

Testing feedback from star clusters in simulations of the Milky Way formation

Gillen Brown  and Oleg Y. Gnedin 

Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA

Accepted 2022 April 25. Received 2022 April 22; in original form 2022 March 1

ABSTRACT

We present a suite of galaxy formation simulations that directly model star cluster formation and disruption. Starting from a model previously developed by our group, here we introduce several improvements to the prescriptions for cluster formation and feedback, then test these updates using a large suite of cosmological simulations of Milky Way mass galaxies. We perform a differential analysis with the goal of understanding how each of the updates affects star cluster populations. Two key parameters are the momentum boost of supernova feedback f_{boost} and star formation efficiency per free-fall time ϵ_{ff} . We find that f_{boost} has a strong influence on the galactic star formation rate, with higher values leading to less star formation. The efficiency ϵ_{ff} does not have a significant impact on the global star formation rate, but dramatically changes cluster properties, with increasing ϵ_{ff} leading to a higher maximum cluster mass, shorter age spread of stars within clusters, and higher integrated star formation efficiencies. We also explore the redshift evolution of the observable cluster mass function, finding that most massive clusters have formed at high redshift $z > 4$. Extrapolation of cluster disruption to $z = 0$ produces good agreement with both the Galactic globular cluster mass function and age–metallicity relation. Our results emphasize the importance of using small-scale properties of galaxies to calibrate subgrid models of star cluster formation and feedback.

Key words: methods: numerical – galaxies: evolution – galaxies: formation – galaxies: star clusters: general – galaxies: star formation.

1 INTRODUCTION

Most stars form in clustered environments (Lada & Lada 2003), and young massive clusters (YMCs) are found in the Milky Way (MW) and other star-forming galaxies. The properties of young clusters are sensitive tracers of the star formation process. Young clusters show a well-defined mass function typically described as a Schechter (1976) function with a low-mass slope of -2 (Bastian 2008; Portegies Zwart, McMillan & Gieles 2010). The cutoff mass scales with the star formation rate of the host galaxy, as does the maximum cluster mass (Larsen 2002).

Globular clusters (GCs) are also ubiquitous within galaxies, as they are found in all nearby galaxies with stellar masses above $10^9 M_{\odot}$ (Brodie & Strader 2006). GCs are typically old, with ages above 10 Gyr (Puzia et al. 2005; Strader et al. 2005), and have sizes of a few parsecs that are consistent with YMCs found in the local universe (Brown & Gnedin 2021b). This naturally leads to the hypothesis that GCs are the surviving subset of a larger population of YMCs that formed at high redshift. However, the mass function of GCs is well characterized by a lognormal distribution with a peak mass of around $2 \times 10^5 M_{\odot}$ (Harris 1991; Jordán et al. 2007), in contrast to the Schechter (1976) function commonly used to describe YMCs. This transformation of the mass function over cosmic time requires a preferential destruction of low-mass clusters (Fall & Zhang 2001; Vesperini et al. 2003; Prieto & Gnedin 2008; Elmegreen 2010; Kruijssen 2015).

The dynamical evolution of clusters results from a combination of stellar evolution, two-body relaxation, tidal truncation, and tidal shocks (Spitzer 1958; Gnedin & Ostriker 1997; Elmegreen 2010; Gieles & Renaud 2016). While stellar evolution and internal two-body relaxation can be well understood by studying isolated clusters, dynamical evolution depends on the tidal field and requires more detailed modelling. Throughout their lifetime, clusters experience tidal interactions with their natal giant molecular clouds (GMCs), the galactic structure, and other GMCs, leading to complex evolution that is not well approximated by idealized models (Meng & Gnedin 2022).

Simulations of galaxy formation are well suited for a detailed investigation of cluster formation and disruption (Renaud, Agertz & Gieles 2017; Pfeffer et al. 2018). By situating clusters within their galactic context, their formation and evolution can be realistically tracked. However, few cosmological simulations have the resolution required to directly resolve cluster formation or disruption, so they must rely on subgrid models (although see Kim et al. 2018; Lahén et al. 2019; Benincasa et al. 2020; Ma et al. 2020; Hislop et al. 2022). As cluster formation is terminated by feedback from the newly formed stars, simulations must self-consistently determine this feedback to obtain reliable properties of star clusters. Prescriptions for stellar feedback, particularly supernova feedback, have undergone many revisions over the years as they are calibrated against observations (Katz 1992; Springel & Hernquist 2003; Stinson et al. 2006; Agertz, Teyssier & Moore 2011; Agertz et al. 2013; Hopkins et al. 2014, 2018; Keller et al. 2014). However, these feedback models are often only tested against galaxy-scale properties, such as the global star formation rate or Kennicutt–Schmidt relation (Schmidt

* E-mail: gillenb@umich.edu

1959; Kennicutt 1998). To properly model star cluster formation, these feedback prescriptions must be calibrated on smaller scales.

In Li et al. (2017) and Li, Gnedin & Gnedin (2018, hereafter L18), our group introduced a suite of cosmological simulations that directly models star cluster formation and disruption. The high spatial resolution of these simulations (3–6 pc) allows us to resolve GMCs where star cluster formation occurs. Star particles are seeded within GMCs and accrete material from their surroundings until feedback from the newly formed stars stops further accretion. The final masses of these star particles are set self-consistently and represent the masses of individual star clusters. These simulations were able to reproduce many aspects of the observed young cluster populations, including the shape of the initial cluster mass function, the total mass of stars contained in bound clusters, the relationship between the maximum cluster mass and the star formation rate surface density, and the formation time-scales of star clusters. Some of the central clusters in satellite galaxies have properties consistent with nuclear star clusters in the local universe (Brown, Gnedin & Li 2018). Star formation sites in the modelled galaxies at high redshift are temporarily contained in giant clumps, which dissolve after ~ 100 Myr (Meng & Gnedin 2020). In addition, these simulations resolve dense irregular structures within the thick galactic disc (Meng & Gnedin 2021), allowing for an accurate calculation of the tidal field and therefore the tidal disruption of clusters (Li & Gnedin 2019; Meng & Gnedin 2022).

While these simulations advanced our modelling of star cluster formation in cosmological simulations, they still had several limitations. First, they reached only redshift $z = 1.5$. This precludes a direct comparison to the GCs of the Milky Way (MW), as the disruption up to $z = 0$ must be estimated. Secondly, these simulations include only one MW-mass galaxy and its satellites, decreasing the statistical power of the results and potentially making the results dependent on the specific initial condition (IC) used.

In this paper, we present the next generation of simulations based on the prescriptions of L18. These simulations use two Local Group-like ICs, with the goal of reaching $z = 0$ with four MW-mass galaxies. In Section 2 we describe improvements to the formation and feedback schemes, then describe the new suite of simulations. This suite includes nine runs using Local Group-like ICs and 20 using the MW-like IC from L18. These runs vary a wide range of feedback and cluster formation parameters, allowing us to explore how different prescriptions affect the resulting cluster properties in Section 3. We perform a differential analysis, systematically exploring each of the parameters we vary. In Section 4, we present an application of these simulations by presenting the redshift evolution of the observable cluster mass function. We discuss remaining uncertainties and compare our results with observations in Section 5, then summarize our results in Section 6.

2 SIMULATION CODE AND SETUP

In this section, we describe the ART code and the properties of the simulations. Throughout this section, we introduce several parameters of the code, which we list for convenience in Table 1.

2.1 The ART code

For our simulations we use the Adaptive Refinement Tree (ART) code (Kravtsov, Klypin & Khokhlov 1997; Kravtsov 1999; Rudd, Zentner & Kravtsov 2008; Li et al. 2017, 2018). The ART code includes many physical processes that are important for modelling the formation of galaxies. Radiative transfer is calculated using an improved version

Table 1. Key parameters of the star formation and feedback prescriptions with the values used in this paper.

Parameter	Value(s)
Molecular fraction threshold for cluster creation	50 per cent
Maximum virial parameter for cluster creation	10
Density threshold for cluster creation and growth	1000 cm^{-3}
Star formation efficiency per freefall time (ϵ_{ff})	1, 10, 100 per cent
Radius of GMC	5 pc
Clumping factor (C_ρ)	3, 10, 30
Stellar IMF range	0.08–50 M_\odot
Stellar mass range for SNII	8–50 M_\odot
Stellar mass range for HN	20–50 M_\odot
Initial supernova fraction ($f_{\text{HN},0}$)	0, 5, 20, 50 per cent
SNII momentum boost (f_{boost})	1, 2, 3, 5
Stellar mass range for AGB	0.08–8 M_\odot
Number of SNIa per unit stellar mass	$1.6 \times 10^{-3} M_\odot^{-1}$

of the Optically Thin Variable Eddington Tensor method (Gnedin & Abel 2001), which has been revised to minimize numerical diffusion (Gnedin 2014). Radiation from both stars and the extragalactic background (Haardt & Madau 2001) are included. A non-equilibrium chemistry network of molecular hydrogen is used to identify star-forming regions within GMCs. It was calibrated using observations in nearby galaxies (Gnedin & Kravtsov 2011) and updated to include line overlap in computing self-shielding of molecular hydrogen (Gnedin & Draine 2014). This chemical network also calculates the ionization states of hydrogen and helium. This model uses the local abundance of all these species to calculate the heating and cooling functions self-consistently, without any assumptions of photoionization equilibrium or collisional equilibrium. The ART code also includes a subgrid-scale (SGS) model for numerically unresolved turbulence developed by Semenov, Kravtsov & Gnedin (2016), which follows the results of the MHD simulations of Padoan, Haugbølle & Nordlund (2012).

A particularly novel aspect of the ART code is the direct modelling of time-resolved star cluster formation (Li et al. 2017, 2018; Li & Gnedin 2019). Star cluster particles are seeded in dense gas, and accrete gas from a surrounding region until feedback from the new cluster terminates gas accretion. This region, which we refer to as the ‘GMC,’ has a radius of 5 pc and is fixed in physical size at all cosmic epochs. With the maximum spatial resolution of our simulations being set in the range of 3–6 pc, the GMC can extend past the central cell, allowing the cluster to accrete gas from neighbour cells. Specifically, the growth rate of a given cluster is

$$\dot{M} = \frac{\epsilon_{\text{ff}}}{t_{\text{ff}}} \sum_{\text{cell}} f_{\text{GMC}} V_{\text{cell}} f_{\text{H}_2} \rho_{\text{gas}}, \quad (1)$$

where ϵ_{ff} is the local star formation efficiency per free-fall time t_{ff} , f_{GMC} is the fraction of cell volume V_{cell} included within the GMC sphere, f_{H_2} is the local mass fraction of molecular gas, and ρ_{gas} is the local total gas density. This mass growth is accumulated at each local time-step, which is typically in range of 10^2 – 10^3 yr. As long as the local gas density is above the threshold, clusters can continue accreting gas. This accretion stops either when it has accreted no material in the last 1 Myr or when it has reached an age of 15 Myr.

To avoid the spurious creation of many small clusters, we impose a threshold such that clusters must have an expected mass (defined as the initial \dot{M} times the maximum allowed formation time of 15 Myr) of at least $6000 M_\odot$. As clusters typically form over a few Myr,

rather than the full 15 Myr, this results in the elimination of small clusters below about 1000 M_{\odot} .

Due to the complex dynamical evolution that occurs throughout the process of cluster formation, not all stars in a given star-forming region will be bound to the fully formed cluster. To model this, star cluster particles include a variable tracking the fraction of mass that is gravitationally bound. This is set at cluster formation (see Section 2.2) and is updated as clusters undergo dynamical disruption throughout their lifetime.

2.2 Updates to the cluster formation modelling

We implement several updates to the ART code to improve the star cluster formation algorithm. In the implementation of L18, a cluster particle is created if the gas density in a cell reaches $n_H > 1000 \text{ cm}^{-3}$ and the local H_2 mass fraction is larger than 0.5, meaning the cell contains mostly dense molecular gas. Here, we introduce an additional criterion based on the local virial parameter of the gas, intended to select gravitationally bound gas. Generally, the virial parameter is

$$\alpha_{\text{vir}} = \frac{5\sigma^2 R}{3GM}, \quad (2)$$

where σ is the local gas velocity dispersion, R is the radius of the sphere we consider, and M is the mass within this sphere. We calculate this locally in any cell meeting the other star formation criteria, assuming a sphere with a diameter equal to the size of the cell ($l = 2R$), giving

$$\alpha_{\text{vir}} = \frac{5\sigma^2}{\pi G \rho_{\text{gas}} l^2}. \quad (3)$$

We use both the turbulent velocity and sound speed when calculating the velocity dispersion ($\sigma^2 = v_{\text{turb}}^2 + c_s^2$), but do not include cell-to-cell velocity differences. We require $\alpha_{\text{vir}} < 10$ to seed star clusters. This threshold is near the typical value for observed GMCs in the MW (Miville-Deschênes, Murray & Lee 2017). Star formation is allowed on the four finest refinement levels.

We also use a new prescription for the initial bound fraction of star clusters, as determined by Li et al. (2019). These authors performed simulations of 80 isolated molecular clouds with a range of mass, size, velocity configuration, and feedback strength. After feedback terminates star formation, they calculate the integrated star formation efficiency ϵ_{int} , which is the fraction of the initial gas mass that formed stars, as well as the fraction of stars that are bound to the final cluster f_{bound} . They then determine the relation between these two parameters:

$$f_{\text{bound}} = \left[\text{erf} \left(\sqrt{\frac{3\epsilon_{\text{int}}}{\alpha_{\star}}} \right) - \sqrt{\frac{12\epsilon_{\text{int}}}{\pi\alpha_{\star}}} \exp \left(-\frac{3\epsilon_{\text{int}}}{\alpha_{\star}} \right) \right] f_{\text{sat}}, \quad (4)$$

where $\alpha_{\star} = 0.48$ and $f_{\text{sat}} = 0.94$ are free parameters the authors fitted. Determining ϵ_{int} in our simulations is not trivial. The initial gas mass when the cluster was seeded is not an accurate representation of the available gas mass, as GMCs accrete material over time. To account for this, we define ϵ_{int} as the ratio of the final stellar mass to the maximum value of the stellar mass plus gas mass at any time during cluster formation:

$$\epsilon_{\text{int}} = \frac{M_{\star, \text{final}}}{\max(M_{\star}(t) + M_g(t))} \quad (5)$$

We then use this directly in equation (4) to calculate the initial bound fraction for each star cluster.

2.3 Cluster disruption modelling

Our model for cluster disruption is unchanged from that described in detail in Li & Gnedin (2019), but we summarize the key points here. At each global time-step of the simulation (the length of the global time-step is typically a few Myr, with a maximum of 50 Myr), we calculate the tidal tensor around all fully formed clusters using the second-order finite difference of the gravitational potential across a $3 \times 3 \times 3$ cell cube centred on the star particle. To determine cluster disruption in runtime, we calculate the three eigenvalues of the tidal tensor $\lambda_1 > \lambda_2 > \lambda_3$, which describe the strength of the tidal field in the direction of their corresponding eigenvectors. We use the maximum of the absolute value of the eigenvalues to determine the dynamical time-scale within the Roche lobe of the cluster:

$$\Omega_{\text{tid}}^2(t) = \frac{\lambda_m}{3} \quad (6)$$

where

$$\lambda_m \equiv \max_i |\lambda_i| \quad (7)$$

We then use it to determine the cluster disruption time-scale:

$$t_{\text{tid}} = 10 \text{ Gyr} \left(\frac{M(t)}{2 \times 10^5 M_{\odot}} \right)^{2/3} \frac{100 \text{ Gyr}^{-1}}{\Omega_{\text{tid}}(t)} \quad (8)$$

Finally, we use this cluster disruption time-scale to decrease the mass bound to each cluster. We track it with the variable f_{dyn} , which describes the fraction of cluster mass bound to the cluster after accounting for dynamical disruption. At the n th global time-step of length dt_n , we update this fraction as follows:

$$f_{\text{dyn}}^{n+1} = \exp(-dt_n/t_{\text{tid}}) f_{\text{dyn}}^n \quad (9)$$

We also output the full tidal tensor for each star particle at each global time-step, allowing us to post-process star cluster disruption and explore how different prescriptions for tidal disruption, including capturing tidal shocks, may change cluster properties.

2.4 Updates to the stellar feedback modelling

2.4.1 Abundances of individual elements

We have implemented runtime tracking of most important individual elements (C, N, O, Mg, S, Ca, Fe) and ejecta of AGB stars. This gives 10 total fields tracking chemical enrichment (C, N, O, Mg, S, Ca, Fe, Z_{SNIIa} , Z_{SNII} , and Z_{AGB}) in both gas and stars. These elements are some of the most abundant in the universe, have reliable yields, and enable comparisons with both gas-phase and stellar abundance measurements at a variety of redshifts. N, O, and S are commonly used to measure gas-phase metallicity (e.g. Kewley & Dopita 2002; Maiolino & Mannucci 2019). Fe, Mg, and Ca are commonly measured in stellar spectra, with Fe representing total metallicity and Mg and Ca being representative α elements (Gallazzi et al. 2005; Kirby et al. 2013; Hayden et al. 2015).

2.4.2 Discrete supernova events

We have updated the supernova (SN) feedback prescriptions in the ART code to include discrete SN explosions at rate calculated from the stellar lifetimes, IMF, and total stellar mass of the particle. Conceptually, we use the stellar lifetimes to calculate the mass range of stars leaving the main sequence during a given time-step, then integrate the IMF over this range to determine the total number of stars leaving the main sequence. We explode

an integer number of these as SN, leaving any fractional SN to accumulate to the next time-step. This leads to only an integer number of SN exploding in a given time-step while also appropriately conserving the total number of SN over the life of the stellar population.

We calculate the number of SN in a given time-step:

$$N_{\text{SN}}(\tau) + N_{\text{SN, leftover}}(\tau + d\tau) = M_*(\tau) \int_{\mathcal{M}(\tau)}^{\mathcal{M}(\tau+d\tau)} \Phi(\mathcal{M}) d\mathcal{M} + N_{\text{SN, leftover}}(\tau), \quad (10)$$

where τ is the age of the stellar population (discussed in more detail in Section 2.4.7), $d\tau$ is the length of the current time-step, M_* is the total mass of the cluster particle, $\mathcal{M}(\tau)$ is the mass of the star leaving the main sequence at age τ , $\Phi(\mathcal{M})$ is the IMF normalized such that $M_* = \int \mathcal{M} \Phi(\mathcal{M}) d\mathcal{M}$, and $N_{\text{SN, leftover}}$ is the fractional number of SN not exploded in the previous time-step. N_{SN} is always an integer value, and $0 \leq N_{\text{SN, leftover}} < 1$. We use a Kroupa (2001) IMF with a mass range of 0.08 to $50 M_\odot$, and use $8 M_\odot$ as the minimum mass to explode as an SN. We use the metallicity-dependent analytic stellar lifetimes from Raiteri, Villata & Navarro (1996).

When SN explode, we inject energy and mass into the surroundings. The mass of different elements is taken directly from the stellar yield tables of Kobayashi et al. (2006). We use the yield for a star of mass $\mathcal{M} = 0.5(\mathcal{M}(\tau) + \mathcal{M}(\tau + d\tau))$, and use the metallicity of the star particle. We linearly interpolate the yield tables in both mass and metallicity to determine the yields at arbitrary stellar masses and metallicities.

2.4.3 Introduction of hypernovae

Hypernovae (HN) are SN explosions with significantly more energy than a typical SN, and may be associated with gamma-ray bursts (e.g. Iwamoto et al. 1998). The Kobayashi et al. (2006) yield tables include stellar yields and energies for HN so we include them in our feedback model. We model both the energy and yields from HN self-consistently. SN with progenitor stellar masses above $20 M_\odot$ are eligible to explode as HN. Each explosion is randomly assigned to be either HN or SN, depending on a metallicity-dependent HN fraction. We use the functional form proposed by Grimmett et al. (2020):

$$f_{\text{HN}} = \max \left(f_{\text{HN},0} \exp \left(-\frac{Z}{0.001} \right), 0.001 \right) \quad (11)$$

These authors suggest that $f_{\text{HN},0} = 0.5$, but we leave it as a free parameter to test how varying it affects galaxy properties. SN explosions always inject $E_{51} \equiv 10^{51}$ ergs of energy, while for HN we use the mass–energy relation from Kobayashi et al. (2006), where the energy ranges from 10 to $30 E_{51}$, with high-mass stars releasing the most energy. We linearly interpolate the energy released by HN for stellar masses between those given in Kobayashi et al. (2006). Increasing f_{HN} significantly changes the energy injected into the simulation. Fig. 1 shows the cumulative energy injected from SN as a function of cluster age. Different lines show different metallicity and therefore different f_{HN} . As HN are only active for stars with masses above $20 M_\odot$, the difference in f_{HN} is apparent at early times, while at later times SN energy injection is the same. As our stellar lifetimes are metallicity-dependent, the age of the onset of SN and the age at which HN end changes as well. Of note, the Raiteri et al. (1996) lifetimes give an onset of SN in this new prescription that is always later than the constant 3 Myr onset adopted by L18.

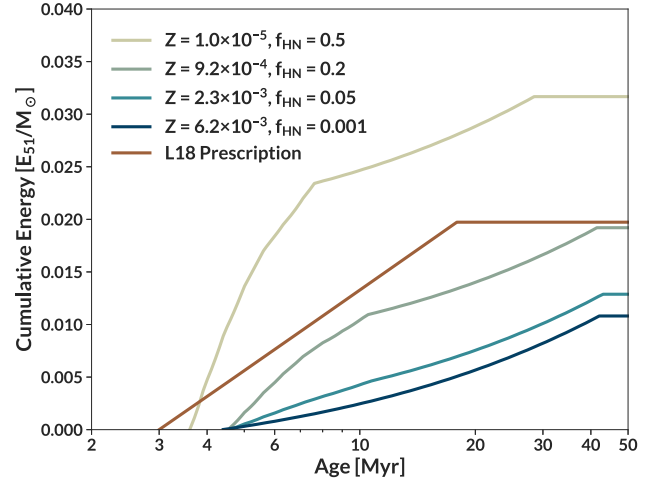


Figure 1. Cumulative energy injected by SN per unit stellar mass in units of $10^{51} \text{ erg } M_\odot^{-1}$ as a function of time since beginning of star formation for different prescriptions. Four lines show the model used in this set of simulations, while the last shows that used by L18. The new prescription is plotted at several metallicities, with HN fractions following equation (11) with $f_{\text{HN},0} = 0.5$. The metallicity-dependent stellar lifetimes also change the time of the onset of SN. The line with $f_{\text{HN}} = 0.1$ per cent is visually indistinguishable from a line with $f_{\text{HN}} = 0$.

2.4.4 Momentum boost

To model SN feedback, we use the prescriptions from Martizzi, Faucher-Giguère & Quataert (2015). They used simulations of inhomogeneous turbulent medium to parametrize the partition of the SN remnant energy into the thermal, kinetic, and turbulent components. The resulting energy and momentum input depend on the ambient gas density and spatial resolution of the simulation. However, their simulations of isolated SN explosions underestimate the effect for star clusters. Cluster-forming regions usually produce a large number of massive stars that undergo simultaneous SN explosion. Gentry et al. (2017) found that such clustering of SN can enhance momentum feedback by an order of magnitude relative to that delivered by an isolated SN. L18 tested a boost to the momentum feedback from SN remnants by a factor $f_{\text{boost}} = 3-10$ and found that the value $f_{\text{boost}} = 5$ can reproduce the galactic star formation history expected from the abundance-matching technique. As f_{boost} is a key parameter of our feedback model, we explore its ideal value in our new simulations below in Section 3.2. The momentum created by stellar particles is distributed spherically to 26 nearest-neighbour cells surrounding the parent cell of the particle, as in Li et al. (2017).

2.4.5 Supernovae type Ia

We have updated the SNIa feedback prescription, implementing discrete SN and a new delay-time distribution (DTD). We use the power-law DTD for field galaxies from Maoz & Graur (2017):

$$\frac{dN_{\text{SNIa}}}{dt} \propto \tau^{-1.13} \quad (12)$$

normalized to produce 1.6×10^{-3} SNIa per M_\odot of stellar mass. Similarly to how we integrate over the IMF to produce the number of SNII, we integrate over the DTD to produce the number of SNIa. We model these as discrete events as we do for SNII, and use the yields from Nomoto & Leung (2018). The feedback from SNIa is modelled simply as an injection of $2E_{51}$ of thermal energy.

2.4.6 AGB feedback

Our final addition to the feedback prescription is chemical enrichment from AGB stars, defined to be the last stages of evolution of stars with masses below $8 M_{\odot}$. The prescription for AGB stars is analogous to that for SNII as described by equation (10). However, we abandon the requirement for integer numbers and simply use the full integral in equation (10). This is justified by the fact that this phase of stellar evolution is not instantaneous like an SN. We use the yields from Ritter et al. (2018). We only inject mass from AGB feedback. We do not inject the energy or momentum, as their wind velocities are small and have little impact on the total feedback budget (Goldman et al. 2017; Hopkins et al. 2018).

We also include two other sources of feedback, which are unchanged from the implementation of L18: radiation pressure from massive stars using the analytical fit by Gnedin (2014), and momentum from stellar winds as an analytical fit to the results of Leitherer, Robert & Drissen (1992).

2.4.7 Timing of cluster feedback

Since our star cluster particles accrete material over time, defining a single age to use in the above feedback prescriptions is not trivial. Without storing the full cluster growth histories, which are prohibitively large, we must make some assumptions. One choice would be to simply use the time t since the star particle was seeded: $\tau_{\text{birth}}(t) = t$. We refer to this as the ‘birth approach’, since it treats all stars as forming at the same time as the first one in the cluster. This prescription is problematic if the cluster has significant star formation after the onset of SN at about 4 Myr. For example, consider some stars formed 6 Myr after the birth of that cluster particle. The birth approach assigns all stars in the cluster an age of 6 Myr, including these newly formed stars with a true age of zero. As these newly formed stars never had an age in the 0–6-Myr range, the feedback they should contribute during that age range is skipped (particularly SN feedback from 4–6 Myr). This prescription also gets the timing of feedback wrong, as the assumption that all the mass of the cluster formed at the initial time is incorrect.

An alternative is to adjust the age based on the mass-averaged time of cluster formation: $\tau_{\text{ave}}(t) = t - t_{\text{ave}}(t)$. This average time for cluster formation is calculated in runtime as

$$t_{\text{ave}}(t) \equiv \frac{\int_0^t t \dot{M}(t) dt}{\int_0^t \dot{M}(t) dt}, \quad (13)$$

where \dot{M} is the cluster star formation rate at time t (L18). This approach, which we refer to as the ‘average approach’, does a much better job of reproducing the total amount of feedback. However, this approach pushes back the onset of SN feedback, allowing some clusters (particularly massive ones) to have unphysically long formation time-scales before their growth is terminated by feedback.

To solve this problem, we introduce a hybrid approach, where we allocate a fraction of cluster feedback to use the birth approach and the rest to use the average approach. Denoting the amount of feedback generally as \mathcal{F} , we set

$$\mathcal{F}_{\text{tot}}(t) = f_{\text{birth}}(t) \mathcal{F}(\tau_{\text{birth}}(t)) + [1 - f_{\text{birth}}(t)] \mathcal{F}(\tau_{\text{ave}}(t)) \quad (14)$$

such that f_{birth} is the fraction of the cluster mass assigned to the birth approach. This hybrid approach gives the best of both worlds, as it gives the correct delay before the first SN explodes while also accurately reproducing the total amount of feedback. Using idealized test cases, we find that clusters with a larger age spread require

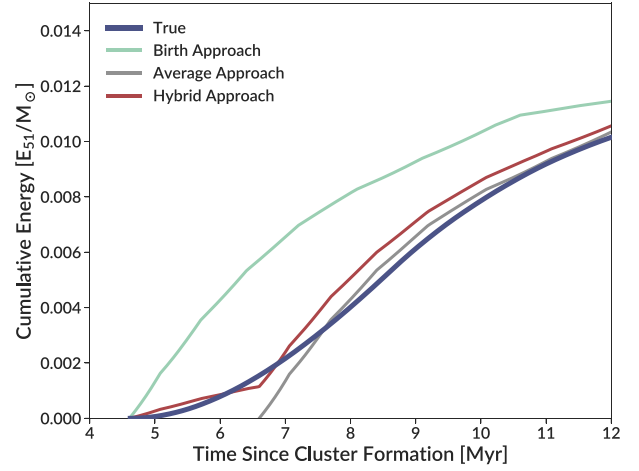


Figure 2. The cumulative energy injected by SN per unit stellar mass as a function of time since beginning of cluster formation for three approaches for the timing of SN. The input star formation history is a constant star formation rate for 4 Myr. The ‘True’ line shows the actual energy injection produced by this stellar population, while the other lines show the energy injection for different ways of treating this star formation history as a simple stellar population, as described in the text. The hybrid approach is a weighted combination of the other two.

a larger f_{birth} . Conceptually, this is because clusters with a large age spread have a larger fraction of their feedback that comes from stars formed away from the mean cluster age. We use the following parametrization:

$$f_{\text{birth}}(t) = \frac{\tau_{\text{spread}}(t)}{20 \text{ Myr}}, \quad (15)$$

where the 20-Myr scale parameter was determined from idealized test cases, and τ_{spread} is the cluster age spread calculated in runtime as

$$\tau_{\text{spread}}(t) \equiv \frac{M(t)}{\langle \dot{M} \rangle} = \frac{M^2(t)}{\int_0^t \dot{M}^2(t) dt} \quad (16)$$

where $\langle \dot{M} \rangle$ is the mass-weighted star formation rate:

$$\langle \dot{M} \rangle = \frac{\int_0^t \dot{M}(t) dM}{\int_0^t dM} = \frac{\int_0^t \dot{M}^2(t) dt}{M(t)} \quad (17)$$

where $dM = \dot{M}(t) dt$. As cluster age spreads are typically a few Myr, this gives no more than 20 per cent of the feedback coming early, with the majority using the average age. Fig. 2 shows an example of this prescription for the feedback from a toy cluster consisting of a 4 Myr period of constant star formation rate. To calculate the true energy injection rate that this toy cluster would be expected to give, we represent it with many simple stellar populations spaced evenly between 0 and 4 Myr. The total energy injection is then the sum of the energy injected by each simple stellar population. We also compute $t_{\text{ave}}(t)$ and $\tau_{\text{spread}}(t)$ to compute the feedback that would result when using the birth, average, and hybrid approaches. In Fig. 2, the delayed onset of SN when using the average approach is clear, as is the increased energy output when assuming all stars formed at the birth of the cluster. As this hybrid approach is a weighted sum of the two other approaches, there is a break in the hybrid approach between 6 and 7 Myr due to the onset of SN in the average approach.

2.5 Hydrodynamics

When updating the code from the version used in L18 to a newer version of the ART code (ART 2.0), we updated the modelling of the energy equation that governs how thermal energy is calculated in the presence of subgrid turbulence. This update more accurately tracks thermal energy in shocks. It has little effect in the disc of the galaxy, as the thermal energy generated by shocks is subdominant to other process that govern energy balance such as heating, cooling, and stellar feedback. However, we find that the circumgalactic medium is affected by this update. In our new runs, there is significantly more hot gas in the halo. This, in turn, leads to less cold gas accreting on to the galaxy, leading to less star formation. The decrease in the amount of cold gas requires changes to the parameters governing star formation and feedback as we describe below. We describe the update to the hydrodynamics in more detail in Appendix A. In our suite of simulations, we used both this updated energy-based approach and the new entropy-conserving scheme of Semenov, Kravtsov & Diemer (2021). These authors found that the entropy-conserving scheme is able to more accurately evolve non-thermal energy components. They ran simulations of an L_* galaxy and found differences between the energy- and entropy-based schemes. However, these differences are much smaller than those we find between the energy-based schemes of L18 and this paper.

2.6 Initial Conditions

In this work, we use three different ICs. One is the IC used by L18, a periodic comoving box of size 4 Mpc that contains a single central galaxy with a total mass of $10^{12} M_\odot$ at $z = 0$, which we refer to as *Isolated MW*. We also use two zoom-in ICs from the ELVIS project (Garrison-Kimmel et al. 2014): *Thelma & Louise* and *Romeo & Juliet*. Both of these ICs contain a Local Group analogue with two MW-mass galaxies, which we describe in more detail below. The *Isolated MW* box is much less computationally expensive to run than the zoom-in runs, so we use it to explore a broader range of parameter space.

Thelma & Louise is a desirable IC as it has qualitative agreement with the accretion histories of the MW and M31. The less massive (MW-like) halo has a quieter accretion history (Hammer et al. 2007), with no significant mergers after $z \approx 5$, while the more massive (M31-like) halo has more mergers at later times as expected from observations (D’Souza & Bell 2018). *Romeo & Juliet* has two galaxies with much quieter merger histories. Including two different sets of ICs allows us to explore how our results vary with galaxy merger histories.

To improve computational performance with the ART code, we modify these zoom-in ICs following the prescription of Brown & Gnedin (2021a). Our ICs have a small zoom region in a large box (50–100 Mpc). This large box size with a small zoom region is difficult for the ART code to parallelize well, so our method decreases the box size and increases the resolution of the root grid. The ICs are created using the MUSIC software (Hahn & Abel 2011), where a white noise field is convolved with the matter power spectrum to produce realistic matter overdensities. We regenerate the original white noise field at higher resolution, then cut out a smaller volume of interest. This smaller white noise cube is then convolved with the matter power spectrum to produce the density within a smaller volume. As the white noise is what seeds the resulting structures, this method reduces the box size while preserving large-scale structure and enforcing periodic boundary conditions. To avoid disturbing the zoom region, the particles from this region are transplanted into the

new box with a velocity offset to match the systemic velocity of this region in the new box. We refer readers to figs 1 and 2 of Brown & Gnedin (2021a) for a visual representation of the method. We find that these modifications improve performance while minimally changing central galaxy properties. Table 2 details some key properties of these ICs, and Fig. 3 shows the halo mass growth of these galaxies in collisionless runs.

We run an initial suite of nine simulations with these zoom-in ICs, varying ϵ_{ff} , $f_{\text{HN}, 0}$, and f_{boost} . Table 3 details the parameters of these runs. We also ran a large suite of 20 simulations on the *Isolated MW* IC varying many aspects of stellar feedback, which are detailed in Table 4.

2.7 Run setup

We keep the spatial resolution of the finest grid level between 3 and 6 physical pc at all times. To accomplish this, we add refinement levels as the simulation progresses. The specific levels and when they are added depend on the IC. In the *Isolated MW* box, we start with nine levels of refinement on the 128^3 root grid, then add levels at $z = 9, 4$, and 1.5 . For *Thelma & Louise*, we allow 11 levels of refinement on the 256^3 root grid, then add additional levels at $z \approx 10.2, 4.6, 1.8$, and 0.41 . *Romeo & Juliet* also starts with 11 levels, but its slightly different box size requires adding levels at $z \approx 9.8, 4.4, 1.7$, and 0.35 .

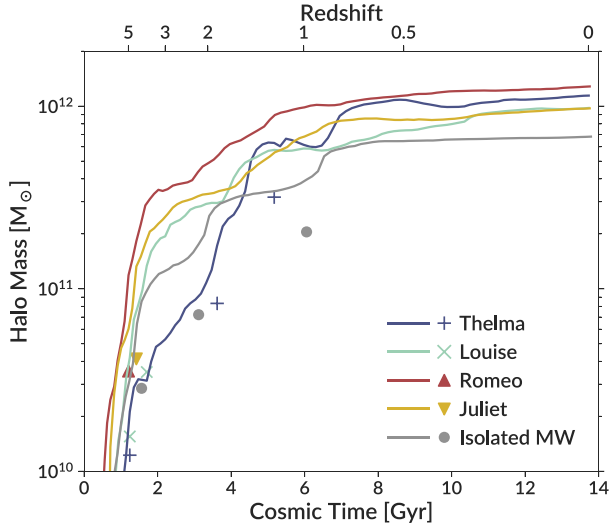
We use three criteria to determine when to refine the grid. In this section, we will illustrate the refinement criteria using specific values from the *Thelma & Louise* IC, but the principles are the same for all ICs. First, we use Lagrangian refinement for both gas and dark matter. Cells are refined when their gas mass exceeds approximately $1.6 \times 10^5 M_\odot$ or dark matter exceeds $3.9 \times 10^6 M_\odot$. The gas refinement is active on all levels, while the dark matter criterion is not active on the four finest levels. We also increase the dark matter mass refinement threshold above that from the simple baryon fraction scaling. These changes are for two reasons. First, the discrete dark matter particles (of mass $1.5 \times 10^5 M_\odot$) do not allow their mass to be distributed evenly, so their distribution cannot be trusted on small scales. Second, we find that there are times when the dark matter criterion will prevent a cell with very small gas mass from derefining. If stellar momentum feedback is imparted on this cell, it will acquire very high velocities due to its small mass, leading to small time-steps and a slower runtime of the simulation. Restricting the levels on which the dark matter Lagrangian criterion is active and increasing the mass threshold for dark matter-triggered refinement mitigates this situation. The final refinement criterion uses a local Jeans length. Cells are refined if their size exceeds twice the Jeans length. This criterion is applied only on the four deepest levels. We find that with these refinement criteria, cell gas masses remain around $2 \times 10^4 M_\odot$. Table 2 includes the typical baryonic cell masses for all ICs.

While we do not record the level on which a star is formed in runtime, we postprocess the outputs to see the levels on which stars can form. In the runs using the Local Group ICs, we find that 15 per cent of the cells that satisfy the star formation criteria are on the highest refinement level with sizes of 3–6 pc, 60 per cent have sizes in the 6–12 pc range, 25 per cent are within 12–24 pc, and a very small fraction are on the fourth level with sizes of 24–48 pc. The lower mass resolution of the *Isolated MW* runs results in the corresponding fractions of 10, 35, 50, and 5 per cent, respectively.

The ART code uses adaptive time-stepping, such that the finest levels have much shorter time-steps than the coarse root grid. For the *Thelma & Louise* runs with $\epsilon_{\text{ff}} = 100$ per cent, the global

Table 2. Description of key properties of the ICs used here. For zoom-in ICs, the mass resolution quantities refer to the zoom region.

Initial Condition	Box size (comoving h^{-1} Mpc)	Dark matter particle mass (M_{\odot})	Typical baryonic cell mass (M_{\odot})	Ω_m	Ω_{Λ}	Ω_b	h
Thelma & Louise	25.0	1.57×10^5	2×10^4	0.266	0.734	0.0449	0.71
Romeo & Juliet	23.12	1.53×10^5	2×10^4	0.31	0.69	0.048	0.68
Isolated MW	4.0	1.0×10^6	4×10^4	0.304	0.696	0.0479	0.681

**Figure 3.** Mass growth of the central haloes from collisionless runs with three ICs. Markers show major mergers with a mass ratio less than 4:1, and are placed at the maximum virial mass of the satellite and the time at which it reached this maximum mass before merging with the central galaxy. Note that Thelma and Isolated MW have major mergers at $z < 2$, while the other three galaxies have quiet merger histories.**Table 3.** The runs using the Local Group ICs included in this simulation suite. z_{last} is the redshift of the last output of each run. All runs use average approach for SN timing, $C_{\rho} = 10$, and the energy-based hydrodynamics scheme. The clumping factor C_{ρ} will be discussed further in Section 3.3.

Initial Condition	ϵ_{ff}	f_{boost}	$f_{\text{HN}, 0}$	z_{last}
Thelma & Louise	1 per cent	5	20 per cent	3.32
Thelma & Louise	10 per cent	5	20 per cent	2.36
Thelma & Louise	100 per cent	1	0 per cent	3.17
Thelma & Louise	100 per cent	3	0 per cent	2.80
Thelma & Louise	100 per cent	5	0 per cent	1.83
Thelma & Louise	100 per cent	5	5 per cent	1.86
Thelma & Louise	100 per cent	5	20 per cent	2.66
Romeo & Juliet	10 per cent	5	20 per cent	2.78
Romeo & Juliet	100 per cent	5	20 per cent	1.87

time-step of the root grid is restricted to be less than 10 Myr. We write outputs at each global time-step. For all other runs, the output spacing is allowed to be at most 50 Myr. The time-step for the finest level is similar for all runs, typically between 100–1000 yr.

3 EFFECTS OF CLUSTER FORMATION AND FEEDBACK MODELLING

In this section, we analyse the large suite of simulations laid out in Tables 3 and 4 to test the implementation of code updates and explore how parameter variation affects our results. We will primarily focus

on the galaxy star formation rate, cluster mass function, and the time-scales of cluster formation. In this section, we exclusively use the particle mass at the end of its star formation episode, which does not account for the initial bound fraction, stellar evolution, or dynamical disruption of a star cluster represented by that particle. We explore those quantities and the observable cluster mass function in Section 4. We also note that when examining star cluster populations, we include all clusters from the central galaxies in the simulations (the one MW-mass galaxy in Isolated MW, and the two galaxies in the Local Group-like environments of Thelma & Louise and Romeo & Juliet). When plotting the star formation rate of these galaxies we plot the two central galaxies in the Local Group-like IC separately, but when plotting cluster properties of a given run we group these two galaxies together.

3.1 Timing of supernova feedback

In Section 2.4.7, we describe how the finite length of cluster formation makes it difficult to create an accurate prescription for the timing of stellar feedback. We ran simulations with the birth approach, the average approach, and the hybrid approach. We also compared these to the feedback model of L18, which has SNe that start earlier (see Fig. 1). We found no significant differences in any galaxy-scale properties between these prescriptions. However, we did find that the cluster formation lifetimes were different between these prescriptions. In particular, the average approach gave significantly longer time-scales for massive clusters. Fig. 4 shows the cumulative distribution of the length of star formation within clusters formed using different timing choices, for the local efficiency $\epsilon_{\text{ff}} = 100$ per cent. Note that the quantity we plot here is the duration of star formation, defined as the age difference between the birth of the cluster and its last accretion event. This is not t_{ave} or τ_{spread} as defined in Section 2.4.7. We use this quantity as it clearly demarcates when feedback ends cluster formation.

With all the approaches to SN feedback, the majority of low-mass clusters have finished their accretion before the onset of SNe at 3–4 Myr, leading to little difference in the durations between our approaches. Such short durations indicate that the other sources of feedback are able to terminate cluster formation before the start of SN feedback (Kruijssen et al. 2019; Grudić et al. 2022). SN feedback remains more relevant for massive clusters.

We do see a difference in the high-mass clusters. The feedback prescriptions of L18 produce the shortest durations of star formation. Among the three new models of determining the timing of SN feedback, the average approach produces clusters with the longest duration, the birth approach gives clusters with the shortest duration, and the hybrid approach is in the middle. As the birth approach has the most early feedback and the average approach has the least early feedback, these results indicate that delaying the start of SN feedback tends to increase the time over which massive clusters can accrete material. This matches what we see in the L18 model, which allows SN feedback begin earlier and stop cluster growth earlier.

Table 4. The runs using the *Isolated* MW ICs included in this simulation suite. In the ‘Hydro Scheme’ column, ‘S21’ refers to the entropy-based scheme of Semenov et al. (2021), ‘Energy’ refers to the updated energy-based scheme, and ‘L18’ is the hydro scheme used in L18. The schemes mentioned in the ‘SN Timing’ column are described in Section 2.4.7. Simulations are grouped by the attribute that is varied, although some simulations are used in multiple subsections. All runs progressed to $z = 1.5$ except for the two runs with $\epsilon_{\text{ff}} < 100$ per cent and $f_{\text{boost}} = 2$, which reached $z \approx 2$.

ϵ_{ff}	f_{boost}	$f_{\text{HN}, 0}$	C_{ρ}	SN timing	Hydro scheme	Other comments
100 per cent	1	0	10	Average	S21	Used in all subsections below
<i>Section 3.1</i>						
100 per cent	1	0	10	Hybrid	S21	
100 per cent	1	0	10	Birth	S21	
<i>Section 3.1.1</i>						
100 per cent	1	0	10	Average	S21	Continuous energy injection from SN
<i>Section 3.2</i>						
100 per cent	1	50 per cent	10	Average	S21	
100 per cent	2	0	10	Average	S21	
100 per cent	3	0	10	Average	S21	
100 per cent	5	0	10	Average	S21	
<i>Section 3.3</i>						
100 per cent	1	0	3	Average	S21	
100 per cent	1	0	3	Average	S21	Changed shielding to Gnedin & Kravtsov (2011)
100 per cent	1	0	30	Average	S21	
<i>Section 3.4</i>						
1 per cent	1	0	10	Average	S21	
10 per cent	1	0	10	Average	S21	
1 per cent	2	0	10	Average	S21	
10 per cent	2	0	10	Average	S21	
<i>Section 3.5</i>						
100 per cent	1	0	10	Average	S21	No virial parameter criterion for star formation
<i>Appendix A</i>						
100 per cent	5	0	10	Average	Energy	Analogous to Local Group runs
100 per cent	5	0	10	Average	Energy	Feedback scheme of L18.
100 per cent	5	0	10	Average	L18	

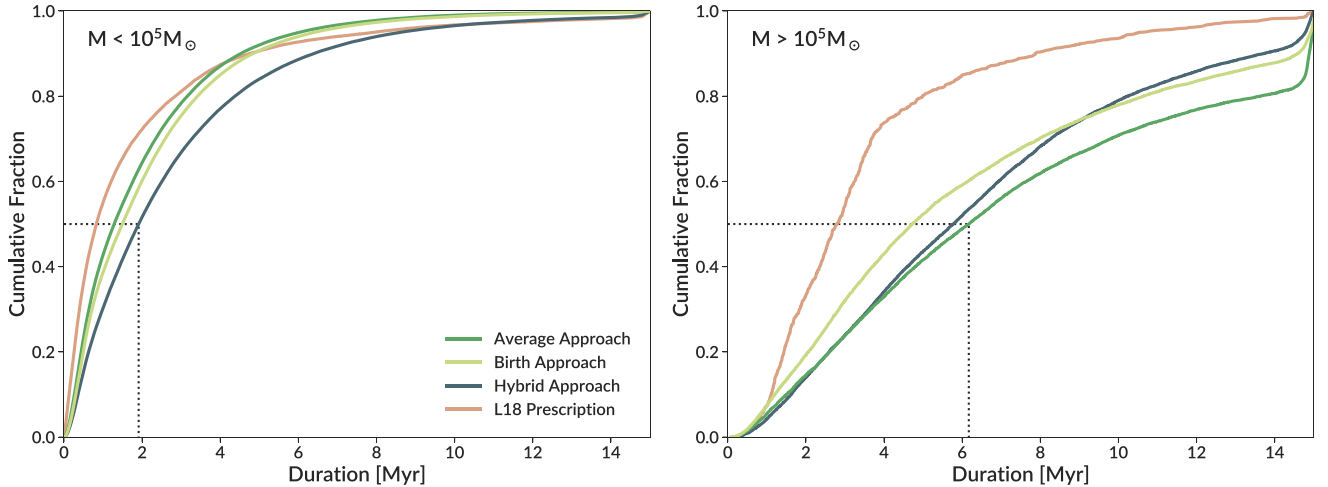


Figure 4. The cumulative distribution of the duration of cluster formation for different approaches to determining the timing of SN feedback, as described in Section 2.4.7. The left-hand panel shows clusters less massive than $10^5 M_{\odot}$, while the right-hand panel shows clusters more massive than $10^5 M_{\odot}$. The dotted line shows the longest median duration of cluster formation. Cluster growth is algorithmically truncated at 15 Myr. Note that here we use a new run with the L18 feedback model, not the L18 simulations themselves. The L18 prescription uses $f_{\text{boost}} = 5$, while all other runs use $f_{\text{boost}} = 1$. All runs use the *Isolated* MW IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{HN}, 0} = 0$, and show all clusters formed before $z = 1.5$.

These trends are also reflected in the integrated star formation efficiency ϵ_{int} , defined in equation (5). Fig. 5 shows the distribution of ϵ_{int} for the runs with variations in the timing of SN feedback. The L18 feedback model has the earliest SN feedback and the lowest

mean value of ϵ_{int} (15 per cent), while the average approach has the latest SN feedback and the highest mean value of ϵ_{int} (35 per cent). Interestingly, the hybrid approach and birth approach are very similar, with mean values at $\epsilon_{\text{ff}} \approx 25$ per cent. This may be because early

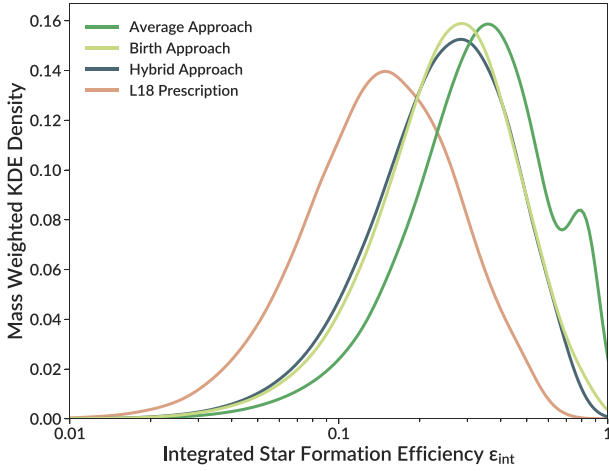


Figure 5. Kernel density estimation for the distribution of integrated star formation efficiency for clusters in the runs with variations in the timing of SN feedback, as described in Section 2.4.7. We use a Gaussian kernel with a width of 0.05 dex. Each curve is normalized to the same area for comparison purposes. Note that here we use a new run with the L18 feedback model, not the L18 simulations themselves. The L18 prescription uses $f_{\text{boost}} = 5$, while all other runs use $f_{\text{boost}} = 1$. All runs use the Isolated MW IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{HN},0} = 0$, and show all clusters formed before $z = 1.5$.

SN feedback (present in both variations to some extent) is important for dispersing gas before it can be accreted by the cluster. While the simulation with the L18 model uses $f_{\text{boost}} = 5$ instead of $f_{\text{boost}} = 1$, other runs varying f_{boost} show no little difference in either the duration of star formation or ϵ_{int} , indicating that the SN timing is responsible.

Despite these differences in the duration of star formation and ϵ_{int} , we see no significant differences in the star particle mass functions.

3.1.1 Discreteness of supernova

In addition to multiple runs with different prescriptions for SN feedback, we also ran one simulation with continuous energy injection from SN. This run uses the Isolated MW IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 1$, and $f_{\text{HN},0} = 0$. The number of SN still follows the IMF integral as in equation (10), but with the modification that we do not require there to be an integer number of SN in each time-step. We find that this change makes little difference to galaxy properties. The star formation rate was not affected, and neither were star cluster properties, including their mass function and age spread.

We note that the similarity between these two runs is despite real differences in how the energy is injected over time. The SN rate changes with time, but is within the range of $(2 - 6) \times 10^{-10} N_{\text{SN}} M_{\odot}^{-1} \text{yr}^{-1}$. Our typical time-steps on the highest refinement levels are below 10^3 yr, so even massive clusters with $M = 10^6 M_{\odot}$ do not have an SN every time-step. Clusters of mass $M = 10^3 M_{\odot}$ have only 10 SN over the ~ 40 Myr time-scale for SN feedback, resulting in significant gaps between SNe. The onset of SN can also be delayed in low-mass clusters, as the decrease in the normalization of the IMF means we need to integrate to lower stellar masses to reach one star (equation 10). These results indicate that the total injected energy and the timing of the onset of SN cause larger differences than does discretizing SN events.

To summarize, we find that different prescriptions to change the onset of SN (without changing the total energy injection) do not affect any galaxy-scale properties, but do affect the properties of

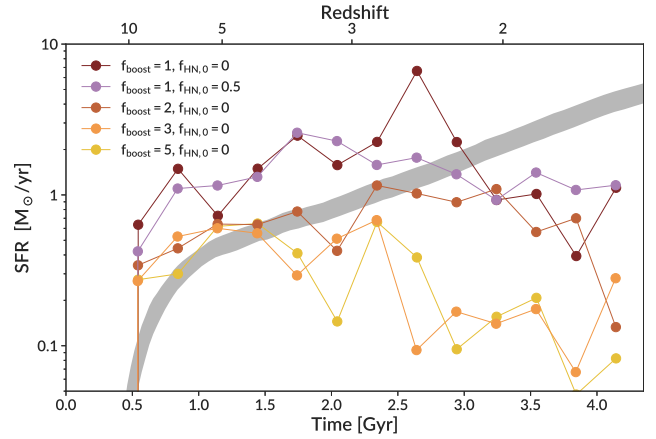


Figure 6. A comparison of the star formation history for the central galaxy in the Isolated MW IC when varying f_{boost} and $f_{\text{HN},0}$. The shaded region shows the expected star formation history as given by UNIVERSEMACHINE. All runs use $\epsilon_{\text{ff}} = 100$ per cent.

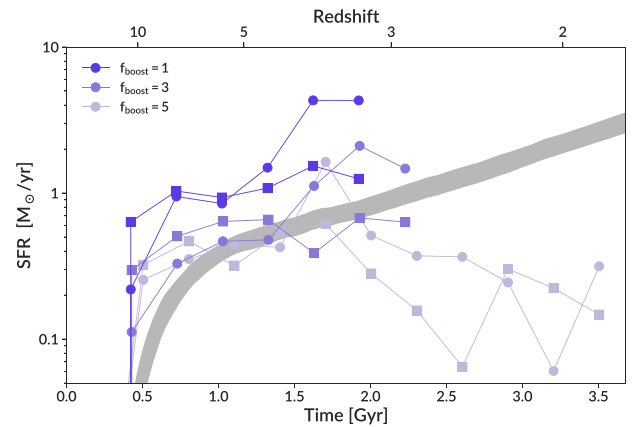


Figure 7. Same as Fig. 6, but for the Thelma & Louise IC and only showing variations in f_{boost} . There are two main galaxies in each run. Circles represent the MW analogue, with squares representing M31. All runs use $\epsilon_{\text{ff}} = 100$ per cent and $f_{\text{HN},0} = 0$.

star clusters. When SN feedback is delayed, massive clusters have longer formation time-scales, and all clusters have higher ϵ_{int} . When comparing discretized SN to continuous energy injection, we find no significant differences.

3.2 Strength of supernova feedback

Our simulations have two main parameters to control the strength of SN feedback: f_{boost} and $f_{\text{HN},0}$. In this section, we explore how those parameters affect our results.

In Figs 6 and 7, we show the impact of these two parameters on the star formation history of the main galaxies. In Fig. 6, we show the star formation history of the single central galaxy of the Isolated MW IC, while in Fig. 7, we show two lines for each run representing the two main galaxies in a Local Group-like environment. We also show the expected star formation history as given by UNIVERSEMACHINE (Behroozi et al. 2019). However, we note that the MW assembly history may be atypical for haloes of its mass, as both the ancient merger of Gaia–Enceladus Sausage and the current infall of the LMC influence its evolution (Evans et al. 2020).

First, we find that f_{boost} has a strong impact on the global star formation rate. Higher values of f_{boost} result in generally lower star formation rates. In the *Isolated MW* runs shown in Fig. 6, the run with $f_{\text{boost}} = 5$ matches the *UNIVERSEMACHINE* prediction well until roughly $z \approx 4$, at which point the star formation rates start to decline significantly. This is similar to what we see for the $f_{\text{boost}} = 5$ run in *Thelma & Louise* in Fig. 7. The star formation rate is reasonable until $z \approx 4$, with a significant decline afterwards. A value of $f_{\text{boost}} = 3$ matches *UNIVERSEMACHINE* more closely in both ICs, but in the *Isolated MW* IC the star formation rate drops off significantly after $z = 3$. The $f_{\text{boost}} = 3$ run using *Thelma & Louise* has only progressed to $z = 2.8$ at the time of writing, so it remains possible that its star formation rate will drop as it did in the *Isolated MW* run. However, we must be careful making direct comparisons between different ICs, as it is likely that they will have different star formation histories. In particular, Santistevan et al. (2020) found that Local Group-like galaxies form earlier than isolated galaxies. They conclude that the denser environment of Local Group-like pairs causes the initial collapse of haloes to happen earlier (Gallart et al. 2015). This leads to more mass forming earlier, and this buildup of stellar mass may affect how feedback affects the galaxy at later times.

A slightly lower value of $f_{\text{boost}} = 2$ matches *UNIVERSEMACHINE* well up to $z \approx 2$ before decreasing greatly. Finally, runs with $f_{\text{boost}} = 1$ have the highest levels of star formation in both ICs. This high level has persisted in *Thelma & Louise* until the last available output, but in *Isolated MW* the star formation rate dropped dramatically starting at $z = 2$. Even this low value of f_{boost} is not able to produce reasonable galactic star formation histories over the full time range spanned by these simulations.

L18 calibrated f_{boost} in their simulations, finding a preferred value of $f_{\text{boost}} = 5$. The difference in our result is due to the changes in hydrodynamics. As described above in Section 2.5, that change led to a decrease in the amount of cold gas that reaches the galaxy. This requires changes to the feedback modelling to compensate. Without decreasing f_{boost} , the galaxies have lower total gas mass and less cold gas, which leads to less molecular gas. Since molecular gas is required by our star formation prescription, this decrease leads to less star formation.

While we find that f_{boost} has a strong impact on the star formation rate, we find that $f_{\text{HN},0}$ does not. In Fig. 6, runs with $f_{\text{boost}} = 1$ have similar star formation histories, regardless of the value of $f_{\text{HN},0}$. While we do not show runs varying $f_{\text{HN},0}$ in Fig. 7 for clarity, runs with $f_{\text{HN},0} = 0, 5$ per cent, and 20 per cent all show similar star formation rates (all using $f_{\text{boost}} = 5$).

This is likely due to the metallicity dependence of the supernova fraction f_{HN} (see equation 11). The value of f_{HN} is highest at low metallicity, but decreases rather quickly with metallicity. Fig. 8 shows the metallicity of stars forming at different times and their f_{HN} . This plot uses the run on the *Isolated MW* IC with maximum $f_{\text{HN},0} = 50$ per cent, yet the quick enrichment means that the bulk of clusters have $f_{\text{HN}} < 10$ per cent. As shown in Fig. 1, this small f_{HN} produces energy injection rates not too dissimilar from $f_{\text{HN}} = 0$. This small change is in contrast to the large changes in momentum feedback that come from varying f_{boost} by a factor of 5, explaining why f_{boost} has a strong impact on galactic properties while $f_{\text{HN},0}$ does not.

While changes in f_{boost} lead to dramatic changes in global galaxy properties, the changes to the cluster mass function are more subtle. Fig. 9 shows the initial cluster mass function for the *Isolated MW* IC when varying f_{boost} and $f_{\text{HN},0}$. We show all clusters formed before $z = 4$, as this higher redshift reduces the differences caused by variations in the star formation rate and includes a higher fraction of low-metallicity clusters where f_{HN} could potentially make a

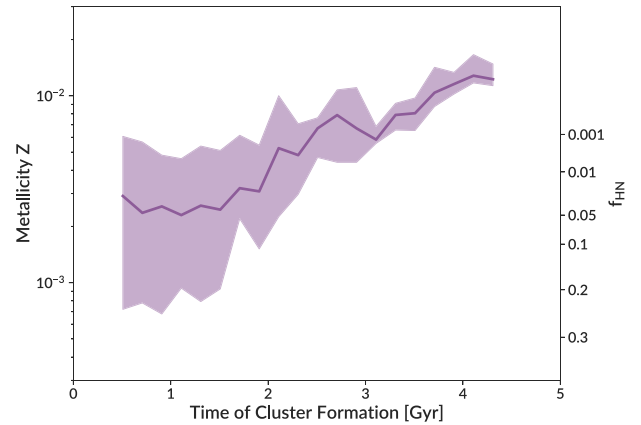


Figure 8. The stellar metallicity Z (not scaled to solar metallicity) of clusters forming across cosmic time in the run using the *Isolated MW* IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 1$, and $f_{\text{HN},0} = 50$ per cent. The shaded region shows the interquartile range at a given age, while the solid line shows the median. The right axis labels shown the supernova fraction at a given metallicity.

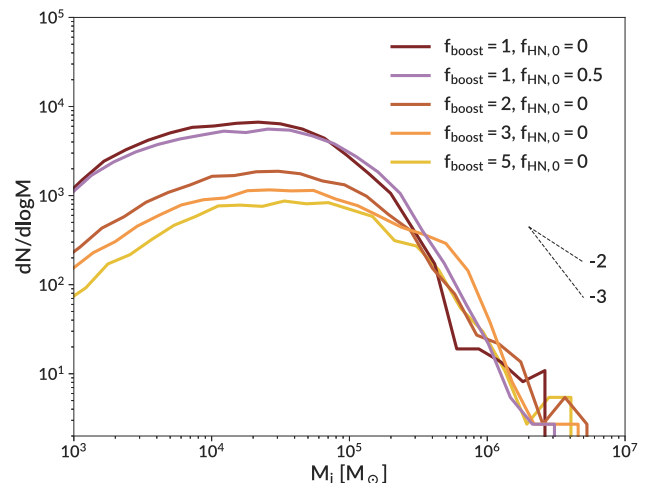


Figure 9. A comparison of the initial cluster mass function for runs with varied f_{boost} and $f_{\text{HN},0}$ at $z = 4$. Black dashed lines indicate power-law slopes of -2 and -3 . The lower limit of the plot corresponds to one cluster. All runs use the *Isolated MW* IC and $\epsilon_{\text{ff}} = 100$ per cent.

difference. The normalization changes reflect the change in total stellar mass. Interestingly, the high mass end is less affected by f_{boost} than the low-mass end. A lower f_{boost} serves to increase the number of low-mass particles without systematically increasing the number of massive clusters or the maximum cluster mass. While not shown in Fig. 9, we see the same trends when examining the runs using the Local Group ICs.

We find little difference in the cluster mass function when changing $f_{\text{HN},0}$. Fig. 9 shows little difference between $f_{\text{HN},0} = 50$ per cent and $f_{\text{HN},0} = 0$ for $f_{\text{boost}} = 1$ for masses below $10^6 M_{\odot}$. However, the run with $f_{\text{HN},0} = 0$ has several clusters with masses above $10^6 M_{\odot}$, while the run with $f_{\text{HN},0} = 50$ per cent does not. There are very few clusters in these mass ranges, so stochasticity may play a role in these results. We also examined the low-metallicity clusters separately, again finding no difference. This is true as well of the runs with the Local Group ICs.

To quantitatively evaluate the shape of the mass functions, we fit them with a power law. As our mass functions do not show a power-law behaviour down to low masses, we restrict our fit to masses above $10^5 M_\odot$ where it is approximately a power-law. Again we note that we are using the particle masses without including the bound fraction, so these results are not directly comparable to observations. Including the bound fraction generally makes the mass function shallower, as high-mass clusters have a higher bound fraction (see Section 4). For $f_{\text{HN},0} = 0$, we find slopes of -2.94 , -2.48 , -2.16 , and -2.31 for $f_{\text{boost}} = 1, 2, 3$ and 5 , respectively. For $f_{\text{HN},0} = 50$ per cent and $f_{\text{boost}} = 1$, we find -2.78 . Lower values of f_{boost} tend to have steeper slopes due to the higher number of low-mass clusters. The $f_{\text{boost}} = 3$ run has the shallowest slope due to the large number of clusters at $\approx 5 \times 10^5 M_\odot$ that deviate from a pure power-law fit and draw the fit toward a shallower slope. This feature becomes less prominent at $z = 1.5$ as more clusters form and fill out the mass function more evenly. We see similar trends in the Local Group runs, where the slope takes values of -2.62 , -2.40 , and -2.22 for $f_{\text{boost}} = 1, 3$, and 5 respectively.

Lastly, we examined the visual appearance of the gas distribution in these galaxies. L18 found that reducing f_{boost} to 3 led to a dramatic increase in star formation and the formation of an axisymmetric disc, while runs with $f_{\text{boost}} = 5$ produced very irregular galaxies (Meng, Gnedin & Li 2019). Here, we find that all of our runs produce irregular galaxies, even with $f_{\text{boost}} = 1$.

In summary, we find that higher values of f_{boost} can greatly decrease the galactic star formation rate by decreasing the number of low-mass clusters that form, without changing the number of massive clusters. Increasing the initial supernova fraction $f_{\text{HN},0}$ has little effect on galaxy properties. The fraction quickly approaches zero as metallicity increases, leading to little difference in the injected energy and momentum.

3.3 Molecular gas prescription

A key ingredient in modelling star formation in our simulations is the amount of molecular gas, as we require a given cell to have a mass fraction of molecular gas greater than 50 per cent to seed a cluster particle.

As discussed in Gnedin & Draine (2014) and appendix A7 of Gnedin & Kravtsov (2011), the clumping factor C_ρ is one of the tunable parameters of the molecular gas model. This factor accounts for the fact that gas is clustered on scales that are not resolved in a given simulation, so H_2 formation would be missed. Larger values of the clumping factor produce more molecular gas at a given surface density. Numerical simulations of turbulent molecular clouds find lognormal density distributions with widths that imply $C_\rho \approx 3$ – 10 (McKee & Ostriker 2007). Gnedin & Kravtsov (2011) and Gnedin & Draine (2014) calibrated the clumping factor in the ART code based on simulations, finding that values in the range of 10 to 30 work well. However, those simulations had lower resolution than our runs. This would imply that our runs should prefer a lower clumping factor, because they are resolving more substructure and leaving less on subgrid scales.

Motivated by the disagreement between our simulations and the observed global galactic star formation history (e.g. Fig. 6), we explored a range of molecular gas prescriptions. We ran simulations with a range of clumping factors, using $C_\rho = 3, 10$, and 30 . L18 used 10, as do all other runs presented in this paper. We also used one run with $C_\rho = 3$ where we changed the prescription for shielding from that of Gnedin & Draine (2014) to that of Gnedin & Kravtsov (2011). The Gnedin & Draine (2014) model includes the effects of line

overlap in the Lyman–Werner bands, increasing self-shielding, which is particularly relevant in low-metallicity environments with less dust shielding. However, both models for self-shielding were calibrated using runs with lower resolution than our runs (>50 pc compared to 3–6 pc) and with a different feedback model. These differences in simulation setup can affect the performance of the H_2 formation model, so we decided to explore both shielding prescriptions. All runs used $f_{\text{boost}} = 1$, $f_{\text{HN},0} = 0$, and $\epsilon_{\text{ff}} = 100$ per cent.

As expected, the only significant differences caused by C_ρ were in the amount of molecular gas. While the mass of molecular gas in each run varies greatly with time, we find a general trend that larger values of C_ρ produce more molecular gas. We see little change in molecular gas masses when changing the shielding prescription. These differences in the amount of molecular gas when changing C_ρ led to some differences in star formation histories. The total stellar mass at $z = 1.5$ for the run with $C_\rho = 3$ is $3 \times 10^9 M_\odot$, while the mass for the run with $C_\rho = 30$ is $6 \times 10^9 M_\odot$. In particular, a higher clumping factor leads to more late-time star formation.

3.4 Star formation efficiency

The local star formation efficiency per free-fall time ϵ_{ff} is a key parameter of our model (see equation 1). As L18 showed, this parameter strongly influences many star cluster properties, particularly the mass function, while not strongly affecting the global galaxy properties. We continue that exploration here.

As ϵ_{ff} controls how fast star particles accrete material, we expect it to be reflected in the duration of cluster formation episodes. We find that to be the case. In particular, we find that runs with low values of ϵ_{ff} often fail to finish forming massive clusters before the algorithmic end to a star formation episode at 15 Myr. For example, in the run using the Isolated MW IC, $\epsilon_{\text{ff}} = 1$ per cent, $f_{\text{boost}} = 1$, and $f_{\text{HN},0} = 0$, only 20 per cent of clusters with masses above $10^5 M_\odot$ finished their formation before it was automatically stopped.

When this time cap is imposed, cluster formation ends even when gas is available to continue accreting on to the cluster. Therefore, we cannot interpret these particles as the end-products of cluster formation. Their masses are not self-consistently determined by their feedback. The masses we obtain are lower limits to the true masses that would have formed over longer time-scales. However, as we will discuss more in Section 5.3, such long age spreads of stars within a single cluster are ruled out by observations. We define runs as having failed cluster formation if more than 50 per cent of clusters with masses above $10^5 M_\odot$ have durations longer than 14 Myr. This applies to all runs with $\epsilon_{\text{ff}} = 1$ per cent and the run using the Isolated MW IC, $\epsilon_{\text{ff}} = 10$ per cent, $f_{\text{boost}} = 1$, and $f_{\text{HN},0} = 0$. While we still include these runs in plots, we indicate the cluster mass ranges where they are unreliable using dashed lines (namely Figs 11, 12, 16, and 19), or use completely dashed lines when mass is not an explicit variable (namely Figs 17, 22, and 24). We defer a full investigation of this failed cluster formation to Section 5.4.

To illustrate the difference in the time-scale of cluster formation, Fig. 10 shows the cumulative distribution of age spread τ_{spread} for runs using the Isolated MW IC. The dependence on ϵ_{ff} is clear. For massive clusters, the median age spread is 8.6 Myr for $\epsilon_{\text{ff}} = 1$ per cent, while it is 2.4 Myr for $\epsilon_{\text{ff}} = 10$ per cent and 0.9 Myr for $\epsilon_{\text{ff}} = 100$ per cent. For $\epsilon_{\text{ff}} = 1$ per cent many clusters have unphysically long age spreads, some longer than 15 Myr. We note that the age spread can be longer than the duration of star formation in some cases, as it is a measure of the variance in the star formation rate rather than simply its length. Atypical star formation histories,

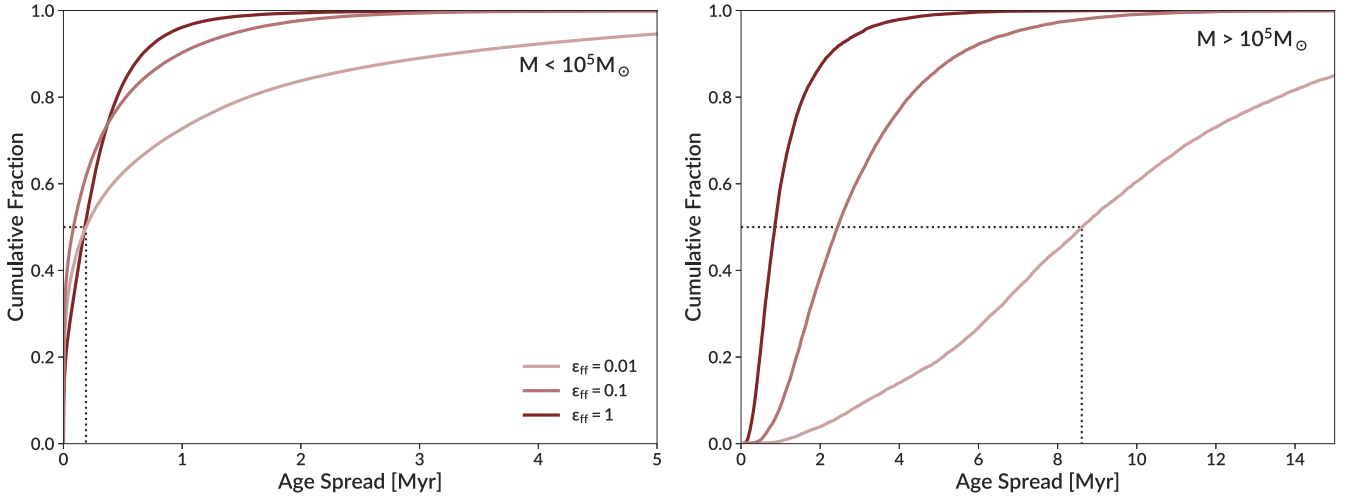


Figure 10. The cumulative distribution of the star particle internal age spread τ_{spread} for the Isolated MW IC runs with varied ϵ_{ff} . Note that this is not the duration of star formation as plotted in Fig. 4, it is the age spread as defined in equation (16) evaluated at the end of cluster formation. The left-hand panel shows clusters less massive than $10^5 M_{\odot}$, while the right-hand panel shows clusters more massive than $10^5 M_{\odot}$. Note the different range spanned by the two panels. The dotted line shows the longest median age spread. All runs use $f_{\text{boost}} = 1$, $f_{\text{HN},0} = 0$, and show all clusters formed before $z = 1.5$.

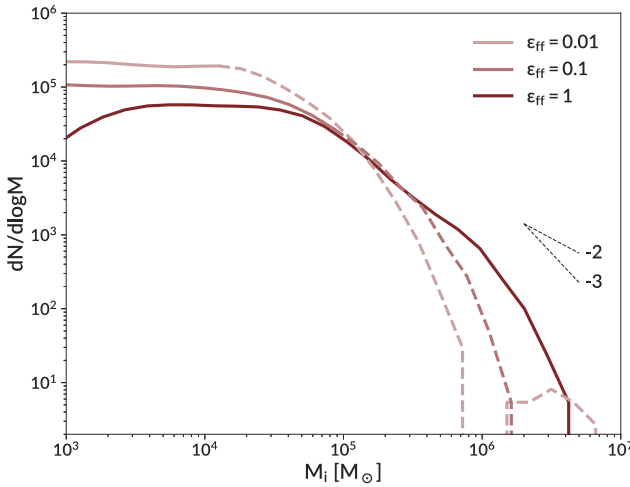


Figure 11. A comparison of the star particle initial mass function for the Isolated MW IC runs with varied ϵ_{ff} . For runs with failed cluster formation, dashed lines indicate the range where more than 50 per cent of clusters have formation durations longer than 14 Myr. Black dashed lines indicate power-law slopes of -2 and -3 . The lower limit of the plot corresponds to one cluster. All runs use $f_{\text{boost}} = 1$, $f_{\text{HN},0} = 0$, and show all clusters formed before $z = 1.5$.

such as one with bursts of star formation at early and late times, can lead to large values of τ_{spread} . There is also a clear mass dependence. Clusters with masses below $10^5 M_{\odot}$ and $\epsilon_{\text{ff}} \geq 10$ per cent have median age spreads less than 0.2 Myr, with all low-mass clusters from those runs having age spreads less than 2 Myr. However, for $\epsilon_{\text{ff}} = 1$ per cent there is a clear tail to long age spreads even among low-mass clusters, with some clusters having age spreads as long as 10 Myr.

We next investigate the effect of ϵ_{ff} on the cluster mass functions. Fig. 11 shows the initial cluster mass function for runs using the Isolated MW IC, and Fig. 12 shows the same for the Local Group runs. Similar trends are seen in both plots. Higher values of ϵ_{ff} lead to more massive clusters and a higher maximum cluster mass, while lower values of ϵ_{ff} produce more low-mass clusters. The exception

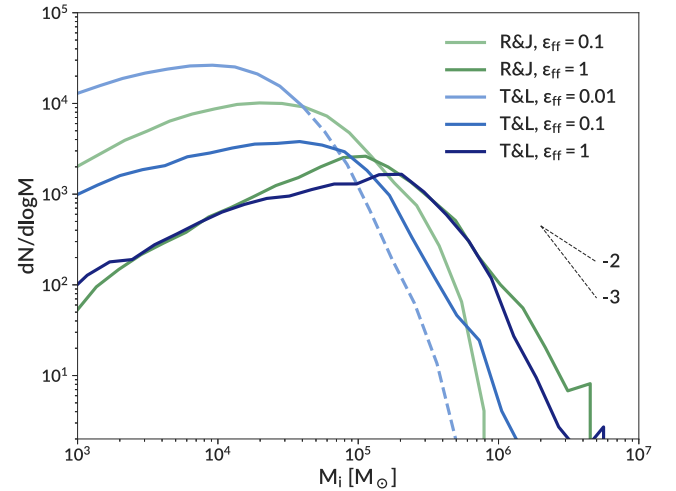


Figure 12. A comparison of the star particle initial mass function for the Local Group ICs with varied ϵ_{ff} . For the $\epsilon_{\text{ff}} = 1$ per cent run with failed cluster formation, dashed lines indicate the range where more than 50 per cent of clusters have formation durations longer than 14 Myr. Black dashed lines indicate power-law slopes of -2 and -3 . The lower limit of the plot corresponds to one cluster. All runs use $f_{\text{boost}} = 5$, $f_{\text{HN},0} = 20$ per cent, and show all clusters formed before $z = 3.3$, the lowest redshift that all simulations have reached.

to this is a handful of very massive clusters that formed in the Isolated MW $\epsilon_{\text{ff}} = 1$ per cent run, leading to a separate hump in the high-mass end of the mass function. This indicates that even with low values of ϵ_{ff} , massive clusters are still possible, although typically rare. We note that we do not see such hump in the Thelma & Louise run with $\epsilon_{\text{ff}} = 1$ per cent.

The slope of the high-mass end of the mass function varies with ϵ_{ff} , with the mass function being shallower for higher values of ϵ_{ff} . As with all calculations of the mass function slope, we restrict our fit to clusters above $10^5 M_{\odot}$. For $\epsilon_{\text{ff}} = 1$ per cent, the slope is between -3.78 and -4.41 for runs on the different ICs, while for $\epsilon_{\text{ff}} = 10$ per cent it is between -2.94 and -3.42 , and for $\epsilon_{\text{ff}} = 100$ per cent it is between -2.25 and -2.60 .

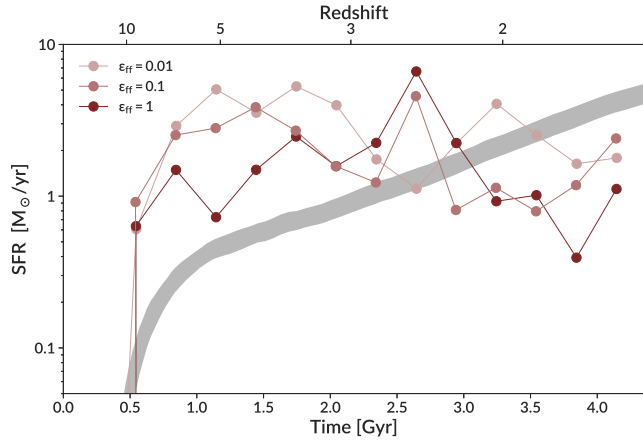


Figure 13. A comparison of the star formation rate of the central galaxy in the `Isolated` MW IC when varying ϵ_{ff} . All runs use $f_{\text{boost}} = 1$ and $f_{\text{HN},0} = 0$.

The exact shape of the mass function is somewhat different between the runs that use the `Isolated` MW IC and those that use the `Local Group` ICs, with the `Local Group` runs having fewer low-mass particles. These `Local Group` runs used $f_{\text{boost}} = 5$, which decreases the number of low-mass clusters compared to lower values of f_{boost} (see Fig. 9). The different redshift of these runs also likely contributes. We find that the majority of massive particles form at very high redshift or in galactic mergers, when the star formation rate is high. This matches what was seen in L18, and agrees with both observations and theoretical expectations (Portegies Zwart et al. 2010; Kruijssen 2014). In more quiescent epochs, high-mass particles do not form, giving proportionally more low-mass particles. As time progresses, more low-mass clusters are likely to form in the `Local Group` runs, possibly making their mass functions more similar to those seen in the `Isolated` MW IC.

While ϵ_{ff} significantly affects cluster properties, it does not change the galactic star formation rate appreciably. Fig. 13 shows the star formation histories of runs when varying ϵ_{ff} while holding $f_{\text{boost}} = 1$ and $f_{\text{HN},0} = 0$ constant. Here, we find that lower values of ϵ_{ff} lead to somewhat higher star formation rates at early times. These star formation rates at $z \approx 5$ are significantly higher than predicted by UNIVERSEMACHINE, and tend to decline with time rather than increase. However, we find opposite trends during the major merger at $z \approx 2.6$, when the high ϵ_{ff} runs show a stronger burst. In the runs using the `Local Group` analogs, the star formation history does not change significantly with ϵ_{ff} .

In summary, we find that ϵ_{ff} does not have a significant impact on the galactic star formation rate, but does strongly influence star cluster properties. In particular, higher values of ϵ_{ff} lead to more massive star clusters and shorter time-scales for cluster formation. These results confirm those found in L18, indicating that they are robust predictions of our simulations.

3.5 Virial criterion

One of the other changes to our star cluster formation prescription was the addition of a criterion restricting star-forming gas to be gravitationally bound (see the beginning of Section 2.2). To investigate the difference this makes in cluster properties, we ran one simulation with the virial criterion turned off. While we find no significant differences in large scale galactic properties, we find differences in the star cluster populations. Fig. 14 shows the mass function for runs with and without the virial criterion. The addition of the virial criterion leads to

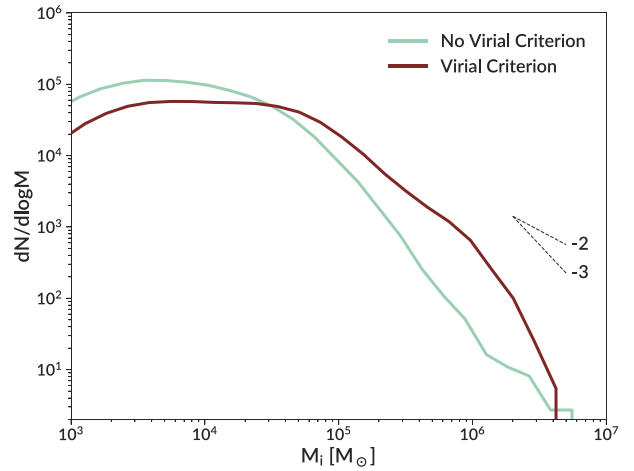


Figure 14. A comparison of the initial cluster mass function for runs with and without the virial criterion for seeding star formation. Black dashed lines indicate power-law slopes of -2 and -3 . Both runs used the `Isolated` MW IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 1$, $f_{\text{HN},0} = 0$, and show all clusters formed before $z = 1.5$.

more high-mass clusters and fewer low-mass clusters. Quantitatively, the power-law slopes of the mass functions for clusters above $10^5 M_{\odot}$ are -2.60 for the run with the virial criterion and -3.30 for the run without it. While the maximum cluster mass is similar between the two runs, there are significantly more clusters with masses above $10^6 M_{\odot}$ when the virial criterion is enabled.

The increase in the number of high-mass clusters is expected, as equation (3) shows lower gas densities lead to higher virial parameters. The cut on the virial parameter prevents these lower density GMCs from forming stars until they accrete more gas and collapse to higher density, leading to more total mass available for star formation. The later onset of star formation also delays stellar feedback, allowing more gas to accrete on to the cluster. These processes shift many low-mass clusters to higher masses, explaining the decrease in the number of low-mass clusters. In addition, as the virial criterion allows more gas accretion on to the GMC, its larger mass becomes more difficult to disperse with feedback, leading to longer durations of star formation. As a consequence of these effects, clusters have higher values of ϵ_{int} when the virial criterion is enabled. In Fig. 15, we show the distribution of ϵ_{int} with and without the virial criterion. Both distributions have widths ≈ 0.25 dex, but the mean value for the run with the virial criterion is significantly higher (35 per cent compared to 21 per cent).

In the run where we did not impose the virial criterion, we output the virial parameter α_{vir} of each cluster as it formed. Using this information, we can post-process the results to see if there are any correlations between the virial parameter and the resulting cluster properties. We find that clusters with $\alpha_{\text{vir}} < 10$ tend to have higher initial masses, higher ϵ_{int} , and higher initial bound fractions than those with $\alpha_{\text{vir}} > 10$. The virial criterion acts in a biased fashion to allow star formation to happen in regions that preferentially lead to higher mass clusters. Additionally, regions with $\alpha_{\text{vir}} > 10$ are able to accrete more material over time until they pass the $\alpha_{\text{vir}} < 10$ threshold, increasing the cluster mass that formed out of a given GMC.

In summary, we find that adding the requirement that star-forming gas have a virial parameter $\alpha_{\text{vir}} < 10$ increases the number of massive clusters, gives clusters a longer formation time-scale, and leads to higher values of ϵ_{int} .

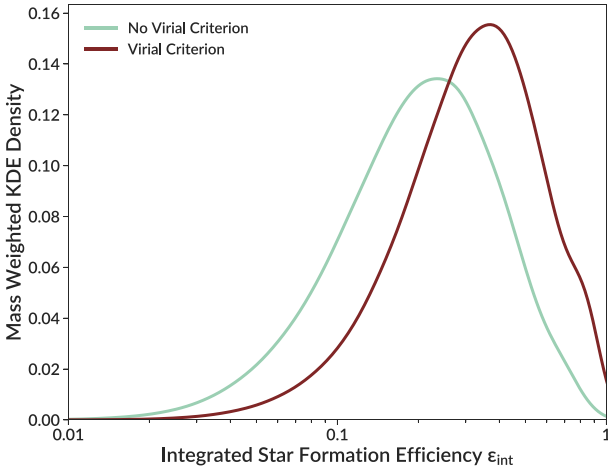


Figure 15. Kernel density estimation for the distribution of integrated star formation efficiency for clusters in the runs with and without the virial criterion. We use a Gaussian kernel with a width of 0.05 dex. Each curve is normalized to the same area for comparison purposes. Both runs used the Isolated MW IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 1$, $f_{\text{HN},0} = 0$, and show all clusters formed before $z = 1.5$.

4 EVOLUTION OF THE CLUSTER MASS FUNCTION

In the previous section, we exclusively used the masses of the star particles at the end of their formation process. As not all stars are gravitationally bound to the newly formed cluster, we must incorporate the initial bound fraction to obtain the observable cluster masses. In addition, the plots in the previous section showed the distributions of initial masses for all clusters formed over the full time spanned by the simulation. This is not observable. In this section, we include the cluster bound fraction and present the instantaneous cluster mass function at a given redshift to allow for more direct comparison with observations. While these are not true mock observations, the results shown here accurately represent the existing cluster populations at a given redshift in our simulations.

We start by examining the cluster initial bound fraction, which is needed to turn raw particle masses into bound cluster masses. Fig. 16 shows the initial bound fraction of clusters as a function of mass. As in L18, we see the trend of higher mass clusters having higher bound fraction. Additionally, runs with higher ϵ_{ff} have higher bound fractions at a given particle mass.

Our prescription for the initial bound fraction (equation 4) makes it solely dependent on the integrated star formation efficiency ϵ_{int} . In Fig. 17, we show the distributions of ϵ_{int} . Runs with lower ϵ_{ff} have lower ϵ_{int} . For a given run, the spread is due to trends with mass, where high-mass clusters have higher ϵ_{int} than low-mass clusters. Quantitatively, the mean value of ϵ_{int} takes values of 1.2, 7.2, and 30 per cent for $\epsilon_{\text{ff}} = 1$, 10, and 100 per cent, respectively. As ϵ_{ff} increases, the widths of these distributions decrease, with values of 0.30, 0.24, and 0.17 dex, respectively.

The trend of higher ϵ_{int} with higher ϵ_{ff} is a direct consequence of ϵ_{ff} controlling the cluster formation rate (equation 1). Higher ϵ_{ff} leads to higher star formation rates, allowing the cluster to accrete more of the gas from its surroundings. This is reflected in the duration of cluster formation in runs with different ϵ_{ff} . A lower value ϵ_{ff} causes clusters to form more slowly. With a slow star formation rate, feedback also starts before the cluster has accreted a significant fraction of the surrounding gas, leading to lower ϵ_{int} . The different time-scales also

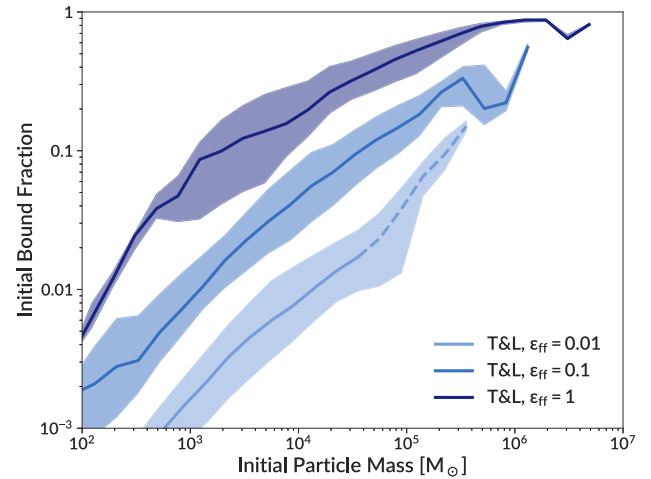


Figure 16. The initial bound fractions for runs using the Thelma & Louise IC and varying ϵ_{ff} . The solid line shows the median, with the shaded region showing the interquartile range of the distribution of the initial bound fraction at that mass. The mass plotted here is the particle mass at the end of cluster formation, not the bound cluster mass, so that the plotted variables are independent. For the $\epsilon_{\text{ff}} = 1$ per cent run with failed cluster formation, dashed lines indicate the range where more than 50 per cent of clusters have formation durations longer than 14 Myr. We only show the Thelma & Louise IC for clarity, but Romeo & Juliet and Isolated MW show the same behaviour. All runs use $f_{\text{boost}} = 5$, $f_{\text{HN},0} = 20$ per cent, and show all clusters formed before $z = 3.3$, the lowest redshift that all simulations have reached.

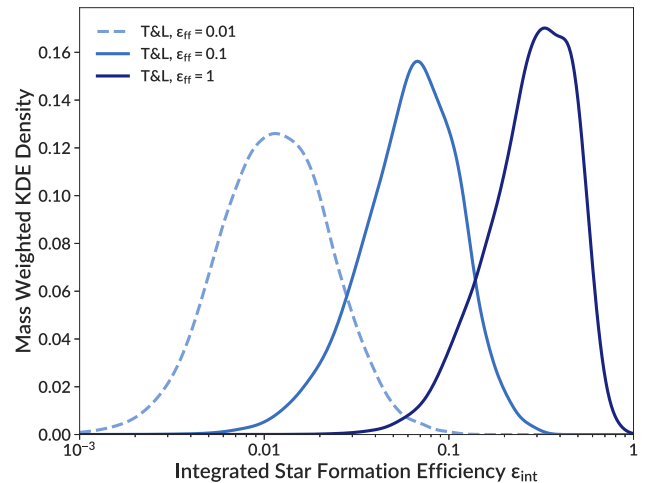


Figure 17. Kernel density estimation for the distribution of integrated star formation efficiency for clusters in the Thelma & Louise IC with variations in ϵ_{ff} . We use a Gaussian kernel with a width of 0.05 dex. Each curve is normalized to the same area for comparison purposes. All runs use $f_{\text{boost}} = 5$, $f_{\text{HN},0} = 20$ per cent, and show all clusters formed before $z = 3.3$, the lowest redshift that all simulations have reached. We plot the $\epsilon_{\text{ff}} = 1$ per cent run with a dashed line as that run had many clusters that failed to finish forming.

likely lead to the change in width of the distributions. As low values of ϵ_{ff} lead to longer time-scales of cluster formation, there is more possibility for variation in the accretion history of the GMC. High values of ϵ_{ff} form quickly, so they are forming mostly out of the gas that was present at cluster birth.

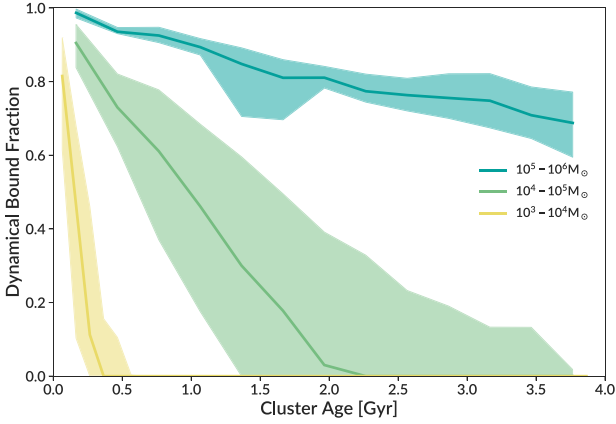


Figure 18. Evolution of the dynamical bound fraction f_{dyn} as a function of cluster age for clusters in different mass ranges. Lines show the median, with the shaded region showing the interquartile range. Clusters are grouped according to their initial bound mass at formation. The plot shows clusters in the central galaxy of the run using the *Isolated* MW IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 1$, $f_{\text{HN},0} = 0$, and shows all clusters formed before $z = 1.5$.

In addition to the initial bound fraction, we also need to account for stellar evolution and dynamical disruption, which both cause clusters to lose mass with time. These processes are calculated in simulation runtime. In general, the mass M_b bound to a cluster at time t can be written as

$$M_b(t) = M_i f_i f_{\text{se}}(t) f_{\text{dyn}}(t), \quad (18)$$

where M_i is the initial particle mass, f_i is the initial bound fraction, $f_{\text{se}}(t)$ accounts for mass loss due to stellar evolution, and f_{dyn} accounts for mass lost due to tidal stripping (Li & Gnedin 2019; Meng & Gnedin 2022). Our feedback scheme self-consistently decreases the stellar mass of the cluster whenever mass is ejected into the ISM, and dynamical disruption is calculated as described in Section 2.3. In Fig. 18, we show the impact of disruption on clusters of different mass, taking as an example the run using the *Isolated* MW IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 1$, and $f_{\text{HN},0} = 0$. We choose a run using the *Isolated* MW IC as it reached a lower redshift, so clusters have more time to disrupt. Clusters with masses below $10^4 M_\odot$ are entirely disrupted within 500 Myr. Clusters of intermediate mass $10^4 - 10^5 M_\odot$ persist for a few Gyr, but do not survive until the present. However, clusters with masses above $10^5 M_\odot$ survive throughout the lifetime of the simulation. Tidal disruption only decreases the mass of these clusters by approximately 20 per cent over the 4 Gyr length of this simulation.

Using these disruption calculations, we now present the mass function of bound clusters at various redshifts. In Fig. 19, we show the mass function of the surviving clusters at $z = 4$ in the Local Group runs with varied ϵ_{ff} . This figure shows trends similar to those seen in Fig. 12, with several trends more exaggerated now that bound cluster mass is included. First, we note similar shapes. Our mass functions have a sharp cutoff at high masses, a peak, and a shallower decrease to low masses. This shape is seen in all runs with $\epsilon_{\text{ff}} \geq 10$ per cent. The position of the peak depends strongly on ϵ_{ff} . For $\epsilon_{\text{ff}} = 100$ per cent it is at approximately $10^5 M_\odot$, while it is closer to $10^4 M_\odot$ for $\epsilon_{\text{ff}} = 10$ per cent. This is due to a combination of three effects. First, as seen in Fig. 12, the initial particle masses are higher for higher values of ϵ_{ff} . Second, higher values of ϵ_{ff} give higher bound fractions, as shown in Fig. 16. The two effects magnify each other, such that higher values of ϵ_{ff} result in cluster mass functions

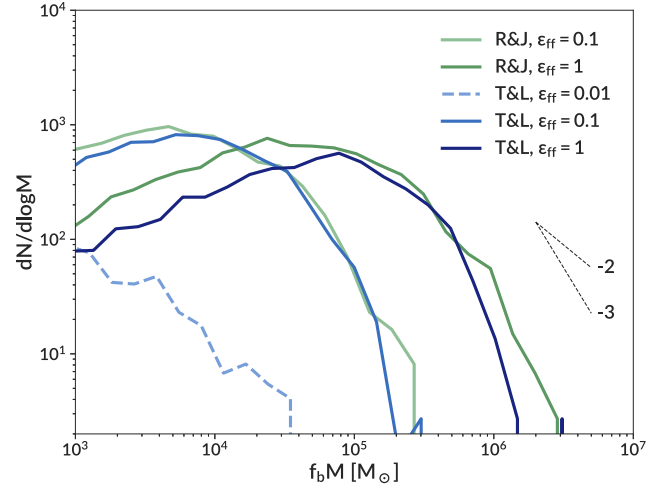


Figure 19. The bound mass function of all clusters present at $z = 4$ using the Local Group ICs for different values of ϵ_{ff} . For the $\epsilon_{\text{ff}} = 1$ per cent run with failed cluster formation, dashed lines indicate the range where more than 50 per cent of clusters have formation durations longer than 14 Myr. Black dashed lines indicate power-law slopes of -2 and -3 . The lower limit of the plot corresponds to one cluster. All runs use $f_{\text{boost}} = 5$ and $f_{\text{HN},0} = 20$ per cent.

that reach to significantly higher masses. The disparity is further increased by the effects of disruption, which preferentially removes low-mass clusters (Fig. 18). These three effects combine to produce dramatically different cluster mass functions when changing ϵ_{ff} . Of note, the $\epsilon_{\text{ff}} = 1$ per cent run has no existing clusters above $10^4 M_\odot$, while the $\epsilon_{\text{ff}} = 10$ per cent runs have no clusters above $3 \times 10^5 M_\odot$.

We also note that, as described in Section 3.2, the low-mass end of the mass function is sensitive to f_{boost} , with higher values of f_{boost} decreasing the number of low-mass clusters. The runs shown in Fig. 19 were all run with $f_{\text{boost}} = 5$. Lower values of f_{boost} would increase the number of low-mass clusters and give it a shape more similar to that seen in the local universe. Similarly, massive clusters tend to form in epochs of intense star formation, while low-mass clusters dominate in more quiescent epochs. As only the *The1ma* IC has any significant mergers after the redshift shown in this plot, we expect there to be more low-mass clusters as time progresses.

In Fig. 20, we show the evolution of the bound cluster mass function from $z = 6$ to 1.9 for the run using the *Romeo & Juliet* IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 5$, and $f_{\text{HN},0} = 20$ per cent. A significant fraction of clusters with masses above $2 \times 10^5 M_\odot$ are in place already at $z = 6$. More massive clusters form by $z = 4$, but we see little change in the massive end of the mass function beyond that redshift. At later epochs low-mass clusters dominate the mass function, particularly increasing the number of clusters around $10^4 M_\odot$. Clusters of low mass that appear in this plot are mainly from recent star formation. As Fig. 18 shows, clusters with masses below $10^5 M_\odot$ disrupt within a few Gyr, and clusters below $10^4 M_\odot$ disrupt within several hundred Myr.

We also show an analytical evolution of star clusters from the last output of this run at $z = 1.9$ to the present, following Li & Gnedin (2019). The prescription for tidal disruption (equations 6–8) depends on the galactic tidal field. In simulation runtime, we calculate it self-consistently, but to extrapolate to $z = 0$ we simply assume a constant value of $\Omega_{\text{tid}} = 175 \text{ Gyr}^{-1}$. This value was chosen to produce the same final number of clusters as are observed in the MW. It results in the disruption of most clusters with masses below $10^5 M_\odot$, and

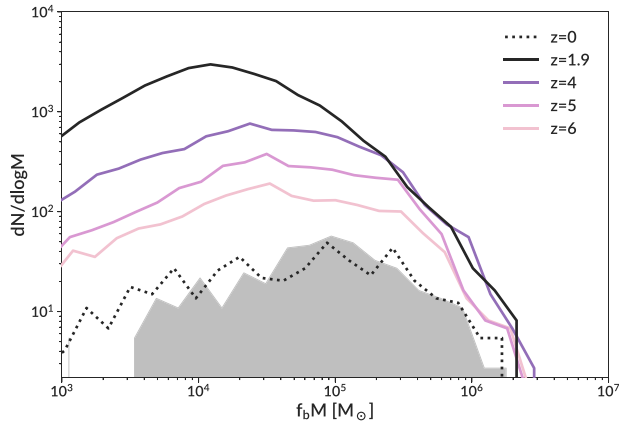


Figure 20. The bound mass function of all clusters present at a range of redshifts. All simulation lines are from the same run that uses the Romeo & Juliet IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 5$, and $f_{\text{HN},0} = 20$ per cent. The dotted line indicates the cluster population analytically evolved from $z = 1.9$ to 0. The shaded region shows the observed mass function of clusters in the MW.

decreases the masses of all surviving clusters, shifting the distribution to lower masses and decreasing the normalization. This calculation also assumes that no new clusters form after $z = 1.9$.

Our chosen value of $\Omega_{\text{tid}} = 175 \text{ Gyr}^{-1}$ is quite high. It is equivalent to a maximum eigenvalue of the tidal tensor $\lambda_m \approx 10^5 \text{ Gyr}^{-2}$. Meng & Gnedin (2022) examined the tidal field for the L18 simulations, finding that clusters experience such strong tidal fields only shortly after their birth. As they migrate away from the high-density star-forming regions, the tidal field decreases significantly to typical values $\lambda_m \approx 3 \times 10^3 \text{ Gyr}^{-2}$, or $\Omega_{\text{tid}} \approx 30 \text{ Gyr}^{-1}$. Choosing this low value of Ω_{tid} would significantly increase the number of low-mass clusters surviving to $z = 0$ in our simulations. However, this analysis was done at $z > 1.5$. The value of the tidal field may increase over time as the galaxy grows. Our adopted value is also similar to that used by Choksi & Gnedin (2019) in an analytic model for cluster formation and destruction. These authors find that $\Omega_{\text{tid}} = 200 \text{ Gyr}^{-1}$ can reproduce several observational results, including the GC mass function at $z = 0$ and the relation between galaxy halo mass and mass of its GC system.

We compare our results with the distribution of masses of the observed MW GCs. We use the V-band absolute magnitude from Harris (1996) along with the luminosity-dependent mass-to-light ratio

$$\frac{M}{L_V} = 1.3 + \frac{4.5}{1 + \exp(2M_V + 21.4)} \quad (19)$$

from Harris, Blakeslee & Harris (2017) to obtain GC masses. We find good agreement between the two mass functions. While we match the normalization by construction through our choice of Ω_{tid} , the similarity of the MF shape to that in the MW system is a test of the model. The maximum cluster mass matches the MW GCs well. We note that the Harris (1996) catalogue includes both *in situ* and *ex situ* clusters in the MW. As the simulation, $z = 0$ result comes from analytic evolution of all clusters in the central galaxies at $z = 1.9$, any later clusters that come in from later mergers would be missed. However, Fig. 20 uses the Romeo & Juliet IC, which has quick early growth with no significant mergers after $z = 1.9$ (Fig. 3). We therefore expect few clusters from later infalling satellites, making a comparison to the full MW population reasonable. We also note that the Romeo & Juliet IC has more massive clusters than

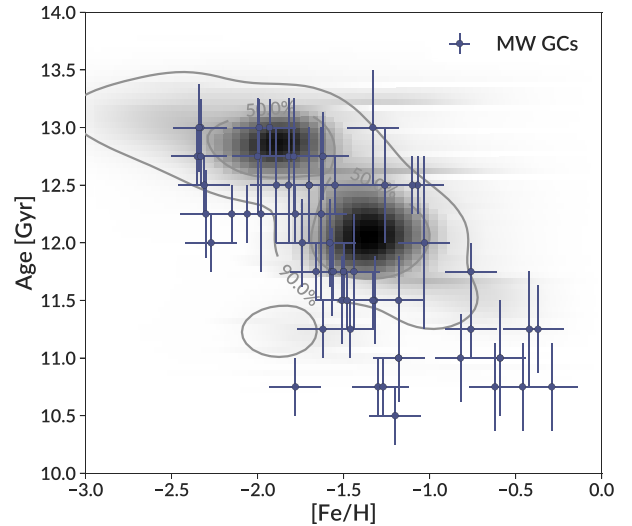


Figure 21. The age–metallicity relation for surviving simulated clusters and MW GCs. Data points show MW GCs from VandenBerg et al. (2013) and Leaman et al. (2013). Grey shaded regions and contours indicate simulated clusters with masses above $3 \times 10^3 M_{\odot}$ at $z = 0$ in the run using the Thelma & Louise IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 5$, and $f_{\text{HN},0} = 0$, with contours enclosing 50 and 90 per cent of the sample. Clusters from both central galaxies are included as there are no systematic differences between the two. The final output of this run corresponds to an age of 10 Gyr, meaning that all regions on the plot are accessible to the simulation.

the Thelma & Louise IC. This is likely because of its quick early growth (Fig. 3), increasing the star formation density at early times and leading to the formation of more massive clusters. For the Thelma & Louise runs, a lower value of Ω_{tid} is required to reproduce the high-mass end of the Galactic GC mass function, leading to too many simulated low-mass clusters.

In the runs with $\epsilon_{\text{ff}} = 100$ per cent, our present-day mass functions have more clusters with masses above $3 \times 10^5 M_{\odot}$ than seen in L18. This is a consequence of our initial mass functions extending to higher masses than in L18. These changes are primarily driven by the addition of the virial criterion. As Fig. 14 shows, the addition of this criterion significantly increases the number of massive clusters. The increase in the number of massive clusters allows us to increase the value of Ω_{tid} from 50 Gyr^{-1} (used by L18) to 175 Gyr^{-1} . In that work higher values of Ω_{tid} would have disrupted too many clusters. In the runs presented in this work, stronger disruption is required to produce an agreement for the massive end of the mass function while reducing the number of low-mass clusters.

Similarly to the mass function at $z = 4$, the mass functions of surviving clusters at $z = 0$ depend strongly on ϵ_{ff} . For all runs with $\epsilon_{\text{ff}} \leq 10$ per cent (not shown), we find no clusters above $4 \times 10^5 M_{\odot}$, and the overall distributions shift to lower masses. That is clearly inconsistent with the observed mass function of MW GCs.

Another important relation found in observations is the age–metallicity relation of MW GCs (e.g. VandenBerg 2000; Marín-Franch et al. 2009; Dotter et al. 2010). Metal-rich clusters form systematically later than metal-poor clusters, as the galaxy enriches its interstellar medium with time. In Fig. 21, we show the age–metallicity relation for simulated clusters that survive to $z = 0$ in the run using the Thelma & Louise IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 5$, and $f_{\text{HN},0} = 0$ and compare to observations of MW GCs presented in VandenBerg et al. (2013) and Leaman, VandenBerg & Mendel (2013). We find broad agreement between the simulated cluster

population and the MW GCs. While the plot shows only one run using the Thelma & Louise IC, we see similar trends in all ICs. We note that the value of Ω_{tid} used in the analytical disruption calculation slightly affects this result. A change in the disruption rate would affect which clusters that survive to the present. In particular, as most of the highest mass clusters form early, increased disruption tends to remove younger, higher metallicity clusters. While the shape of the age–metallicity relation changes little, the distribution of clusters within it does.

5 DISCUSSION

5.1 Timing of supernova feedback

In Section 2.4.7, we discussed several prescriptions for the timing of SN, then in Section 3.1 we examined how these prescriptions affect the properties of star clusters. We find that later SN feedback leads to longer time-scales for cluster formation and higher values of the integrated star formation efficiency. In this formalism, we assume that there is no difference in the formation time of low- and high-mass stars within a cluster. Individual stars of all masses have the same age. However, this assumption may be incorrect. For example, using a simulation of a star cluster forming out of a $2 \times 10^4 M_{\odot}$ GMC, Grudić et al. (2022) find that massive stars ($m > 10 M_{\odot}$) finish accreting 1 Myr later than the average star. Padoan et al. (2020) find a similar result using a simulation of $2 \times 10^6 M_{\odot}$ of gas in a $(250 \text{ pc})^3$ box with several star-forming regions. The delay in massive star formation in turn delays the onset of feedback. While our simulations account for the stellar age spread within the cluster when determining the timing of SN, they do not account for this systematic delay in the formation of individual stars.

As shown in Fig. 2, our hybrid approach to the timing of SN feedback approximates well the delay in SN feedback due to the age spread of the stars, so it is our preferred model for future simulations. However, it may need to be further refined to account for the delay in massive star formation. In particular, one possible approach would be to calibrate a subgrid model for the timing of cluster feedback to the results of GMC-scale simulations such as in Grudić et al. (2022). Further delays in the onset of massive star feedback may increase the time-scales of cluster formation and the integrated star formation efficiency, but these effects are likely to be small compared to the effects of other parameters, namely ϵ_{ff} .

5.2 Strength of stellar feedback

In Figs 6 and 7, we showed how the star formation rate of the central galaxies in our simulations changed when varying f_{boost} . We found that $f_{\text{boost}} = 5$ produces too little star formation in the current simulation setup. In the Isolated MW runs we find that $f_{\text{boost}} = 1$ –2 matches the UNIVERSEMACHINE predictions fairly well, as do $f_{\text{boost}} = 1$ –3 in the Thelma & Louise runs. As we discuss more in Appendix A, updates to the hydrodynamics are primarily responsible for the change in preferred values of f_{boost} . Such low values of f_{boost} are unexpected. L18 calibrated f_{boost} , finding $f_{\text{boost}} = 5$ to be their preferred value. Numerical tests in Semenov, Kravtsov & Gnedin (2017) have also shown that values of $f_{\text{boost}} \approx 5$ best account for numerical losses of momentum as an SN shell moves across the simulation grid. Theoretical grounds for $f_{\text{boost}} > 1$ also exist, with Gentry et al. (2017) finding that clustered SN can enhance momentum feedback by up to an order of magnitude relative to an isolated SN.

We also note that all of these runs, even with $f_{\text{boost}} = 1$, show a large decrease in the star formation rate at $z < 2$, in conflict with

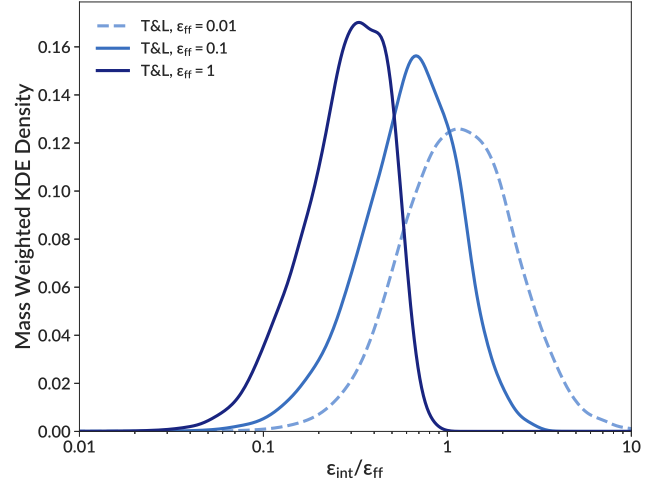


Figure 22. The distribution of $\epsilon_{\text{int}}/\epsilon_{\text{ff}}$ for different values of ϵ_{ff} in the Thelma & Louise IC. All runs used $f_{\text{boost}} = 5$, $f_{\text{HN},0} = 20$ per cent, and show all clusters formed before $z = 3.3$, the lowest redshift that all simulations have reached. We plot the $\epsilon_{\text{ff}} = 1$ per cent run with a dashed line as that run had many clusters that failed to finish forming.

the abundance matching expectation. Both the hydrodynamics and feedback models have been updated to be more physically realistic than those used in L18, but produce worse agreement in the star formation histories. This may indicate that there is additional relevant physics that needs to be included in our simulation.

Our model assumes that all stars above $8 M_{\odot}$ explode as SN. However, this assumption may not hold. Simulations of SN find that some progenitors collapse directly to a black hole without an SN explosion (Heger et al. 2003; Horiuchi et al. 2014; Pejcha & Thompson 2015). If we were to include such scenarios in the feedback scheme, it would decrease the total energy and momentum from SN. We would therefore require a higher value of f_{boost} to obtain reasonable star formation rates. Additionally, changing the minimum progenitor mass for SN makes a large difference in the energy injected by SN (Keller & Kruijssen 2022). We assume $M_{\text{min}} = 8 M_{\odot}$, but this value is uncertain. Increasing it would decrease the number of SN, again requiring a higher f_{boost} to compensate.

5.3 Constraints on star formation efficiency

Fig. 17 shows the distribution of ϵ_{int} for the runs varying ϵ_{ff} . While we find a clear trend that decreasing ϵ_{ff} decreases ϵ_{int} , we can also examine the ratio $\epsilon_{\text{int}}/\epsilon_{\text{ff}}$. Fig. 22 shows this ratio for the Thelma & Louise runs, which can be directly compared with Fig. 8 of L18. For all values of ϵ_{ff} we consider, we find higher values of $\epsilon_{\text{int}}/\epsilon_{\text{ff}}$ than did L18. Two of the changes discussed in Section 3 are responsible. First, the modified SN feedback prescription delays SN feedback compared to L18, which results in higher ϵ_{int} (Fig. 5). Second, the introduction of the virial criterion leads to higher ϵ_{int} at a given ϵ_{ff} (Fig. 15). Combined, these two effects shift our distributions of $\epsilon_{\text{ff}}/\epsilon_{\text{int}}$ to higher values.

We still see the same trend with ϵ_{ff} as did L18, where higher values of ϵ_{ff} lead to smaller $\epsilon_{\text{int}}/\epsilon_{\text{ff}}$ ratios. Quantitatively, the mean value of this ratio drops from 1.15 to 0.57 to 0.30 for $\epsilon_{\text{ff}} = 1, 10$, and 100 per cent, respectively. Conceptually, this ratio is proportional to the number of freefall times over which the cluster accreted material. As discussed in Section 3.4, lower values of ϵ_{ff} lead to longer formation time-scales, in accordance with this result.

As we have discussed throughout Section 3, the duration of a star formation episode is sensitive to cluster feedback and formation prescriptions. While difficult to constrain precisely, current observations indicate that the age spread within clusters is less than ≈ 6 Myr (see the compilation of age data in L18). These age spreads can still be significantly larger than t_{ff} . One example is the Orion Nebula Cluster (ONC), where star formation appears to have occurred over several freefall times (Da Rio, Tan & Jaehnig 2014; Caldwell & Chang 2018; Kounkel et al. 2018). In contrast, simulations of individual molecular clouds generally show star formation ending after one t_{ff} (e.g. Grudić et al. 2022). Our simulated age spreads, shown in Fig. 10, are consistent with the observations for $\epsilon_{\text{ff}} \geq 10$ per cent. We see a strong mass trend, but even for massive clusters the vast majority have age spreads smaller than 6 Myr. However, our results rule out $\epsilon_{\text{ff}} = 1$ per cent, which has unphysically long age spreads for clusters of all masses.

The shape of the initial cluster mass function is another key observable. YMCs in the MW and nearby galaxies are found to follow the functional form of Schechter (1976), with a power-law slope of -2 at the low-mass end (Portegies Zwart et al. 2010). Our mass functions have a positive power-law slope at low mass, peak at a mass that depends on ϵ_{ff} ($10^5 M_{\odot}$ for $\epsilon_{\text{ff}} = 100$ per cent), then decline in a manner consistent with a power law. In essence, our simulations are missing low-mass clusters. While our cluster formation algorithm only seeds clusters if they have an expected mass of $6 \times 10^3 M_{\odot}$, runs with $\epsilon_{\text{ff}} = 100$ per cent show the increasing mass function above this mass. This may indicate that $\epsilon_{\text{ff}} = 100$ per cent forms stars too efficiently, leading to too few low-mass clusters. However, $\epsilon_{\text{ff}} \leq 10$ per cent results in too few massive clusters, with no clusters projected to reach $z = 0$.

Lastly, ϵ_{ff} has been measured in observations with several methods (Evans, Heiderman & Vutisalchavakul 2014; Usero et al. 2015; Lee, Miville-Deschênes & Murray 2016; Ochsendorf et al. 2017; Utomo et al. 2018). While the observations have somewhat different medians, uncertainties, and intrinsic scatter in ϵ_{ff} , a value of $\epsilon_{\text{ff}} \approx 1$ per cent is typical. However, we find that this value does not produce reasonable star cluster properties in our simulations. The time-scales of cluster formation reach our algorithmically imposed limit of 15 Myr. Such time-scales are in conflict with observations. Low values of ϵ_{ff} also produce few massive clusters. Even a value of $\epsilon_{\text{ff}} = 10$ per cent produces few clusters with high enough mass to reach $z = 0$ as GCs. Our simulations prefer higher values of ϵ_{ff} . Among the runs presented here, $\epsilon_{\text{ff}} = 100$ per cent produced the most realistic cluster properties, as it did in L18.

To compare with observations more directly, we postprocess the simulations to calculate an effective value of ϵ_{ff} in a way analogous to how it is derived in observations. First, we identify clusters that are actively forming in several simulation snapshots. Within a sphere of radius r centred on the cluster, we calculate the inferred value of ϵ_{ff} as

$$\bar{\epsilon}_{\text{ff}}(r) = \frac{\bar{t}_{\text{ff}}(r) \dot{M}(< r)}{M_{\text{gas}}(< r)}, \quad (20)$$

where $\bar{t}_{\text{ff}} \equiv \sqrt{3\pi/32G\bar{\rho}}$ is calculated using the mean density $\bar{\rho}$ within the sphere. In the rest of this section, we will use $\bar{\epsilon}_{\text{ff}}$ to refer to the inferred value from this procedure, while ϵ_{ff} will refer to the value used in runtime of the simulation. To calculate \dot{M} , we use a procedure analogous to that used in studies that determine ϵ_{ff} by counting young stellar objects (YSOs) to determine the star formation rate within a cloud (Evans et al. 2014; Heyer et al. 2016; Ochsendorf et al. 2017). These studies use YSOs to estimate the mass of recently formed stars, then divide it by the lifetime of the YSO phase typically set to a fixed

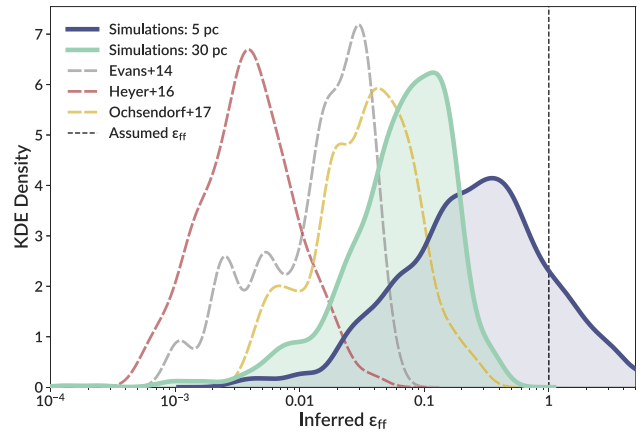


Figure 23. Kernel density estimation of the distribution of ϵ_{ff} inferred from post-processing simulated star clusters and from observations. We use a normalized Gaussian kernel with a width of 0.15 dex. The simulated clusters are from the run with the Isolated MW IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 1$, and $f_{\text{HN},0} = 0$. The 5 pc line shows $\bar{\epsilon}_{\text{ff}}$ as inferred from the region actively participating in cluster formation, while the 30 pc line shows the value inferred for larger star-forming complexes.

time of 0.5 Myr. As we do not store the full accretion histories of simulated clusters, we cannot directly obtain the star formation rate over the last 0.5 Myr. Instead, we approximate it with the average star formation rate over the relevant time-scale:

$$\dot{M} = \frac{M}{\max(\tau_{\text{spread}}, 0.5 \text{ Myr})}, \quad (21)$$

where M is the current mass of the actively forming cluster. For clusters with large age spreads this prescription gives the average star formation rate, while for clusters with short age spreads this matches the rate inferred observations assuming an 0.5 Myr time-scale. We choose to use the cluster age spread rather than the full duration as it more accurately reflects the time-scale over which the bulk of cluster formation happens. The total \dot{M} within a given sphere is the sum of \dot{M} from all actively forming clusters in the sphere.

This calculation of $\bar{\epsilon}_{\text{ff}}$ involves significant averaging both in time and space, compared to the application in simulation runtime. A typical local time-step at the highest refinement levels is 100–1000 yr, orders of magnitude shorter than even 0.5 Myr. Therefore, the finite difference calculation of the star formation rate \dot{M} from equation (1) is a much closer approximation to the true derivative than equation (21). Considering spheres of radius $r > 5$ pc also introduces averaging of the stellar and gas mass on a larger scale than our adopted GMC radius. Both of these effects tend to shift $\bar{\epsilon}_{\text{ff}}$ to smaller values than the input ϵ_{ff} .

In Fig. 23, we show the distribution of values of $\bar{\epsilon}_{\text{ff}}$ calculated for two choices of the averaging radius: 5 and 30 pc. As there are few clusters actively forming in any given snapshot, we use all snapshots from $z = 9$ –1.5 in the run using the Isolated MW IC, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 1$, and $f_{\text{HN},0} = 0$, giving a sample of 748 actively forming clusters. The radius of 5 pc matches the GMC sphere actively participating in star formation. The inferred values peak at around 30 per cent with large scatter but are significantly below the simulation input $\epsilon_{\text{ff}} = 100$ per cent. The procedure to infer $\bar{\epsilon}_{\text{ff}}$ uses the cluster formation time-scale to average the star formation rate, which creates the wide spread and systematic shift. In addition, this procedure calculates t_{ff} and M_{gas} at one instant, which may not reflect typical conditions over the course of the cluster’s growth.

Considering a larger sphere radius of 30 pc adds also spatial averaging. For an isolated cluster, increasing the size of the sphere would simply include more surrounding gas without increasing \dot{M} , leading to smaller inferred values of $\bar{\epsilon}_{\text{ff}}$. However, we find that clusters often form in larger star-forming complexes with many clusters in close proximity of each other. Our choice of 30 pc corresponds to the typical radius of these star-forming complexes. Fig. 23 shows that these complexes have a peak value of $\bar{\epsilon}_{\text{ff}} \approx 10$ per cent, with less scatter than the values inferred on 5 pc scales. The lower mean value is due to the inclusion of more gas not participating in star formation, while the decreased scatter comes from averaging together multiple clusters within each region.

This exercise shows that the inferred values of $\bar{\epsilon}_{\text{ff}}$ are a factor of 10 lower than the simulation input. Still, for this run typical $\bar{\epsilon}_{\text{ff}}$ are higher than those seen in observations. In Fig. 23 we include observations from Evans et al. (2014), Heyer et al. (2016), and Ochsendorf et al. (2017), which all use the YSO method but do so on different scales. Evans et al. (2014) and Heyer et al. (2016) use clumps with typical radii of a few pc, while Ochsendorf et al. (2017) use star-forming complexes with radii around 40 pc. Even with these differences of scale, all studies measure mean values of ϵ_{ff} consistent with ~ 1 per cent. However, we note that we cannot make a direct comparison between these observations and our inferred values of $\bar{\epsilon}_{\text{ff}}$. Each ingredient of the calculation of ϵ_{ff} has systematic differences. The mass of recently formed stars is calculated differently, as we do not directly model the number of observable YSOs in each cluster. The time-scales for calculating the star formation rate are also different, as many of our clusters have τ_{spread} longer than the 0.5 Myr used in observations. Lastly, detailed modelling of CO and HCN abundances and ionization states is needed to calculate M_{gas} exactly as is done in observations. To resolve these differences would require a further analysis in the simulation runtime. Nevertheless, Fig. 23 demonstrates that the discrepancy with observations is substantially smaller than appears from a straightforward comparison with the simulation input.

5.4 Failed cluster formation

In Section 3.4, we showed that in some runs with low ϵ_{ff} , clusters fail to finish formation before it is automatically ended at 15 Myr. In this section, we investigate the reasons for these failed clusters.

We find that no runs with the high value of $\epsilon_{\text{ff}} = 100$ per cent have failed cluster formation, all runs with the low value of $\epsilon_{\text{ff}} = 1$ per cent fail, and among the runs with the intermediate value $\epsilon_{\text{ff}} = 10$ per cent, only the run using the *Isolated* MW IC and $f_{\text{boost}} = 1$ failed. All other runs with $\epsilon_{\text{ff}} = 10$ per cent used higher f_{boost} and did not fail. In total, 4 of our 29 runs experience failed cluster formation.

These trends are due to an interplay between ϵ_{ff} and f_{boost} . When ϵ_{ff} is low, cluster formation progresses slowly, leaving significant amounts of gas. We find low values of ϵ_{int} for low ϵ_{ff} (Fig. 17), meaning that at the end of cluster formation, only a small fraction of gas has been turned into stars. This applies in the midst of cluster formation, too. We examine the gas densities of the host cells of clusters as they form and find that for lower values of ϵ_{ff} there is more gas near the cluster at a given time after the beginning of cluster formation, meaning that GMCs are more massive with low values of ϵ_{ff} . In addition, the slower star formation with low ϵ_{ff} leads to fewer stars to provide feedback. When SNe begin, they must first disperse the gas within the cluster. Higher values of f_{boost} make this process more efficient. Therefore, higher values of f_{boost} lead to shorter time-scales for cluster formation when ϵ_{ff} is low. In contrast,

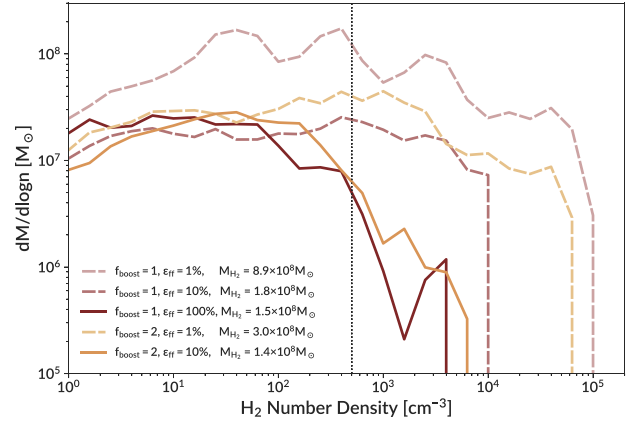


Figure 24. The distribution of cell mass-weighted molecular gas densities for different combinations of ϵ_{ff} and f_{boost} in the *Isolated* MW IC. All runs used $f_{\text{H},0} = 0$ and show the gas within 10 kpc of the galactic centre. Dashed lines indicate runs with failed cluster formation. Runs with $f_{\text{boost}} = 1$ are shown at $z = 1.56$, while those with $f_{\text{boost}} = 2$ are shown at $z = 2.57$. The dotted line at 500 cm^{-3} indicates the minimum density for star formation.

when $\epsilon_{\text{ff}} = 100$ per cent, clusters consume a high fraction of the gas within their GMC. SN feedback of any f_{boost} is able to clear the smaller amounts of gas that remain.

In Fig. 24, we illustrate this effect by presenting instantaneous distribution of the molecular gas density within galaxies with different combinations of ϵ_{ff} and f_{boost} . Runs with failed cluster formation have distributions that extend to higher densities than runs without failed cluster formation. As feedback cannot terminate star formation, gas continues to accrete on to the GMC, increasing the density. Of particular note is the gas at densities higher than the star formation threshold. The total gas number density must be greater than 10^3 cm^{-3} with a molecular fraction of 0.5, giving a minimum molecular number density of 500 cm^{-3} . Above this threshold, the failed runs have significantly more gas than runs that successfully terminate star formation.

The prescription for SN feedback also contributes to why these runs had failed cluster formation while the runs of L18 did not. In the 15 Myr time-scale for cluster formation, the prescription of L18 injects significantly more energy than the updated model (see Fig. 1). For low supernova fractions typical of most clusters, the new prescription injects only 26 per cent of the energy of L18 within the first 15 Myr, increasing to about 50 per cent after 40 Myr. This is exacerbated by the lower f_{boost} used in the updated runs. The total momentum injected by SN feedback during the 15 Myr of cluster formation can be more than an order of magnitude lower than in L18. For low values of ϵ_{ff} , this results in SN feedback being unable to disperse the GMC.

The timing of SN also contributes to failed cluster formation. Compared to L18, SN start later in the new prescription due to the stellar lifetimes chosen (Fig. 1). These runs also use the average approach for determining cluster feedback timing, as they were run before the hybrid approach was finalized. This average approach further delays the onset of SN (Fig. 2). In addition, we find that runs with lower ϵ_{ff} have later average times of cluster formation, t_{ave} , meaning that SN is delayed even further in these runs. These delays in the onset of SN gives the gas outside the GMC more time to accrete on to the GMC, leading to higher gas masses that SN feedback then needs to disperse. This combines with the effect described in the previous paragraphs to make GMCs more difficult

to disperse for lower ϵ_{ff} , further explaining why we find that f_{boost} affects the time-scales of cluster formation for low ϵ_{ff} but not for $\epsilon_{\text{ff}} = 100$ per cent.

To summarize our understanding of why massive star clusters fail to finish forming when $\epsilon_{\text{ff}} \leq 10$ per cent, lower values of ϵ_{ff} turn gas into stars at a slower pace. At a given time after the beginning of cluster formation this results in fewer stars, and therefore less stellar feedback, embedded in a more massive GMC. The onset of SN is delayed compared to L18 due to our choice of stellar lifetimes (see Fig. 1), and then is delayed further after accounting for the age spread within the cluster. This allows more material to accrete on to the GMC, making it even more difficult for feedback to disperse. Once SN feedback starts, the updated feedback prescription injects less momentum than L18. The new prescription has fewer SN and is further exacerbated if low values of f_{boost} are chosen. Although lower values of f_{boost} produce more reasonable star formation rates for $z > 2$, these low values fail to provide enough feedback to disperse GMCs when ϵ_{ff} is low. This may indicate that another source of feedback is needed at early times to help disperse GMCs or that the combination of $\epsilon_{\text{ff}} \leq 10$ per cent and $f_{\text{boost}} = 1$ is ruled out by our simulations.

6 CONCLUSIONS

We have described improvements to the implementation of star cluster formation and feedback in the ART code. We introduced a new criterion for the seeding of cluster particles, requiring the star-forming gas to be gravitationally bound. We also implemented a new prescription for the initial bound fraction of clusters based on simulations of individual GMCs. We added runtime tracking of C, N, O, Mg, S, Ca, and Fe, with enrichment coming from SNIa, SNI, stellar winds, and AGB stars. We updated the SN feedback prescriptions significantly. We now implement SN as discrete events, with rates based explicitly on the stellar lifetimes and IMF. We also explored effects of hypernovae, which inject more energy and have different elemental yields. Lastly, we improved our prescription for the timing of SN to account for the age spread of stars within a cluster.

With these code updates, we ran 20 simulations using the IC from L18 and 9 simulations using two Local Group-like ICs from the ELVIS project. These runs have a range of parameters, including variations in ϵ_{ff} , $f_{\text{HN},0}$, f_{boost} , and the timing of SN feedback. We explored how these parameters affect the properties of galaxies as well as the populations of star clusters within them. Our results are summarized as follows.

(i) Delaying the onset of SN (without changing the total energy injection) results in longer formation time-scales for massive clusters and higher ϵ_{int} (Figs 4 and 5), but does not significantly change the galaxy star formation rate.

(ii) Higher values of the momentum boosting factor for SN greatly decrease the galactic star formation rate (Figs 6, 7). While no value of f_{boost} can reproduce the abundance matching expectation for the full redshift range explored here ($z > 1.5$), we find that the range $1 < f_{\text{boost}} < 3$ produces reasonable agreement for $z > 2$. Higher values of f_{boost} decrease the total stellar mass by decreasing the number of low-mass clusters that form, without changing the number of massive clusters (Fig. 9).

(iii) The hypernova fraction $f_{\text{HN},0}$ makes little difference to galaxy or cluster properties (Figs 6 and 9). The strong decrease in f_{HN} with metallicity (equation 11) results in limited change in the total energy injected by SN (Figs 1 and 8).

(iv) The local star formation efficiency per freefall time does not have a strong impact on the galactic star formation rate (Fig. 13). However, it strongly changes cluster properties. Higher values of ϵ_{ff} lead to more massive clusters (Figs 11 and 12), shorter time-scales for cluster formation (Fig. 10), higher initial bound fractions (Fig. 16), and higher ϵ_{int} (Fig. 17).

(v) Adding the virial parameter criterion to require star-forming gas be gravitationally bound produces more high-mass clusters (Fig. 14), longer time-scales for cluster formation, and higher ϵ_{int} (Fig. 15).

(vi) In runs with low values of ϵ_{ff} , we find a population of clusters that fail to finish forming after 15 Myr. Low values of ϵ_{ff} form stars slowly, leaving massive GMCs that are difficult for feedback to disperse, especially with low values of f_{boost} .

(vii) We present the evolution of the observable mass function of clusters at various redshifts (Fig. 20). Most massive clusters form at high redshifts when the star formation density is high, with low-mass clusters dominating in quiescent epochs.

(viii) We analytically extrapolate the dynamical disruption of clusters from the last available output to $z = 0$ (Fig. 20). We can match the observed mass function of MW GCs by assuming a high value for the cluster disruption rate. The surviving clusters also match the age–metallicity relation of MW GCs (Fig. 21).

(ix) Among the values of ϵ_{ff} we explored, only $\epsilon_{\text{ff}} = 100$ per cent can match the MW GC mass function. Runs with $\epsilon_{\text{ff}} = 1$ per cent produces clusters with unphysically long age spreads (Fig. 10), and runs with $\epsilon_{\text{ff}} = 10$ per cent produce too few high-mass clusters (Figs 11, 12, and 19).

This exploration emphasizes the importance of well-calibrated subgrid models for modelling star clusters in simulations of galaxy formation. Some modelling choices, such as the optimal value of ϵ_{ff} , whether to enforce a virial criterion when seeding star clusters, and different prescriptions for the timing of SN feedback all affect the resulting cluster populations without significantly impacting global galaxy properties. A successful model of star formation and feedback in simulations must be able to reproduce not only galaxy-scale properties, but also the small-scale properties of individual star clusters.

ACKNOWLEDGEMENTS

We thank Vadim Semenov, Hui Li, Eric Bell, Mateusz Ruszkowski, and Gus Evrard for helpful discussions. We also thank the referee for suggestions that improved the paper. GB and OG were supported in part by the US National Science Foundation through grant 1909063 and by NASA through grant *HST*-AR-16614.001-A.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Agertz O., Teyssier R., Moore B., 2011, *MNRAS*, 410, 1391
- Agertz O., Kravtsov A. V., Leitner S. N., Gnedin N. Y., 2013, *ApJ*, 770, 25
- Bastian N., 2008, *MNRAS*, 390, 759
- Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, *MNRAS*, 488, 3143
- Benincasa S. M. et al., 2020, *MNRAS*, 497, 3993
- Brodie J. P., Strader J., 2006, *ARA&A*, 44, 193
- Brown G., Gnedin O. Y., 2021a, *New Astron.*, 84, 101501

- Brown G., Gnedin O. Y., 2021b, *MNRAS*, 508, 5935
- Brown G., Gnedin O. Y., Li H., 2018, *ApJ*, 864, 94
- Caldwell S., Chang P., 2018, *MNRAS*, 474, 4818
- Choksi N., Gnedin O. Y., 2019, *MNRAS*, 486, 331
- D'Souza R., Bell E. F., 2018, *Nat. Astron.*, 2, 737
- Da Rio N., Tan J. C., Jaehnig K., 2014, *ApJ*, 795, 55
- Dotter A. et al., 2010, *ApJ*, 708, 698
- Elmegreen B. G., 2010, *ApJ*, 712, L184
- Evans N. J., Heiderman A., Vutisalchavakul N., 2014, *ApJ*, 782, 114
- Evans T. A., Fattahi A., Deason A. J., Frenk C. S., 2020, *MNRAS*, 497, 4311
- Fall S. M., Zhang Q., 2001, *ApJ*, 561, 751
- Gallart C. et al., 2015, *ApJ*, 811, L18
- Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, *MNRAS*, 362, 41
- Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Lee K., 2014, *MNRAS*, 438, 2578
- Gentry E. S., Krumholz M. R., Dekel A., Madau P., 2017, *MNRAS*, 465, 2471
- Gieles M., Renaud F., 2016, *MNRAS*, 463, L103
- Gnedin N. Y., 2014, *ApJ*, 793, 29
- Gnedin N. Y., Abel T., 2001, *New A*, 6, 437
- Gnedin N. Y., Draine B. T., 2014, *ApJ*, 795, 37
- Gnedin N. Y., Kravtsov A. V., 2011, *ApJ*, 728, 88
- Gnedin O. Y., Ostriker J. P., 1997, *ApJ*, 474, 223
- Goldman S. R. et al., 2017, *MNRAS*, 465, 403
- Grimmett J. J., Karakas A. I., Heger A., Müller B., West C., 2020, *MNRAS*, 496, 4987
- Grudić M. Y., Guszejnov D., Offner S. S. R., Rosen A. L., Raju A. N., Faucher-Giguère C.-A., Hopkins P. F., 2022, *MNRAS*, 512, 216
- Haardt F., Madau P., 2001, Clusters of Galaxies and the High Redshift Universe Observed in X-rays. preprint ([arXiv:astro-ph/0106018](https://arxiv.org/abs/astro-ph/0106018))
- Hahn O., Abel T., 2011, *MNRAS*, 415, 2101
- Hammer F., Puech M., Chemin L., Flores H., Lehnert M. D., 2007, *ApJ*, 662, 322
- Harris W. E., 1991, *ARA&A*, 29, 543
- Harris W. E., 1996, *AJ*, 112, 1487
- Harris W. E., Blakeslee J. P., Harris G. L. H., 2017, *ApJ*, 836, 67
- Hayden M. R. et al., 2015, *ApJ*, 808, 132
- Heger A., Fryer C. L., Woosley S. E., Langer N., Hartmann D. H., 2003, *ApJ*, 591, 288
- Heyer M., Gutermuth R., Urquhart J. S., Csengeri T., Wienen M., Leurini S., Menten K., Wyrowski F., 2016, *A&A*, 588, A29
- Hislop J. M., Naab T., Steinwandel U. P., Lahén N., Irodotou D., Johansson P. H., Walch S., 2022, *MNRAS*, 509, 5938
- Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, *MNRAS*, 445, 581
- Hopkins P. F. et al., 2018, *MNRAS*, 480, 800
- Horiuchi S., Nakamura K., Takiwaki T., Kotake K., Tanaka M., 2014, *MNRAS*, 445, L99
- Iwamoto K. et al., 1998, *Nature*, 395, 672
- Jordán A. et al., 2007, *ApJS*, 171, 101
- Katz N., 1992, *ApJ*, 391, 502
- Keller B. W., Kruijssen J. M. D., 2022, *MNRAS*, 512, 199
- Keller B. W., Wadsley J., Benincasa S. M., Couchman H. M. P., 2014, *MNRAS*, 442, 3013
- Kennicutt Robert C. J., 1998, *ARA&A*, 36, 189
- Kewley L. J., Dopita M. A., 2002, *ApJS*, 142, 35
- Kim J.-h. et al., 2018, *MNRAS*, 474, 4232
- Kirby E. N., Cohen J. G., Guhathakurta P., Cheng L., Bullock J. S., Gallazzi A., 2013, *ApJ*, 779, 102
- Kobayashi C., Umeda H., Nomoto K., Tominaga N., Ohkubo T., 2006, *ApJ*, 653, 1145
- Kounkel M. et al., 2018, *AJ*, 156, 84
- Kravtsov A. V., 1999, PhD thesis, New Mexico State University
- Kravtsov A. V., Klypin A. A., Khokhlov A. M., 1997, *ApJS*, 111, 73
- Kroupa P., 2001, *MNRAS*, 322, 231
- Kruijssen J. M. D., 2014, *Class. Quantum Gravity*, 31, 244006
- Kruijssen J. M. D., 2015, *MNRAS*, 454, 1658
- Kruijssen J. M. D. et al., 2019, *Nature*, 569, 519
- Lada C. J., Lada E. A., 2003, *ARA&A*, 41, 57
- Lahén N., Naab T., Johansson P. H., Elmegreen B., Hu C.-Y., Walch S., 2019, *ApJ*, 879, L18
- Larsen S. S., 2002, *AJ*, 124, 1393
- Leaman R., VandenBerg D. A., Mendel J. T., 2013, *MNRAS*, 436, 122
- Lee E. J., Miville-Deschênes M.-A., Murray N. W., 2016, *ApJ*, 833, 229
- Leitherer C., Robert C., Drissen L., 1992, *ApJ*, 401, 596
- Li H., Gnedin O. Y., 2019, *MNRAS*, 486, 4030
- Li H., Gnedin O. Y., Gnedin N. Y., Meng X., Semenov V. A., Kravtsov A. V., 2017, *ApJ*, 834, 69
- Li H., Gnedin O. Y., Gnedin N. Y., 2018, *ApJ*, 861, 107 (L18)
- Li H., Vogelsberger M., Marinacci F., Gnedin O. Y., 2019, *MNRAS*, 487, 364
- Ma X., Quataert E., Wetzel A., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2020, *MNRAS*, 498, 2001
- Maiolino R., Mannucci F., 2019, *A&A Rev.*, 27, 3
- Maoz D., Graur O., 2017, *ApJ*, 848, 25
- Marín-Franch A. et al., 2009, *ApJ*, 694, 1498
- Martizzi D., Faucher-Giguère C.-A., Quataert E., 2015, *MNRAS*, 450, 504
- McKee C. F., Ostriker E. C., 2007, *ARA&A*, 45, 565
- Meng X., Gnedin O. Y., 2020, *MNRAS*, 494, 1263
- Meng X., Gnedin O. Y., 2021, *MNRAS*, 502, 1433
- Meng X., Gnedin O. Y., 2022, *MNRAS*, preprint ([arXiv:2201.09826](https://arxiv.org/abs/2201.09826))
- Meng X., Gnedin O. Y., Li H., 2019, *MNRAS*, 486, 1574
- Miville-Deschênes M.-A., Murray N., Lee E. J., 2017, *ApJ*, 834, 57
- Nomoto K., Leung S.-C., 2018, *Space Sci. Rev.*, 214, 67
- Ochsendorff B. B., Meixner M., Roman-Duval J., Rahman M., Evans Neal J. I., 2017, *ApJ*, 841, 109
- Padoan P., Haugbølle T., Nordlund Å., 2012, *ApJ*, 759, L27
- Padoan P., Pan L., Juvela M., Haugbølle T., Nordlund Å., 2020, *ApJ*, 900, 82
- Pejcha O., Thompson T. A., 2015, *ApJ*, 801, 90
- Pfeffer J., Kruijssen J. M. D., Crain R. A., Bastian N., 2018, *MNRAS*, 475, 4309
- Portegies Zwart S. F., McMillan S. L. W., Gieles M., 2010, *ARA&A*, 48, 431
- Prieto J. L., Gnedin O. Y., 2008, *ApJ*, 689, 919
- Puzia T. H., Kissler-Patig M., Thomas D., Maraston C., Saglia R. P., Bender R., Goudfrooij P., Hempel M., 2005, *A&A*, 439, 997
- Raiteri C. M., Villata M., Navarro J. F., 1996, *A&A*, 315, 105
- Renaud F., Agertz O., Gieles M., 2017, *MNRAS*, 465, 3622
- Ritter C., Herwig F., Jones S., Pignatari M., Fryer C., Hirschi R., 2018, *MNRAS*, 480, 538
- Rudd D. H., Zentner A. R., Kravtsov A. V., 2008, *ApJ*, 672, 19
- Santistevan I. B., Wetzel A., El-Badry K., Bland-Hawthorn J., Boylan-Kolchin M., Bailin J., Faucher-Giguère C.-A., Benincasa S., 2020, *MNRAS*, 497, 747
- Schechter P., 1976, *ApJ*, 203, 297
- Schmidt M., 1959, *ApJ*, 129, 243
- Semenov V. A., Kravtsov A. V., Gnedin N. Y., 2016, *ApJ*, 826, 200
- Semenov V. A., Kravtsov A. V., Gnedin N. Y., 2017, *ApJ*, 845, 133
- Semenov V. A., Kravtsov A. V., Diemer B., 2021, *ApJS*, preprint ([arXiv:2107.14240](https://arxiv.org/abs/2107.14240))
- Spitzer Lyman J., 1958, *ApJ*, 127, 17
- Springel V., Hernquist L., 2003, *MNRAS*, 339, 289
- Stinson G., Seth A., Katz N., Wadsley J., Governato F., Quinn T., 2006, *MNRAS*, 373, 1074
- Strader J., Brodie J. P., Cenarro A. J., Beasley M. A., Forbes D. A., 2005, *AJ*, 130, 1315
- Usero A. et al., 2015, *AJ*, 150, 115
- Utomo D. et al., 2018, *ApJ*, 861, L18
- VandenBerg D. A., 2000, *ApJS*, 129, 315
- VandenBerg D. A., Brogaard K., Leaman R., Casagrande L., 2013, *ApJ*, 775, 134
- Vesperini E., Zepf S. E., Kundu A., Ashman K. M., 2003, *ApJ*, 593, 760

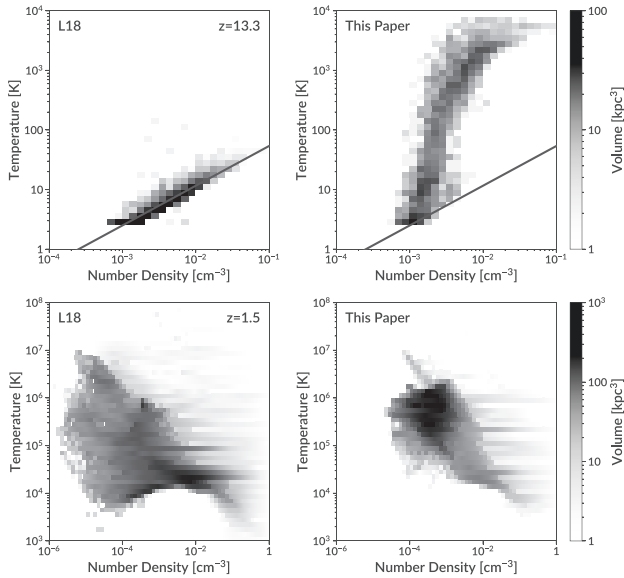


Figure A1. Heatmap showing the temperature and density of gas within the virial radius of the largest halo. In each panel, the shading shows the volume of gas at the given temperature and density. The left-hand column shows a run using the hydrodynamic scheme of L18, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 5$, and $f_{\text{HN},0} = 0$, while the right column shows the run with the updated energy-based hydrodynamics scheme, $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 1$, and $f_{\text{HN},0} = 0$. The top row show these runs at $z = 13.3$ before any stars formed, while the bottom row shows the runs at $z = 1.5$. In the top panels, the red line shows the expected behaviour for pure adiabatic compression. The code version of L18 exactly follows this line, while the updated version has extra heating from proper treatment of shocks.

APPENDIX A: HYDRODYNAMICS

When updating from the version of the ART code used in L18, we changed the model of how internal energy is calculated in the presence of subgrid turbulence. The hydro solver independently tracks total energy, thermal energy, and energy of unresolved subgrid turbulence. The thermal energy and subgrid turbulence are assumed to evolve adiabatically (other than energy injection from sources such as stellar feedback). As these are calculated independently, there is no initial restriction for the sum of thermal, kinetic, and turbulent energies to equal the total. As the adiabatic assumption is not always correct for thermal energy (particularly in shocks), the new version calculates the thermal energy as $E_{\text{th}} = E_{\text{tot}} - E_{\text{kinetic}} - E_{\text{turb}}$. This energy synchronization allows for shocks to transfer energy from kinetic to thermal, as should happen. The adiabatic assumption is only used in cases where the gas is highly supersonic, such that

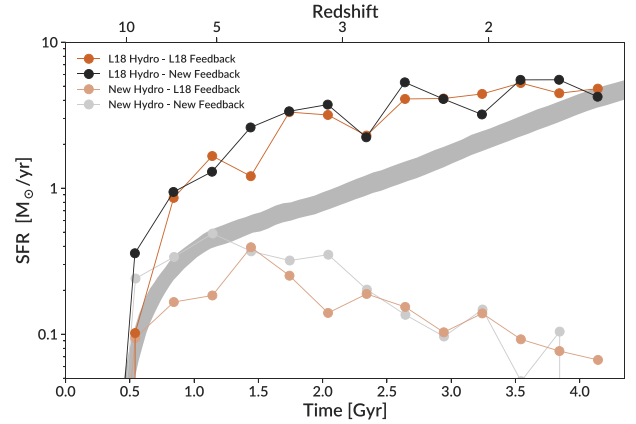


Figure A2. The star formation history for galaxies in our test runs with varying hydrodynamics and feedback. We compare the feedback model of L18 to the feedback model presented in this paper. All runs use $\epsilon_{\text{ff}} = 100$ per cent, $f_{\text{boost}} = 5$, and $f_{\text{HN},0} = 0$. We compare to the UNIVERSEMACHINE model (Behroozi et al. 2019). The change in hydrodynamics is solely responsible for the change in star formation rate, while our updates to feedback have little effect.

$E_{\text{tot}} \approx E_{\text{kinetic}}$. In this case, the subtraction would be susceptible to numerical errors, so we revert to the adiabatic assumption. In the old version of the code, which *always* relied on the adiabatic assumption, shocks were not treated properly and energy that should have been transferred from kinetic to thermal was simply lost. This is visualized in the top row of Fig. A1, where we show the phase diagram of gas within the virial radius at $z = 13.3$ before stars have formed. The hydrodynamic scheme of L18 follows what is expected for pure adiabatic compression, while the new scheme shows gas being heated by virial shocks.

While the newer version of the code is better physically motivated, it significantly changed the structure of modelled galaxies. We find large differences in temperature distributions of the gas. The bottom panel of Fig. A1 shows the phase diagram of gas within the virial radius at $z = 1.5$. Here, the run with the updated hydrodynamics has significantly more hot, low-density gas in the halo. This hot gas prevented cold gas from accreting on to the disc, effectively reducing star formation. We show this star formation in Fig. A2 using test runs that vary both the hydro and feedback schemes. We test the stellar feedback model presented in this paper as well as the model used by L18. All runs use $f_{\text{boost}} = 5$, yet runs with the new hydro scheme produce dramatically lower star formation rates.

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