The Astrophysical Journal, 990:135 (24pp), 2025 September 10 https://doi.org/10.3847/1538-4357/adf19d


# ΛCDM Star Clusters at Cosmic Dawn: Stellar Densities, Environment, and Equilibrium

Claire E. Williams[1,2], Smadar Naoz[1,2], William Lake[1,2], Blakesley Burkhart[3,4], Federico Marinacci[5,6], Mark Vogelsberger[7], Naoki Yoshida[8,9,10], Shyam H. Menon[3,4], Avi Chen[3], and Angela Adamo[11]
[1] Department of Physics and Astronomy, UCLA, Los Angeles, CA 90095, USA; clairewilliams@astro.ucla.edu
[2] Mani L. Bhaumik Institute for Theoretical Physics, Department of Physics and Astronomy, UCLA, Los Angeles, CA 90095, USA
[3] Department of Physics and Astronomy, Rutgers, The State University of New Jersey, 136 Frelinghuysen Road, Piscataway, NJ 08854, USA
[4] Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA
[5] Department of Physics & Astronomy "Augusto Righi," University of Bologna, via Gobetti 93/2, 40129 Bologna, Italy
[6] INAF, Astrophysics and Space Science Observatory Bologna, Via P. Gobetti 93/3, I-40129 Bologna, Italy
[7] Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8] Department of Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan
[9] Kavli Institute for the Physics and Mathematics of the Universe (WPI), UT Institute for Advanced Study, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
[10] Research Center for the Early Universe, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan
[11] Astronomy Department, Stockholm University & Oskar Klein Centre, Roslagstullsbacken 21, Stockholm, SE-10691, Sweden
*Received 2025 February 21; revised 2025 June 20; accepted 2025 July 16; published 2025 September 3*

## Abstract

The James Webb Space Telescope (JWST) has opened a window on many new puzzles in the early Universe, including a population of high-redshift star clusters with extremely high stellar surface density, suggesting unique star formation conditions in the Universe's early evolution. We study the formation and evolution of these first star clusters and galaxies using an AREPO cosmological simulation box designed to resolve the intricate environments of the smallest halos hosting Population III star clusters at $z \geqslant 12$. Our approach, which prioritizes baryonic structure identification through a friends-of-friends algorithm, provides new insights into early star cluster formation and delivers predictions directly relevant to observations. We investigate the dynamical properties of these first star clusters and use numerical and analytical methods to understand the populations of virialized and nonvirialized systems. Our findings indicate that high-$z$ star clusters in a feedback-free regime can achieve extreme surface densities, consistent with the systems detected by JWST. These results imply that JWST may have the opportunity to uncover stellar systems at high redshift whose dynamical state preserves evidence of the hierarchical structure formation process.

*Unified Astronomy Thesaurus concepts:* Star clusters (1567); Hydrodynamical simulations (767); High-redshift galaxies (734); James Webb Space Telescope (2291); Population III stars (1285)

## 1. Introduction

The earliest phase of galaxy formation is driven by the hierarchical assembly of dark matter halos and the formation of star clusters from the primordial gas of the early Universe (e.g., M. Vogelsberger et al. 2020). The James Webb Space Telescope (JWST) is revealing both the products of this process—early galaxies at extremely high redshift (e.g., M. Castellano et al. 2022; N. J. Adams et al 2023; S. L. Finkelstein et al. 2023; M. Franco et al. 2024)—and, in a few lensed scenarios, the building blocks themselves (e.g., A. Adamo et al. 2024; L. Mowla et al. 2024; E. Vanzella et al. 2023). These observations probe the first starlight to light up our Universe at "Cosmic Dawn," and offer an opportunity to test models of cosmology, star formation, and galaxy buildup.

In the local, low-redshift context, the conversion of molecular gas clouds into stars is well-understood to be an inefficient process, limited by various feedback processes that prevent total collapse (e.g., C. F. McKee & E. C. Ostriker 2007). However, at high redshift, JWST observations reveal stellar systems of extremely high density. For example, JWST observations find a striking overabundance of compact, blue galaxies (e.g., H. Atek et al. 2023; S. L. Finkelstein et al. 2023, 2024; F. Cullen et al. 2024; Y. Harikane et al. 2025), with resolved sizes implying surface densities consistent with compact elliptical galaxies, ultracompact dwarfs, and superstar clusters in the local context (e.g., J. F. W. Baggen et al. 2024; C. M. Casey et al. 2024). Below the scale of galaxies, gravitational lensing effects provide a view of individual star clusters at extremely high redshifts: in particular the "Firefly Sparkle" at $z = 8.3$ (L. Mowla et al. 2024) and the "Cosmic Gems" at $z = 10.1$ (A. Adamo et al. 2024). Furthermore, observations at intervening redshifts reveal small clusters (i.e., E. Vanzella et al. 2022, 2023; S. Fujimoto et al. 2025; K. E. Whitaker et al. 2025), as well as clumps within lensed galaxies (e.g., U. Meštrić et al. 2022; A. Claeyssens et al. 2023), providing an evolving picture of star clusters at various cosmological epochs, including before, during, and after reionization. The Cosmic Gems are observed to be gravitationally bound and colocated in a 70 pc region, while the Firefly Sparkle system is seen in a roughly 1 kpc region and the clusters are likely at least marginally bound as well (A. Adamo et al. 2024; L. Mowla et al. 2024). These observed clusters are $\sim 10^5 - 10^6 M_\odot$ and 1 pc in size. Thus, when individual high-redshift stellar clusters are observed at the parsec scale, the stellar component is seen to be arranged in extremely clumpy, high-density configurations with surface densities exceeding $10^4 - 10^5 M_\odot\,\mathrm{pc}^{-2}$. Thus, it seems that the

conditions for star formation in the first billion years must have differed from the local paradigm.

In the theoretical picture, there are many unique conditions at Cosmic Dawn that differentiate this epoch from standard star formation. For one, the first generation of stars or "Population III" (Pop III) forms from pristine gas, devoid of all metals (e.g., N. Yoshida et al. 2006). The nature of these first stars, particularly their initial mass function, is not yet constrained, with some models and simulations predicting high mass cutoffs upwards of $10^3 M_\odot$ (e.g., S. Chon et al. 2022). Consequently, the impact of feedback processes that may result from such an extreme population is also unknown. However, high surface densities observed in the JWST data thus far already hint at a different mode of star formation during this period, one which allows for highly dense clusters to form.

On the scales of individual star clusters, high-resolution star formation simulations can resolve individual stars and their formation environment given a set of physical assumptions (e.g., W. Lake et al. 2025; S. Chon et al. 2022). To understand the environment and context of galaxy formation, a cosmological perspective is needed. Large volume cosmological simulations can resolve the large scale evolution of early structures, but are typically limited to $\gtrsim 10^6 M_\odot$ particle mass resolution and $\geqslant 0.2-1$ kpc spatial scales. (e.g., EAGLE: J. Schaye et al. 2015; M. Schaller et al. 2015; BLUETIDES: Y. Feng et al. 2016; IllustrisTNG: A. Pillepich et al. 2018; THESAN: R. Kannan et al. 2022). To understand the small scales, many studies use a zoom-in framework, often building a more statistical sample via the combination of multiple zoom ins (e.g., FIRSTLIGHT: D. Ceverino et al. 2017; Y. Nakazato et al. 2024; SERRA: V. Gelli et al. 2020; M. Kohandel et al. 2020; A. Pallottini et al. 2022; FIRE: X. Ma et al. 2018). Several studies have dedicated the resolution and statistical sample size to specifically study the formation of high-redshift galaxy clumps and star clusters (e.g., L. Mayer et al. 2025; Y. Nakazato et al. 2024). The question has also been explored analytically (e.g., at somewhat larger scales: M. Boylan-Kolchin 2025). Here, we investigate a high-resolution simulation of a small cosmological volume $(2.5\,\mathrm{Mpc}\,h^{-1})^3$ with the AREPO code. This box contains thousands of dark matter halos above a minimum resolved halo mass of $M_{\rm DM} = 2.3 \times 10^5\, h^{-1} M_\odot$, resolving baryonic structure above $M_{\rm b} \sim 2 \times 10^4 M_\odot$.

Even the higher resolution simulation suites usually rely on subgrid models, tuned to observations and/or higher resolution simulations, to incorporate the conversion of gas into stars and the desired feedback effects (e.g., B. D. Smith et al. 2017; F. Marinacci et al. 2019). Additionally, in order to make relevant comparisons to observables, it is necessary to group the simulated particles into structures corresponding to galaxies and clusters. Many halo- and subhalo-finding algorithms exist in the literature (see e.g., A. Knebe et al. 2011, 2013). The simulations studied in this work were run using the AREPO simulation code (R. Weinberger et al. 2020), and thus we employ the built-in friends-of-friends (FOF) algorithm, which is described in detail in V. Springel et al. (2001). We test the method of primarily searching for stars and gas—mimicking an observer that has no knowledge of the underlying dark matter field—to identify early star clusters and protogalaxies. This method is most similar to the clump finding method of Y. Nakazato et al. (2024), although we do not differentiate between young and old stars for the purpose of structure detection.

In this work, we investigate the properties of star clusters and early low-mass galaxies in a simulation suite in AREPO. We characterize several populations of star clusters by their virialization and suggest that in JWST's earliest observing epochs, some systems still preserve the dynamical evidence of the first structure formation. Our results imply that the stellar components of the high-$z$ structure are frequently not virialized, and this assumption should not be used to estimate properties in this epoch. Additionally, our star clusters are consistent with extremely high surface density, as observed in the Cosmic Gems, Sunrise, and Sunburst clusters at high redshift. Because no feedback was included in this simulation box, this implies that these clusters may have formed in a scenario where feedback is weak with respect to gravity (e.g., A. Dekel et al. 2023), as explored in recent studies that show high star formation efficiency in high surface density cases (e.g., M. Y. Grudić et al. 2018; J.-G. Kim et al. 2018; S. H. Menon et al. 2023).

This paper is organized as follows. In Section 2, we describe the simulation runs, FOF algorithms, and the methods used to test for virialization and boundedness. In Section 3, we present our findings, comparing the number counts using the different FOF runs. We identify two populations of virialized clusters in Section 4, and use analytical timescale arguments to understand their formation. With these numerical and analytical tools in hand, we investigate the overall environment and properties of star clusters in Section 5, comparing them to JWST observations. A discussion and summary of the work are presented in Section 6. In this study, we assume a ΛCDM cosmology, with $\Omega_\Lambda = 0.73$, $\Omega_{\rm m} = 0.27$, $\Omega_{\rm b} = 0.044$, $\sigma_8 = 1.7$, and $h = 0.71$.

## 2. Simulations and Identification of Star Clusters

### 2.1. Cosmological Simulation

For this study, we use the AREPO cosmological simulation code to run a $(2.5\ \mathrm{Mpc}\ h^{-1})^3$ volume simulation with $768^3$ particles from $z = 200$ to $z = 12$. Our simulation does not include any feedback effects nor the primordial baryon-dark matter streaming velocity. We note that the simulation is the same as the nonstream velocity ($0\sigma_{\rm bc}$) run from W. Lake et al. (2023b) and C. E. Williams et al. (2024).

In AREPO, hydrodynamics is discretized with a Voronoi mesh that is allowed to move with the fluid flow (V. Springel 2010). Our minimum simulated Voronoi cell volume is 0.645 $\mathrm{pc}^3$, giving a minimum gas cell size of ∼0.86 pc. The simulation gravitational softening length results in a minimum particle resolution of 1 pc at the time of the last snapshot. The simulation uses the GRACKLE chemistry and cooling library (B. D. Smith et al. 2017; G. Chiaki & J. H. Wise 2019) to explicitly include nonequilibrium molecular hydrogen cooling and the associated radiative cooling. This includes molecular hydrogen cooling, HD cooling, and chemistry for 15 primordial species. These are: $e^-$, H, $H^+$, He, $He^+$, $He^{++}$, $H^-$, $H_2$, $H_2^+$, D, $D^+$, HD, $HeH^+$, $D^-$, and $HD^+$. For molecular hydrogen, the cooling rate includes both rotational and vibrational transitions.

Star formation is implemented according to the method of F. Marinacci et al. (2019). When gas mesh cells exceed the Jeans mass on the scale of the cell, they are eligible to form stars.

**Table 1**
Summary of the Structure Finding Algorithms Used in this Work

| Category | Description | Primary Particle(s) | Secondary Particle(s) |
|---|---|---|---|
| DM-centric | DM primary † (standard) | Dark matter | Stars and gas |
| ⋯ | Dark matter and baryons primary | Dark matter, stars, and gas | None |
| Baryon-centric | Stars and gas primary† | Stars and gas | None |
| ⋯ | Star primary | Stars | None |

**Note.** † indicates the algorithm used as the standard for the category in the analysis after Section 3, where only the difference between the DM-centric and the baryon-centric methods are considered. In other words, when we refer to "DM-centric" and "baryon-centric" algorithms in those sections, we mean the DM primary and the stars and gas primary runs, respectively.

Additionally, star formation is restricted to only gravitationally bound regions—so the virial parameter of the cell must be negative. Once these collapse criteria are met, the creation of star particles occurs stochastically, by converting a gas cell into a collisionless star particle with a probability determined by the local star formation rate. The mass of each star particle is equal to the mass of its progenitor gas cell. The collapse timescale is set to the gravitational dynamical time of the gas cell ($t_{\rm dyn} = \sqrt{3\pi/(32G\rho_{\rm g})}$). The local star formation rate of each gas cell is computed as $\dot{M} = \epsilon M_{\rm gas}/t_{\rm dyn}$, where $M_{\rm gas}$ is the mass of the gas mesh cell and $\epsilon$ is an efficiency factor set to 0.01.

### 2.2. Baryonic-centered Structure Finder

In AREPO's FOF algorithm (V. Springel et al. 2001), simulation particles are grouped together using a linking-length criterion, where the distance between particles in a prospective group is compared with a multiple of the mean interparticle separation. Traditionally, dark matter is used as the primary linking particle type to identify halos that contain galaxies. Each halo is associated with its gravitational radius, often chosen to be the radius at which the halo density falls below 200 times the mean or critical density at that redshift. In the case of AREPO's algorithm, a secondary particle FOF algorithm can then be run to identify the group's other particle types, such as gas and/or stars, providing the stellar mass, gas mass, and star formation rate of galaxies in the simulation box (R. Weinberger et al. 2020). This method aligns with the typical picture of ΛCDM galaxies—a small baryonic component of gas and stars sitting at the center of a large, spherical, and cuspy dark matter halo (e.g., R. Barkana & A. Loeb 2001). Since our cosmological simulation does not incorporate the stream velocity, stellar structures should not form outside of dark matter halos in the nominal picture of structure formation. This assumption can therefore reliably track the concentrations of stellar light that would be "observed" by a telescope under the condition that all significant clusters of stars reside in resolved, approximately spherical dark matter halos.

Of course, electromagnetic observatories such as JWST cannot directly study the underlying dark matter field and in this work we study a context where this assumption breaks down: the earliest epoch of star cluster formation in a high-resolution cosmological simulation. This environment is challenging for a dark-matter-focused FOF method because these first structures form in low-mass minihalos, which are subject to frequent mergers, fragmentation, and accretion onto larger structures. Star clusters may reside in small halos, which are not the subhalos of a larger host halo in the box. Although most cosmological simulations target dark matter for their catalogs, a method searching for baryonic particles is necessary for many simulation contexts. For example, Y. Nakazato et al. (2024) use a "clump finder" to identify stellar clumps in mock images of the FirstLight simulation. In zoom-in simulations, satellite galaxies are sometimes found using search criteria targeted at maxima in the stellar density field (e.g., V. Gelli et al. 2020). In addition, previous works investigating the cosmic streaming velocity searched for gas particles independently of dark matter in order to find structures that were removed from their parent halos by the cosmic steam velocity (e.g., C. Popa et al. 2016; Y. S. Chiou et al. 2018, 2019, 2021; W. Lake et al. 2021, 2023a, 2023b; C. E. Williams et al. 2023, 2024).

We ran a set of FOF structure-finding algorithms targeting different particle types on the simulation snapshot outputs at $z = 12$, 13, 14, and 15 with AREPO's default halo-finding algorithm, which uses a FOF technique (see V. Springel 2005, for details). Table 1 lists all the runs. With this method, the user chooses a simulated particle type (typically dark matter (DM)) to serve as the "primary," and the FOF outputs a catalog of halos. The user can optionally select additional particle types (typically stars and/or gas) to serve as the "secondary," and a secondary stage of the algorithm will associate these particles with the existing halos. In this work, we ran the standard FOF algorithm that uses dark matter particles as the primary particle type and baryons as the secondary, in addition to three sets of runs with combinations of particles as the primary particle type: DM+stars+gas, stars +gas, and stars only. Figure 1 shows the dark matter density in three example regions where the structures identified by these algorithms are overlaid. The $R_{200}$ of halos detected using the DM-primary, gas+stars secondary method are shown in black, while the radii of objects detected by the stars+gas method are shown in green. Green "X" markers denote the center of mass of baryon-centric objects which do not lie within the $R_{200}$ of a DM-detected halo. From this region, we can see that these methods—which trace concentrated structures in different particle types—will give different catalogs of halos/galaxies.

As we discuss further in Section 3, there is a similar behavior among algorithms that use DM as one of the primary types, as well as among those that do not. However, important differences exist between these two categories. Thus, we refer to "DM-centric" and "baryon-centric" structure-finding algorithms. As denoted in Table 1, when a particular run is not specified, we use the DM primary, stars+gas secondary as the default "DM-centric" data set, and the stars+gas run as the default "baryon-centric" data set (because it includes the added information of the gas particles).

For DM-centric runs, we enforce a resolution cutoff of 300 DM particles. For baryon-centric runs, we enforce a resolution cutoff of 100 star particles. We test our results to ensure

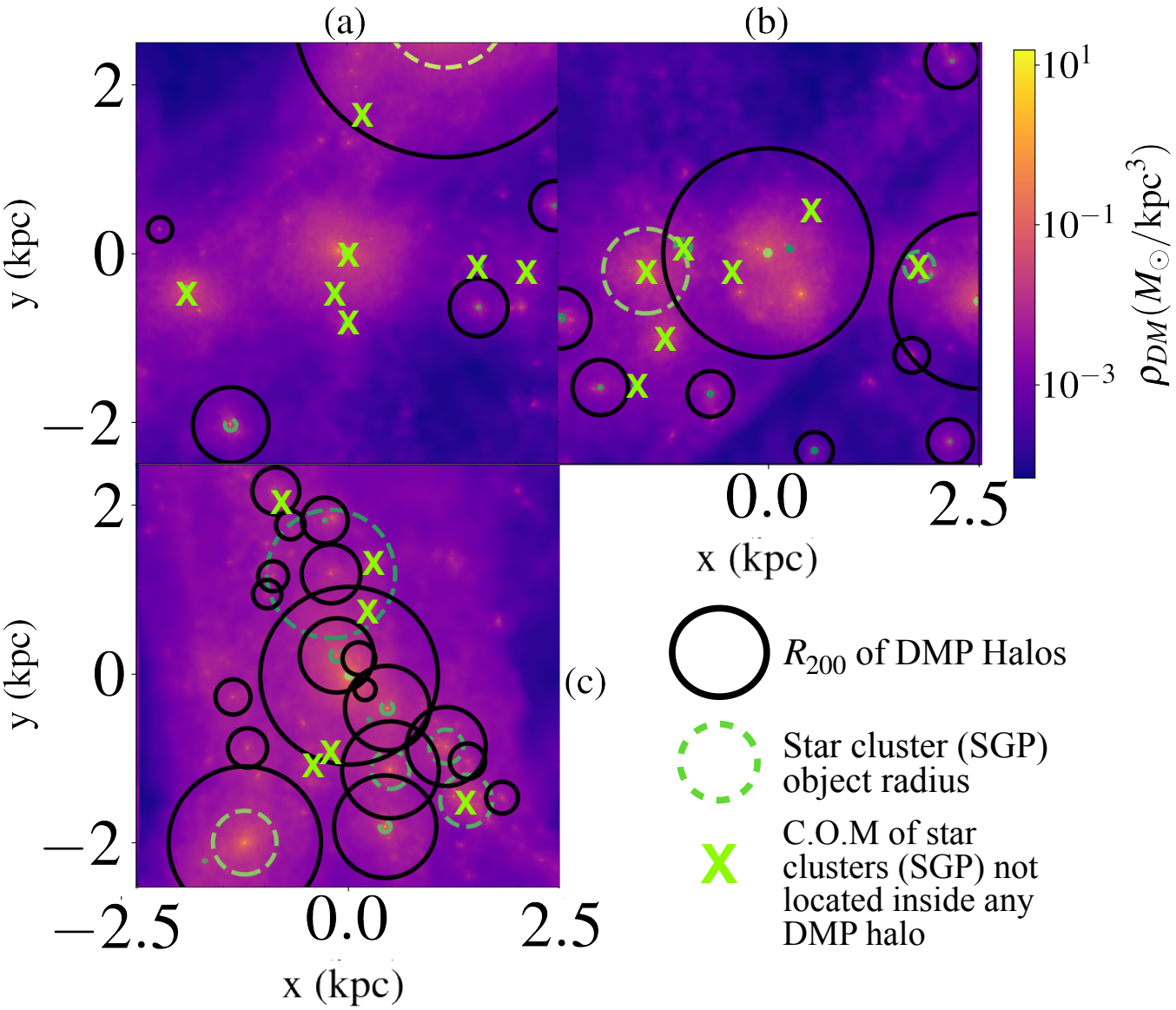


**Figure 1.** Example systems where DM primary fails to characterize the stellar component. The projected density of dark matter is shown in a 10 kpc region of the simulations. Black circles show $R_{200}$ of dark matter halos as detected by the dark matter primary FOF algorithm. The green circles show stars and gas primary (SGP) FOF objects, colored by stellar mass. For nonvirialized objects, the FOF radius is shown as a dashed line. For objects that are virialized, the virial radius is shown as a solid line. An "X" marker is plotted where a star cluster is outside of any DM halo. Three cases are shown. (a) Several baryonic clusters exist in a region where the underlying dark matter distribution does not meet the DM FOF criteria. (b) A star cluster is accreting onto a larger DM halo, and its host is not picked up as a separate halo. (c) A highly dense environment at the intersection of gaseous and dark matter filaments contains many halos and star clusters. Some star clusters lie outside of the halo $R_{200}$.

convergence of the structure-finding algorithm by varying the linking length. The results of that investigation are presented in Appendix A. A pressure floor is not needed for AREPO because the gas cells use a Voronoi mesh method. In order to assure that our smallest structures are not artificial fragments, we have checked the simulation box to ensure that the Jeans mass is resolved by at least 10 cells. Finally, we provide the typical number of star particles in the clusters of most importance to the analysis presented here in Table 2. For high densities, our objects are resolved with a minimum of ∼800 particles, with many systems containing $10^3$–$10^4$ particles. For virialized objects, the typical particle counts are lower, between $10^2$ and $10^3$. High density systems often constitute some of the more massive systems in our box. The full distribution of particle counts in virialized objects can be found in Appendix A.

### 2.3. Boundedness and Virialization

For baryon-centric structure-finding runs without dark matter, we check to ensure boundedness to exclude potential numerical artifacts. First, among the $N$ star particles in the group, which each have mass $m_i$, position $\boldsymbol{r_i} = (x_i, y_i, z_i)$, and velocity $v_i = |\boldsymbol{v}_i|$ we calculate the energy of the star particles:

$$E_* = \sum_i^N \frac{m_i v_i^2}{2} - \sum_i^N \sum_{j \neq i}^N \frac{G m_i m_j}{|\boldsymbol{r_i} - \boldsymbol{r_j}|} \,. \tag{1}$$

If $E_* < 0$, the star particles are certainly bound, and we include that structure in our further analysis. For groups that

**Table 2**
Summary of Typical Numbers of Star Particles in the Objects of Particular Interest to this Analysis—High Density and Virialized Systems

| Parameter | Statistic | $z = 12$ | $z = 13$ | $z = 14$ |
|---|---|---|---|---|
| Number of star particles (top 10% of clusters by density) | Mean | 10,546 | 9478 | 7970 |
| … | Median | 2110 | 1919 | 1827 |
| … | Minimum | 770 | 852 | 784 |
| Number of star particles (virialized star clusters) | Mean | 4018 | 3611 | 3613 |
| … | Median | 243 | 229 | 249 |
| … | Minimum | 101 | 101 | 101 |

**Note.** The stars + gas primary run is shown for the three output redshifts. The mean, median, and minimum number of stars for each category is given at each redshift. The statistics were rounded to the nearest integer. For the stars-only run (not shown in the table), the calculated values are similar to within a factor of <10%.

fail this criterion, we perform an additional calculation incorporating nearby dark matter particles to see if the stars are bound to an underlying dark matter halo. First, we take all dark matter particles within a sphere surrounding the center of mass of the object with a radius equal to the galactocentric distance of the most distant star particle. Among this collection of $N$ stars and $M$ dark matter particles, we calculate the total energy among the stars and dark matter:

$$E_{*,\mathrm{DM}} = E_* + \sum_{i}^{M} \frac{m_{dm,i} v_{dm,i}^2}{2} - \sum_{i}^{M} \sum_{j \neq i}^{M} \frac{G m_{dm,i} m_{dm,j}}{|\boldsymbol{r}_{dm,i} - \boldsymbol{r}_{dm,j}|} - \sum_{i}^{M} \sum_{j}^{N} \frac{G m_{dm,i} m_{*,j}}{|\boldsymbol{r}_{dm,i} - \boldsymbol{r}_{*,j}|}. \quad (2)$$

If this quantity is negative, we consider the structure to be bound. Structures failing this additional criterion are excluded from the analysis as unresolved numerical artifacts.

For bound structures, we check whether the system is virialized. We use the criterion $1.5 < -U/K < 2.5$ to determine virialization, where $U$ is the potential energy and $K$ is the kinetic energy. This threshold is rather broad and chosen somewhat arbitrarily to encompass a number of systems that are virialized or very nearly virialized. However, it is important to ensure that our further analysis does not depend on our choice of cutoff values. In Appendix A, we provide the distribution of virial ratios present among our star clusters and show that similar object distributions result in whether we increase or decrease the threshold values used.

For the structures that we find to be virialized, we calculate the radius $R_{\mathrm{vir}}$. For this we take the radius for which $|U|/K = 2$. If a structure is not virialized, we characterize its radius as the $R_{\mathrm{max}}$, the galactocentric distance of the most distant star particle. This method is summarized in Figure 2, which depicts a flow chart of our process.

Table 2 shows the typical number of star particles contained in our virialized objects. We find virial ratios consistent with equilibrium all the way down to our resolution threshold. Although our resolution threshold of 100 star particles should be enough to ensure that we can resolve the dynamical status of clusters, in principle, it is possible that some of the

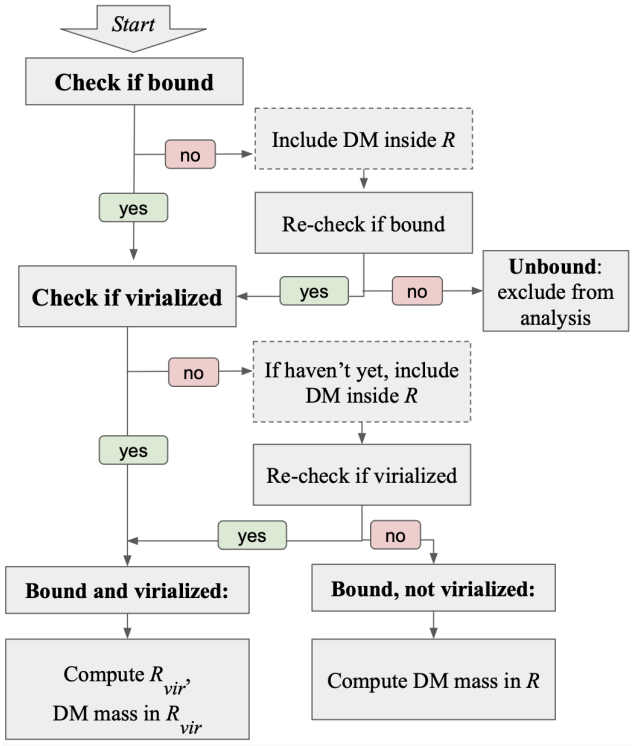


**Figure 2.** Flow chart schematic showing the method for checking boundedness and virialization and computing DM mass within $R$.

structures closest to this threshold that we identify as virialized could actually be unrelaxed or even unbound. We investigate the robustness of our conclusions with respect to the lowest-mass objects in Appendix A. We find that the scatter in virial ratios does not depend on the number of particles at the low-mass end of our systems and, additionally, the population-level trends that we focus on in the remainder of this paper emerge even if we increase our resolution cutoff by a factor of five.

### 2.4. Association of DM Halos

In traditional algorithms, the DM halo mass is computed for the DM-primary structure using a common formalism, such as the mass within a radius $R_{200}$ corresponding to a region enclosing a density 200 times the mean (or critical) density in the Universe. This process is used in our DM-centric runs because it is straightforward to calculate $M_{200}$ within $R_{200}$ when DM is associated with the primary. However, running a structure-finding algorithm on the stars or gas particles/cells without associating DM as a primary or secondary particle complicates the process of calculating the halo mass. Here, we use two methods to understand the dark matter environment of our simulated structures. First, we search the halo catalog provided by the DM-centric method, and for each star cluster identified by the baryon-centric algorithm, we find the nearest DM halo. However, we additionally seek to associate the DM particles that would be found "observationally," near stars or gas observed electromagnetically. This may not correlate with the halos in the DM halo catalog, as evidenced by Figure 1. This limits us to the dark matter particles in the region of the baryonic particles. Thus, we also compute the mass of dark matter contained within a sphere centered at the center of mass of radius equal to that identified by the method of Section 2.3 (either the virial or maximum radius, depending on the

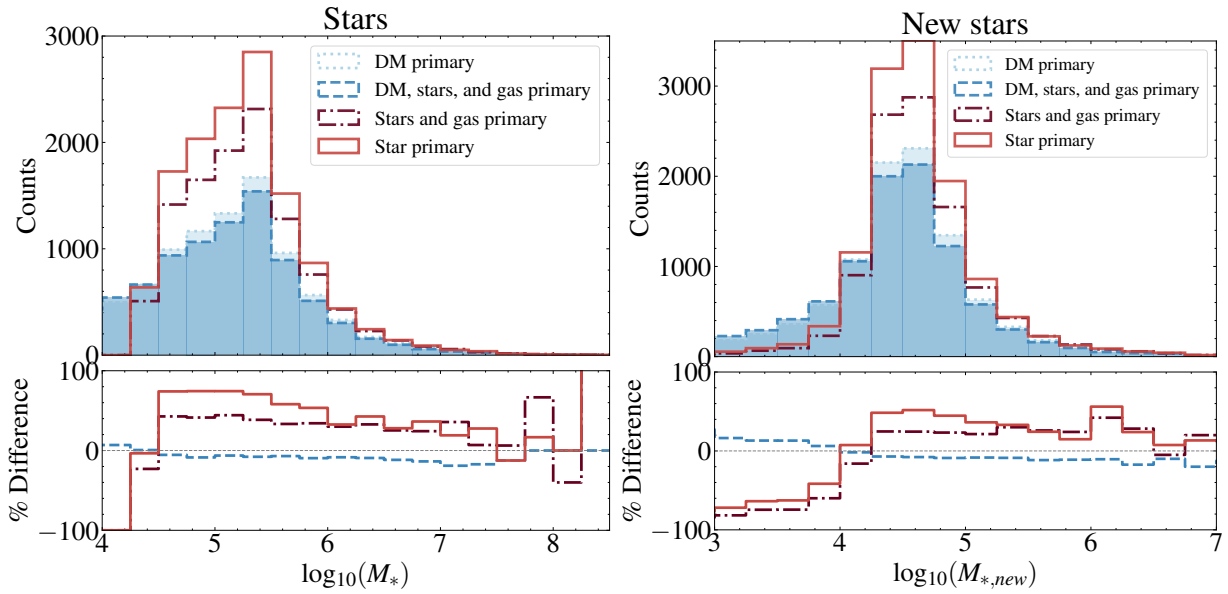


**Figure 3.** Top left panel: Histogram of object counts by stellar mass at $z = 12$ using algorithms centered on the particles listed in Table 1. Bottom left panel: Percent difference between each run and the standard DM-primary, stars+gas method ($f = N/N_{\rm DMP}$). The gray horizontal line shows $f = 1$. Blue shades are runs that include DM—the DM primary (dotted) and the stars, gas, and DM primary (dashed). The red shades are baryonic runs—the star primary (solid) and the stars and gas primary (dotted–dashed). Top right panel: Same as the corresponding left panel, but objects are instead binned by newly formed stellar mass at $z = 12$. Bottom right panel: Percent difference between each run and the standard DM-primary, stars+gas secondary algorithm ($f = N/N_{\rm DMP}$). The gray horizontal line shows $f = 1$.

virialization state). In Section 5, we compare these two probes of DM environment.

In the next sections, we have split our major results into three parts. First, in Section 3, we compare the results of the baryon-centric and DM-centric structure-finding methods. Second, in Section 4, we describe the virialization state of the simulated structures using a simple analytic timescale model. Finally, in Section 5, we use our methodology combined with the analytical understanding to paint a picture of early star clusters: their environment and properties. There, we analyze our systems in relation to the JWST star clusters with high surface density.

## 3. Comparison of Algorithms Targeting DM versus Baryonic Particles

The left-hand panel of Figure 3 shows the number of galaxies binned by stellar mass found in our $z = 12$ snapshot according to the different structure-finding algorithms. The bottom panel shows the fractional comparison with the standard DM-primary, stars+gas secondary algorithm. From this figure, it is clear that simply counting and binning the number of galaxies in a simulation snapshot depends strongly on the particle type used for the structure-finding algorithm. In Appendix B, we derive the star cluster mass function for the box from these number counts. Since this volume uses an increased $\sigma_8$ parameter, the provided mass function reflects that of very high-density peaks at this epoch and not the Universe on average.

Variations in the number of galaxies of 50% or more are present between the runs. This demonstrates that a system can form enough stars to represent a small galaxy by observational classification (i.e., its stars may be detected by a sufficiently sensitive observatory), but a dark-matter-focused algorithm may not find it. Alternatively, the dark matter primary (DMP) algorithm may discover this structure but misclassify its stellar mass and star particles, since a merging system or a complex filamentary system may contain a stellar component at the edge. The DM+stars+gas primary (where baryonic and DM particles are combined as the primary particle type) was designed to combat this case and add weight to the systems with clustered baryonic components. However, as evidenced by Figure 3, the DM+stars+gas primary number counts trace the dark matter only version to within a few percent at all masses.

A relevant parameter for these galaxies is the number of newly formed stars in the snapshot, which is directly used to calculate the star formation rate (e.g., W. Lake et al. 2024; C. E. Williams et al. 2024). Here, we calculate the "newly formed stellar mass," ($M_{*,\rm new}$):

$$M_*, \text{ new} = \sum_i^{N_{\rm new}} m_i, \tag{3}$$

for each object, where the object contains $N_{\rm new}$ stars whose particle index did not appear in the previous snapshot and $m_i$ is the mass of each new star particle. We show histograms of the counts by newly formed stars in the right-hand panel of Figure 3, and it is clear that the number at each mass varies significantly for this parameter as well. In the left panel of Figure 3, our resolution limit of 100 star particles is clearly visible in the sharp cutoff of the low-mass end of the distributions for the baryon-focused methods. The DMP can distinguish arbitrarily small numbers of star particles since its resolution cutoff is limited to dark matter particles. For the new stellar mass distributions, both methods can detect any number of newly formed star particles (indeed, systems appear with zero—these are not included in the distributions). At high masses (where the newly formed stellar mass is $> \sim$few$\times 10^4 M_\odot$), the same trend seen on the right panel is present: more systems are found by the baryon-focused methods. At low masses, the resolution cutoff comes into play.

Again, the DMP can distinguish an arbitrarily low number of star particles. However, the baryon-focused method's number counts are fewer than the dark-matter focused number counts because they are missing systems with less than 100 star particles. We emphasize that the discrepancy at high mass is most relevant for JWST predictions. Our resolution cutoff, preventing the baryon-focused method from identifying systems with less than 100 new stars, results in us missing systems with extremely faint apparent magnitude, which would be unrealistic to consider for detection by JWST.

The discrepancy in $M_{*,\mathrm{new}}$ is especially concerning given that the number counts by newly formed stars are used in the determination of the star formation rate and UVLF for these simulations.[12]

To understand why the various structure-finding algorithms produce such number counts, let us consider again the visualizations of example systems in Figure 1. This figure illustrates scenarios where the DM and baryonic methods disagree. In scenario (a), the central star cluster is obviously located at a density peak in the dark matter distribution. However, this object is not identified by the DM primary algorithm. Thus, a relatively massive star cluster/early galaxy is not counted by the DM-centric run. In case (b), a star cluster is located near $R_{200}$ of a larger halo. It seems to be undergoing a merger and, again, is colocated with a DM density peak that is not identified as a halo. In cases where a merger is ongoing, the dark matter halo may be disrupted by a larger neighbor and thus is missed in the DM search. However, the star cluster remains and would be observationally detected by a theoretical observatory of sufficient sensitivity. In case (c), we see a highly complex region at the intersection of several filaments. In this case, it is clear that the approximation of spherical halos may not be appropriate. Here, the dark matter and gas in the region are located along filaments streaming into the central halo. This could be observationally explored through studies of galaxy morphology (e.g., V. Pandya et al. 2024). In this complex region, star clusters have become spatially decoupled from their host halos, and thus several clusters are found outside of the DM primary.

Note that in all three boxes, several star clusters are present that are either outside any dark matter halo (marked with an "X") or located toward the edges of a central halo. Inside a halo, using the baryon-centric methodology allows us to identify individual clumps, which can give a more accurate accounting of the surface density. We will discuss the significance of the surface density in comparison to JWST observations in Section 5.2.

The figure also identifies some of the common scenarios and, to further quantify how frequently these occur in the simulation box, we consider systems and their distance from host halos. Systems with large distances to the closest halo may lie in the outskirts of a large halo or may be in a diffuse or disrupted halo. If we drop the dark matter resolution threshold to 32 particles, we only identify <10 star clusters in each catalog that are colocated with a detected but subresolution halo. We then count the number of objects whose center of mass is located outside the $R_{200}$ of the nearest neighbor and find 2396 systems in the stars + gas catalog at $z = 12$. Of these systems that are not associated with any dark matter halo, we find that 132 have an enclosed mass greater than 300 dark matter particles. Thus, there are 2396 objects whose host halos are either diffuse, disrupted, or undergoing a merger such that the star cluster is removed. For 132 of these objects, we can confidently identify that the halo is too diffuse to meet identification criteria, given that a large mass of dark matter is colocated. This accounting of the environment will be explored further in Section 5.

[12] Without an indication of the underlying halo mass, it is difficult to accurately correct the runs for the increased clustering parameter $\sigma_8$ used in this study. So, by this method, it is challenging to predict the UVLF for the simulated box.

**Table 3**
Virial Fractions in Each Run at Output Redshifts

| Object Category | $f_{\mathrm{vir}}(z = 12)$ | $f_{\mathrm{vir}}(z = 13)$ | $f_{\mathrm{vir}}(z = 14)$ | $f_{\mathrm{vir}}(z = 14)$ |
|---|---|---|---|---|
| DM primary | 0.779 | 0.769 | 0.765 | 0.746 |
| Stars and gas primary | 0.160 | 0.144 | 0.168 | 0.169 |
| Star primary | 0.174 | 0.158 | 0.156 | 0.153 |

**Note.** Results include all halos with $\log_{10}(M_*) > 4.25$.

Our analysis of these algorithms suggests that there is a population of relatively massive, bound star clusters that are hosted in halos that are undetected by the traditional dark matter-based structure-finding algorithm. Potentially, they have halos that are too small to be considered by our threshold of 300 DM particles or, alternatively, the DM halos that host them do not meet other criteria of the halo-finding algorithm (such as being sufficiently dense). We note that simulations including the primordial cosmic streaming velocity between dark matter and baryons, which serves to separate gas from dark matter and generate more diffuse halos (e.g., W. Lake et al. 2023b; C. E. Williams et al. 2023), may suffer even more from this issue. Another group of star clusters may be in the process of accretion or fragmentation, and thus be missed as they straddle the boundary between the $R_{200}$ of multiple halos.

## 4. Dynamical Properties of Star Clusters

In isolation, we expect that long-lived star clusters will tend toward equilibrium. In this section, we investigate the virial state of star clusters in our simulation. We use simple dynamical arguments to estimate the timescales for virialization at various masses and use this to understand the simulated structures.

### 4.1. Distribution of Virialized Systems

As described in Section 2, for runs with only baryonic (stellar or gas) particles as the primary particle type, one concern is that the structure-finding algorithm may pick up unbound structures that are numerical artifacts. As detailed in that section, we checked all objects for boundedness based on their total energy. During that process, we also tested whether or not systems are virialized. We find that even after including DM particles in the immediate region of the stars, the majority of objects are not virialized. The top row in Table 3 shows the virial fraction in each run, i.e., the number of virialized objects ($N_{\mathrm{vir}}$) per total number ($N_{\mathrm{tot}}$):

$$f_{\mathrm{vir}} = \frac{N_{\mathrm{vir}}}{N_{\mathrm{tot}}}. \quad (4)$$

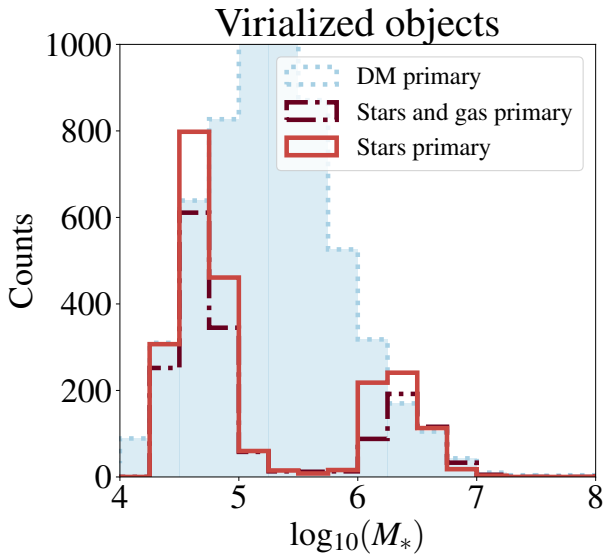


**Figure 4.** Number of virialized objects at $z = 12$ by stellar mass in solar masses. Here, we consider the star primary (red solid line), star + gas primary (dark red dotted–dashed line), and DM primary (light blue dotted line).

Around 80% of systems detected through the DM primary are virialized. However, as shown in this table, for baryon-primary objects, the virial fraction varies slightly but remains around 16% for all redshifts. While systems are theoretically expected to tend toward equilibrium over time, the process of virialization may be disrupted by external factors, such as mergers, accretion, and tides. The fact that a greater fraction of the baryonic-centric systems are not virialized hints that the systems that are not detected through the DM primary algorithm are systems undergoing disruptive processes, such as mergers, which additionally cause nondetection of the host halo.

To investigate this behavior, in Figure 4 we plot the distribution of virialized objects by their stellar mass at $z = 12$. For the DM-centric detected structures (blue histogram), the distribution is similar to the overall distribution of halos as in Figure 3. Indeed, for halos with $M_* > 10^5 M_\odot$, a Kolmogorov–Smirnov test cannot reject the null hypothesis that the virialized and overall distributions are the same ($p$-values between 0.17 and 0.54 for $12 \leqslant z \leqslant 15$). In other words, a roughly constant fraction of objects at most masses are virialized. However, the baryonic-centered runs (shown in red in the figure) show a bimodal distribution, with a dearth of virialized objects between $10^5 M_\odot$ and $10^6 M_\odot$.[13] This bimodality is found in every redshift where a snapshot was taken in the simulation (see Appendix D). Since the overall halos (traced by the dark matter) are virialized, this suggests processes affecting the conversion of gas to stars or the stellar systems themselves. The following two subsections investigate these processes in detail.

### 4.2. Early Gas Conditions

To uncover the origin of the two populations of virialized objects, we start by considering the state of the gas at slightly higher redshift. This baryonic matter is the material that must condense and cool to form the star clusters seen at our final redshifts. We take a snapshot from the simulation at $z = 25$, and run only the standard DM-primary, stars+gas secondary algorithm. For these early halos, we plot the average gas temperature versus average number density in Figure 5. Here, we see that the halos fall into two typical densities, roughly independent of mass and temperature at low masses. Since nonequilibrium molecular hydrogen cooling has been incorporated into the simulation, we suggest that the characteristic gas density achieved by these halos results from their progress through primordial molecular hydrogen tracks. If this is the case, we suggest that the resulting temperatures and densities should correspond with typical molecular cooling tracks as seen in other works (e.g., N. Yoshida et al. 2006; T. H. Greif et al. 2010; K. Omukai et al. 2010; S. Hirano et al. 2018; W. Lake et al. 2025). To test this, we present a comparison to these studies in the bottom panel of Figure 5. Here, we plot the core density, which is the quantity used in the literature. In this figure, there is good agreement and the two populations correspond to a buildup along the cooling tracks. In other words, primordial molecular hydrogen cooling seeds two characteristic densities into the early gas, according to each halo's evolution along the cooling track. To ensure that splitting based on the mean density will also result in a split based on the core density achieved through molecular cooling, we plot our objects' mean versus core density and find a strong correlation, suggesting that this criterion captures the underlying physical behavior.

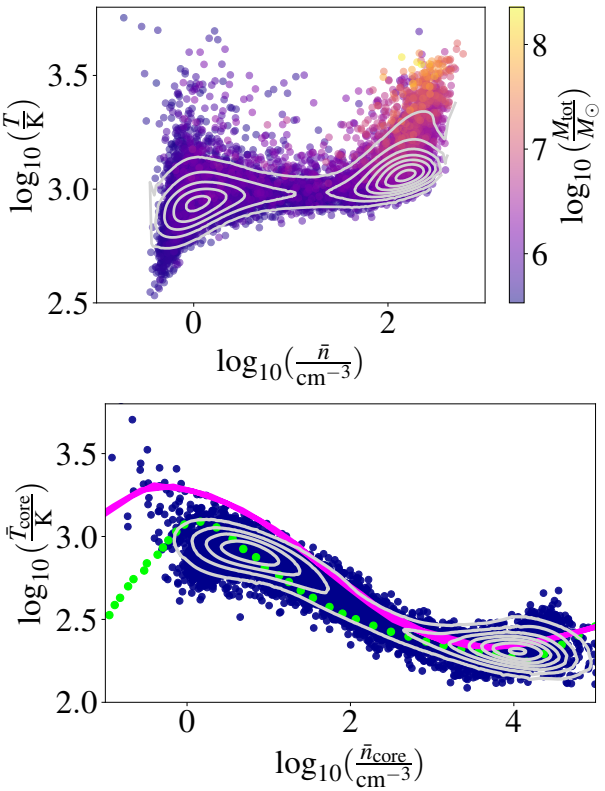


**Figure 5.** Top panel: Logarithm of average temperature vs. average number density at $z = 25$ in DM halos. Contours show the 2D density of points in the parameter space. The color bar depicts total halo mass ($M_{\mathrm{DM}} + M_{\mathrm{b}}$). Bottom panel: Same as top panel, but the relation is plotted for the core gas cell in each halo instead of the mean. Overplotted are the results of K. Omukai et al. (2010) (green dotted) and N. Yoshida et al. (2006) (pink solid).

[13] For these virialized objects, the null hypothesis that they arise from the same distribution as the overall population fails with $p$-values below $10^{-120}$ for $12 \leqslant z \leqslant 15$.

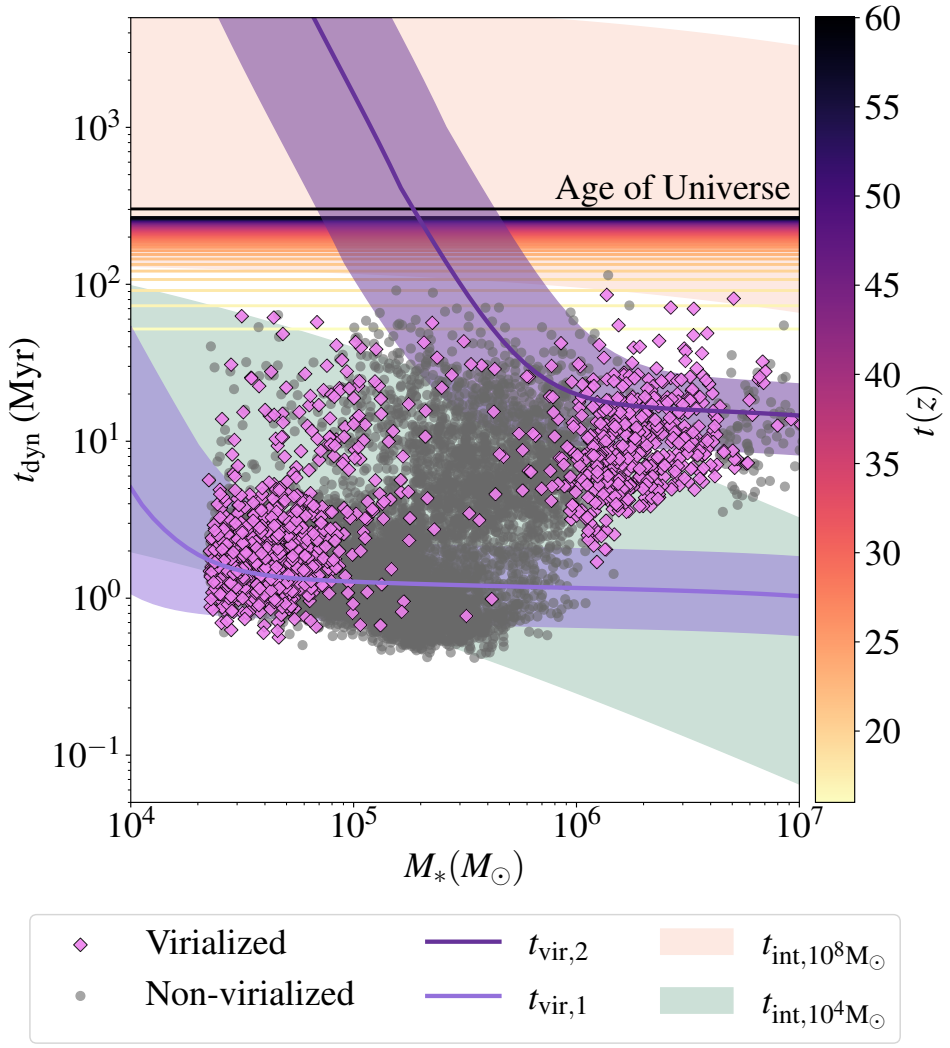


**Figure 6.** Analytic virialization and interaction timescales compared to dynamical timescales of Star+Gas primary objects at $z = 14$. $t_{\rm vir,1}$ is the virialization timescale for systems with density $\rho_1$, and $t_{\rm vir,2}$ corresponds to $\rho_2$ (see Table 4). Shaded regions around these curves denote the range of values associated with the simulated parameters. The dynamical timescale is calculated from the star particles' orbital velocity. Similar results are obtained if the gas density or gas cell orbital velocities are used to trace the dynamical timescale. The color bar shows the time since a variety of redshifts in Myr. The solid black line shows the age of the Universe at $z = 14$. The red region shows the range of interaction timescales for a $10^8 M_\odot$ halo, and the green region shows the interaction range of timescales for a $10^4 M_\odot$ halo.

We postulate that the two gas densities present already in the DM halos at $z = 25$ may relate to the two virialized populations we find at lower redshift. In the following section, we estimate two characteristic densities for an analytical investigation corresponding to the peak density of each of these two populations.

### 4.3. Timescale Investigation

Given the two characteristic densities present in halos from high redshift in the simulation, we understand the state of virialization of simulated objects in the context of a simple timescale argument. Specifically, in Figure 6, we consider each star cluster dynamical timescale $t_{\rm dyn} = r/v_c$, where $r$ is the cluster radius and $v_c$ is the mean circular velocity of the star particles.[14] This timescale is plotted as a function of mass. As depicted, we can identify two branches of virialized systems associated with two separate mass distributions. We find the densities corresponding to the two peaks of the average density distribution for this analysis (see Figure 5 top panel). These are $n_1 = 1.89 \times 10^2\,{\rm cm}^{-3}$ and $n_2 = 9.48 \times 10^{-1}\,{\rm cm}^{-3}$. In terms of mass density, $\rho_1 = 3.16 \times 10^{-22}\,{\rm g\ cm}^{-3}$ and $\rho_2 = 1.58 \times 10^{-24}\,{\rm g\ cm}^{-3}$.

Then, to understand this behavior, we consider the dynamical processes governing the clusters' evolution.

Our analytical method to uncover the underlying dynamics is to compare the minimum timescale for virialization to the range of typical timescales for interaction with nearby halos. The minimum timescale for virialization is a function of mass and density: $t_{\rm vir}(M,\ \rho)$ (see below for a detailed expression). The timescale for interaction is additionally a function of the mass of the nearby halo: $t_{\rm int}(M,\ \rho,\ M_{\rm DM})$ (again, derived below). There are three relevant scenarios:

1. $t_{\rm vir}(M,\ \rho,\ ) > t_{\rm Universe}$: If the virialization timescale is longer than the age of the Universe (the simulation run time), we do not expect an object to be able to virialize in our simulation.
2. $t_{\rm int}(M,\ \rho,\ M_{\rm DM}) < t_{\rm vir}(M,\ \rho) < t_{\rm Universe}$: If interactions with nearby halos occur much more frequently than the

[14] The dynamical timescale was calculated using the star particles' orbital velocity. We repeated this analysis with other tracers of the dynamical time, such as gas density and gas cell velocity, and similar results were obtained.

virialization timescale, we do not expect virialization because clusters may be disrupted or merged with nearby objects before they reach equilibrium.

3. $t_{\rm vir}(M, \rho) < t_{\rm int}(M, \rho, M_{\rm DM}) < t_{\rm Universe}$: If the interaction timescale is longer than the virialization timescale, then the system has time to virialize before being disrupted again, and we expect to find it in a state of virialization.

These timescales are used to explain the bifurcation of the clusters' virialization properties and are derived in more detail below. We first describe the behavior followed by the relevant equation.

The minimum virialization timescale is plotted in purple for our two characteristic densities in Figure 6. Additionally, we show a range of reasonable values for the interaction timescale in the orange and green shaded regions (there is a large range of environments in the simulation). The green region corresponds to halos of only $M_{\rm DM} \sim 10^4 M_\odot$. The orange shaded region corresponds to halos of $M_{\rm DM} \sim 10^8 M_\odot$ (see Appendix C.1 Figure 16 for more values). We show the age of the Universe as a horizontal black line, as well as the time since various redshifts as horizontal lines corresponding to the values shown on the color bar. The bifurcation in dynamical timescales of the simulated points corresponds to the two populations by density (see Figure 18 in Appendix C.2), and the data points lie roughly along the two tracks for analytic virialization timescales.

From this, we can use the criteria above to correctly predict whether a population at a given mass will be virialized. We note that the correct interaction timescale is that corresponding to a halo mass that would cause a major merger ($\sim > 10\% M_*$). First, our model correctly shows that low-density objects ($\rho_2$, $t_{\rm vir,2}$) should not be virialized below $M_* \sim 10^6 M_\odot$. However, those that do virialize should stay virialized, since the interaction time for very massive halos is much longer than their virialization time. For high-density objects, our model also explains the mix of virialized and nonvirialized systems. For example, a roughly $10^4 M_\odot$ cluster of high density should very frequently be virialized, since the interaction timescales (even for similarly low-mass halos) are typically longer. Meanwhile, a $10^{5.5} M_\odot$ cluster may often be found in a disrupted state, since the interaction timescale for halos 10% −100% of its size dips below its virialization timescale.

*Derivation of the Virialization Timescale*. Here, we describe in detail how we analytically estimate the minimum timescale for star cluster virialization (the purple lines in Figure 6). To do this, we assume that the baryonic matter must undergo two stages: first, gas must cool such that gravity may overcome gas pressure and form stars; and second, the stars must undergo violent relaxation, such that virialization occurs. Inspired by the methods of W. Lake et al. (2023a), we estimate the timescale for molecular cooling with a two-phase model: first, the gas cools isochorically, until the Jeans mass falls below the total mass, and then the gas collapses to form stars on the cooling timescale (defined in A. T. P. Schauer et al. 2021). In this model, our cooling timescale is:

$$t_{\rm cool} = \frac{k_{\rm B} T_{\rm collapse}}{n_{\rm H}(\gamma - 1)\Lambda(T_{\rm collapse}) f_{\rm H2}} + t_{\rm iso}, \tag{5}$$

where $T_{\rm collapse}$ is the temperature of collapse after the Jeans mass is exceeded, $n_{\rm H}$ is the hydrogen number density, $f_{\rm H2}$ is the molecular hydrogen fraction, $\gamma$ is the adiabatic constant, $k_{\rm B}$ is the Boltzmann constant, and $\Lambda(T)$ is the temperature and density dependent molecular cooling rate. We use the rates provided by D. Galli & F. Palla (1998):

$$\Lambda(T) = \frac{\Lambda({\rm LTE})}{1 + [n^{\rm cr}/n({\rm H})]}. \tag{6}$$

Again following W. Lake et al. (2023a), the initial temperature is taken to be 500 K or the Jeans temperature, whichever is higher:

$$T_{\rm J} = \frac{3{\rm G}\mu m_{\rm H}}{5k_{\rm B}} \left(\frac{36M^2\rho}{\pi^2}\right)^{\frac{1}{3}}. \tag{7}$$

If the Jeans mass is not initially reached, the isochoric cooling timescale in Equation (5) is:

$$t_{\rm iso} = \int_{T_{\rm Collapse}}^{T_{\rm init}} \frac{{\rm k_B}}{n_{\rm H}(\gamma - 1)\Lambda(T) f_{\rm H2}} dT. \tag{8}$$

The cooling timescale is plotted in blue in Appendix C.1 Figure 16 for $\rho_1$ and $\rho_2$. For high densities, the cooling timescale on the order of Myr or well below for most of the masses considered. For the lower density case, the cooling timescale exceeds the age of the Universe at these redshifts for masses below $\sim 10^5 M_\odot$.

Once a star cluster has formed, the next phase of virialization is violent relaxation. This requires a rapidly changing gravitational potential, such that the orbits of stars experience significant perturbation on dynamical timescales. A detailed estimate would require an understanding of the complex environment surrounding each star cluster. Here, we seek to instead provide a minimum timescale for virialization, in order to understand whether or not we should expect many clusters to virialize at each mass and density. In other words, this estimate assumes that the star cluster undergoes virialization immediately following its formation. To place this lower bound, $t_{\rm vir}$, our "virialization timescale" is simply:

$$t_{\rm vir} = t_{\rm cool} + t_{\rm dyn}. \tag{9}$$

This sum is what is plotted in purple in Figure 6. We estimate the dynamical timescale for mass $M$ and density $\rho$ by assuming a uniform density. In this case,

$$t_{\rm dyn} = \frac{r}{v_c}, \tag{10}$$

where $v_c$ is the circular velocity of a halo of constant density with mass $M_{\rm tot} = M_{\rm g} + M_* + M_{\rm DM}$. For this estimate, we numerically fit $f_{\rm g} = M_*/M_{\rm g}$ and $M_*/M_{\rm DM}$ to match our simulations (see Appendix C.2). This lends the dynamical timescale a very weak dependence on mass. In Appendix C.1 Figure 16, the dynamical timescales for $\rho_1$ and $\rho_2$ are shown in gray.

The overall $t_{\rm vir}$, (Equation (9)), is plotted in Figure 6 for both densities as a purple line. For low masses, the virialization timescale is dominated by the gas cooling, whereas for high masses, the timescale is set by the dynamical time. Already, without including any disruption processes, we see from the two purple lines in Appendix C.1 Figure 16 that the low-density gas should not virialize within the age of the Universe for systems below $M_* \sim 10^{5.5} M_\odot$. This aligns well with the trough of virialization in Figure 4. However, this does not explain why high-density systems, if they exist for higher masses, do not achieve virialization for all masses. For this reason, we invoke the interaction timescale, derived below.

*Derivation of the Interaction Timescale*. The most important process disrupting virialization for the purposes of our estimation here is the interaction with the surrounding environment.[15] Since our simulations are designed to follow the first epoch of galaxy formation, when the hierarchical merger mechanism is building up larger galaxies, we expect interactions to be quite common. Accretion of external star clusters may cause virialized clusters to move out of equilibrium. However, the degree and timescale of disruption depends on the mass ratio of the interaction. Major mergers, or interactions with other structures of similar mass, should be expected to severely impact the virial status of a system (N. Yoshida et al. 2003). Meanwhile, accretion of a halo whose mass is orders of magnitude smaller may not be enough to significantly disrupt equilibrium. We note that since this is a gravitational process, all halos (not just star cluster hosts) should be included in the calculation.

A simple rate calculation provides the timescale for a nearby halo to enter within the virial radius of the dark matter halo hosting the star cluster in question:

$$t_{\rm int} \sim \frac{1}{n_{\rm DM}(M)\,\sigma_{\rm int}\,v_{\rm group}}. \tag{11}$$

Here, $n_{\rm DM}(M)$ is the number density of dark matter halos of mass $M$ in the region, $\sigma_{\rm int}$ is the cross section of interaction, and $\sigma_{\rm int} = \pi(r_{\rm group} + r_{\rm perturber})^2$ and $v_{\rm group}$ are the group velocity of the star cluster with respect to the bulk flow. We estimate the radii in the cross-section $\sigma_{\rm int}$ using the analytical formula for virial radii as a function of mass and redshift from R. Barkana & A. Loeb (2001):

$$r_{\rm vir} = 0.784\left(\frac{M}{10^8 h^{-1} M_\odot}\right)^{1/3}\left(\frac{\Omega_m \Delta_c}{\Omega_m^z 18\pi^2}\right)^{-1/3} \times \left(\frac{1+z}{10}\right)^{-1} h^{-1}{\rm kpc}\,. \tag{12}$$

The other quantities in Equation (11) are estimated from a reasonable numerical range in the simulations. For the number density of halos, we provide an upper and lower bound using two numerical estimates drawn from our simulations. Appendix C.2 Figure 17 shows the mean number of dark matter halos within a radius of 10 ckpc of each star primary object. For most low-mass objects, there are around 6–7 dark matter halos in the vicinity. Assuming these are evenly distributed, we use the mean $N = 6$ (see Table 4) to estimate the number density in Equation (11) for an interaction with the local environment. This figure likely overestimates the interaction time because it assumes that the nearby halos are uniformly distributed in the region. However, we know that structure is highly clustered, and often many halos are located along a filament of the cosmic web (see Figure 1). Thus, as a lower bound, we estimate the density from the nearest dark matter halo, calculating the timescale for that interaction. This "nearest neighbor" estimate provides an upper limit. It is most accurate for low-mass halos, which have the highest overall number density. Thus, structures are most likely to have a low-mass halo as their nearest neighbor. Figure 19 in Appendix C.2 shows histograms of the mass of the nearest neighbor, velocity with respect to the bulk flow, and the number of nearby neighbors. We find the average group velocity of the star clusters to be 26 km s$^{-1}$ with respect to the bulk flow, with a standard deviation of 16.6 km s$^{-1}$. We use the standard deviation to represent the typical relative velocity difference between two nearby neighbors.

[15] Tidal heating was also investigated as a process that may possibly disrupt the virialization state of a cluster. For the systems considered in this paper, the diffusive tidal heating timescale was several orders of magnitude longer than the age of the Universe at this epoch.

**Table 4**
Derived Quantities from Simulated Results used in Analytical Calculations of Timescales

| Parameter | Value |
|---|---|
| $\rho_1$ | $3.16^{+6.83}_{-2.16} \times 10^{-22}$ g cm$^{-3}$ |
| $\rho_2$ | $1.58^{+0.954}_{-3.43} \times 10^{-24}$ g cm$^{-3}$ |
| $\bar{v}_{\rm group}$ | 26.3 km s$^{-1}$ |
| $\sigma_{\rm group}$ | 16.6 km s$^{-1}$ |
| $\bar{N}_{\rm DM}$ | 6.06 ± 3.89 |

**Note.** Further data is presented in Appendix C.2.

Finally, we must modify the number density to account for the variation of number density with mass. To do this, we use the R. K. Sheth & G. Tormen (1999) number density, and compute:

$$n(M) = f_\delta M(M) n_{\rm est},\ {\rm sim}, \tag{13}$$

where $f_\delta M$ is the fraction of halos at mass $M$,

$$f_\delta M(M) = \frac{N(>(M+\delta M)) - N(>M)}{N_{\rm tot}}. \tag{14}$$

Using these quantities, we plot $t_{\rm int}$ in Appendix C.1 Figure 16. The left panel of the plot shows the upper bound interaction timescale with the local environment as a series of lines colored by interacting halo mass. The inverse dependence on mass reflects the larger cross section of larger halos, while larger interacting halo mass leads to longer timescales because of their lower number density. In the right panel of the figure, the lower bound nearest neighbor timescale is shown, also colored by interacting halo mass. These timescales are much shorter because the nearest halo is much closer than the average halo.

The interactions between low-mass halos at high redshift have been studied on similar scales with full merger trees in other work, especially the context of early massive black hole formation (e.g., R. Fernandez et al. 2014; J. H. Wise et al. 2019; J. A. Regan et al. 2020; J. Regan 2023). For example, J. Regan (2023) finds that for halo mass of $>10^6 M_\odot$ at $z = 20$, the rapid assembly of halos in overdense regions causes the analytic molecular hydrogen cooling time to underestimate the timescale according to which the associated gas cells cool in the simulation This dynamical heating is a similar process to the dynamical disruption we understand to be pulling halos. Multiplying by a rough fraction of $M^*/M_{\rm halo} \sim 10^{-(\Gamma-2)}$ and allowing for some growth by $z = 12$, these halos could represent a similar population to those we find to host nonvirialized clusters at $z = 12$ due to interaction in our analytical model. Thus, we tentatively suggest a basic picture of consistency with our work. Of course, in further investigation, it would be interesting to construct a simulation with a full halo merger tree to probe the specific dynamical histories of our systems and trace the time evolution of virialization.

## 5. Early Star Clusters

Now, we combine our methodology that identifies dense clusters of stars with the numerical and analytical understanding of the processes governing equilibrium to provide an overall explanation of the nature of Cosmic Dawn star clusters in our simulation box.

### 5.1. Environment

In the previous sections, we have shown that the population of low-mass clusters and galaxies deviates from a simple theoretical model where clusters of stars reside in spherical or ellipsoidal halos. From our analysis, it seems that the role of environment in shaping their distribution is central. In this section, we further investigate the population in our box with respect to their local environment.

First, we plot the stellar mass of baryon-centric-detected objects as a function of DM mass enclosed within their radius $R$ (see Section 2).

Virialized objects emerge as a track with higher $M_*/M_{\rm DM}(r < R)$ ratio than the nonvirialized population. Since the radius considered in this calculation for virialized objects is the virial radius $R_{\rm vir}$, which may be considerably smaller than $R_{\rm max}$, it follows that less dark matter is enclosed. The two tracks converge at low masses, suggesting that $R_{\rm vir} \sim R_{\rm max}$ for these structures. However, at larger masses, the divergence in Figure 7 suggests that objects have a diffuse stellar component at greater distances than the virial radius of the central cluster. Additionally, underplotted in gray are the relations from the DM-centric structure catalog. Generally, these converge with the baryon-centric run for the intermediate (and high) masses for the nonvirialized case, suggesting that the radius of virialized objects is similar to the dark matter $R_{200}$ at those masses.

Similar trends are present in the lower panel of Figure 7, which compares the DM mass of the nearest halo from the dark matter-centric run to the dark matter mass within the radius of the stellar object. Many nonvirialized objects lie along the 1:1 curve, again suggesting that the closest dark matter halo is indeed the host halo of the object, and that the stars in nonvirialized systems represent a more diffused component of similar radius to the dark matter. The virialized objects lie mostly above the 1:1 line, suggesting that they have tight virial radii that do not align with the DM $R_{200}$. From the color bar, which denotes the distance to the nearest dark matter halo, we can see that these virialized structures are indeed colocated with a dark matter halo, but in the enclosed mass measurement we are only tracing its core.

Another reason structures may be deficient in dark matter is that they represent a fragmented or recently accreted system, and thus they lie at some distance from the central dark matter cusp. The color bar indicates these objects lie at a large distance from the nearest halo in the DM catalog. In some cases, objects lie below the 1:1 curve. Nearly all of these systems are located at a great distance from the nearest halo (colored yellow), compared to the hypothetical separation from a host halo, which would lie nearly on top of the star cluster. However, their location on the $x$-axis shows that there is a large mass of dark matter in the region. This reveals the population mentioned previously in Section 2: star clusters that live in halos not identified by the DM-centric algorithm, which would be missed by a DM-primary only study. Of course, these may coincidentally lie elsewhere in the plot as well. From our visualizations in

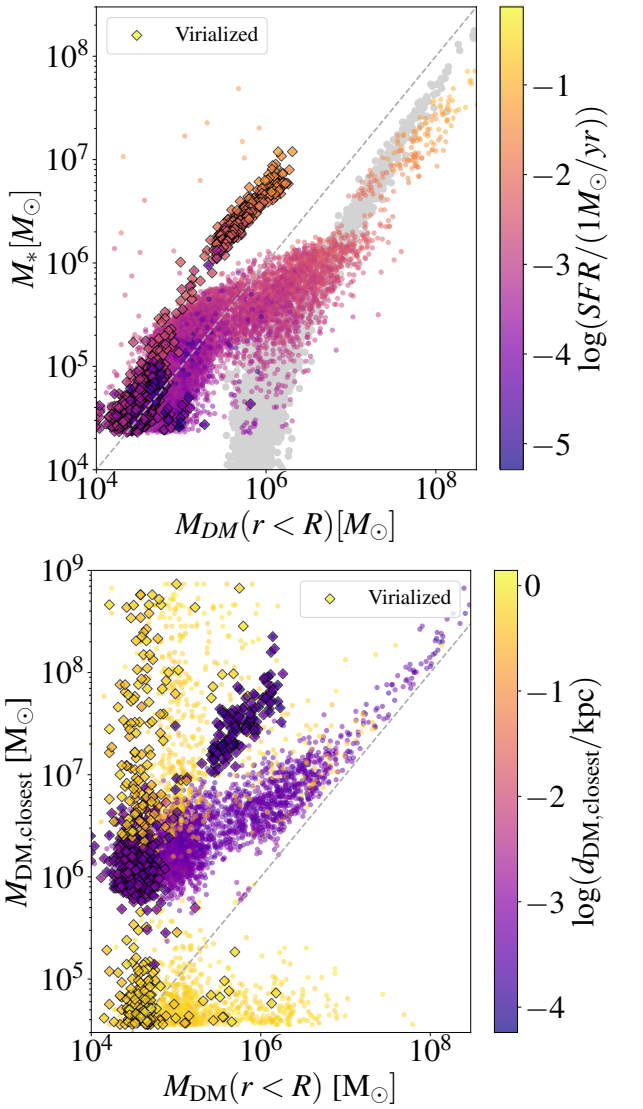


**Figure 7.** Top panel: Stellar mass vs. dark matter mass within $R_{\rm obj}$ at $z = 12$. $R_{\rm obj}$ is defined as the radius given from the FOF algorithm or the virial radius, if the object is virialized (see Section 2). The color bar gives the logarithm of the star formation rate. Points shown are for the gas + stars primary. Bottom panel: Mass of the closest dark matter halo in the DM FOF runs vs. the enclosed DM mass within the object radius for the star + gas primary run at $z = 12$. Points are colored by the logarithm of the distance to the closest halo. In both panels, the gray dashed line shows the 1:1 relation. In the lower panel, all DM objects with particle number $N_{\rm DM} > 32$ (i.e., the entire catalog) were included when searching for the nearest neighbor, in order to identify which systems may have missed our resolution threshold.

Figure 1, we can see that this is especially frequent for *virialized* objects, which are typically found in the core of halos—the dark matter-centric run has not picked up their host halo. Based on this visual analysis, we postulate that almost all the virialized objects that appear to have a large distance to their nearest halo instead are hosted by nondetected halos. This emphasizes why a subhalo algorithm method—such as using AREPO's built-in subhalo flag (which uses the SubFind algorithm) to identify fragmented clusters—might still be unable to identify all the star clusters in the simulation box.

We present a summary of the number of virialized objects in the various populations in Table 5. This table shows the total number of objects and the number of virialized objects that satisfy the condition that their distance to the nearest neighbor

**Table 5**
Comparison of Star Clusters Located Outside of Dark Matter Focused Halo Catalog, Split in Columns by All Objects and the Subset of Virialized Objects

| Population | All | Virialized |
|---|---|---|
| $d > R_{200,\mathrm{DM}}$ | 2164 | 371 |
| $d > R_{200,\mathrm{DM}}$, $M_{\mathrm{DM}} > M_{\mathrm{DM,res}}$ | 137 | 10 |
| $\bar{M}_{\mathrm{DM}}$ enclosed in $r_*$, max for those with $d > R_{200,\mathrm{DM}}$ | $8.0 \times 10^4 M_\odot$ | $3.4 \times 10^4 M_\odot$ |

**Note.** First row: Objects located outside the virial radius of a detected host halo. Second row: Objects located outside of the virial radius of a detected host halo, and whose enclosed dark matter mass within $r_{\mathrm{max}}$ is greater than the nominal resolution limit of 300 particles. Third row: The average enclosed dark matter mass for objects in the first category. The counts shown are for the gas + stars, $z = 12$ catalog.

halo is greater than the virial radius of that halo. These may be objects outside of any halo or in a nondetected halo. The table additionally shows the number of objects whose enclosed dark matter mass is greater than the resolution threshold of the dark matter catalog (300 particles). This is a stringent criterion that can identify massive host halos that are disrupted enough as to be nondetected. Because the enclosed mass typically underestimates the total halo mass (see, e.g., Figure 7), this strict cutoff is likely a strong underestimate.

### 5.2. Comparison with JWST Observations

As mentioned in Section 2, an accurate accounting of the density of star clusters requires a search for baryonic particles. With the structures in hand, we continue with an estimation of their compactness. For the purposes of the analytic arguments, we assumed a constant density, using $r_{\mathrm{max}}$ to estimate density (see Figure 8, bottom left panel). However, to estimate the compactness and cuspiness of these stellar systems, we also compute the half mass radius $r_{\mathrm{h}}$. This is the distance of the particle which encloses half the mass of the star cluster. Comparing $r_{\mathrm{max}}$ to $r_{\mathrm{h}}$ in Figure 8, we see that the objects are certainly not of uniform density. This contributes to our simulated timescales scattering outside of the error bars in Figure 6. Thus, to calculate the surface density of the inner star cluster, which is most likely to be found by high-redshift observations given the steep relation between surface brightness and redshift, we use $r_{\mathrm{h}}$ (see Figure 8 right panel).

In Figure 9, we plot the surface density of star clusters in the Stars+Gas primary run against their half mass radius. Gray lines show the constant density relation for a given mass. We overplot recent JWST observations in the large neon points—clusters from the Firefly Sparkle, Cosmic Gems, Sunrise, and Sunburst lensed arcs (E. Vanzella et al. 2022a, 2023; A. Adamo et al. 2024; L. Mowla et al. 2024). (We are limited by resolution on the left side of the plot. This is denoted by the red region, which shows radii smaller than the scale of gas cells in our simulation, whereas the green region is well resolved.) Our points generally lie in the same region of parameter space as these unusual, high surface density systems. This is emphasized by the gray contours, showing local young massive clusters (YMCs; T. M. Brown et al. 2014) and local globular clusters (GCs; M. Gieles et al. 2011), which fall in the lowest surface density regions of the plot. While some of our low surface density points lie in this region, we expect that inclusion of feedback processes would be required to reproduce the densities of these populations. The agreement between our simulated points and the high-$z$ systems leads us to suggest that within ΛCDM, star formation at the observed densities is possible given a weak-feedback scenario, although further modeling is needed to draw a direct connection between our simulations and the observed population. We note that had we only investigated our simulations from a DM halo perspective, taking the surface density across the entire halo, our simulations would not agree with these points.

We additionally compare our simulated star clusters at $z = 12$ to a sample of star clusters observed with JWST (E. Vanzella et al. 2022, 2023; A. Claeyssens et al. 2023, 2025; A. Adamo et al. 2024; S. Fujimoto et al. 2024; L. Mayer et al. 2025; L. Mowla et al. 2024; M. Messa et al. 2025). The observed cluster surface density increases with redshift. The median of our simulated data set aligns slightly above the median observed by JWST at $z > 8$. Our systems have a lower density tail that reaches a much lower surface density than the whiskers of the observed distribution. However, we note that these low-density systems would almost certainly not be observed at high redshift. Furthermore, these low-density systems may be the most susceptible to feedback processes, and may not form in a full-feedback simulation.

## 6. Discussion and Conclusions

In this work, we present star clusters in a cosmological simulation, grouped using structure-finding methods focused on multiple particle types. We compare the properties of clusters grouped by the typical dark matter-centric halo-finding algorithm with a version focused on gas and star particles aimed at mimicking an observational approach.

In general, the star clusters produced by our simulation are of comparable size and surface density to observed systems at high redshift, although we do not have the resolution to accurately model clusters below the pc scale. An important caveat is that although our simulation does not include stellar feedback, feedback processes should be important in shaping the distribution of clusters that form. For this reason, we propose that these simulations are an upper limit for the most extreme star formation permitted for Λ CDM Pop III stars. Our simulated systems may have similar properties to observed systems which form in high star formation efficiency cases. Using the methods available here, it is difficult to separate the role of feedback from other physical processes in shaping the high stellar densities in this era. High densities may be attributed to weak feedback, efficient cooling, small sizes of baryonic objects, and possibly other factors. Notably, in the standard ΛCDM cosmology, the evolution of dark matter halos naturally (and unavoidably) provides high densities at high redshifts (e.g., M. Boylan-Kolchin 2025). We speculate that the inclusion of feedback may have several effects. First, low-density systems may be disrupted, removing the low-density tail of systems (seen in the violin plot of Figure 10). Furthermore, the injection of kinetic energy and momentum provided by feedback could increase the virial radius of clusters, thus playing an important role in their internal properties. Additionally, it is possible that the most massive and high-density clusters (nonvirialized systems that lie around the $10^7 M_\odot$ line in Figure 9) may be prevented from reaching such high masses. This could potentially move these systems toward slightly lower mass in the figure, perhaps driving them to emulate the Cosmic Gems systems. However, given recent studies suggesting that at $\sigma > 10^3 M_\odot\,\mathrm{pc}^{-2}$ the efficient conversion of gas into stars is inevitable (e.g., M. Y. Grudić et al. 2018;

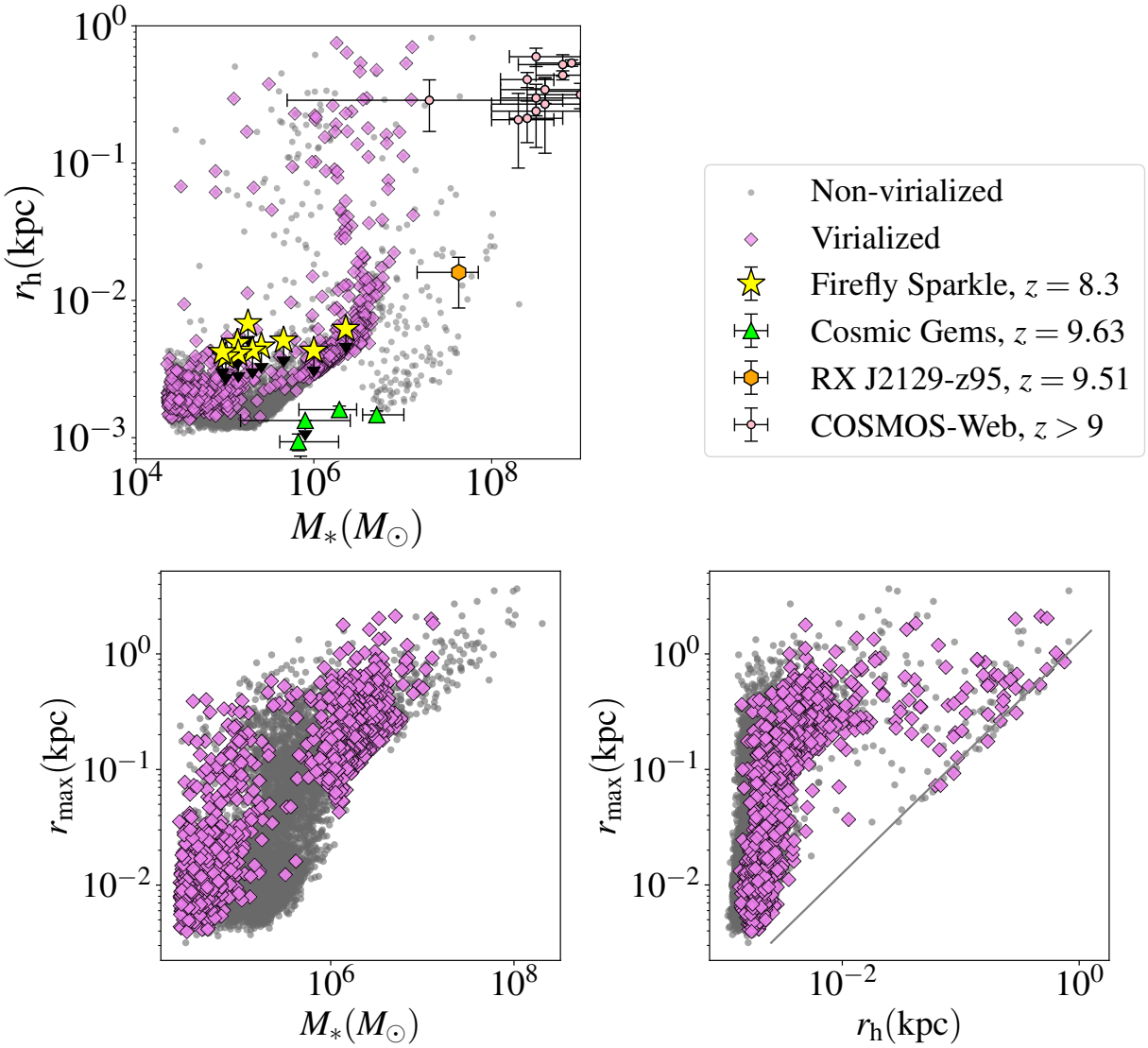


**Figure 8.** Top panel: Half mass radius vs. stellar mass. Gray points are nonvirialized objects, and violet diamonds are virialized objects. Bottom left panel: Maximum radius vs. stellar mass. Bottom right panel: Maximum radius vs. half mass radius. The gray line shows the expected relation for a sphere of uniform density. All points shown are for stars+gas primary FOF run, at $z = 12$. Compact systems from high-redshift observations are overplotted on the top panel according to the legend in the top right (H. Williams et al. 2023; A. Adamo et al. 2024; M. Franco et al. 2024; L. Mowla et al. 2024; A. Adamo et al. 2025, in preparation).

S. H. Menon et al. 2023), we suggest that this remains uncertain. Finally, if strong feedback processes were implemented, this could lead to a reduced occurrence of star clusters outside of halos, thus lessening the discrepancy between the DM halo-finding algorithm and the baryonic version. Thus, we highlight that additional feedback simulations or modeling should be used to determine the role of feedback to regulate various elements of the population, including abundance (especially at the low-mass/low-density end) and continued survival.

Additional physical processes at play at these high redshifts may shape the population in ways not considered by this work. For example, the primordial baryon-dark matter streaming velocities in ΛCDM should delay star formation, advect gas from primordial halos, and shape the rotation and geometry of collapsing gas clouds in many regions of the Universe (e.g., S. Hirano et al. 2023; C. E. Williams et al. 2023). This simulation should accurately represent regions of the Universe with a $0\sigma_{bc}$ or very weak fluctuation, but in high-$v_{\rm bc}$ regions, we acknowledge that dense star clusters may struggle to form. Furthermore, our simulation does not include the effects of magnetic fields on collapsing gas clouds. A more detailed investigation would account for the role of magnetic fields to support primordial gas as it collapses into star clusters (potentially also receiving a boost from strong stream velocities, e.g., S. Naoz & R. Narayan 2013).

Clusters at Cosmic Dawn serve as the building blocks of early galaxies and may later evolve into galactic nuclei or GCs. While we do not explore the chemical properties of these clusters or their evolution to lower redshifts in this work, their potential role in shaping early galaxy populations remains significant. Lensed, high-redshift clusters are often inferred to have young ages (A. Adamo et al. 2024, e.g.,), particularly when compared to the time elapsed between our simulations at $z = 12$. These observed Population II systems likely represent a later generation of stars, whereas our simulated clusters correspond to an earlier population present at $z \geqslant 12$. If the weak-feedback, low-stream-velocity scenario explored here applies at high redshifts, the clusters we study may influence broader properties of the early

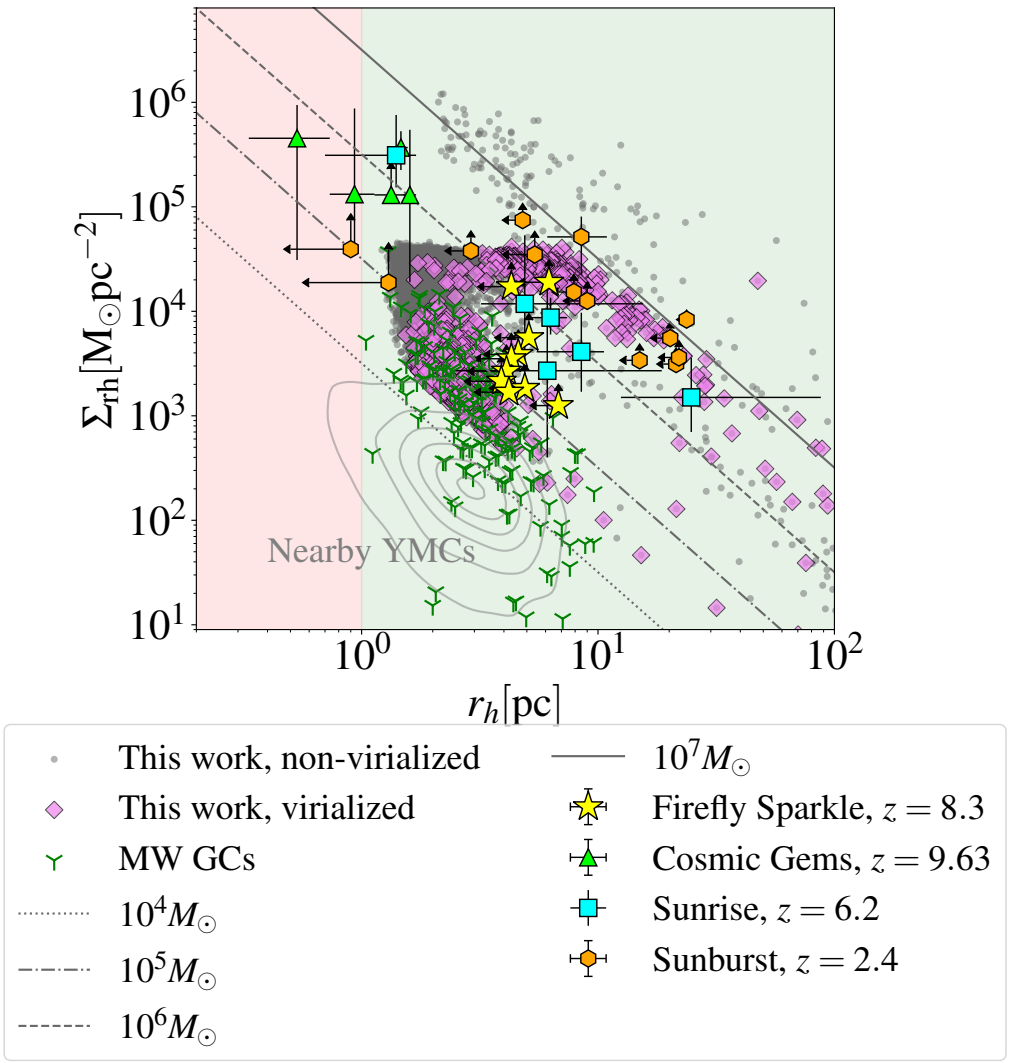


**Figure 9.** Surface density within half mass radius vs. half mass radius of simulated objects, colored by the logarithm of the baryonic density. Gray lines show the uniform density relations for various masses. The points shown are star+gas primary objects at $z = 12$. Additionally, JWST high-redshift clusters are shown for the Firefly Sparkle (L. Mowla et al. 2024), Cosmic Gems (A. Adamo et al. 2024, and private communication), Sunrise (E. Vanzella et al. 2023), and Sunburst (E. Vanzella et al. 2022a) lensed arcs. Local Milky Way GCs (M. Gieles et al. 2011) are shown in green, while the gray contours show local YMCs in nearby galaxies ( T. M. Brown et al. 2014). The green region on the right shows systems that are resolved in the simulation, whereas the red shaded region signifies the parameters space that cannot be explored in our simulation due to resolution.

galaxy population—even if individual clusters are not directly observed at $z = 12$. Through mergers, accretion, and continued growth, these systems become constituent components of JWST's highest redshift galaxies. The extreme stellar densities of these early clusters could contribute to unusual chemical enrichment patterns (e.g., C. Charbonnel et al. 2023). Moreover, S. H. Menon et al. (2025) demonstrate that high surface density star-forming clumps create conditions favorable for Lyman continuum photon escape during the lifetimes of massive stars, suggesting that these clusters may play a role in driving cosmic reionization.

Finally, these dense star clusters serve as ideal sites for early black hole formation in the intermediate mass regime. If the recoil kick experienced during a black hole merger is lower than the system's escape velocity (which may often be the case given the low spin reported by LIGO/Virgo; e.g., B. P. Abbott et al. 2019; R. Abbott et al. 2023a, 2023b), the remnant will remain in the cluster and has the chance to interact and merge again (e.g., C. L. Rodriguez et al. 2019). Through a gravitational runaway process, this may result in intermediate mass black holes (IMBHs) of $10^{3-5} M_\odot$ (e.g., S. F. Portegies Zwart & S. L. W. McMillan 2002; S. Hirano et al. 2014; Y. Sakurai et al. 2017, 2019). Additionally, if these clusters merge into a larger protogalaxy, dynamical friction could facilitate the formation of early overmassive supermassive black holes (e.g., A. Dekel et al. 2025; L. Mayer et al. 2025). This population of IMBHs at high redshift would provide sites of dynamical interaction that could lead to stellar mass binary black hole mergers and extreme mass ratio inspirals (e.g., S. J. Aarseth 2012; S. Naoz et al. 2022; M. Arca Sedda et al. 2023; Z. Xuan et al. 2024), as well as IMBH mergers and wandering black holes through the interaction of star clusters with larger halos and galaxies (see e.g., J. E. Greene et al. 2020). This would imply many promising avenues of detection through next generation gravitational wave detectors.

*Summary*. Our main results are as follows:

1. When tracking star clusters, rather than dark matter halos, a discrepancy is found in the number counts by

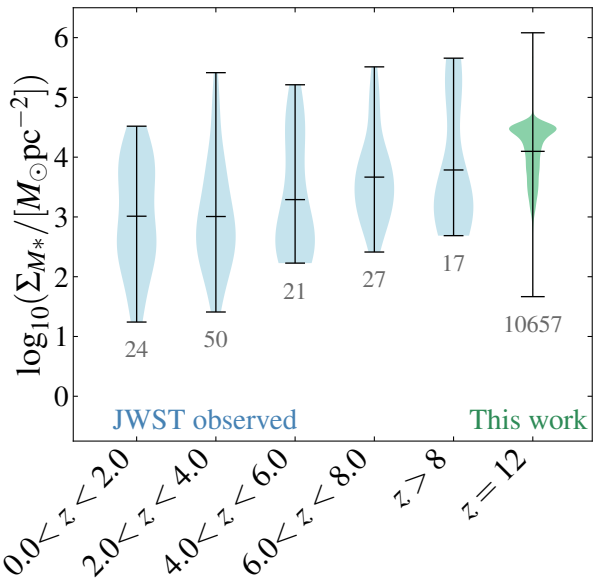


**Figure 10.** Violin plots of surface density of JWST observed star clusters (blue, left side) compared to our simulated sample (green, right side). The mean, minimum, and maximum of each distribution are shown in the whiskers. For observed systems, clusters with effective radius <25 pc were selected. The simulated structures were filtered for those with $M_* > 10^{4.368} M_\odot$, the minimum mass of the observed sample, and radius <25 pc, to match the cutoff in assembling the observed sample. Below each distribution, the gray number is the number of objects included in the distribution. The sample includes clusters from L. Mowla et al. (2024), A. Claeyssens et al. (2025), M. Messa et al. (2025), A. Adamo et al. (2024) and private communication; E. Vanzella et al. (2022b), A. Claeyssens et al. (2023), E. Vanzella et al. (2023), and S. Fujimoto et al. (2024).

stellar mass and star formation rate up to a factor of several at all masses. In these cases, studies should consider which method most benefits the observable quantity of interest. Often, the star primary or star and gas primary algorithm will identify concentrations of starlight, which are useful for observational comparison.

2. We study the virialization of star clusters in our simulation in detail, identifying the processes that lead systems at a given mass and initial density to equilibrate or not. For high-redshift clusters and galaxies, achieving virialization erases the system's initial conditions in the hierarchical process of galaxy buildup. We show that many systems at the redshifts considered here are not virialized.
3. In particular, we find two populations of virialized clusters originating from the tracks of molecular cooling of primordial metal-free gas. They correspond to roughly $10^4$–$10^5 M_\odot$ and $10^6$–$10^7 M_\odot$. We show how hierarchical mergers disrupt these objects, providing an analytical framework for virialization and disruption.
4. We use our star and gas primary cluster searches to explore the stellar surface density of high-redshift clusters. Our results are consistent with JWST's highest surface density clusters. Since the simulation suite investigated here does not include stellar feedback, this could imply that these JWST clusters formed in a low-feedback mode. For example, our results would be consistent with "feedback-free" star clusters (e.g., A. Dekel et al. 2023). In any case, the clusters shown here represent a theoretical upper limit for extreme star formation permitted within ΛCDM Pop III stellar systems.

In future investigations, we plan to work toward an improved physical picture by including more relevant physical effects, as discussed above. This will allow us to determine the sensitivity of the findings presented in this paper to these potentially disrupting effects. Moreover, in addition to this improved physical modeling, we hope to extend the observational comparison through the synthesis of mock observables.

This work highlights the importance of detailed investigation of ΛCDM simulations to uncover the relationship between simulated structures and observed systems and hints that a weak-feedback mode may be present in the early Universe.

## Acknowledgments

The authors thank the anonymous reviewer for their feedback and suggestions to improve this work. C.E.W. acknowledges the support of the National Science Foundation Graduate Research Fellowship, the University of California, Los Angeles (UCLA), and the UCLA Center for Developing Leadership in Science Fellowship. C.E.W., W.L., S.N., Y.S.C, B.B., F.M., and M.V. thank the support of NASA grant Nos. 80NSSC20K0500 (9-ATP19-0020) and 80NSSC24K0773 (ATP-23- ATP23-0149) and the XSEDE/ACCESS AST180056 allocation, as well as the UCLA cluster Hoffman2 for computational resources. B.B. also thanks the Alfred P. Sloan Foundation and the Packard Foundation for support. N.Y. acknowledges financial support from JSPS International Leading Research 23K20035. F.M. acknowledges support by the European Union—Next Generation EU within PRIN 2022 project n.20229YBSAN—Globular Clusters in Cosmological Simulations and in Lensed fields: From Their Birth to the Present Epoch. Simulation runs for this work were performed on the Anvil Cluster (X. C. Song et al. 2022). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under grant No. DGE-2034835. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This work used computational and storage services associated with the Hoffman2 Cluster, which is operated by the UCLA Office of Advanced Research Computing's Research Technology Group.

## Data Availability

A catalog of the star clusters and dark matter halos used in this work is available for public use: arepo-clusters[16] with a copy deposited to Zenodo: DOI: 10.5281/zenodo.15392999 (C. E. Williams 2025a). The code to postprocess the FOF runs is also available on GitHub: process-fof[17] with a copy deposited to Zenodo: DOI: 10.5281/zenodo.15701631 (C. E. Williams 2025b).

*Software*: astropy (Astropy Collaboration et al. 2013, 2018), yt (M. J. Turk et al. 2011), matplotlib (J. D. Hunter 2007), numpy (C. R. Harris et al. 2020), scipy (P. Virtanen et al. 2020).

## Appendix A
## Convergence Tests

*Convergence of Simulation FOF Algorithm*. In order to ensure convergence and assess the robustness of our structure-finding algorithm we test for convergence by varying the

[16] https://www.astro.ucla.edu/~clairewilliams/cluster-catalog
[17] https://github.com/astro-claire/process-fof.git

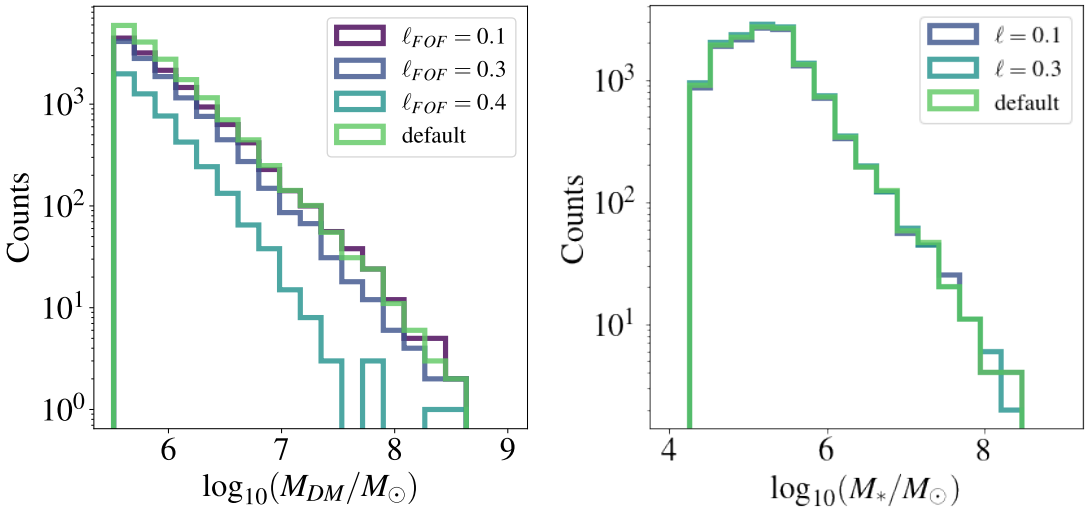


**Figure 11.** Left: Counts of objects by DM mass in DM primary FOF run with varying linking lengths ($\ell_{\rm FOF}$). Right: Counts of objects by stellar mass in the star primary run, colored by three versions where the linking length is varied.

linking length used in the FOF algorithm. During the particle linking stage, two particles are identified as group members if they are closer to one another than $\ell_{\rm FOF}$ times the mean interparticle separation. By default, we set $\ell_{\rm FOF} = 0.2$.

We reran our DM-primary, stars+gas secondary structure-finding algorithm with three additional values of the linking length: $\ell_{\rm FOF} = 0.1$, 0.3, and 0.4. The counts by dark matter halo mass are plotted in the left panel of Figure 11. The simulated number of objects is consistent for variations within 0.1 of our value, and starts to diverge when the linking length is varied by a factor of 2. The variations between these convergence runs are much smaller than the error found between the DM-centric and baryon-centric methods, so we can confirm that our simulated discrepancies are not due to convergence issues. We also show the results of a similar convergence testing exercise for the star primary algorithm in the right panel of Figure 11. As demonstrated by the agreement between the histograms, the star primary run also exhibits good convergence despite variation of the linking-length parameter.

*Choice of Virialization Ratio*. Given that our choice of virialized range from $1.5 < -U/K < 2.5$ is somewhat arbitrary, in this appendix we seek to test whether or not the analysis of our population of virialized objects is sensitive to the choice of virialization ratios. First, we show the distribution of virial ratios in Figure 12. There is no obvious population of virialized objects; objects lie in a smooth distribution with a peak slightly above virialized. The vertical solid lines show our choice of the definition of virialized for the above analysis. We then plot the same distributions as shown in Figure 4, using two different cuts on the data, shown in the dotted and dashed lines in Figure 12. As sown in Figure 13, these are a looser definition of $1.25 < -U/K < 2.75$ (dashed) and a stricter definition of $1.75 < -U/K < 2.25$ (dotted). From the figure, which shows histograms by stellar mass of virialized objects, we can see that the presence and location of the two populations of objects does not depend on the choice of virial ratio cutoff.

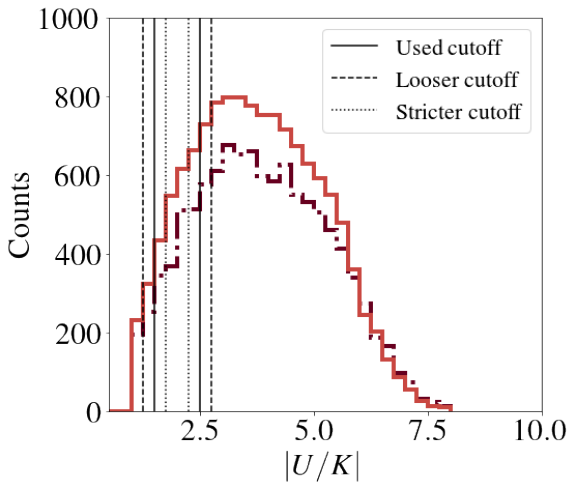


**Figure 12.** Histogram of virial ratios in the baryonic-focused catalogs. The line convention denoting the two baryonic FOF runs is the same as Figure 3 and the subsequent figures, with the solid line denoting the star primary and the dotted–dashed line denoting the stars and gas primary.

In order to have confidence in our determination of the virialization or nonvirialization of stellar clumps, we inspect our systems and virialized systems to ensure that low-mass systems do not suffer from resolution effects. In particular, the lowest-mass star clusters near our resolution limit of 100 star particles may be particularly susceptible to resolution errors. First, we test for a dependence of the scatter in virial ratios on the number of particles included in the system. In the left panel of Figure 14, we show the virial ratios versus the number of particles per object for the stars + gas primary catalog at $z = 12$. The orange line shows the standard deviation (which is also depicted by the blue-shaded region around the mean). Focusing on the low-mass end, the scatter remains roughly constant at a value of $\sim$1 over an order of magnitude in particle number. We further inspect the distribution of virialized objects in the right panel of Figure 14, which shows the subset of objects in the run that are virialized, binned by

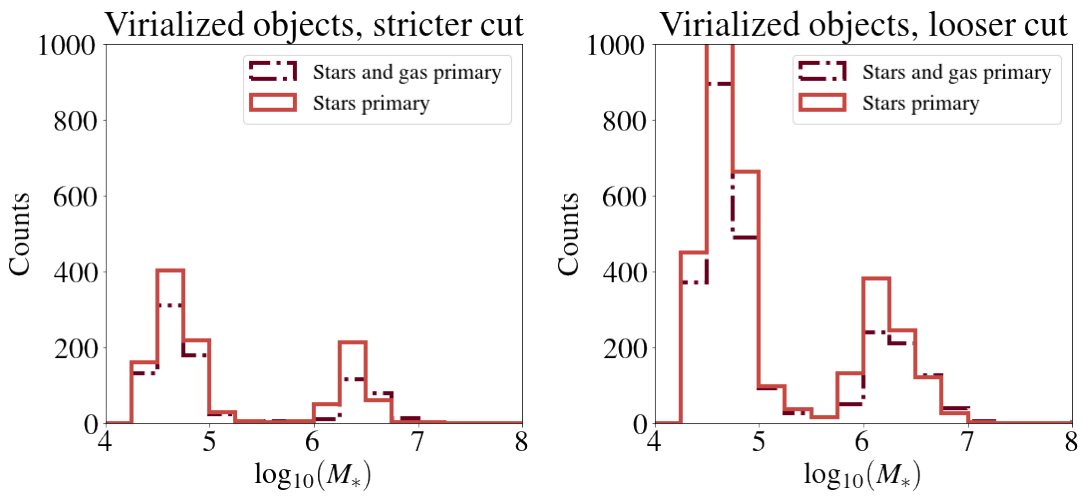


**Figure 13.** Same as Figure 4, but using the looser and stricter virial ratio cutoffs shown in Figure 12 by the dashed and dotted lines. The left panel shows the stricter cut ($1.75 < -U/K < 2.25$), and the right panel shows the looser cut ($1.25 < -U/K < 2.75$). The line convention denoting the two baryonic FOF runs is the same as Figure 3 and the subsequent figures.

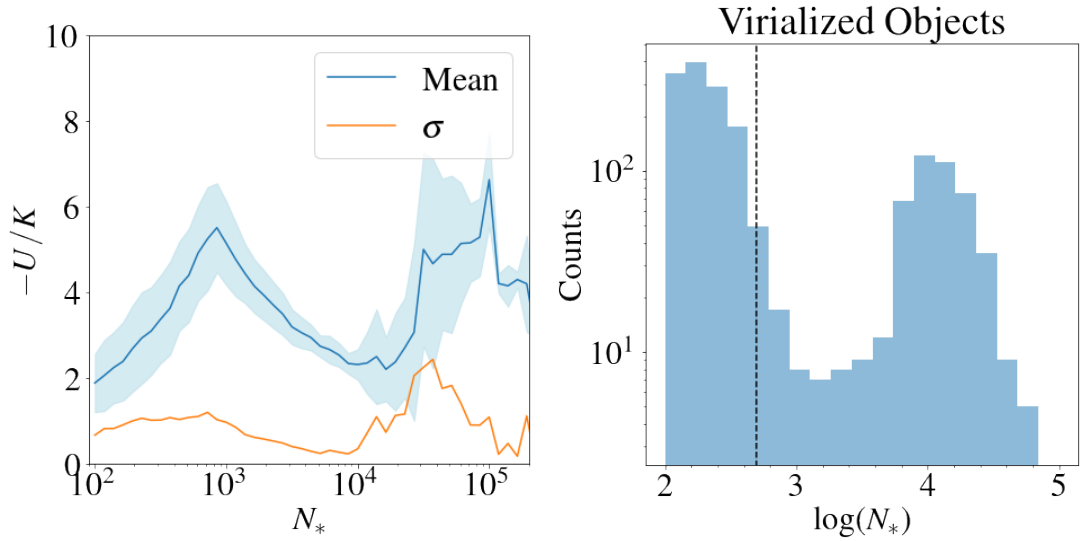


**Figure 14.** Left: Mean virial ratio vs. number of star particles per object for the stars + gas primary run at $z = 12$ (blue). The orange curve denotes the standard deviation ($\sigma$). Additionally, the shaded region shows the one $\sigma$ boundaries of the distribution. Right: Histogram of virialized star clusters binned by the logarithm of the number of star particles. Lowest-mass clusters have $10^2$ particles, our resolution limit. The dashed vertical line shows $N_* = 500$ for reference.

their number of star particles. In particular, we want to ensure that even if the population of the very lowest $N_*$ objects contains some unrelaxed systems, the trend of two virialized populations is robust. To guide the eye, we include a vertical line at $N_* = 500$. From this figure, we can see that although the bulk of the low-mass peak has particle number $N_* < 500$, the second peak begins to emerge even for systems with 500 $-1000$ star particles, which should be well resolved. Thus, we find a similar result even if the resolution threshold is increased by a factor of five.

## Appendix B
## Cluster Mass Function

In this appendix, we provide the cumulative cluster mass function for structures in our simulation box. Since we have increased $\sigma_8$ in our simulation (see discussion in W. Lake et al. 2023b, Section 2), this mass function is only applicable to high-density peaks in the Universe. Therefore, our structures are more numerous than the average patch of the Universe with similar volume. The mass function is given in Figure 15.

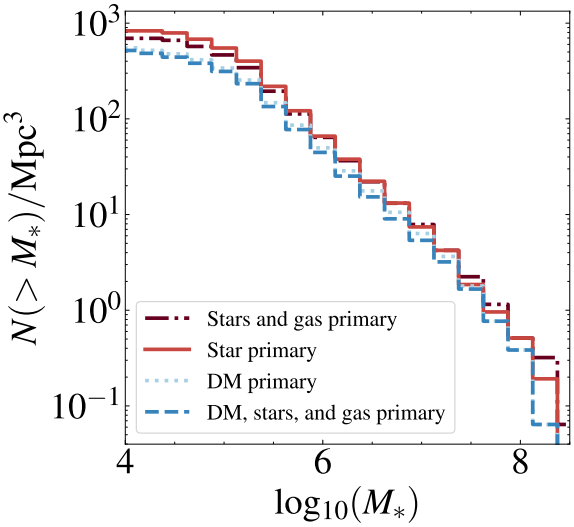


**Figure 15.** Cumulative mass function of star clusters in the simulation. The volume has not been corrected for increased $\sigma_8$, so the provided data represents a very high -ensity peak in the Universe. Results are presented at $z = 12$. The line styles follow the convention defined in Figure 3.

## Appendix C
## Ancillary Data for Dynamical Investigation

### *C.1. Analytic Timescales*

In Figure 16, we show a more detailed comparison between the timescale for virialization and the timescales for interaction. This figure shows how the upper and lower limit error bars in Figure 6 were derived. In both panels, the virialization timescale is shown in purple, as in Figure 6, along with the cooling and dynamical time for each density (see Equation (9)). The black horizontal line shows the age of the Universe. The color bar denotes the mass of the interacting halo (used in the number density calculation, Equation (13)).

On the left, we show upper limit case for interaction—this is if we use the average density of halos in the local region to calibrate the number density. This is almost certainly an

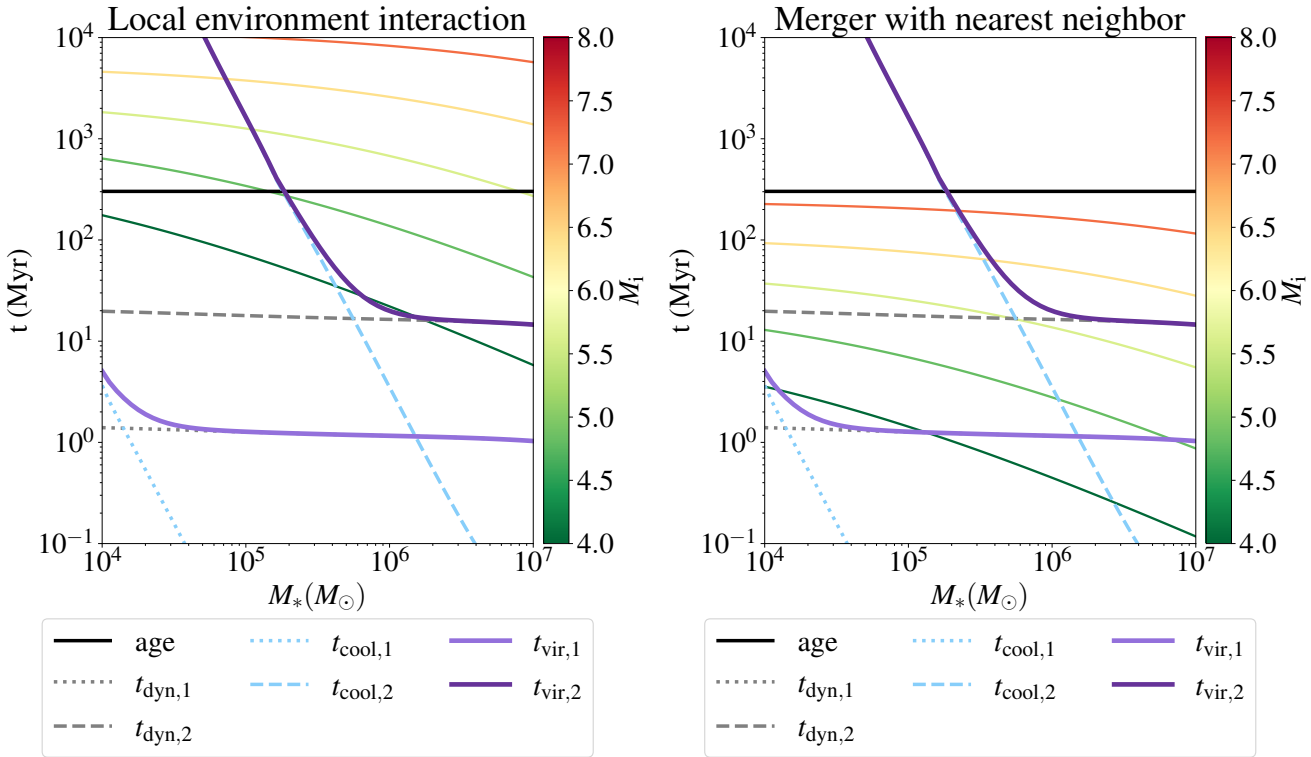


**Figure 16.** Comparison of analytic timescales for virialization (purple lines) and interaction (tan line). The blue lines show the $H_2$ cooling timescale for the two typical densities discussed in Section 4.2. The gray lines show the dynamical timescales for those densities. The solid black line is the age of the Universe at $z = 14$. On the left-hand side, the color bar separates lines of interaction timescale by interacting halo mass, with the timescale calculated for the average nearby environment (see Equation (11) and discussion in text). On the right-hand side, the interaction timescales are instead calculated as if the interacting halo mass was at the distance of the nearest neighbor in the simulation, providing a lower limit.

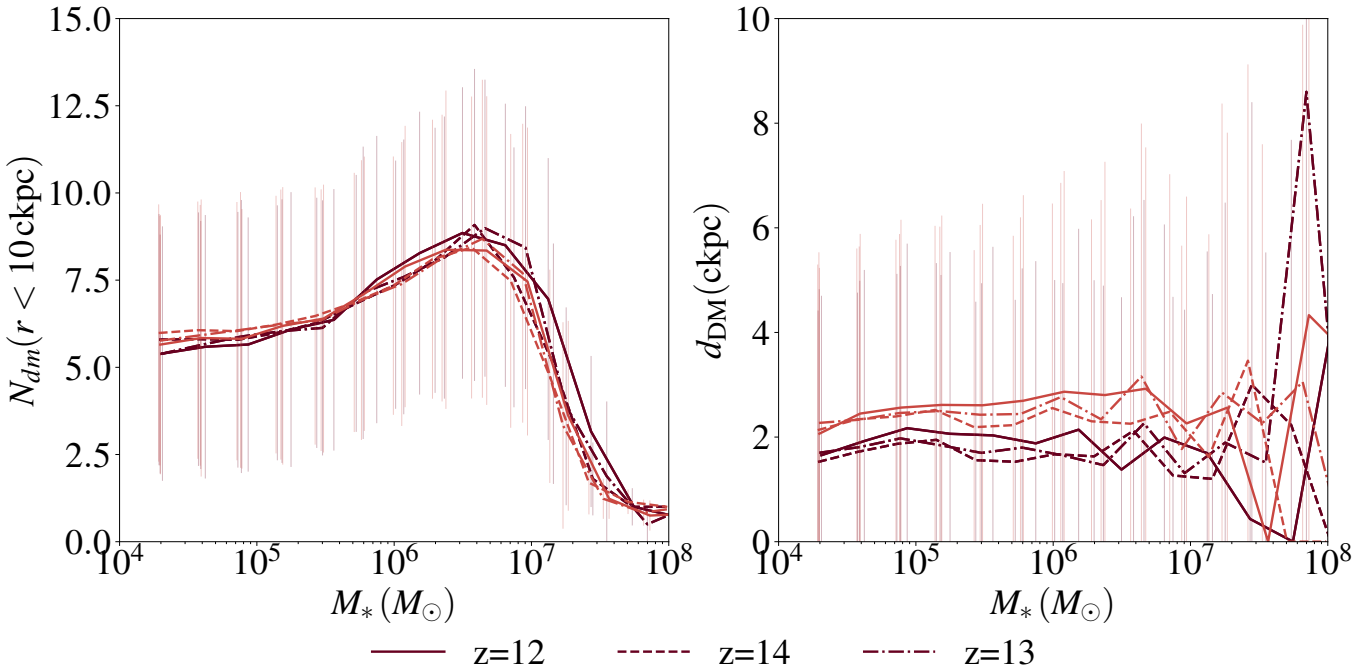


**Figure 17.** Left panel: Number of DM neighbors within a radius of 10 comoving kpc as a function of stellar mass for various redshifts. The line styles denote the various redshifts of the simulation snapshots, while the lighter red is star primary and the black is star + gas primary (all redshifts and FOF runs are similar). The error bars show one standard deviation, $\sigma$, above and below the mean in each bin. Right panel: Mean distance to the nearest dark matter halo.

overestimate for most systems, since here we assume uniformly distributed halos while real systems are highly clustered. Thus, to provide a lower limit for the interaction timescale, the timescale in the right-hand panel is calibrated to the interaction time of the nearest neighboring halo. This overestimates the interaction timescale in most cases because there is not a constant supply of neighbors at the distance of the closest neighbor and many system's nearest neighbor is much farther than the simulation box average (see Figure 17).

### *C.2. Numerically Derived Quantities: Data and Fitting Functions*

Here, we provide the simulated data used to calibrate the analytic timescales from our investigation.

Figure 17 shows simulated data relating to the dark matter environment of baryon-centric structures. In the left panel, the average number of dark matter halos from the DM-centric catalog located within 10 ckpc of each baryon-centric object is plotted as a function of stellar mass. On the right, the distance to the nearest DM halo is plotted again as a function of stellar mass for the baryon-primary objects. In both panels, the error bars denote one standard deviation above and below the mean in each bin.

Figure 18 gives the average gas density as a function of stellar mass for objects detected in the stars+gas primary catalog. The two populations of virialized objects are highlighted. This average gas density is computed using the mass contained in gas cells within a sphere of radius equal to the maximum galactocentric star particle distance. Reassuringly, the densities estimated for the analytic argument from the high-redshift DM halos (see Table 4) fall roughly within the two typical gas densities in the simulation at this later time.

Next, we show additional histograms of quantities used for the interaction timescale in Figure 19. From left to right, the

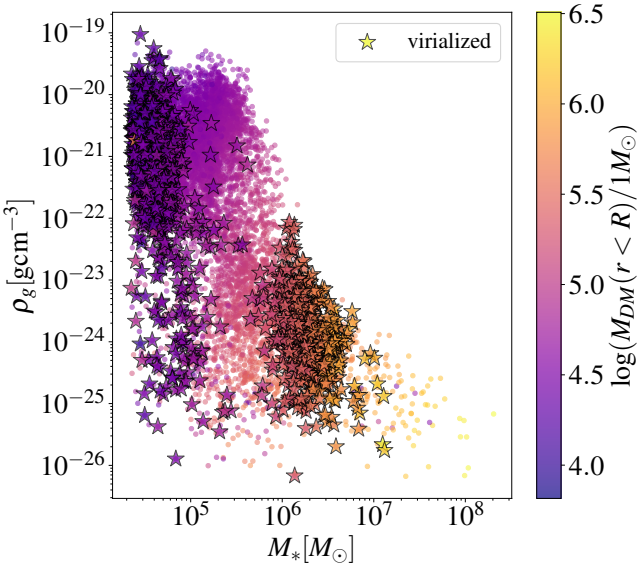


**Figure 18.** Mean gas density vs. stellar mass for simulated objects in the stars +gas primary run at $z = 12$. The color bar shows the dark matter mass within the radius of the object.

panels show the mass of the closest dark matter object, the group velocity with respect to the bulk flow, and the number of dark matter neighbors within 10 ckpc. The populations are split into nonvirialized and virialized. In general, the virialized and nonvirialized populations trace each other. However, there is a dearth of virialized objects with a large number of DM neighbors, and a smaller fraction of virialized objects with the most massive DM neighbors. This aligns with the analytical expectation—a large number of neighbors or an extremely massive neighbor is more likely to disrupt a system's state of equilibrium.

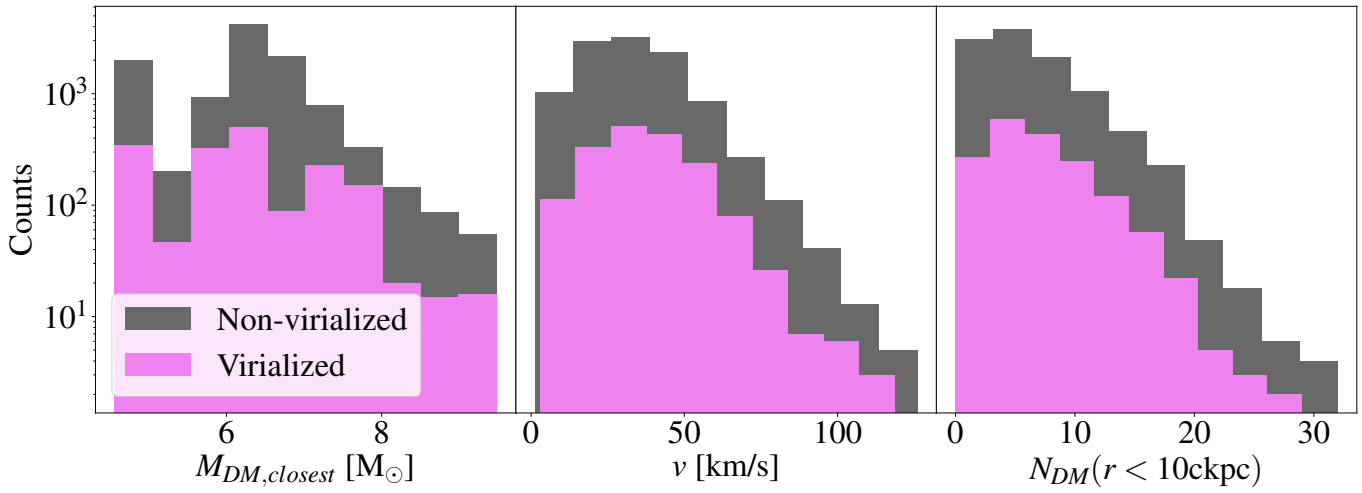


**Figure 19.** Left panel: Mass of nearest neighbor in solar masses. Central panel: Group velocity with respect to the bulk flow. Right panel: Number of DM halos within 10 ckpc, as traced by the DMP FOF run. Gray histograms represent the nonvirialized objects, and violet represent the virialized objects. All results shown are for the stars+gas primary run at $z = 12$. Mean values are listed in Table 4.

*$M_*$–$M_{\rm DM}$ relation*. We find a good fit to the simulated data with $R^2 > 0.91$ up to $M_* \leqslant 10^7 M_\odot$ using the following fit:

$$\begin{aligned}\log M_{\rm DM} = &-0.01584(\log M_*)^4 \\ &+ 0.308989(\log M_*)^3 - 2.03908(\log M_*)^2 \\ &+ 5.61641 \log M_* + 0.19979,\end{aligned} \tag{C1}$$

where $\log M_* = \log 10\left(\frac{M_*}{1M_\odot}\right)$ and $\log M_{\rm DM} = \log 10\left(\frac{M_{\rm DM}}{1M_\odot}\right)$.

*$M_*$–$M_{\rm g}$ relation*. Additionally, we use the following fit to model the stellar mass to gas mass relation:

$$\begin{aligned}\log M_{\rm g} = &-0.020562(\log M_*)^4 \\ &+ 0.402479(\log M_*)^3 - 2.717265(\log M_*)^2 \\ &+ 7.688866 \log M_* - 2.688908,\end{aligned} \tag{C2}$$

where $\log M_{\rm g} = \log 10\left(\frac{M_{\rm g}}{1M_\odot}\right)$. This provides $R^2 > 0.83$ for all masses below $M_* \leqslant 10^7$ These fits are used to estimate the halo and gas mass as a function of stellar mass for the purpose of deriving the virialization timescale.

## Appendix D
## Results at Various Redshifts

In this appendix, we provide results at $13 \leqslant z \leqslant 15$. Figures 20 and 21 give the overall histograms of objects by stellar mass as in Figure 3. Figure 22 shows the virialized population by stellar mass at $z = 13$, 14, and 15, as in Figure 4. The trends discussed in Section 3 are consistent at all the redshifts investigated here.

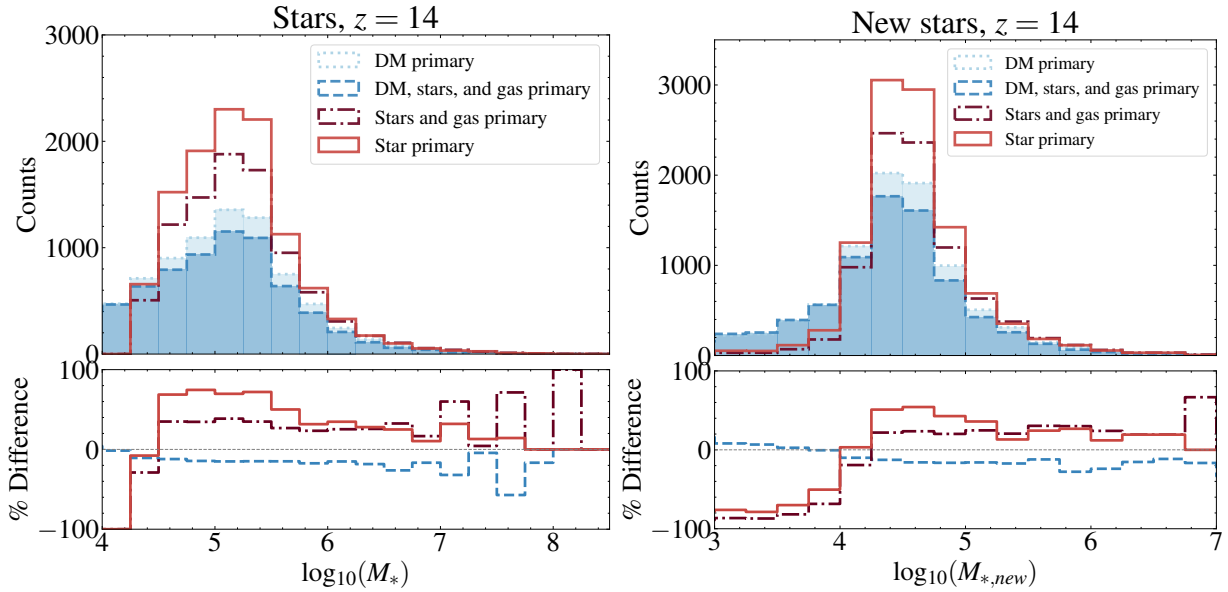


**Figure 20.** Top left panel: hHtogram of object counts by stellar mass at $z = 14$ using algorithms centered on the particles listed in Table 1. Bottom left panel: Percent difference between each FOF run and the standard DM-primary FOF ($f = N/N_{\rm DMP}$). Gray horizontal line shows $f = 1$. Blue shades are runs that include DM—the DM primary (dotted) and the stars, gas, and DM primary (dashed). The red shades are baryonic runs—the star primary (solid) and the stars and gas primary (dotted–dashed). Top right panel: Histogram of object counts by newly formed stellar mass at $z = 12$ using FOF primaries listed in Table 1. Bottom right panel: Percent difference between each FOF run and the standard DM-primary FOF ($f = N/N_{\rm DMP}$). Gray horizontal line shows $f = 1$.

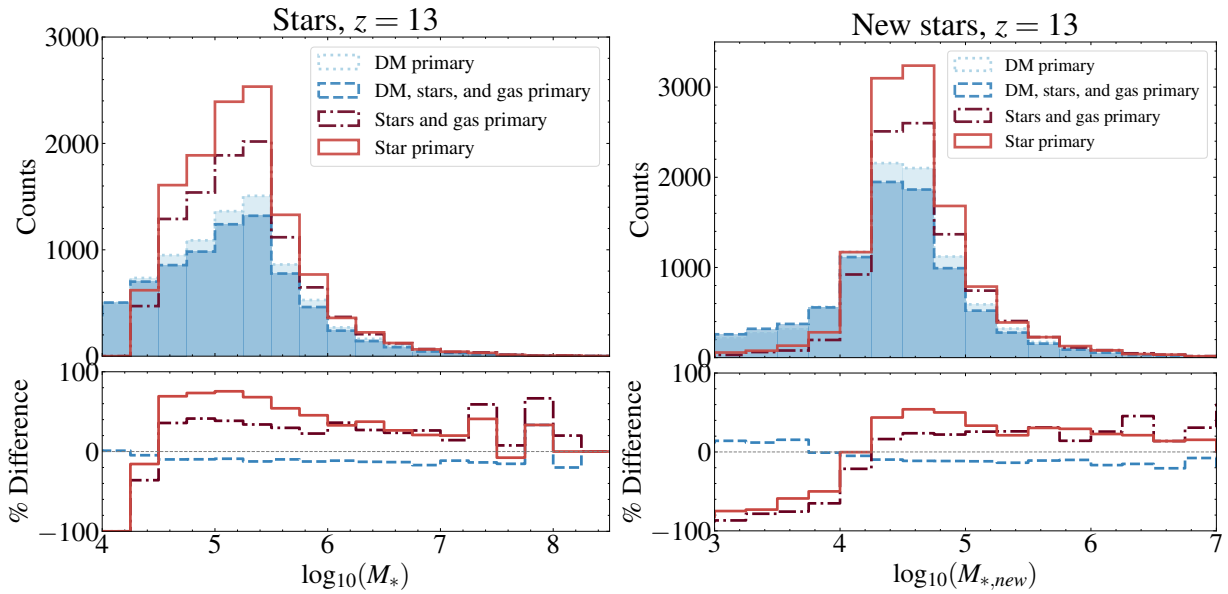


**Figure 21.** Top left panel: Histogram of object counts by stellar mass at $z = 13$ using algorithms centered on the particles listed in Table 1. Bottom left panel: Percent difference between each FOF run and the standard DM-primary FOF ($f = N/N_{\rm DMP}$). Gray horizontal line shows $f = 1$. Blue shades are runs that include DM—the DM primary (dotted) and the stars, gas, and DM primary (dashed). The red shades are baryonic runs—the star primary (solid) and the stars and gas primary (dotted–dashed). Top right panel: Histogram of object counts by newly formed stellar mass at $z = 12$ using FOF primaries listed in Table 1. Bottom right panel: Percent difference between each FOF run and the standard DM-primary FOF ($f = N/N_{\rm DMP}$). Gray horizontal line shows $f = 1$.

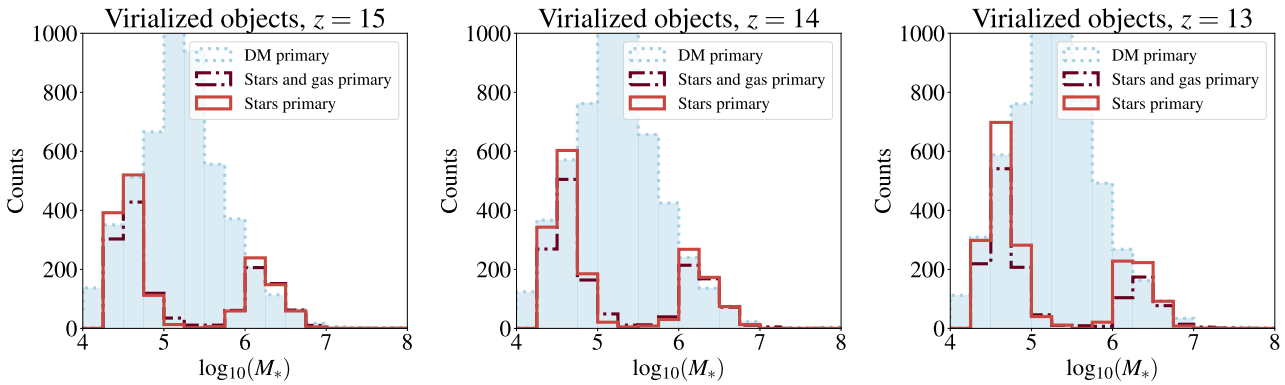


**Figure 22.** Number of virialized objects at $z = 15$, 14, and 13 (from left to right) by stellar mass in solar masses. Light blue is the DM primary (dotted). The red shades are baryonic runs—the star primary (solid) and the stars and gas primary (dotted–dashed).

## ORCID iDs

Claire E. Williams https://orcid.org/0000-0003-2369-2911
Smadar Naoz https://orcid.org/0000-0002-9802-9279
William Lake https://orcid.org/0000-0002-4227-7919
Blakesley Burkhart https://orcid.org/0000-0001-5817-5944
Federico Marinacci https://orcid.org/0000-0003-3816-7028
Mark Vogelsberger https://orcid.org/0000-0001-8593-7692
Naoki Yoshida https://orcid.org/0000-0001-7925-238X
Shyam H. Menon https://orcid.org/0000-0001-5944-291X
Avi Chen https://orcid.org/0000-0002-8859-7790
Angela Adamo https://orcid.org/0000-0002-8192-8091

## References

Aarseth, S. J. 2012, MNRAS, 422, 841
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019, ApJ, 882, L24
Abbott, R., Abbott, T. D., Acernese, F., et al. 2023a, PhRvX, 13, 041039
Abbott, R., Abbott, T. D., Acernese, F., et al. 2023b, PhRvX, 13, 011048
Adamo, A., Bradley, L. D., Vanzella, E., et al. 2024, Natur, 632, 513
Adams, N. J., Conselice, C. J., Ferreira, L., et al. 2023, MNRAS, 518, 4755
Arca Sedda, M., Naoz, S., & Kocsis, B. 2023, Univ, 9, 138
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Atek, H., Shuntov, M., Furtak, L. J., et al. 2023, MNRAS, 519, 1201
Baggen, J. F. W., van Dokkum, P., Brammer, G., et al. 2024, ApJL, 977, L13
Barkana, R., & Loeb, A. 2001, PhR, 349, 125
Boylan-Kolchin, M. 2025, MNRAS, 538, 3210
Brown, T. M., Tumlinson, J., Geha, M., et al. 2014, ApJ, 796, 91
Casey, C. M., Akins, H. B., Shuntov, M., et al. 2024, ApJ, 965, 98
Castellano, M., Fontana, A., Treu, T., et al. 2022, ApJL, 938, L15
Ceverino, D., Glover, S. C. O., & Klessen, R. S. 2017, MNRAS, 470, 2791
Charbonnel, C., Schaerer, D., Prantzos, N., et al. 2023, A&A, 673, L7
Chiaki, G., & Wise, J. H. 2019, MNRAS, 482, 3933
Chiou, Y. S., Naoz, S., Burkhart, B., Marinacci, F., & Vogelsberger, M. 2019, ApJL, 878, L23
Chiou, Y. S., Naoz, S., Burkhart, B., Marinacci, F., & Vogelsberger, M. 2021, ApJ, 906, 25
Chiou, Y. S., Naoz, S., Marinacci, F., & Vogelsberger, M. 2018, MNRAS, 481, 3108
Chon, S., Ono, H., Omukai, K., & Schneider, R. 2022, MNRAS, 514, 4639
Claeyssens, A., Adamo, A., Messa, M., et al. 2025, MNRAS, 537, 2535
Claeyssens, A., Adamo, A., Richard, J., et al. 2023, MNRAS, 520, 2180
Cullen, F., McLeod, D. J., McLure, R. J., et al. 2024, MNRAS, 531, 997
Dekel, A., Sarkar, K. S., Birnboim, Y., Mandelker, N., & Li, Z. 2023, MNRAS, 523, 3201
Dekel, A., Stone, N. C., Chowdhury, D. D., et al. 2025, A&A, 695, A97
Feng, Y., Di-Matteo, T., Croft, R. A., et al. 2016, MNRAS, 455, 2778
Fernandez, R., Bryan, G. L., Haiman, Z., & Li, M. 2014, MNRAS, 439, 3798
Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, ApJL, 946, L13
Finkelstein, S. L., Leung, G. C. K., Bagley, M. B., et al. 2024, ApJL, 969, L2
Franco, M., Akins, H. B., Casey, C. M., et al. 2024, ApJ, 973, 23
Fujimoto, S., Naidu, R. P., Chisholm, J., et al. 2025, The Astrophysical Journal, 989 (1) 46
Fujimoto, S., Ouchi, M., Kohno, K., et al. 2024, arXiv:2402.18543
Galli, D., & Palla, F. 1998, A&A, 335, 403
Gelli, V., Salvadori, S., Pallottini, A., & Ferrara, A. 2020, MNRAS, 498, 4134
Gieles, M., Heggie, D. C., & Zhao, H. 2011, MNRAS, 413, 2509
Greene, J. E., Strader, J., & Ho, L. C. 2020, ARA&A, 58, 257
Greif, T. H., Glover, S. C. O., Bromm, V., & Klessen, R. S. 2010, ApJ, 716, 510
Grudić, M. Y., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2018, MNRAS, 475, 3511
Harikane, Y., Inoue, A. K., Ellis, R. S., et al. 2025, ApJ, 980, 138
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hirano, S., Hosokawa, T., Yoshida, N., et al. 2014, ApJ, 781, 60
Hirano, S., Shen, Y., Nishijima, S., Sakai, Y., & Umeda, H. 2023, MNRAS, 525, 5737
Hirano, S., Yoshida, N., Sakurai, Y., & Fujii, M. S. 2018, ApJ, 855, 17
Hunter, J. D. 2007, CSE, 9, 90
Kannan, R., Smith, A., Garaldi, E., et al. 2022, MNRAS, 514, 3857
Kim, J.-G., Kim, W.-T., & Ostriker, E. C. 2018, ApJ, 859, 68
Knebe, A., Knollmann, S. R., Muldrew, S. I., et al. 2011, MNRAS, 415, 2293
Knebe, A., Libeskind, N. I., Pearce, F., et al. 2013, MNRAS, 428, 2039
Kohandel, M., Pallottini, A., Ferrara, A., et al. 2020, MNRAS, 499, 1250
Lake, W., Grudić, M. Y., Naoz, S., et al. 2025, ApJL, 985, L6
Lake, W., Naoz, S., Chiou, Y. S., et al. 2021, ApJ, 922, 86
Lake, W., Naoz, S., Burkhart, B., et al. 2023a, ApJ, 943, 132
Lake, W., Naoz, S., Marinacci, F., et al. 2023b, ApJL, 956, L7
Lake, W., Williams, C. E., Naoz, S., et al. 2024, ApJ, 973, 115
Ma, X., Hopkins, P. F., Garrison-Kimmel, S., et al. 2018, MNRAS, 478, 1694
Marinacci, F., Sales, L. V., Vogelsberger, M., Torrey, P., & Springel, V. 2019, MNRAS, 489, 4233
Mayer, L., van Donkelaar, F., Messa, M., Capelo, P. R., & Adamo, A. 2025, ApJL, 981, L28
McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565
Menon, S. H., Burkhart, B., Somerville, R. S., Thompson, T. A., & Sternberg, A. 2025, ApJ, 987, 12
Menon, S. H., Federrath, C., & Krumholz, M. R. 2023, MNRAS, 521, 5160
Messa, M., Vanzella, E., Loiacono, F., et al. 2025, A&A, 694, A59
Meštrić, U., Vanzella, E., Zanella, A., et al. 2022, MNRAS, 516, 3532
Mowla, L., Iyer, K., Asada, Y., et al. 2024, Natur, 636, 332
Nakazato, Y., Ceverino, D., & Yoshida, N. 2024, ApJ, 975, 238
Naoz, S., & Narayan, R. 2013, PhRvL, 111, 051303
Naoz, S., Rose, S. C., Michaely, E., et al. 2022, ApJ, 927, L18
Omukai, K., Hosokawa, T., & Yoshida, N. 2010, ApJ, 722, 1793
Pallottini, A., Ferrara, A., Gallerani, S., et al. 2022, MNRAS, 513, 5621
Pandya, V., Zhang, H., Huertas-Company, M., et al. 2024, ApJ, 963, 54
Pillepich, A., Springel, V., Nelson, D., et al. 2018, MNRAS, 473, 4077
Popa, C., Naoz, S., Marinacci, F., & Vogelsberger, M. 2016, MNRAS, 460, 1625
Portegies Zwart, S. F., & McMillan, S. L. W. 2002, ApJ, 576, 899
Regan, J. 2023, OJAp, 6, 12
Regan, J. A., Wise, J. H., Woods, T. E., et al. 2020, OJAp, 3, 15
Rodriguez, C. L., Zevin, M., Amaro-Seoane, P., et al. 2019, PhRvD, 100, 043027
Sakurai, Y., Yoshida, N., & Fujii, M. S. 2019, MNRAS, 484, 4665
Sakurai, Y., Yoshida, N., Fujii, M. S., & Hirano, S. 2017, MNRAS, 472, 1677

Schaller, M., Dalla Vecchia, C., Schaye, J., et al. 2015, MNRAS, 454, 2277
Schauer, A. T. P., Bromm, V., Boylan-Kolchin, M., Glover, S. C. O., & Klessen, R. S. 2021, ApJ, 922, 193
Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521
Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119
Smith, B. D., Bryan, G. L., Glover, S. C. O., et al. 2017, MNRAS, 466, 2217
Song, X. C., Smith, P., Kalyanam, R., et al. 2022, in Practice and Experience in Advanced Research Computing 2022: Revolutionary: Computing, Connections, You, PEARC '22 (New York, NY: Association for Computing Machinery), 1
Springel, V. 2005, MNRAS, 364, 1105
Springel, V. 2010, MNRAS, 401, 791
Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726
Turk, M. J., Smith, B. D., Oishi, J. S., et al. 2011, ApJS, 192, 9
Vanzella, E., Castellano, M., Bergamini, P., et al. 2022a, A&A, 659, A2
Vanzella, E., Castellano, M., Bergamini, P., et al. 2022b, ApJL, 940, L53
Vanzella, E., Claeyssens, A., Welch, B., et al. 2023, ApJ, 945, 53
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Vogelsberger, M., Marinacci, F., Torrey, P., & Puchwein, E. 2020, NatRP, 2, 42
Weinberger, R., Springel, V., & Pakmor, R. 2020, ApJS, 248, 32
Whitaker, K. E., Cutler, S. E., Chandar, R., et al. 2025, arXiv:2501.07627
Williams, C. E. 2025a, Astro-claire/arepo-clusters: First Catalog Release (v2), doi:10.5281/zenodo.15392998
Williams, C. E. 2025b, Astro-claire/process-fof: Referee Version 1 Release (v2), doi:10.5281/zenodo.15701630
Williams, C. E., Lake, W., Naoz, S., et al. 2024, ApJL, 960, L16
Williams, C. E., Naoz, S., Lake, W., et al. 2023, ApJ, 945, 6
Williams, H., Kelly, P. L., Chen, W., et al. 2023, Sci, 380, 416
Wise, J. H., Regan, J. A., O'Shea, B. W., et al. 2019, Natur, 566, 85
Xuan, Z., Naoz, S., Kocsis, B., & Michaely, E. 2024, ApJ, 965, 148
Yoshida, N., Abel, T., Hernquist, L., & Sugiyama, N. 2003, ApJ, 592, 645
Yoshida, N., Omukai, K., Hernquist, L., & Abel, T. 2006, ApJ, 652, 6