z \approx 1$. This shape of the age-metallicity distribution is one of the robust predictions of our model.

Hints of the age-metallicity relation were already present in the MG10 model (see their Figure 8). Here we quantify it with larger samples of clusters, multiple independent realizations of the mass assembly history, and better galactic scaling relations. Figure 5.15 compares the model relation with the existing age measurements of the Galactic GCs, collected by Leaman et al. (2013). We also add clusters from three nearby early-type galaxies with photometrically derived ages Georgiev et al. (2012). Despite the large scatter and individual observational errors, the model trend is supported by these data remarkably well.

A thorough interpretation of this plot requires further study. It is likely that the turnover of the age-metallicity relation from the old metal-poor clusters to the younger metal-rich clusters occurs at different metallicity in galaxies of different mass. For example, in the MG10 model tuned for the Galactic GCs, the turnover is around [Fe/H] ≈ -0.8, whereas in our current model tuned for massive elliptical galaxies it is around [Fe/H] ≈ -0.4. In addition, the ages of the extragalactic clusters are determined with a different method and different fidelity than those of the Galactic GCs. Nevertheless, the emerging age-metallicity relation of GCs is tantalizing and invites further accurate measurements of cluster ages in extragalactic systems.

5.7 Summary

We have constructed a model of cluster formation, incorporating the halo merger trees from the MM-II cosmological N-body simulation, to investigate the origin of GC systems in massive early-type galaxies. We include the empirical galactic scaling relations, such as the stellar mass-halo mass relation, stellar mass-gas mass relation, and stellar MMR. These come either from direct observations or from the empirical abundance matching technique. We test the scenario in which clusters are formed as a result of major mergers of gas-rich galaxies. By matching the masses of our selected halos with the galaxies in the Virgo cluster, we compare the metallicity distributions of modeled and observed GCs and thus constrain the model parameters. We have also tested alternative models in order to examine the sensitivity of our results to various adopted prescriptions. Our main conclusions are listed below:

- Our fiducial model can successfully reproduce both the number and the metallicity distribution of GCs within a large range of halo masses from $2 \times 10^{12}\,M_{\odot}$ to $6.9 \times 10^{13}\,M_{\odot}$. The metallicity distribution appears to have a bimodal shape, and the metallicities of the blue and red peaks are consistent with those observed

in the Virgo galaxies.

- The fiducial model requires a minimum merger ratio of 1:3 to trigger cluster formation. This ratio is consistent with the theoretical expectation of a major merger.

- A detailed analysis of the formation history of GCs reveals that the bimodality arises from different merger epochs and host galaxy masses: the metal-rich population is produced by late mergers between massive halos, while the metal-poor population is produced by early mergers among less massive halos.

- The model predicts a robust age-metallicity relation of GCs, which can be falsified by further observations. While the bulk of metal-poor clusters are very old, the metal-rich clusters are progressively younger, by up to 5 Gyr.

- When the evolution of the galaxy MMR with cosmic time is turned off, the model GC metallicity distribution shifts to higher [Fe/H] and the bimodal distribution disappears. This suggests that the evolution of MMR is necessary in our model.

- The evolution of the cold gas fraction within galaxies at high redshift is largely unconstrained by current observations. We use different methods to parameterize this evolution and find that the best-fitting model results for the GC number and metallicity distributions are insensitive to the details of the adopted prescriptions, within the considered range.

- We also challenged our major merger scenario and tested an alternative starburst scenario, which required a minimum sSFR to trigger cluster formation. Because of the smooth behavior of the average sSFR derived from the abundance matching, the alternative model fails to reproduce the observed metallicity distribution as well as the merger model.

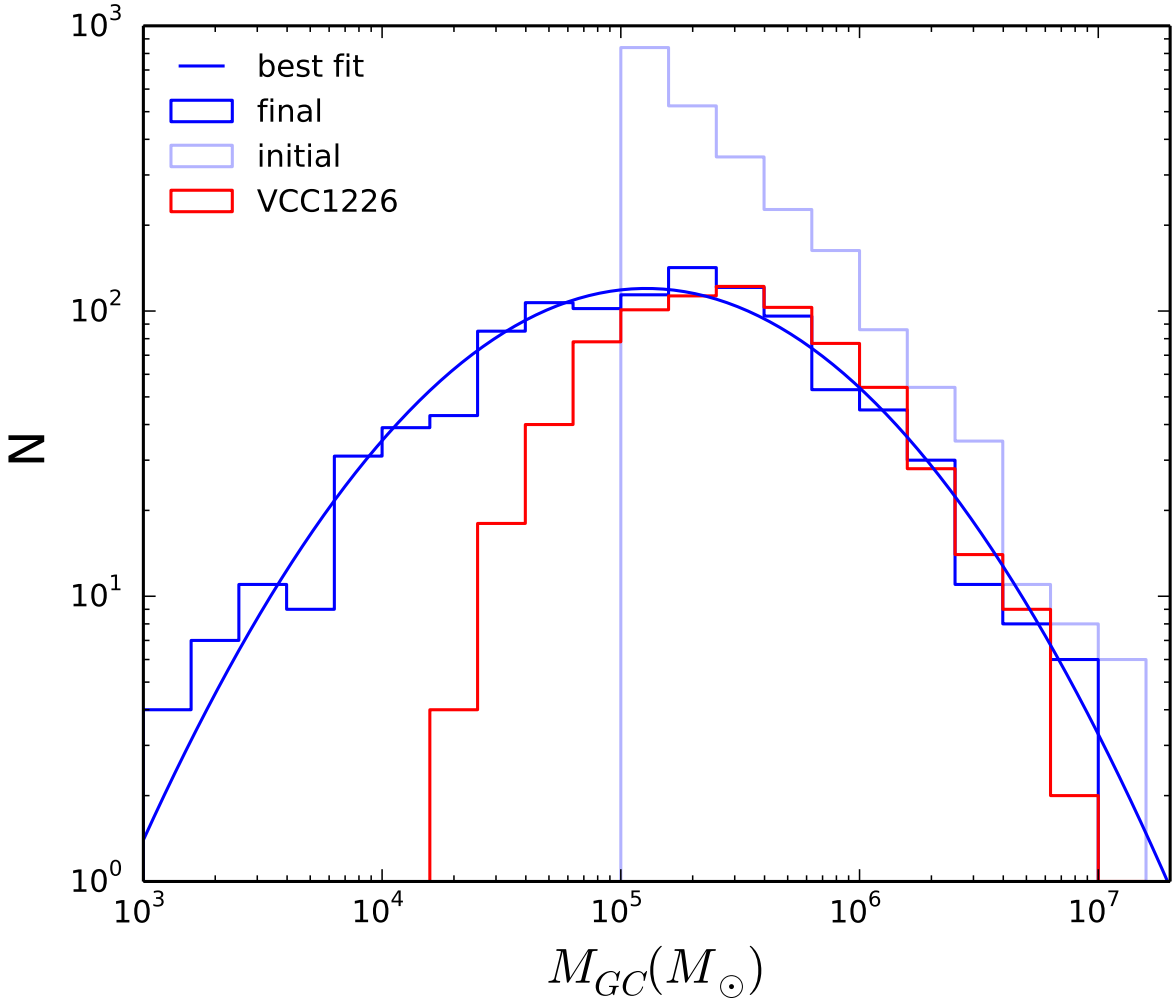

Figure 5.8: Dynamical evolution of the GC mass function from an initial power law (light blue histogram) to the current peaked distribution (blue histogram, with the overplotted log-normal fit), in the fiducial Model 1 for a halo of $6.9 \times 10^{13}\, M_\odot$. The mass function of the GC system in VCC 1226 is shown for comparison (red histogram). A sharp drop-off at low mass is likely due to incompleteness of the observed sample. The KS test comparison of the model and data shows that they are consistent; $p_{KS} \approx 2\%$ for clusters with $M > 10^5\, M_\odot$.

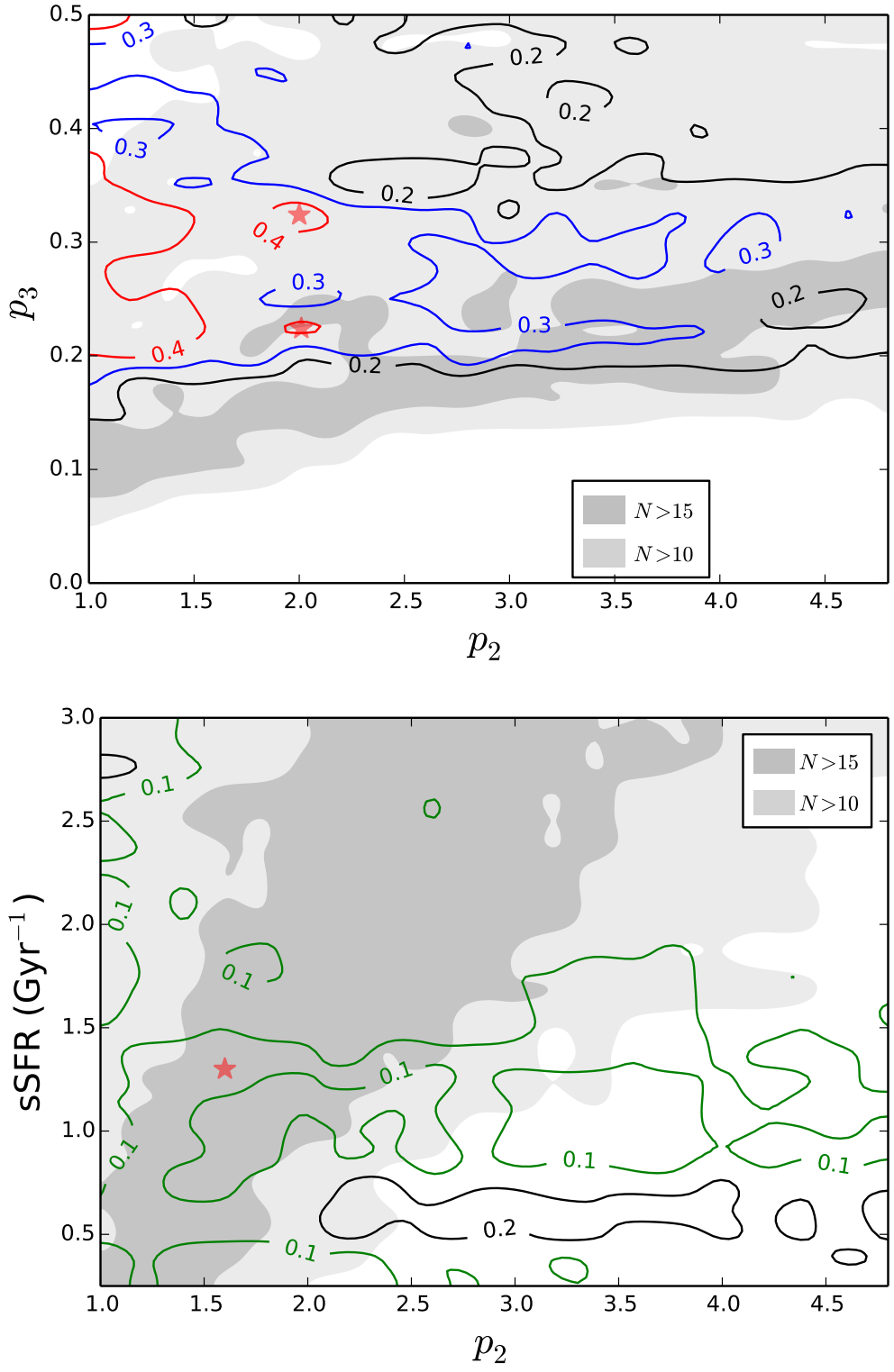

Figure 5.9: Same as Figure 5.5, but for the alternative Model 3 (top panel) and Model 4 (bottom panel). Best-fit parameters for these models are given in Table 5.3.

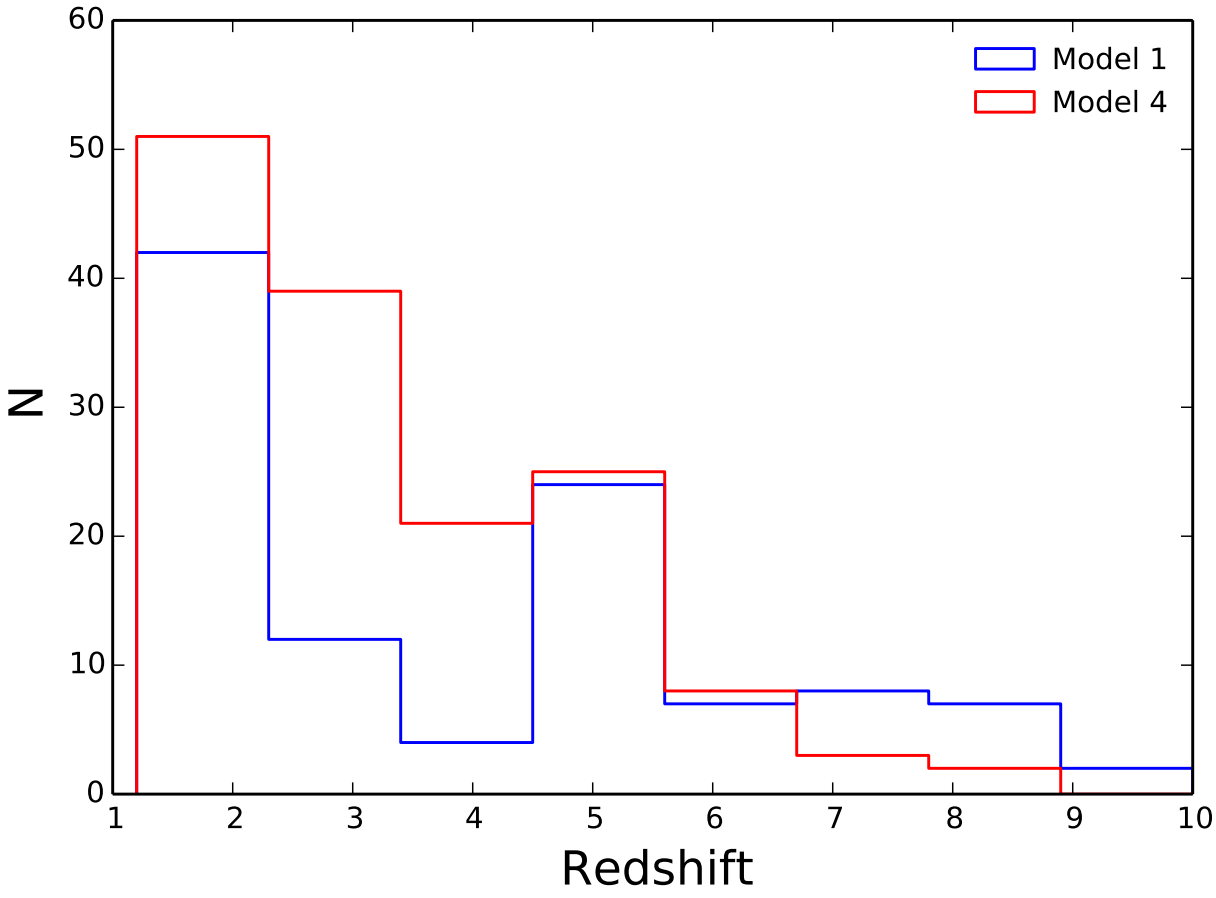

Figure 5.10: Distribution of GC formation redshift in Model 1 and Model 4 within a $2 \times 10^{12}\, M_{\odot}$ halo.

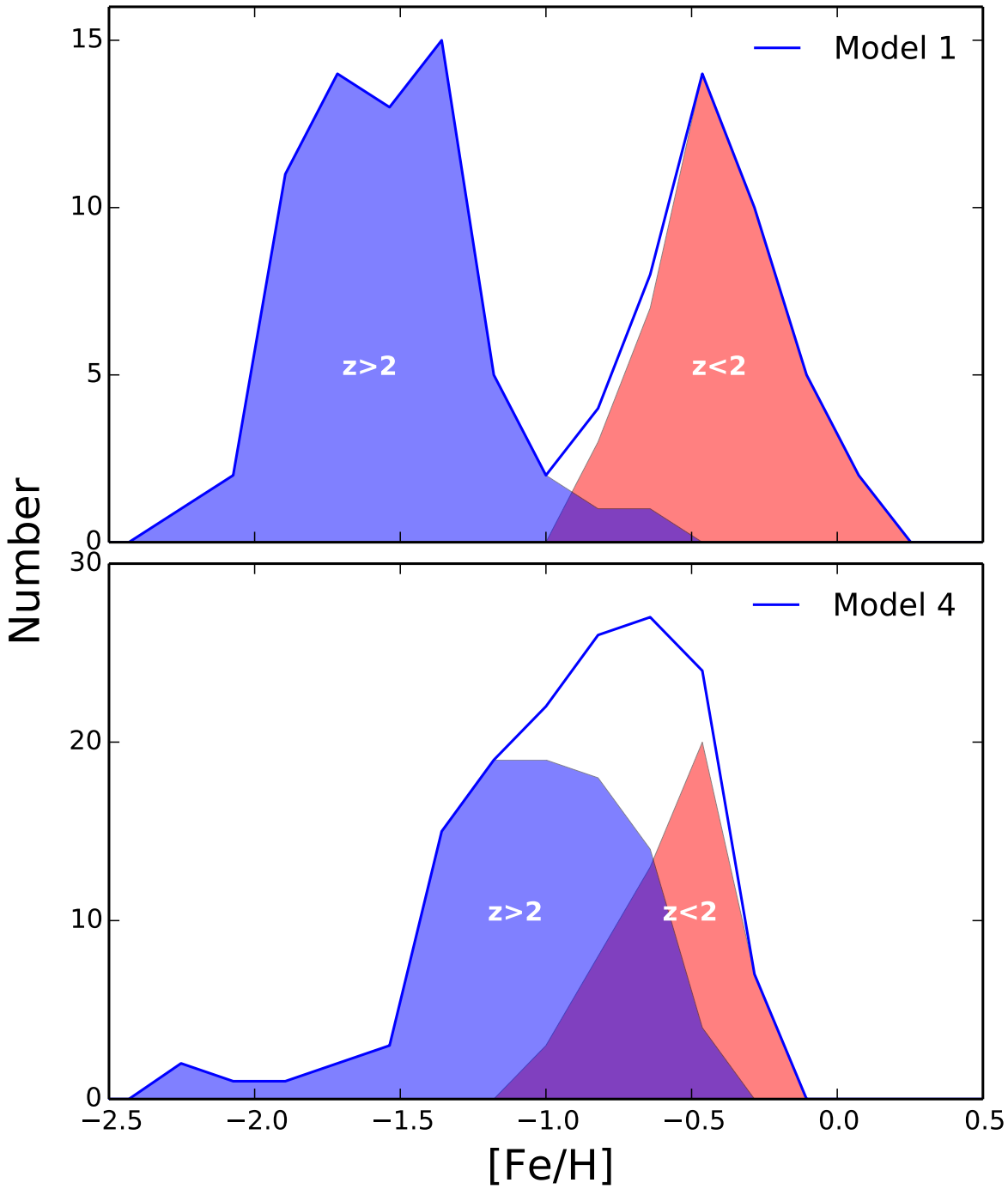

Figure 5.11: Metallicity distribution of GCs within a $2 \times 10^{12}\, M_{\odot}$ halo for Model 1 (*upper panel*) and Model 4 (*lower panel*). The samples are split into two groups based on formation redshift: $z < 2$ (*red shaded*) and $z > 2$ (*blue shaded*).

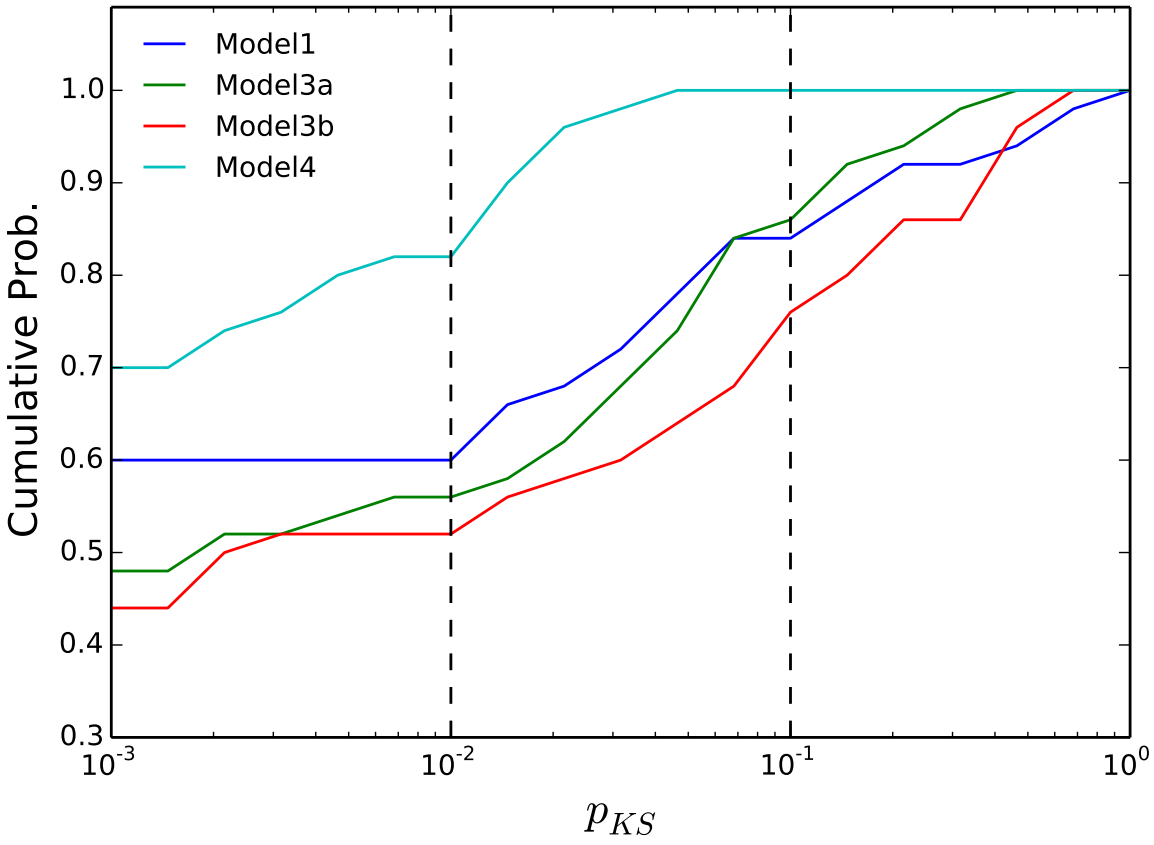

Figure 5.12: Cumulative distribution of p-values of the KS test for the metallicity distribution, for all best-fitting models. The vertical scale is related to the "goodness" parameter as $1 - G_{p_{KS}}$. Lower lines have higher "goodness" of fit.

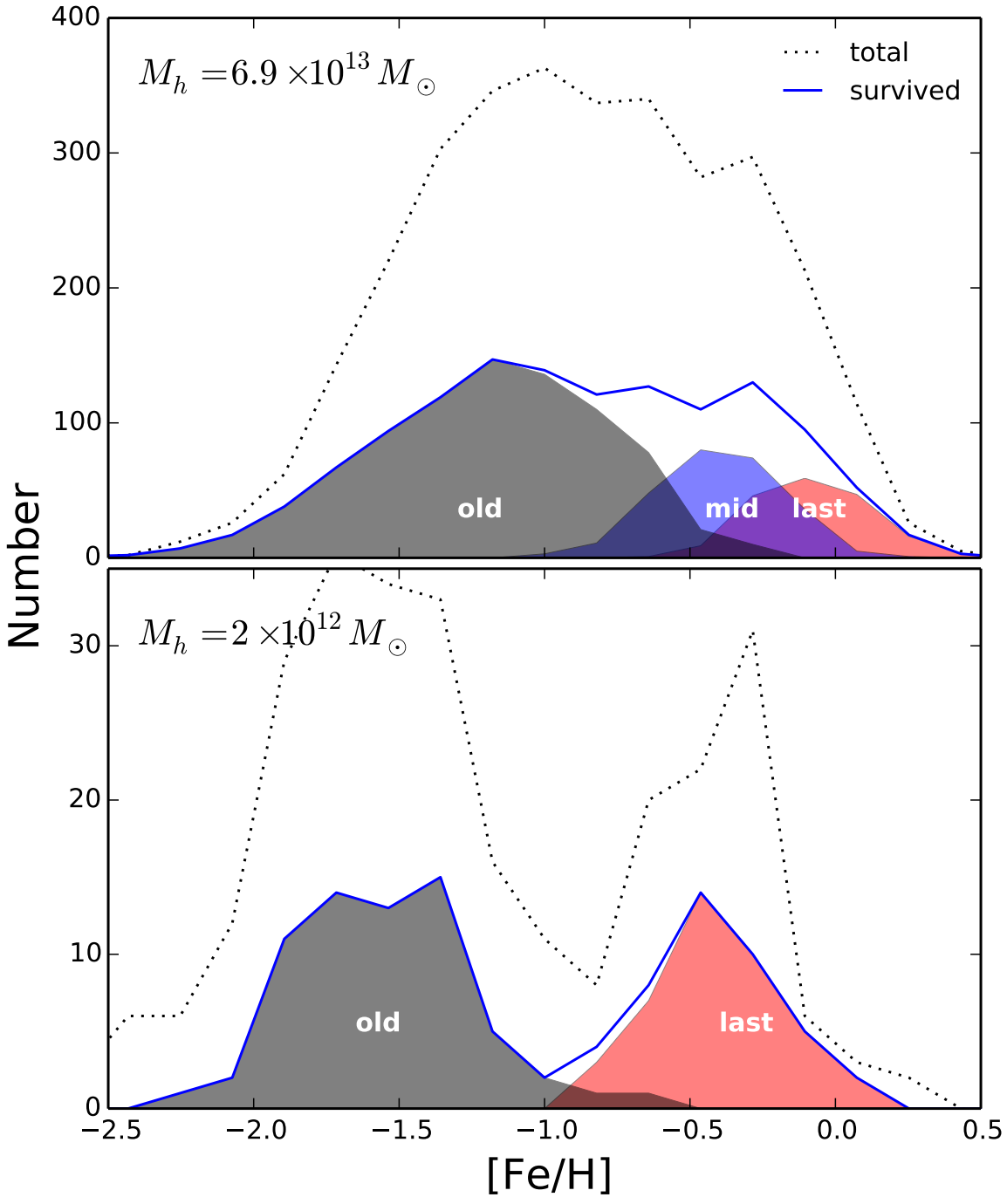

Figure 5.13: Metallicity distributions of the total (dotted lines) and survived (blue solid lines) GC systems within $6.9 \times 10^{13} M_\odot$ halo (upper panel) and $2 \times 10^{12} M_\odot$ halo (lower panel). The distributions are also split by the merger events that produced the clusters: late mergers (red shaded), intermediate mergers (blue shaded), and early mergers (gray shaded).

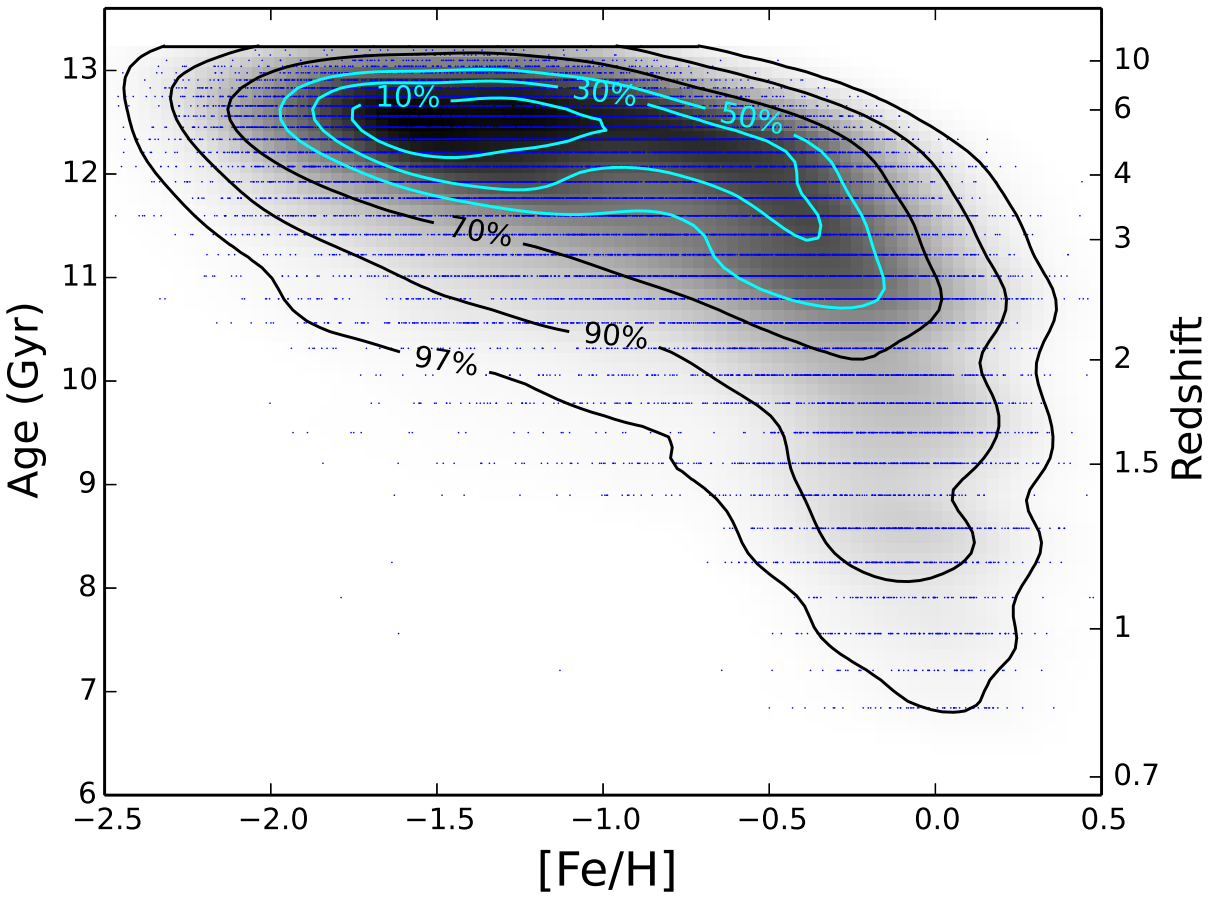

Figure 5.14: Age-metallicity distribution of model clusters from all 20 halos, with the best-fit parameters of Model 1. Each blue dot represents one model GC. The labeled percentage levels represent the fraction of GCs enclosed within the corresponding contour. Redshift corresponds to the cluster formation epoch.

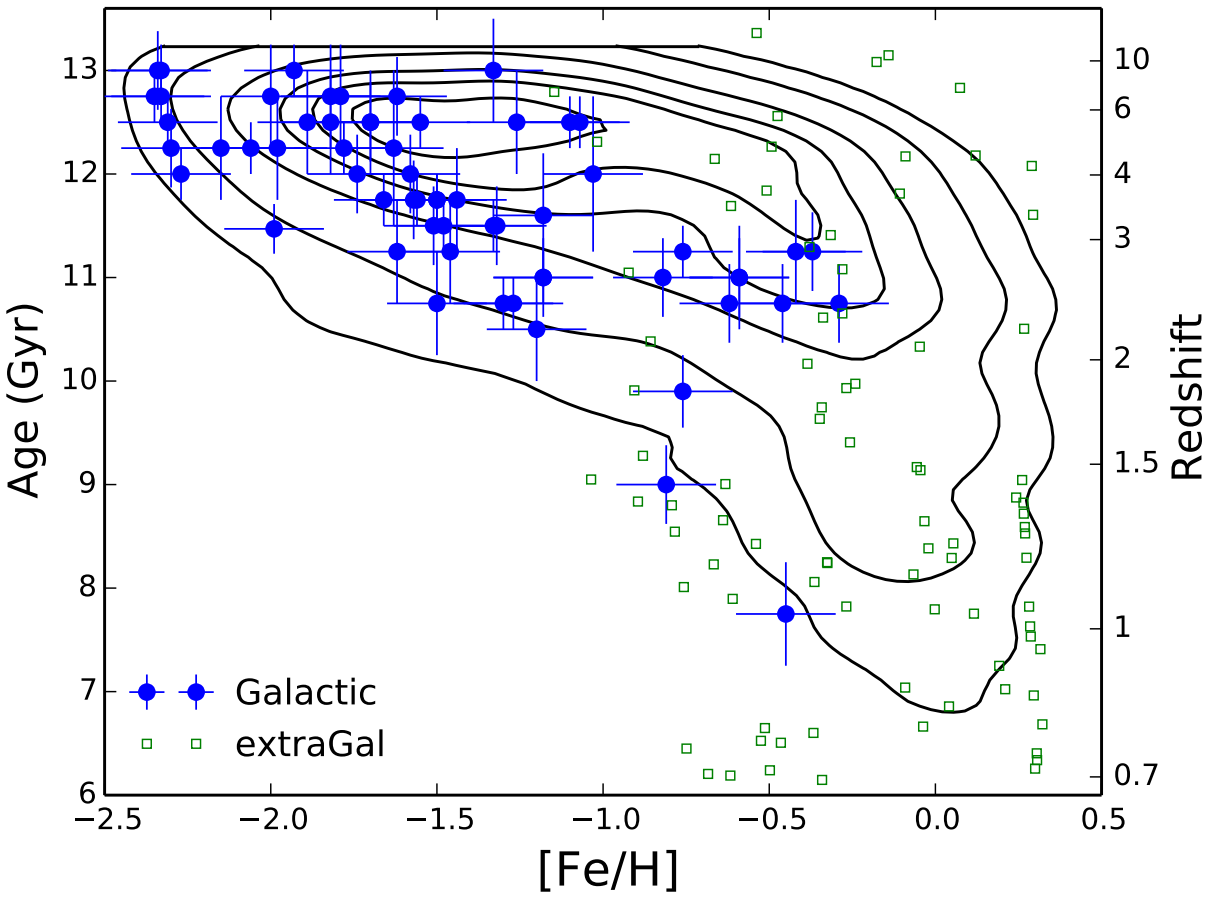

Figure 5.15: Age-metallicity relation of Galactic GCs (filled circles with error-bars; from Leaman et al. 2013 and other sources, see the text) and extragalactic GCs (open squares; from Georgiev et al. 2012). Overlaid contours are the same as in Figure 5.14, for our model clusters.

CHAPTER VI

Summary and Future Works

Let me first draw your attention back to the current issues on the sub-grid models in cosmological simulations I mention in Chapter I: lack of a reliable prescription for star formation and lack of methods to calibrate the subgrid models on scales of star-forming regions. These two issues are responded at the same time with my new cluster formation prescription.

In my thesis, I develop a new implementation of star formation in cosmological simulations, continuous cluster formation (CCF), by considering star clusters as a unit of star formation, inspired by observations that most stars form in clusters. In CCF, a cluster particle grows its mass through gas accretion within a star-forming sphere. The accretion is terminated by its own feedback, thus the final mass is set self-consistently. I also introduce a prescription that describe the initial bound fraction of individual star clusters, f_i, to estimate the mass fraction that is remind bound to the cluster when it emerges from the giant molecular clouds (GMCs). I implement CCF in the Eulerian gasdynamics and N-body Adaptive Refinement Tree code. I perform a series of high resolution cosmological simulations of a Milky-Way-sized galaxy with 3D radiative transfer, non-equilibrium chemical network, and H_2 formation and destruction.

I find that the global properties of the galaxies, such as the galaxy morphology

and SFH, are strongly affected by the strength of feedback. To match the SFH from the abundance matching result of a Milky Way-sized galaxy, the momentum boosting factor of the SNR feedback is tightly constrained to be in the range $f_{\rm boost} = 3 - 10$. On the other hand, these global properties are not sensitive to the choice of star formation efficiency $\epsilon_{\rm ff}$, thus cannot be used to constrain the parameters of star formation prescription. Fortunately, the value of $\epsilon_{\rm ff}$ shows a dramatic effect on the properties of modeled star clusters, which can be used to calibrate $\epsilon_{\rm ff}$ on a scale that is compatible to the size of GMCs.

We show that our cluster formation model leads to a large variation of the integral star formation efficiency even when $\epsilon_{\rm ff}$ is kept constant. The range of this variation spans about two orders of magnitude, similar to recent observations of star formation efficiency within GMCs. This indicates that the cluster formation prescription is a realistic way of model star formation in the simulations and captures the detail gas density evolution in star-forming regions with continuous gas infall and stellar feedback. I find the initial bound fraction, f_i, increases strongly with cluster masses, and this trend is independent of the global galactic environment. However, f_i shows a strong positive correlation to $\epsilon_{\rm ff}$. This correlation leads to another positive correlation between the maximum cluster mass and $\epsilon_{\rm ff}$. I find that the cluster initial mass function is best described by a power-law with an exponential cutoff. The cutoff mass scales with the star formation rate of the host galaxies, suggesting that cluster formation depends strongly on galactic environment. The power-law slope of the CIMF depends on the value of $\epsilon_{\rm ff}$. I find runs with $\epsilon_{\rm ff} = 0.5 - 1.0$ have a CIMF with a slope similar to -2, best matches the observed slopes. I measure the integrated cluster fraction and find it correlates with star formation rate surface density. The normalization of this correlation depends strongly on $\epsilon_{\rm ff}$, with $\epsilon_{\rm ff} = 0.5 - 1.0$ best matches the observational measurements. Finally, I find a clear trend that cluster formation timescale is shorter with higher $\epsilon_{\rm ff}$: $\tau_{\rm ave} \propto \epsilon_{\rm ff}^{-1/2}$. Since clusters in most runs have cluster formation

timescale that is shorter than 3 Myr, they are all within the range of the observed age spread for young star cluster. One exception is the SFEturb run, in which $\sim 30\%$ of massive clusters have a timescale larger than 6 Myr, which is not consistent with observations. Based on various diagnostics such as CIMF, integrated cluster fraction, and cluster formation timescale, I conclude that $\epsilon_{\rm ff} = 0.5 - 1.0$ is preferred under the current setup.

Beside the cluster formation prescription, I implement a new algorithm to calculate the strength of the tidal field around each cluster particle along its orbits and model its mass loss via stellar evolution, internal evaporation, and tidal disruption. This is the first time that cluster disruption under realistic galactic environment is estimated in the simulation runtime. With high spatial resolution of the simulations, the tidal interaction between clusters and dense structures on the gaseous disk can be captured. I study the dynamical evolution of the massive star clusters across cosmic time and characterize the evolution of its metallicity distribution. I find the strength of the tidal field decreases as cluster particles orbiting away from the central, dense part of the galaxy. Tidal disruption in our simulations is crucial to the mass loss of clusters during the first Gyr. With the combination of tidal disruption and internal evaporation, the shape of the bound cluster mass function changes from power-law to log-normal across cosmic time. The obtained log-normal mass function is similar to the observed mass function of GCs, suggesting that young massive clusters formed at high redshifts are promising candidates of the progenitor of GCs. Interestingly, the dynamical evolution preferentially removes clusters with median metallicity and generates a dip at [Z/H] $= -1.5 - -1.0$. Thus a bimodal shape metallicity distribution is emerged for survival clusters, very similar to the metallicity distribution of the Milky Way GC populations.

The main issue of full cosmological simulations is that these simulations are extremely time consuming. One shortcut is to apply an phenomenological model of the

formation and evolution of GCs to the hierarchical structure formation framework. Following this philosophy, I construct a semi-analytical model of GC formation and evolution onto the halo merger trees in Millennium-II simulations. We test the scenario in which clusters are formed as a result of major mergers of gas-rich galaxies. We compare the metallicity distributions of modeled and observed GCs and thus constrain the model parameters. The fiducial model successfully reproduce both the number and the metallicity distribution of GCs within a large range of halo masses from $2 \times 10^{12}\,M_{\odot}$ to $6.9 \times 10^{13}\,M_{\odot}$. The metallicity distribution appears to have a bimodal shape, and the metallicities of the blue and red peaks are consistent with those observed in the Virgo galaxies. The bimodality arises from different merger epochs and host galaxy masses: the metal-rich population is produced by late mergers between massive halos, while the metal-poor population is produced by early mergers among less massive halos. The model predicts a robust age-metallicity relation of GCs, which can be falsified by further observations. I find that, while the bulk of metal-poor clusters are very old, the metal-rich clusters are progressively younger, by up to 5 Gyr.

6.1 Future Works

There are many immediate extension to my current work on modeling star cluster formation and evolution in cosmological simulations. The most obvious one is to run simulations in a larger simulation volume that contains galaxies of different masses and types. With the large volume, I can study various global properties of galaxies with my realistic star formation prescription and compare the similarities and differences to simulations that use old prescriptions, especially on the pc-scale structures. More importantly, I will study the properties of model clusters in different host galaxies with different environments, for example the changes of CIMF in galaxies with different masses. Another interesting properties of star clusters is their

spatial distribution. I will calculate the two-point correlation function of the young star clusters across the galactic disks. I will explore how this correlation function depends on the star formation and feedback parameters. I will also study whether this correlation function varies for different host galaxy mass or merger status. This is another potential observable that can be used to constrain subgrid models in the simulations.

For the tidal disruption implementations, I only show some preliminary results from one of my simulations. In the near future, I will compare the CIMF evolution for different runs. I will study the unique physical conditions that lead to the formation and survival of most massive clusters. As I show in Chapter 4, early disruption of star clusters is crucial in our simulations. I will analyze the disruption rate of clusters at early time when the clusters are still residing in the gaseous disk. I will examine how disruption rate depends cluster masses, which is still hotly debated in observations (Bastian et al., 2012). I will study the metallicity distribution of survival clusters in the simulations, especially the origin of its apparent bimodal shape. This study is attempt to solve to mystery of GC formation that is puzzling for a long time.

The disruption prescription used in my simulations is calibrated based on the simulations of dynamical evolution of individual star clusters with N-body code. These simulations typically assume a cluster orbiting around a static, analytical galactic potential. This simplification ignores either the fine structure of the ISM or the hierarchical structure formation of the galaxy. With my current cosmological simulations and numerical treatments on estimating the tidal intensity, I am now able to construct the time evolution of the tidal field for a given clusters along its lifetime. I will connect this time-dependent field to the N-body code and study the dynamical evolution of star clusters in the real galactic environment. I will study the effects of different types of tides to the mass loss and the radius of the cluster. I will calibrate a new mass loss relation based on the N-body results that can be used in future

cosmological simulations.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aarseth, S. J. 1963, MNRAS, 126, 223

Aarseth, S. J., & Binney, J. 1978, MNRAS, 185, 227

Adamo, A., Kruijssen, J. M. D., Bastian, N., Silva-Villa, E., & Ryon, J. 2015, MNRAS, 452, 246

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

—. 2016, ApJ, 824, 79

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

Agertz, O., Teyssier, R., & Moore, B. 2011, MNRAS, 410, 1391

Alpher, R. A., & Herman, R. C. 1950, Reviews of Modern Physics, 22, 153

Ashman, K. M., & Zepf, S. E. 1992, ApJ, 384, 50

Aumer, M., White, S. D. M., Naab, T., & Scannapieco, C. 2013, MNRAS, 434, 3142

Ballesteros-Paredes, J., & Hartmann, L. 2007, Rev. Mexicana Astron. Astrofis., 43, 123

Barnes, J., & Hut, P. 1986, Nature, 324, 446

Bastian, N. 2008, MNRAS, 390, 759

Bastian, N., Adamo, A., Gieles, M., et al. 2012, MNRAS, 419, 2606

Bate, M. R., Bonnell, I. A., & Price, N. M. 1995, MNRAS, 277, 362

Baumgardt, H., & Makino, J. 2003, MNRAS, 340, 227

Begeman, K. G. 1989, A&A, 223, 47

Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013a, ApJ, 762, L31

—. 2013b, ApJ, 770, 57

Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013c, ApJ, 762, 109

Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013d, ApJ, 763, 18

Behroozi, P. S., Zhu, G., Ferguson, H. C., et al. 2015, MNRAS, 450, 1546

Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJS, 149, 289

Bernardi, M., Meert, A., Sheth, R. K., et al. 2013, MNRAS, 436, 697

Bertschinger, E. 1998, ARA&A, 36, 599

Bigiel, F., Leroy, A., Walter, F., et al. 2008, AJ, 136, 2846

Bigiel, F., Leroy, A. K., Walter, F., et al. 2011, ApJ, 730, L13

Bond, J. R., Kofman, L., & Pogosyan, D. 1996, Nature, 380, 603

Booth, C. M., Agertz, O., Kravtsov, A. V., & Gnedin, N. Y. 2013, ApJ, 777, L16

Bothwell, M. S., Maiolino, R., Kennicutt, R., et al. 2013, MNRAS, 433, 1425

Boylan-Kolchin, M., Ma, C.-P., & Quataert, E. 2008, MNRAS, 383, 93

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009a, MNRAS, 398, 1150

—. 2009b, MNRAS, 398, 1150

Briceno, C., Hartmann, L. W., Stauffer, J. R., et al. 1997, AJ, 113, 740

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

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

Cen, R., & Ostriker, J. P. 1992, ApJ, 399, L113

Ceverino, D., Klypin, A., Klimek, E. S., et al. 2014, MNRAS, 442, 1545

Chandrasekhar, S. 1943, ApJ, 97, 255

Cho, J., Sharples, R. M., Blakeslee, J. P., et al. 2012, MNRAS, 422, 3591

Coles, P., & Lucchin, F. 2002, Cosmology: The Origin and Evolution of Cosmic Structure, Second Edition, 512

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

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

Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, ApJ, 647, 201

Cullen, F., Cirasuolo, M., McLure, R. J., Dunlop, J. S., & Bowler, R. A. A. 2014, MNRAS, 440, 2300

Dalla Vecchia, C., & Schaye, J. 2012, MNRAS, 426, 140

De Marchi, G., Paresce, F., & Pulone, L. 2007, ApJ, 656, L65

Dotter, A., Sarajedini, A., & Anderson, J. 2011, ApJ, 738, 74

Evans, II, N. J. 1999, ARA&A, 37, 311

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

Evans, N. W., An, J., & Walker, M. G. 2009, MNRAS, 393, L50

Evrard, A. E. 1988, MNRAS, 235, 911

Fall, S. M., & Rees, M. J. 1985, ApJ, 298, 18

Federrath, C. 2015, MNRAS, 450, 4035

Federrath, C., Banerjee, R., Clark, P. C., & Klessen, R. S. 2010, ApJ, 713, 269

Federrath, C., & Klessen, R. S. 2012, ApJ, 761, 156

Feldmann, R. 2013, MNRAS, 433, 1910

Feldmann, R., & Gnedin, N. Y. 2011, ApJ, 727, L12

Ferrarese, L., Côté, P., Jordán, A., et al. 2006, ApJS, 164, 334

Forbes, D. A., & Bridges, T. 2010, MNRAS, 404, 1203

Forbes, D. A., Pastorello, N., Romanowsky, A. J., et al. 2015, MNRAS, 452, 1045

Frenk, C. S., White, S. D. M., Bode, P., et al. 1999, ApJ, 525, 554

Friedmann, A. 1922, Zeitschrift fur Physik, 10, 377

Gallazzi, A., Bell, E. F., Zibetti, S., Brinchmann, J., & Kelson, D. D. 2014, ApJ, accepted, arXiv:1404.5624, arXiv:1404.5624

Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, MNRAS, 362, 41

Gamow, G. 1946, Physical Review, 70, 572

Geller, M. J., & Huchra, J. P. 1989, Science, 246, 897

Gentry, E. S., Krumholz, M. R., Dekel, A., & Madau, P. 2017, MNRAS, 465, 2471

Georgiev, I. Y., Goudfrooij, P., & Puzia, T. H. 2012, MNRAS, 420, 1317

Geyer, M. P., & Burkert, A. 2001, MNRAS, 323, 988

Gieles, M., Larsen, S. S., Bastian, N., & Stein, I. T. 2006, A&A, 450, 129

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

—. 2016, ArXiv e-prints, arXiv:1603.07729

Gnedin, N. Y., & Abel, T. 2001, New A, 6, 437

Gnedin, N. Y., & Kravtsov, A. V. 2010, ApJ, 714, 287

—. 2011, ApJ, 728, 88

Gnedin, N. Y., Kravtsov, A. V., & Rudd, D. H. 2011, ApJS, 194, 46

Gnedin, O. Y., Maccarone, T. J., Psaltis, D., & Zepf, S. E. 2009, ApJ, 705, L168

Gnedin, O. Y., Ostriker, J. P., & Tremaine, S. 2014, ApJ, 785, 71

Goddard, Q. E., Bastian, N., & Kennicutt, R. C. 2010, MNRAS, 405, 857

Goodwin, S. P. 1997, MNRAS, 284, 785

Goodwin, S. P., & Bastian, N. 2006, MNRAS, 373, 752

Gott, III, J. R., Jurić, M., Schlegel, D., et al. 2005, ApJ, 624, 463

Goudfrooij, P., & Kruijssen, J. M. D. 2014, ApJ, 780, 43

Governato, F., Willman, B., Mayer, L., et al. 2007, MNRAS, 374, 1479

Governato, F., Brook, C., Mayer, L., et al. 2010, Nature, 463, 203

Governato, F., Zolotov, A., Pontzen, A., et al. 2012, MNRAS, 422, 1231

Groth, E. J., Peebles, P. J. E., Seldner, M., & Soneira, R. M. 1977, Scientific American, 237, 76

Grudić, M. Y., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2017, ArXiv e-prints, arXiv:1612.05635

Guedes, J., Callegari, S., Madau, P., & Mayer, L. 2011, ApJ, in press; arXiv:1103.6030, arXiv:1103.6030

Gunn, J. E., & Gott, III, J. R. 1972, ApJ, 176, 1

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

Guth, A. H., & Pi, S.-Y. 1982, Physical Review Letters, 49, 1110

Haardt, F., & Madau, P. 2001, in Clusters of Galaxies and the High Redshift Universe Observed in X-rays, ed. D. M. Neumann & J. T. V. Tran

Harris, W. E. 1996, AJ, 112, 1487

Harris, W. E. 2001, in Saas-Fee Advanced Course 28: Star Clusters, ed. L. Labhardt & B. Binggeli, 223

Harris, W. E., & Harris, G. L. H. 2002, AJ, 123, 3108

Harris, W. E., Whitmore, B. C., Karakla, D., et al. 2006, ApJ, 636, 90

Hartmann, L. 2009, Accretion Processes in Star Formation: Second Edition (Cambridge University Press)

Hartmann, L., Ballesteros-Paredes, J., & Heitsch, F. 2012, MNRAS, 420, 1457

Heiderman, A., Evans, II, N. J., Allen, L. E., Huard, T., & Heyer, M. 2010, ApJ, 723, 1019

Hennebelle, P., & Chabrier, G. 2013, ApJ, 770, 150

Herbig, G. H., Vrba, F. J., & Rydgren, A. E. 1986, AJ, 91, 575

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19

Hollyhead, K., Bastian, N., Adamo, A., et al. 2015, MNRAS, 449, 1106

Holtzman, J. A., Faber, S. M., Shaya, E. J., et al. 1992, AJ, 103, 691

Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, MNRAS, 445, 581

Hopkins, P. F., Narayanan, D., & Murray, N. 2013, MNRAS, 432, 2647

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

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

Hummels, C. B., & Bryan, G. L. 2012, ApJ, 749, 140

Jeans, J. H. 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1

Jiang, C. Y., Jing, Y. P., & Lin, W. P. 2010, A&A, 510, A60

Johnson, L. C., Seth, A. C., Dalcanton, J. J., et al. 2016, ApJ, 827, 33

—. 2017a, ApJ, 839, 78

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

Jordán, A., Peng, E. W., Blakeslee, J. P., et al. 2009, ApJS, 180, 54

Kainulainen, J., Federrath, C., & Henning, T. 2014, Science, 344, 183

Katz, N. 1992, ApJ, 391, 502

Katz, N., Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19

Keller, B. W., Wadsley, J., Benincasa, S. M., & Couchman, H. M. P. 2014, MNRAS, 442, 3013

Keller, B. W., Wadsley, J., & Couchman, H. M. P. 2015, MNRAS, 453, 3499

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

Kereš, D., Katz, N., Fardal, M., Davé, R., & Weinberg, D. H. 2009, MNRAS, 395, 160

Kim, J.-h., Abel, T., Agertz, O., et al. 2014, ApJS, 210, 14

Kim, J.-h., Agertz, O., Teyssier, R., et al. 2016, ApJ, 833, 202

Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, ApJ, 779, 102

Klypin, A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, submitted, arXiv:1002.3660, arXiv:1002.3660

Klypin, A. A., & Shandarin, S. F. 1983, MNRAS, 204, 891

Knebe, A., Green, A., & Binney, J. 2001, MNRAS, 325, 845

Kolb, E. W., & Turner, M. S. 1990, The early universe.

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18

Kravtsov, A., Vikhlinin, A., & Meshscheryakov, A. 2014, ApJ submitted, arXiv: 1401.7329, arXiv:1401.7329

Kravtsov, A. V. 1999, PhD thesis, New Mexico State University

—. 2003, ApJ, 590, L1

Kravtsov, A. V., & Gnedin, O. Y. 2005, ApJ, 623, 650

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

Kritsuk, A. G., Lee, C. T., & Norman, M. L. 2013, MNRAS, 436, 3247

Kroupa, P. 2001, MNRAS, 322, 231

Kruijssen, J. M. D. 2012, MNRAS, 426, 3008

—. 2014, Classical and Quantum Gravity, 31, 244006

—. 2015, MNRAS, 454, 1658

Kruijssen, J. M. D., & Mieske, S. 2009, A&A, 500, 785

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

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

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

Krumholz, M. R., McKee, C. F., & Klein, R. I. 2004, in Astronomical Society of the Pacific Conference Series, Vol. 323, Star Formation in the Interstellar Medium: In Honor of David Hollenbach, ed. D. Johnstone, F. C. Adams, D. N. C. Lin, D. A. Neufeeld, & E. C. Ostriker, 401

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

Lada, C. J., & Lada, E. A. 2003, ARA&A, 41, 57

Lada, C. J., Lombardi, M., & Alves, J. F. 2010, ApJ, 724, 687

Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, A&A, 521, L53

Lara-López, M. A., Hopkins, A. M., López-Sánchez, A. R., et al. 2013, MNRAS, 434, 451

Larsen, S. S. 2002, AJ, 124, 1393

—. 2009, A&A, 494, 539

Larsen, S. S., Brodie, J. P., Huchra, J. P., Forbes, D. A., & Grillmair, C. J. 2001, AJ, 121, 2974

Larson, R. B. 1981, MNRAS, 194, 809

Leaman, R., VandenBerg, D. A., & Mendel, J. T. 2013, MNRAS, 436, 122

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

Li, C., & White, S. D. M. 2009, MNRAS, 398, 2177

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

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

Lieberz, P., & Kroupa, P. 2017, MNRAS, 465, 3775

Lin, C. C., & Shu, F. H. 1964, ApJ, 140, 646

Loeb, A., & Furlanetto, S. R. 2013, The First Galaxies in the Universe

Ma, X., Kasen, D., Hopkins, P. F., et al. 2015, MNRAS, 453, 960

Mac Low, M.-M., & Klessen, R. S. 2004, Reviews of Modern Physics, 76, 125

Magdis, G. E., Daddi, E., Sargent, M., et al. 2012, ApJ, 758, L9

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

Mather, J. C., Cheng, E. S., Eplee, Jr., R. E., et al. 1990, ApJ, 354, L37

May, A., & van Albada, T. S. 1984, MNRAS, 209, 15

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

Miller, G. E., & Scalo, J. M. 1979, ApJS, 41, 513

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution

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

Moustakas, J., Zaritsky, D., Brown, M., et al. 2011, ApJ submitted, arXiv: 1112.3300, arXiv:1112.3300

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

Muratov, A. L., Kereš, D., Faucher-Giguère, C.-A., et al. 2015, MNRAS, 454, 2691

Murray, N. 2011, ApJ, 729, 133

Murray, N., & Chang, P. 2015, ApJ, 804, 44

Naab, T., & Ostriker, J. P. 2016, ArXiv e-prints, arXiv:1612.06891

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563

Navarro, J. F., & White, S. D. M. 1993, MNRAS, 265, 271

Nomoto, K., Iwamoto, K., Nakasato, N., et al. 1997, Nuclear Physics A, 621, 467

Nomoto, K., Tominaga, N., Umeda, H., Kobayashi, C., & Maeda, K. 2006, Nuclear Physics A, 777, 424

Ostriker, J. P., & Peebles, P. J. E. 1973, ApJ, 186, 467

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

Padoan, P., & Nordlund, Å. 2011, ApJ, 730, 40

Papastergis, E., Cattaneo, A., Huang, S., Giovanelli, R., & Haynes, M. P. 2012, ApJ, 759, 138

Peebles, P. J. E. 1970, AJ, 75, 13

—. 1984, ApJ, 277, 470

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

Peng, E. W., Jordán, A., Côté, P., et al. 2006, ApJ, 639, 95

—. 2008, ApJ, 681, 197

Penzias, A. A., & Wilson, R. W. 1965, ApJ, 142, 419

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A15

—. 2016, A&A, 594, A13

Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, ARA&A, 48, 431

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425

Prieto, J. L., & Gnedin, O. Y. 2008, ApJ, 689, 919

Read, J. I., Agertz, O., & Collins, M. L. M. 2015, ArXiv e-prints, arXiv:1508.04143

Rees, M. J., & Ostriker, J. P. 1977, MNRAS, 179, 541

Renaud, F., Agertz, O., & Gieles, M. 2017, MNRAS, 465, 3622

Renaud, F., Bournaud, F., & Duc, P.-A. 2015, MNRAS, 446, 2038

Roberts, M. S., & Rots, A. H. 1973, A&A, 26, 483

Roškar, R., Teyssier, R., Agertz, O., Wetzstein, M., & Moore, B. 2014, MNRAS, 444, 2837

Rubin, V. C., Ford, Jr., W. K., & Thonnard, N. 1980, ApJ, 238, 471

Rudd, D. H., Zentner, A. R., & Kravtsov, A. V. 2008, ApJ, 672, 19

Salem, M., & Bryan, G. L. 2014, MNRAS, 437, 3312

Scannapieco, C., Tissera, P. B., White, S. D. M., & Springel, V. 2008, MNRAS, 389, 1137

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521

Schmidt, W., Almgren, A. S., Braun, H., et al. 2014, MNRAS, 440, 3051

Semenov, V., Kravtsov, A., & Gnedin, N. 2017, ArXiv e-prints, arXiv:1704.04239

Semenov, V. A., Kravtsov, A. V., & Gnedin, N. Y. 2015, ArXiv e-prints, arXiv:1512.03101

—. 2016, ApJ, 826, 200

Silva-Villa, E., & Larsen, S. S. 2011, A&A, 529, A25+

Smith, R., Fellhauer, M., Goodwin, S., & Assmann, P. 2011, MNRAS, 414, 3036

Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, ApJ, 396, L1

Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51

Somerville, R. S., & Primack, J. R. 1999, MNRAS, 310, 1087

Spitzer, L. 1987, Dynamical Evolution of Globular Clusters (Princeton: Princeton University Press)

Springel, V. 2010, ARA&A, 48, 391

Springel, V., Frenk, C. S., & White, S. D. M. 2006, Nature, 440, 1137

Springel, V., & Hernquist, L. 2003a, MNRAS, 339, 289

—. 2003b, MNRAS, 339, 312

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Stadel, J., Potter, D., Moore, B., et al. 2009, MNRAS, 398, L21

Stinson, G., Seth, A., Katz, N., et al. 2006, MNRAS, 373, 1074

Stinson, G. S., Brook, C., Macciò, A. V., et al. 2013, MNRAS, 428, 129

Sun, W., de Grijs, R., Fan, Z., & Cameron, E. 2016, ApJ, 816, 9

Tacconi, L. J., Neri, R., Genzel, R., et al. 2013, ApJ, 768, 74

Teyssier, R. 2015, ARA&A, 53, 325

Tonini, C. 2013, ApJ, 762, 39

Toomre, A. 1964, ApJ, 139, 1217

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

Uhlig, M., Pfrommer, C., Sharma, M., et al. 2012, MNRAS, 423, 2374

Urquhart, J. S., Moore, T. J. T., Csengeri, T., et al. 2014, MNRAS, 443, 1555

Usher, C., Forbes, D. A., Brodie, J. P., et al. 2012, MNRAS, 426, 1475

VandenBerg, D. A., Brogaard, K., Leaman, R., & Casagrande, L. 2013, ApJ, 775, 134

Vanderbeke, J., West, M. J., De Propris, R., et al. 2014, MNRAS, 437, 1734

Vesperini, E., & Heggie, D. C. 1997, MNRAS, 289, 898

Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., et al. 2009, ApJ, 692, 1060

Vogelsberger, M., Genel, S., Springel, V., et al. 2014, Nature, 509, 177

Vutisalchavakul, N., Evans, II, N. J., & Heyer, M. 2016, ApJ in press (arXiv/1607.06518), arXiv:1607.06518

Wetzel, A. R., Hopkins, P. F., Kim, J.-h., et al. 2016, ArXiv e-prints, arXiv:1602.05957

White, S. D. M. 1978, MNRAS, 184, 185

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341

Whitmore, B. C. 2004, in ASP Conf. Ser. 322: The Formation and Evolution of Massive Young Star Clusters, 419 (astro–ph/0403709)

Whitmore, B. C., Zhang, Q., Leitherer, C., et al. 1999, AJ, 118, 1551

Woosley, S. E., & Weaver, T. A. 1995, ApJS, 101, 181

Worthey, G. 1994, ApJS, 95, 107

Zahid, J., Dima, G., Kudritzki, R., et al. 2014, ApJ, submitted, arXiv:1404.7526, arXiv:1404.7526

Zinn, R. 1985, ApJ, 293, 424

Zuckerman, B., & Evans, II, N. J. 1974, ApJ, 192, L149

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110