UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

Università degli Studi di Trieste

XXXVIII CICLO DEL DOTTORATO DI RICERCA IN

FISICA

Ente finanziatore: Istituto Nazionale di Astrofisica (INAF)

Protoclusters and High-z Clusters: Connecting Simulations and Observations

---

SETTORE SCIENTIFICO-DISCIPLINARE: ASTRONOMIA E ASTROFISICA

DOTTORANDA:
Michela Esposito

COORDINATORE:
Prof. Francesco Longo

SUPERVISORE DI TESI:
Dott.ssa Veronica Strazzullo

CO-SUPERVISORE DI TESI:
Prof. Stefano Borgani

Anno Accademico 2024-2025

ABSTRACT

Protoclusters of galaxies are regions of the distant Universe ($z \gtrsim 2$) that are expected to collapse into galaxy clusters by the present time. These regions typically consist of loose associations of galaxies, spanning tens of physical megaparsecs and comprising multiple distinct halos that will eventually merge into a single virialized cluster. Studying these primordial structures is crucial, as their typical redshift range $2 \lesssim z \lesssim 4$ – corresponding to the "cosmic noon" – marks a period of intense activity , including star formation, galaxy mergers, and accretion onto supermassive black holes (SMBHs). These processes shape galaxy evolution and possibly impact the emerging intracluster medium (ICM), contributing to the ultimate fate of these systems as the passive galaxy clusters filled with hot ambient gas commonly observed at $z \leqslant 1$.

Over the past ~ 30 years since the discovery of the first protoclusters, a variety of approaches to identify such structures have produced a highly heterogeneous picture of the early assembly and evolution of galaxy clusters. This diversity is enhanced by rapid transformations driven by fast evolutionary processes during the cosmic noon. To interpret the complexity associated with such transformational processes within a cosmological context, cosmological hydrodynamical simulations are essential. In this context, a detailed comparison between simulations and observations is crucial to assess the capability of simulations to properly describe the processes driving the formation and evolution of galaxy populations within the most extreme cosmic environments, and to extract from observations physical insight into cluster formation.

In this Thesis, I analyzed the latest version of the `DIANOGA` cosmological hydrodynamical simulations of galaxy clusters, produced with the `OpenGADGET-3` code, at high resolution. These simulations implement a galaxy formation model that describes star formation and Supernova feedback, black hole growth and active galactic nucleus (AGN) feedback, radiative cooling of gas, stellar evolution and chemical enrichment. I also analyzed a "cosmological box" – a representative average region of the Universe – simulated with the same setup, serving as a reference for field galaxies. The simulations have been calibrated to reproduce the stellar mass - black hole mass relation in the local Universe. In the nearby Universe, at $z = 0$, such simulations provide a good match to the galaxy stellar mass function (GSMF) both in clusters and in the field, albeit with an overproduction of galaxies with $M_* \gtrsim 10^{11}\ M_\odot$, indicating somewhat inefficient AGN feedback in the high-mass end, and a too steep low-mass end, attributed to the too mild Supernova feedback in this regime. A comparison of the simulated and observed ICM thermodynamic properties reveals also some discrepancies possibly connected to the implementation of AGN feedback:

the electron density profiles are too flat and the temperature values are too high in simulated cluster cores ($\sim 0.01 \times R_{500}$). Besides these discrepancies in specific regimes, the simulations broadly agree with observations on fundamental properties of the galaxy population and of the ICM.

The first part of this Thesis focuses on the galaxy populations within protoclusters at $z \sim 2$, the epoch when the first massive virialized clusters ($\sim 10^{14}$ $M_{\odot}$) emerge from the cosmic web and observational data on protoclusters are most abundant. I investigated how the environment affects the assembly of stellar mass in cluster galaxies, their star formation and its suppression, by studying specifically the GSMF, the star-forming main sequence (MS), and the fraction of quenched (i.e., non–star-forming) galaxies. The comparison with observations shows that the integrated star formation history and the stellar mass assembly (as captured by the GSMF) of DIANOGA protoclusters are consistent with observations. However, simulated star formation rates (SFRs) at $z \sim 2$ are ~0.3 - 0.5 dex lower than observed, across all environments (both protoclusters and field). The quenched galaxy fraction displays similarly high diversity as in observed clusters and protoclusters at comparable redshift. Within the simulations, environmental effects are evident in all diagnostics: protoclusters host an excess of massive galaxies, show suppressed SFRs, and have a higher fraction of quenched galaxies compared to the field, particularly within their most massive halos. As part of this analysis, I have processed simulations with the radiative transfer code SKIRT-9 to compute optical/NIR spectra for simulated galaxies. The goal was to address the impact of observational techniques and biases on the comparison with simulation results.

In the second part of the Thesis, I analyzed the thermodynamic properties of the hot diffuse gas component permeating the massive halos of protoclusters, focusing on the build-up of the ICM during the transition from diffuse protocluster structures to the first established clusters at $1.5 \leqslant z \leqslant 3$. I studied the contributions of different gas phases to the total gas budget in the most massive halos ($M_{200} > 10^{13}$ $M_{\odot}$), finding that the hot diffuse component – progenitor of the ICM or of the intragroup medium (IGrM) – already dominates the gas content across all probed halos at these redshifts. The second most abundant component is cold, high-density gas that fuels star formation; its contribution declines with increasing halo mass, as this gas was consumed earlier in more massive halos. Tracing the thermalized ICM at $z = 1.5$ back in time, I find that a large fraction of it was accreted from the intergalactic medium (IGM) between $z = 3$ and $z \sim 2$. By $z \sim 2$, this hot component is largely already in place in halos with $M_{200} \gtrsim (2\text{–}3) \times 10^{13}$ $M_{\odot}$. The radial profiles of thermodynamic properties (density, temperature, pressure, entropy) show good agreement with observational data for massive clusters at $z \sim 1.5$. At higher redshifts, the comparison with the limited available observations indicates central gas densities that are too low and temperatures that are too high in the simulations – consistent with the discrepancies seen at lower redshift. The few observations available for this comparison include the first detection of extended X-ray emission from thermal bremsstrahlung due to diffuse hot gas in a system at redshift $z \gtrsim 3$.

For this comparison, I produced synthetic X-ray maps with the code `SMAC`. Additionally, I include as an appendix a complementary project on simulation post-processing: the development and testing of a deep learning–based algorithm for the identification of substructures in cosmological simulations. This method achieves significantly shorter computing times with respect to traditional algorithms, thus making it potentially a valid and convenient alternative to traditional methods based on the hierarchical identification of bound substructures.

CONTENTS

List of Figures viii

List of Tables xxvii

1 Introduction 1

2 Galaxy clusters and protoclusters 7

2.1 Galaxy clusters in the nearby Universe 7

2.1.1 Search for low-redshift clusters 8

2.1.2 The population of cluster galaxies 9

2.1.3 Intracluster Medium 18

2.2 Protoclusters and high-redshift clusters 23

2.2.1 Identification of high-redshift clusters and protoclusters 27

2.2.2 Galaxy populations in protoclusters and high-redshift clusters 31

2.2.3 The first observations of proto-ICM 35

3 The DIANOGA simulations of galaxy clusters 41

3.1 GAlaxies with Dark matter and Gas intEracT (GADGET) . . . 41

3.1.1 Gravity 41

3.1.2 Hydrodynamics 42

3.1.3 Additional physics 44

3.2 DIANOGA simulations 50

3.2.1 Properties of `DIANOGA` simulations at z=0 53

3.2.2 Post-processing of `DIANOGA` simulations 60

4 Galaxy properties in simulated protoclusters 65

4.1 Identification of protoclusters 65

4.1.1 Definition of galaxy properties 69

4.2 Properties of protocluster halos 70

4.3 Galaxy stellar mass function 72

4.4 Star-forming main sequence 76

4.5 Quenched galaxy fractions 83

4.5.1 UVJ diagrams 87

4.6 Impact of apertures on stellar masses of simulated galaxies . 90

4.7 Effects of averaging SFRs . . . 91
4.8 Summary and conclusions . . . 91
5 Diffuse gas across the protocluster to cluster transition 97
5.1 Gas composition in high-z halos . . . 97
5.1.1 The build-up of ICM . . . 102
5.2 Comparison with observations . . . 105
5.2.1 Thermodynamic profiles . . . 106
5.2.2 X-ray surface brightness at z=3 . . . 110
5.3 Summary and conclusions . . . 113
6 Conclusions and future perspectives 117
6.1 Properties of protocluster galaxies . . . 117
6.2 Gas properties across the protocluster to cluster transition . . 118
6.3 Summary and conclusions . . . 120
6.4 Future perspectives for protocluster studies . . . 120
6.4.1 New and future observational facilities . . . 120
6.4.2 Ongoing work and future perspectives from simulations 121
A Substructure identification with Deep Learning 127
A.1 Introduction . . . 127
A.2 Methods . . . 130
A.2.1 Simulations . . . 130
A.2.2 Convolutional neural networks . . . 131
A.2.3 Basic structure of a CNN . . . 132
A.2.4 `U-Net`, data set, and training . . . 135
A.3 Results . . . 139
A.3.1 `SubDLe` . . . 139
A.3.2 Stellar `SubDLe` . . . 143
A.3.3 Testing performance against galaxy properties and environment . . . 149
A.3.4 Testing the robustness of the stellar `SubDLe` . . . 154
A.3.5 Testing with a larger training set . . . 155
A.4 Summary and conclusions . . . 156
B Basics of cosmology 161
B.1 Homogeneous expanding Universe . . . 161
B.1.1 Friedmann equations . . . 162

B.2 Statistical properties of the perturbations 164
B.2.1 Power spectrum of the perturbations 165
C Linear evolution of cosmic perturbations 167
C.1 Newtonian theory of small perturbations 167
C.2 Collisionless gas 170
C.3 Free-streaming damping 172
D Numerical methods for cosmological simulations 173
D.1 N-body codes 173
D.1.1 Direct Integration Codes 174
D.1.2 Particle-Mesh scheme 175
D.1.3 Tree codes 177
D.1.4 Hybrid schemes 178
D.1.5 time step criterion 179
D.1.6 Initial conditions 180
D.2 Numerical Hydrodynamics 181
D.2.1 Smoothed Particle Hydrodynamics 182

Bibliography 189

LIST OF FIGURES

Figure 1.1 Dark matter field in the Millennium Run dark matter-only simulation (Springel et al. 2005b) at $z = 0$, in a ΛCDM cosmology. 2

Figure 2.1 Composite image of X-ray and optical views of Abell 1689 at z=0.18. The purple halo shows the distribution of the ICM observed by Chandra X-ray Observatory. The underlying optical image was taken with Hubble Space Telescope. Image credits: X-ray: NASA/CXC/MIT/E.-H Peng et al; Optical: NASA/STScI. 8

Figure 2.2 Color-magnitude diagram of galaxies in the Coma Supercluster, from the Sloan Digital Sky Survey (SDSS). Different colors mark the morphological type according to Hubble's classification: red for early-type galaxies, green for bulge-dominated spirals and blue for disk galaxies, classified though inspection of SDSS images and spectra. Reproduced from Gavazzi et al. (2010). 10

Figure 2.3 GSMFs of cluster galaxies, from a sample of 21 clusters at 0.5<z<0.7 (top-left panel, van der Burg et al. 2018), compared with field galaxies at the same redshift (top-right panel). The GSMFs are shown for all galaxies and for quiescent (quenched) and star-forming populations separately, classified according to U-V and V-J colors. The bottom panels show the fractions of quenched and star-forming galaxies as a function of stellar mass in the two coeval environments. Reproduced from van der Burg et al. (2018). . 12

Figure 2.4 Left panel: quenched fraction as a function of stellar mass in clusters and groups at $z \sim 1$ (Balogh et al. 2016, yellow and red points, respectively), and clusters at $z \sim 1.6$ (green shaded region; Lee-Brown et al. 2017), compared to coeval field samples from UltraVISTA (dashed-dotted line at $z \sim 1$ and blue shaded region at $z \sim 1.6$) and SDSS local clusters (dotted line). Right panel: the environmental quenching efficiency in massive clusters (and two protoclusters) is shown as a function of redshift, color coded according to the stellar mass limit of each sample. Observations are taken from Quadri et al. (2012), Cooke et al. (2016), Nantais et al. (2017), van der Burg et al. (2018, 2020), Pintos-Castro et al. (2019), Zavala et al. (2019), and Strazzullo et al. (2019). Filled symbols and regions contain multiple (proto)clusters, while open symbols are single clusters. Gray hatches and circles mark ϵ_{env} measured within $0.5 \times R_{200}$; all the others are within R_{200}. Reproduced from Alberts & Noble (2022). 13

Figure 2.5 Mean SFR of star-forming galaxies at 0.02<z<0.085, as a function of stellar mass and environment – represented by galaxy overdensity – from the SDSS. Reproduced from Peng et al. (2010). 15

Figure 2.6 Star-forming main sequence of cluster galaxies at $0.04 < z < 0.07$ (black triangles), in 6 different cluster-centric radial bins (from top-left to bottom-right), and of field galaxies at $0.02 < z < 0.09$ (magenta circles). Blue and red symbols show the median SFRs in field and clusters, respectively. Reproduced from Paccagnella et al. (2016). 16

Figure 2.7 Galaxy stellar mass function (left panel) and quenched fractions as a function of stellar mass (right panel) at $z = 0$ from Hydrangea simulations (Bahé et al. 2017). The GSMF is shown in three differently colored lines, representing galaxy samples matching the selections in Vulcani et al. (2011, green) and Wang & White (2012, black), as well as the full population of satellites (blue). Quenched fractions are shown for satellite galaxies in three different halo mass bins, from Hydrangea (solid black, green and purple lines) and, in the lowest halo mass bin and in the field, from EAGLE simulations (black dotted line and blue lines, respectively; see Bahé et al. 2017 for details). Simulations are compared to observations from Wetzel et al. (2012) and Geha et al. (2012). Reproduced from Bahé et al. (2017). 17

Figure 2.8 Temperature (left panel) and entropy (right panel) profiles of 12 clusters at $0.05 \lesssim z \lesssim 0.09$ from the X-COP sample (Eckert et al. 2017), from Ghirardini et al. (2019). Solid lines show analytic fits to the data for the entire sample ("ALL"), as well as for the cool-core ("CC") and non-cool-core ("NCC") populations separately, overplotted on individual data points. Shaded regions indicate the 1σ uncertainty on the best-fit relations, while the dashed lines show the intrinsic scatter. The dashed and dotted vertical lines mark the positions of R_{500} and R_{200}, respectively. Temperature and entropy profiles are normalized to the self-similar predictions at R_{500}. The black solid line in the right panel shows the theoretical entropy profile from Voit et al. (2005). Adapted from Ghirardini et al. (2019). 19

Figure 2.9 Radial profiles of entropy (in units of keV cm^2) of a galaxy cluster of mass $M_{500} = 10^{15}\ M_\odot h^{-1}$ at $z = 0$, simulated with GADGET-3 (Springel 2005). Different lines refer to different simulations of the same cluster, as indicated: a non-radiative version and three radiative simulations including *i*) only cooling and star formation, *ii*) also the effect of SN feedback, and *iii*) also the effect of AGN feedback. A power-law entropy profile, as predicted by non-radiative models, is also shown. The vertical dotted line marks the position of R_{500} of the cluster. Reproduced from Kravtsov & Borgani (2012). 21

Figure 2.10 Dark matter density maps tracing the evolution of the most massive cluster ($M_{200} = 8 \times 10^{14}\ M_\odot$ at $z = 0$) in the Millennium-II N-body simulation from $z \sim 6$ to $z = 0$ (from top to bottom), shown at three different comoving scales: 100, 45, 15 cMpc h^{-1} (from left to right). Reproduced from Boylan-Kolchin et al. (2009). 24

Figure 2.11 Evolution of the effective radius (introduced by Chiang et al. 2013, see text) of (proto)clusters, grouped in bins of $z = 0$ mass. Lines (errorbars) mark the median values (16th-84th percentile ranges). Solid and dotted lines show results for different assumed cosmological parameters. Reproduced from Chiang et al. (2013) 25

Figure 2.12 Evolution of the overdensity in the number of massive halos ($M_{200} > 10^{11}$ $M_\odot$; top) and star-forming galaxies (SFR>1 $M_\odot yr^{-1}$; bottom) in simulatd protocluster regions, on scales of $(15 \text{ Mpc})^3$ and $(25 \text{ Mpc})^3$, in the left and right panels, respectively. The full protocluster population (black line) is also shown split by halo mass at $z = 0$ (colored lines). For reference, the gray line shows the same measurement in random regions in the simulation. Adapted from Chiang et al. (2013). 26

Figure 2.13 Smoothed image of the X-ray emission in the 0.5 - 3 keV band of XLSSC 122 at $z = 2$, measured from XMM-Newton. Nearby point sources are masked out. The cyan circles mark the R_{500} and R_{200} radii. Reproduced from Mantz et al. (2018). 29

Figure 2.14 Composite image of the Spiderweb protocluster in multiple wavelengths. The optical image is from Hubble Space Telescope and overlaid are: the emission from the Spiderweb galaxy radio jet, measured with the Karl G. Jansky Very Large Array (VLA, in red); the extended Lyman-α nebula observed with FORS1 at Very Large Telescope (VLT, in pink); and the SZ signal from the Atacama Large Millimeter/-Submillimeter Array (ALMA, light blue). The white diamonds denote all the spectroscopically confirmed member galaxies from Tozzi et al. (2022a). Reproduced from Di Mascolo et al. (2023). 30

Figure 2.15 Star-forming main sequence of galaxies in the protoclusters USS1558 at $z = 2.53$ (orange circles) and Spiderweb at $z = 2.16$ (blue circles), from Pérez-Martínez et al. (2024a, 2023), compared with coeval field galaxies from Wisnioski et al. (2019, empty circles) and the main sequence fit from Speagle et al. (2014, black solid line) at z=2.53, with its 3σ scatter (shaded area). Reproduced from Pérez-Martínez et al. (2024a). 31

Figure 2.16 Galaxy stellar mass function in the CLJ1001 (proto)cluster at z=2.5, compared with other protoclusters at $z \sim 2 - 2.5$. Blue (red) solid line shows the GSMF for star-forming (quenched) galaxies in CLJ1001. The blue dashed line shows the GSMF of Hα emitters only, for a fair comparison with the GSMFs in Spiderweb and in USS1558 from Shimakawa et al. (2018b,a, brown dashed and green dotted lines, respectively). All the GSMFs shown are re-scaled to match the normalization of CLJ1001. Reproduced from Sun et al. (2024). 32

Figure 2.17 Left panel: quenched galaxy fraction (for $M_* > 10^{11}$ $M_\odot$) as a function of cluster-centric radius in the CLJ1001 protocluster ($z = 2.5$), compared with the clusters ClJ1449 ($z = 2$) and JKCS 041 ($z = 1.8$), as well as the field at $z = 2.5$. Reproduced from Wang et al. (2016). Right panel: quenched fraction as a function of stellar mass in the 14 protocluster candidates from Edward et al. (2024, in green), compared with field galaxies (in blue). Reproduced from Edward et al. (2024). 33

Figure 2.18 Multiwavelength view of XLSSC 122: Hubble Space Telescope image in the background; XMM-Newton smoothed contours in white; SZ flux measured with ALMA in blue. The red contours represent the best model reconstruction of the SZ signal and the circles mark cluster member galaxies (in cyan for star-forming and dusty galaxies). Reproduced from van Marrewijk et al. (2024). 36

Figure 2.19 Radial temperature profiles of the proto-ICM in the Spiderweb protocluster. Different colors show the profiles for different assumptions of the n_d parameter that accounts for the unknown surface brightness in the innermost region contaminated by the radio galaxy. The solid black line shows the temperature value constrained by Tozzi et al. (2022b) by assuming an isothermal profile, with 1σ error shown by the shaded region. Reproduced from Lepore et al. (2024). 37

Figure 2.20 Smoothed Chandra X-ray image (map of counts) in the 0.5-2.0 keV rest-frame band of the central halo of MQN01, after subtraction of the contribution of the central QSO. The solid (shaded) black dot marks the 1 (2) arcsec radius around the QSO center. The magenta contours are Lyman-α isophotes and the red cross marks the position of a companion QSO detected with ALMA. Reproduced from Travascio et al. (2025b) 38

Figure 3.1 Evolution of a `DIANOGA` cluster from $z = 2$ (top panels), to $z = 1$ (middle panels) and then to $z = 0$ (bottom panels). From left to right, the panels show the rendering of the DM, gas and stars distributions. 52

Figure 3.2 Galaxy stellar mass function at $z = 0$ in simulations, compared with observational data. The top panel shows the prediction for galaxies in the cosmological box (with Poissonian uncertainties), compared with the GSMF in the COSMOS field at $0.2 < z < 0.5$, from Muzzin et al. (2013) and Weaver et al. (2023). The bottom panel shows the GSMF in DIANOGA clusters, excluding the BCGs. Star symbols are the total number of galaxies in all clusters divided by the number of clusters and the shaded region shows the 16th-84th percentile range among the 12 clusters. These are compared with the observed GSMF of WINGS clusters ($0.04 \leqslant z \leqslant 0.07$, Vulcani et al. 2011). Galaxies are selected within $0.6{\times}R_{200}$ to match the galaxy sample in almost all WINGS clusters. The observed GSMF in clusters is shown above the mass completeness limit of $10^{9.8}$ $M_\odot$. 56

Figure 3.3 Radial profiles of thermodynamic properties of the ICM in DIANOGA clusters, compared with observations from Ghirardini et al. (2019). For DIANOGA simulations, the solid line represents the median profile among the clusters and the shaded area is the 16th-84th percentile range. The solid lines corresponding to the data are the global functional forms fitted to the observational data, while dotted lines mark the intrinsic scatter on the profiles. The top-left panel shows the comparison between the radial profiles of number density of free electrons. The top-right panel shows the temperature profiles, where two possible definitions are used in simulations, mass-weighted (MW) and spectroscopic-like (SL) temperature. The bottom-left and right panels show the radial profiles of pressure and entropy of free electrons, respectively. 59

Figure 4.1 Projected DM density maps of three simulated clusters at $z = 0$ (left column), with total masses $M_{200} = 2 \times 10^{15}$ $M_\odot$ (top), 9×10^{14} $M_\odot$ (middle), and 2×10^{14} $M_\odot$ (bottom), and their $z = 2.2$ progenitors (right column). The cyan circle at $z = 0$ traces the R_{200} radius of each cluster, the one at $z = 2.2$ traces the PC region as defined in the text. DM particles at $z = 2.2$ that collapse within the R_{200} radius by $z = 0$ are plotted in cyan. Densities are computed in pixels of size 15 kpc in the top panels and 10 kpc in the middle and bottom panels. 66

Figure 4.2 Extension of the PC regions (in units of R_{200} of the main progenitor halos) at $z = 2.2$ as a function of the mass, M_{200}, of the descendant clusters at $z = 0$ (red diamonds). The downward arrows show, for each progenitor, the distance from the boundary of the high-resolution region (see main text). 68

Figure 4.3 Projected stellar mass density maps of two DIANOGA PC regions, which are the progenitors of clusters of total mass M_{200} equal to 5×10^{14} $M_\odot$ and 2×10^{15} $M_\odot$ at $z = 0$ (left and right panel, respectively). In each panel, the larger red dashed circle corresponds to the PC region as defined in the text, while the smaller circles show the R_{200} of halos in these regions with $M_{200} > 10^{13}$ $M_\odot$. Densities are computed in pixels of size 15 (10) kpc in the right (left) panel. 69

Figure 4.4 Mass M_{200} of massive halos in PC regions as a function of velocity dispersion σ_v along three orthogonal lines of sight (represented by circles, squares and diamonds) as traced by galaxies with stellar masses larger than 10^9 $M_\odot$ and lying within the R_{200} of each halo. Each point is color coded according to the halo R_{200}. The symbols with red contours mark the central halos of DIANOGA PCs. The dashed curve shows the relation between dynamical masses and galaxy velocity dispersions calibrated in simulations by Saro et al. (2013) up to $z = 1.2$, and extrapolated to $z = 2.2$. Star symbols represent velocity dispersions and dynamical mass estimates within the core of the Spiderweb PC and for the two subgroups of USS1558 from Shimakawa et al. (2014). The cross symbol corresponds to CLJ1001 (Wang et al. 2016). The pentagon symbol shows M_{200} for the Spiderweb PC, from Sunyaev-Zeldovich effect measurements with ALMA (Di Mascolo et al. 2023), the triangle represents the cluster XLSSC 122 (Mantz et al. 2018; Willis et al. 2020; Noordeh et al. 2021) and the downward triangle represents CLJ1449 (Gobat et al. 2013). The latter three points are shown at $\sigma_v < 0$ for illustrative purposes, since velocity dispersions measured consistently within the estimated radii are not available. 71

Figure 4.5 Galaxy stellar mass functions in the PC environment compared with the cosmological box. The red line is the median of the GSMFs in PC cores, with and without central galaxies (solid and dashed lines, respectively; I note that the dashed line stops at $M_* \sim 10^{11} M_\odot$ because the median GSMF in PC cores in the higher mass bins is zero when central galaxies are excluded). The green line shows the median GSMF in the less dense, diffuse PC regions, masking out all PC cores. The normalizations have been rescaled by the factors reported in the legend. The shaded regions define the 16th-84th percentile range spanned by the different PCs (or PC cores). The black line is the GSMF in the cosmological box, with Poissonian uncertainties. 73

Figure 4.6 Comparison between the GSMF in `DIANOGA` PCs and in observations of PCs at $z \sim 2-2.5$ (Spiderweb, CLJ1001, 14 PCs in the COSMOS field; Pannella et al. in prep; Sun et al. 2024; Edward et al. 2024, respectively). For each set of observations, a simulated GSMF has been produced to match the selection of galaxies as closely as possible. For each selection, GSMFs have been extracted along three orthogonal lines of sight, convolved with Gaussian errors with scatter 0.3 dex (see text), and the solid lines (shaded regions) show the median GSMFs (16th-84th percentile ranges, respectively) across all realizations, in the same color as the corresponding data. The GSMF from Edward et al. (2024) has been arbitrarily re-scaled by a factor of 10 to match the normalization in simulations, since the extension of these PCs along the line of sight in unconstrained and, thus, it cannot be replicated in the simulations. 74

Figure 4.7 Distributions of sSFRs of galaxies in the `DIANOGA` PC regions (red line) and in the cosmological box (blue line) at $z = 2.2$ compared with the threshold separating star-forming from quenched galaxies, at this redshift, from Franx et al. (2008). Galaxies with SFR=0 have been artificially assigned the minimum non-zero sSFR, corresponding to 10^{-14} yr^{-1}. 77

Figure 4.8 Star-forming MS of galaxies in the DIANOGA PCs (orange line) and in the cosmological box (green line). The solid orange and green lines show the median SFRs of all star-forming galaxies within each stellar mass bin while the shaded areas mark the 16th-84th percentile intervals. The star symbols show the proto-BCGs identified in the simulations, color coded according to the mass M_{200} of the main progenitor halo of each PC. Black diamonds, green and red dots are data for the Spiderweb PC from Pannella et al. (in prep.) and from Pérez-Martínez et al. (2023) and Pérez-Martínez et al. (2024b), respectively. The blue line is the best fit for the observed field star-forming MS by Schreiber et al. (2015) at $z = 2.2$. The red bar shows the interval of estimated SFRs for the Spiderweb galaxy from FIR measurements by Seymour et al. (2012) and Drouart et al. (2014). Simulated SFRs are averaged over 200 Myr (see text). For readability, error bars on observational data have been omitted. . 78

Figure 4.9 Environmental dependence of the simulated star-forming MS at $z = 2.2$ as a function of stellar mass. Results are shown for the whole sample of PC galaxies, as well as separating PC galaxies located within/outside R_{200} of halos with $M_{200} > 1.5 \times 10^{13}\ M_{\odot}$ (PC cores/extended PCs), compared to galaxies in the cosmological box. The solid lines and shaded regions show the MS offset, and its intrinsic scatter, between pairs of environments as indicated in the legend (see text for details). 80

Figure 4.10 Cold gas fractions (left panel) and depletion times (right panel) in star-forming galaxies in the DIANOGA PCs (orange) and in the cosmological box (green) as a function of stellar mass. Solid lines (shaded regions) show median values (16th-84th percentile ranges, respectively). For comparison, I show observations at $z \sim 1.5 - 3$ from Magdis et al. (2012), Liu et al. (2019), Sanders et al. (2023), Mérida et al. (2024), Leroy et al. (2024) and Pérez-Martínez et al. (2025). In the latter, the two higher-mass values correspond to massive galaxies ($\log_{10}(M_* \ [M_{\odot}]) > 10.5$) that do and do not host AGNs (8 and 12 galaxies, respectively), at median masses $M_* \sim 10^{10.9}$ and $M_* \sim 10^{11}\ M_{\odot}$, respectively. 82

Figure 4.11 Environmental dependence of the fraction of cold gas (left panel) and of the cold gas depletion time (right panel) at $z = 2.2$ as a function of stellar mass. Here I consider four environments: the cosmological box, the whole PCs, the low-density extended PCs and the PC cores. The solid lines define the offsets between the cold fractions and the depletion times in the PC environments with respect to the cosmological box, as defined in the legend, resulting from 1000 Monte Carlo realizations (see text). The shaded regions represent the intrinsic scatter on the offsets. 83

Figure 4.12 Fraction of quenched galaxies in PC cores as a function of the core mass M_{200} in the DIANOGA PCs at $z = 2.2$. The dots are median quenched fractions across different cores in each mass bin and the shaded areas around them are the 16th-84th percentile ranges. The horizontal bands show the fraction of all quenched galaxies in the coeval cosmological box (its width is the 1σ confidence interval calculated following Cameron 2011), and the 16th-84th percentile interval of quenched fractions in the PC extended regions (masking out the PC cores), across the 14 different simulations, independently of their host halo mass. Results are shown for galaxies with stellar masses in the range $10^9 < M_*/M_\odot < 10^{10}$ and $M_* > 10^{10}\ M_\odot$ separately. 85

Figure 4.13 Comparison of quenched fractions in the most massive simulated PC cores (with $M_{200} \geqslant 5 \times 10^{13}$ $M_{\odot}$) and observations at $z \sim 2-2.5$, including all galaxies with stellar masses larger than the completeness limit (M_{lim}) of each observational study. I note that measurements in this figure are not homogeneous in terms of the size of the considered region around the PC center (see text). The cyan dots show the median quenched fractions among selected halos, while shaded dark (light) regions are the 16th-84th percentile (minimum-maximum) ranges. The cyan diamonds, with binomial population uncertainties (Cameron 2011), are quenched fractions extracted from the cosmological box. For comparison with the observations in CLJ1001, XLSSC 122 and ClJ1449, galaxies in PCs were selected above a limiting stellar mass and within a physical aperture around halo centers, matching those adopted in Wang et al. (2016), Noordeh et al. (2021) and Strazzullo et al. (2018), respectively (see text for details). For comparison with the cosmological box, I also show the quenched fractions in the field from COSMOS/UltraVISTA catalog (Muzzin et al. 2013) and from the CANDELS/3D-HST survey (Skelton et al. 2014) at $2 < z < 2.5$. 86

Figure 4.14 Simulated UVJ diagrams for galaxies within seven massive halos in `DIANOGA` PC regions at $z = 2.2$, within a 700 kpc aperture, produced with `SKIRT-9`. Intrinsic colors are connected to reddened colors by gray arrows that end with star symbols color coded according to the sSFRs of galaxies. The left panel shows the diagram obtained by tracing dust in the diffuse ISM through a fixed dust-to-metal ratio. The right panel shows the diagram produced with the model for MCs described in Sect. 3.2.2.1. Typical cuts adopted in UVJ diagrams to separate quenched and star-forming galaxies (Williams et al. 2009; Whitaker et al. 2011, 2012; Muzzin et al. 2013) are shown as black lines, while the red line defines a custom-defined boundary to separate these populations in the `DIANOGA` simulations. The colored arrows show the reddening vectors for typical attenuation laws (Allen 1976; Prevot et al. 1984; Bouchet et al. 1985; Calzetti et al. 2000). 88

Figure 4.15 Properties of a galaxy characterized by a steep reddening vector. The left panel shows the distribution of stars, color coded according to their age, while the middle panel shows the distribution of gas particles color coded according to their mass in metals, both within 0.25 kpc from the center of the galaxy along the line of sight. The right panel shows the galaxy star formation history. 89
Figure 4.16 Ratio of `SubFind` and aperture stellar masses, as a function of aperture mass for galaxies in `DIANOGA` PCs, separated in centrals and satellites, and in the cosmological box at $z = 2.2$. Aperture masses are estimated within a two-dimensional projected radius equal to 8.5 kpc. The bottom panel shows the median (16th-84th percentiles) of the mass ratios of galaxies plotted in the top panel, in solid lines (shaded areas, respectively). 92
Figure 4.17 Two-dimensional histograms of the ratios between instantaneous SFR (SFR_{gas}) and the SFR averaged over 200 Myr (SFR_{200}) of galaxies in `DIANOGA` PCs (left panel) and in the cosmological box (right panel) at $z = 2.2$, as a function of stellar mass. Proto-BCGs are plotted separately. 93
Figure 4.18 Distribution of the instantaneous (averaged) sSFRs for galaxies with $\mathsf{SFR}_{200} = 0$ ($\mathsf{SFR}_{gas} = 0$), in the left (right, respectively) panel. The dashed lines show the average sSFR of main-sequence galaxies in the cosmological box (black) and in the PCs (red). 93
Figure 5.1 Phase diagram of the gas in the region within $5{\times}R_{500}$ (left panel) and R_{500} (right panel) of a cluster of mass $M_{200} = 1.2 \times 10^{14}\ M_{\odot}$ at $z = 2.2$. The solid black lines define the separation between different phases of the gas. The discontinuity at $\rho \sim 10^{6.5}$ $M_{\odot}\mathrm{kpc}^{-3}$ corresponds to the transition to the star-forming phase in the SH03 model for star formation. 98
Figure 5.2 Projected mass map of the gas components, as well as of DM and stars, in the region spanning 4 Mpc per side around a simulated cluster of mass $M_{200} = 1.2 \times 10^{14}\ M_{\odot}$ at $z = 2.2$. The different gas components (condensed phase, IGM, WHIM, ICM) are shown in the upper- and middle-row panels. The bottom-row panels show the distribution of DM mass (left panel) and of stellar mass (right panel). In all panels, the dashed (dotted) circle defines R_{500} ($5{\times}R_{500}$) of the central halo. 99

Figure 5.3 Fraction of gas mass in four different phases (condensed, IGM, WHIM and ICM), within R_{500} of each halo, as a function of the halo mass. From left to right the halos are selected at $z = 3$, 2.2 and 1.5. 100
Figure 5.4 Median mass fractions of different gas components (condensed, IGM, WHIM and ICM) lying within R_{500} of massive halos, as a function of halo mass. For each component, the different lines show the median fraction at different redshifts ($z = 1.5$, 2.2 and 3), with shaded areas showing the corresponding 16th-84th percentile range. 101
Figure 5.5 Gas components across the ICM build-up: the left (right) panel shows, as a function of the halo mass at $z = 1.5$, what mass fraction of gas belonging to the ICM at $z = 1.5$ is in different phases – condensed, IGM, WHIM and ICM – at $z = 3$ (2.2, respectively). . 102
Figure 5.6 Median values (solid lines; and 16th-84th percentile ranges, shaded areas) of the fractions (shown in Fig. 5.5) of the different components of the $z = 1.5$ ICM precursor, as a function of the $z = 1.5$ halo mass. The figure shows the fractions of condensed (top-left panel), IGM (top-right panel), WHIM (bottom-left panel) and ICM (bottom-right panel) phases, at $z = 3$ and 2.2. 103
Figure 5.7 Variation of the mass fraction of the condensed, IGM and WHIM components (from left to right, as indicated) of the $z = 1.5$ ICM precursor gas, shown against the variation of fraction of gas that is already in the ICM phase, between $z = 3$ and 2.2. The arrows, color coded according to the halo mass at $z = 1.5$, start at fractions at $z = 3$ and point to fractions at $z = 2.2$. 103
Figure 5.8 Free electron number density (left panel) and temperature (right panel) profiles of simulated halos with $M_{200} \geqslant 10^{13}\ M_{\odot}$, color coded by their mass, in comparison with observations from Ghirardini et al. (2021) of seven clusters at $z = 1.2 - 1.7$. All the temperature profiles are normalized by the self-similar T_{500}. The hatched area marks the region where the observed profiles are not resolved. 106

Figure 5.9 Pressure (left panel) and entropy (right panel) profiles of simulated halos with $M_{200} \geqslant 10^{13}$ $M_\odot$, color coded with their mass, in comparison with observations from Ghirardini et al. (2021) of seven clusters at $z = 1.2 - 1.7$. All the pressure and entropy profiles are normalized by the self-similar P_{500} and K_{500}, respectively. The hatched areas mark the regions where the observed profiles are not resolved. 107

Figure 5.10 Free electron density (left panel) and pressure (right panel) profiles of simulated halos with $M_{200} \geqslant 10^{13}$ $M_\odot$ at $z = 2.2$, color coded with their mass, and compared with observations from Lepore et al. (2024) of the Spiderweb protocluster at $z = 2.2$ (shaded grey region). The range shown for the observed density profile represents its variation when assuming different values of the n_d parameter that accounts for the unknown surface brightness in the innermost region contaminated by the radio galaxy. The range shown for the pressure profile of the Spiderweb protocluster represents the statistical uncertainty around the best-fit pressure profile. 108

Figure 5.11 Temperature (left panel) and entropy (right panel) profiles of simulated halos with $M_{200} \geqslant 10^{13}$ $M_\odot$ at $z = 2.2$, color coded with their mass, compared with observations from Lepore et al. (2024) of the Spiderweb protocluster. The range shown for the observations represents variations associated with the assumptions of different values of the n_d parameter that accounts for the unknown surface brightness in the innermost region contaminated by the radio galaxy.109

Figure 5.12 X-ray surface brightness maps of the 12 most massive halos in the `DIANOGA` simulations at $z = 3$, modeled with `SMAC`. From the top-left to the bottom-right panel, the halo mass is decreasing, from $M_{200} = 6 \times 10^{13}$ to 2×10^{13} $M_\odot$. All panels are on the same spatial scale as indicated in the top-left one. 111

Figure 5.13 X-ray surface brightness profiles in the 0.5–2 keV band for halos at $z \sim 3$ (and Spiderweb at $z = 2.2$). The solid black line shows the median profile of the 12 selected DIANOGA halos at $z = 3$ (see text), derived from SMAC X-ray luminosity maps. Dashed and dotted lines indicate the 16th–84th and minimum–maximum ranges, respectively. Red dots mark Chandra X-ray data for MQN01 at $z = 3.25$. The data point below 15 kpc (gray shaded area) is extrapolated from the best-fit model of the data, while beyond 30 kpc only upper limits are available. For comparison, the X-ray surface brightness profile of the Spiderweb system from Lepore et al. (2024), corrected for differences in the redshift-dependent intrinsic energy band, is shown in blue. Reproduced from Travascio et al. (2025b). 112

Figure 5.14 Electron density profiles of $M_{200} > 10^{13}$ $M_\odot$ DIANOGA halos at $z = 3$, compared with the best-fit model to observational data in MQN01 at $z = 3.25$. The grey shaded area represents the full range covered by simulated halos. The red solid line is the best fit to the data, with the shaded area defining the 68% confidence interval. For comparison, the density profile of Spiderweb from Lepore et al. (2024) is shown in blue. 113

Figure 6.1 Correlation between integrated specific star formation rates of satellite galaxies in massive halos at $z = 1.5$ and timing of build-up of their ICM. The latter is quantified by the fraction of ICM identified at $z = 1.5$, that was already thermalized by $z = 2.2$ (left panel) and $z = 3$ (right panel). Here, the sSFRs are calculated by considering all satellites within R_{500} and with stellar mass $M_* > 10^{10}$ $M_\odot$, so as to exclude galaxies whose evolution might be affected by resolution in the densest environment (see Sect. 4.4). Stellar masses and SFRs are calculated within a 8.5 kpc radius around SubFind centers, with the SFR being averaged over the past 200 Myr. Solid lines (shaded areas) show the median (16th-84th percentile range of) sSFR, color coded according to the value of the median mass in three halo mass intervals. 123

Figure A.1 Schematic representation of a convolution between a 5×5 grid and a 3×3 filter, with depth 1. The filter is slid across the grid and the dot-product is performed at each alignment (represented by the yellow subset of the grid in the figure), until all the grid is covered and the 3×3 output grid is completed. 133

Figure A.2 Convolution between a $64 \times 32 \times 3$ input grid with two $5 \times 5 \times 3$ filters gives a $60 \times 28 \times 2$ grid. Note that the depth of the filter and that of the grid must be the same, while the depth of the output grid is given by the number of filters. 134
Figure A.3 Example of the assignment scheme of labels adopted in the training set of the 3D `U-Net` implemented in `SubDLe`. Here red particles trace the substructures to be identified, and blue particles are part of a background. In the upper panel, one particle at most is present in each grid pixel, thus the substructures (associated with the label "1" in the right-hand grid, whereas the background is labeled as "0") are simply associated with pixels containing one red particle. In the lower panel, closer to the real identification problem, more than one particle falls in each grid pixel, and the assignment of labels is performed with a majority vote. 137
Figure A.4 Projected map of the active pixels associated with the substructures identified by the first step performed by `SubDLe` (through a 3D `U-Net`), in a run on the whole cluster `CL-5` at $z = 0$. The blobs of active pixels need to be grouped together to identify subhalos, since at this stage all subhalos are associated with the same class label. 139
Figure A.5 Projected map of particles belonging to subhalos of the cluster `CL-2` at $z = 0$. The left and right panels show the subhalos identified by `SubFind` and `SubDLe`, respectively. 140
Figure A.6 Distribution of pairs of `SubFind` and `SubDLe` subhalos as a function of the distance between their centers, in physical kiloparsecs, for the identification performed by the all-particle `SubDLe` (in blue) and stellar `SubDLe` (in orange). The vertical dashed lines mark the gravitational softening used in the gravitational force calculations for the DM particles (in blue) and for star particles (in orange). 142

Figure A.7 Scatter plot of all particles (DM, gas, and stars) belonging to a subhalo identified by `SubDLe` (top-row panels) and the matching subhalo identified by `SubFind` (bottom-row panels), projected in 2D. Three typical cases are shown: in the left column, a subhalo with an extension of few tens of kiloparsecs, perfectly identified by `SubDLe`; in the central column a larger subhalo, spanning a distance of about 1 Mpc, where only the central region is identified by `SubDLe`; and in the right column, a system of two close substructures that are combined into one subhalo in `SubDLe` catalog due to their short distance, compared with the density grid resolution. 143

Figure A.8 Projected distribution of star particles belonging to different galaxies identified by `SubFind` (left panel) and by stellar `SubDLe` (right panel) in the cluster `CL-2` at $z = 0$. 145

Figure A.9 Distribution of pairs of `SubFind` and `SubDLe` subhalos as a function of the ratio between the mass of common particles ($M_{SF/SD}$) and the mass given by `SubFind` (M_{SF}), as a tracer of the completeness of the mass reconstruction performed by `SubDLe`, for the run carried on the mass distribution of all particles (DM, gas, and stars), in orange, and for the run of stellar `SubDLe`, in blue, in the cluster `CL-2` at $z = 0$. . . 145

Figure A.10 Distribution of the ratio $M_{SF/SD}/M_{SF}$ in all galaxies identified by the stellar `SubDLe` and matched to a `SubFind` substructure for all the 12 clusters analyzed at $z = 0$. The mean value, across all 12 clusters, is traced by the black dashed line, while the orange dashed line represents the mean value for the all-particle run of `SubDLe` on cluster `CL-2` (associated with the orange histogram shown in Fig. A.9). 148

Figure A.11 Scaling of the percentage of resolved `SubFind` galaxies in the 12 clusters analyzed at $z = 0$ (`CL-[2-13]`), identified by the stellar `SubDLe`, with cluster properties: total virial mass (upper panel), mean inter-galaxy separation within one virial radius (middle panel), and median galaxy stellar mass (lower panel), color coded with the total number of `SubFind` galaxies (the ones deemed resolved enough, with at least 100 star particles). The circles represent clusters that required a treatment with sparsely distributed sub-grids (with $M_{th} = 0$) and the diamonds represent the three smaller clusters that did not require it. 149

Figure A.12 Dependence of the stellar SubDLe completeness (defined as the fraction of SubFind galaxies identified by the stellar SubDLe) on SubFind stellar masses (left panels) and on 3D distance from the cluster center (in units of virial radii, right panels), for all the clusters analyzed at $z = 0$ (see Table A.2). In the top panels I show, with solid lines, the results for all the analyzed clusters, while dashed and dotted lines represent clusters analyzed with a single grid and a double grid ($M_{th} = 0$), respectively. In the bottom-left panel different colors of the dashed curves correspond to different radial bins. In the bottom-right panel different colors correspond instead to different intervals of the stellar mass of the galaxies. 152

Figure A.13 Comparison of the distributions of stellar masses in the simulated galaxy clusters at $z = 0$, as predicted by the subhalo catalogs produced by SubFind (red dots) and by the stellar SubDLe (blue diamonds). The points show the median counts in each stellar mass bin among the 12 different clusters, while the error bars mark the 25th and 75th percentiles. 153

Figure A.14 Comparison of the performance of SubDLe on the same cluster, with different d grid resolutions. Left panel: Distributions of values of $M_{SF/SD}/M_{SF}$ for SubDLe galaxies matched to SubFind galaxies. Results are shown for the cluster CL-13 at $z = 1$, for SubDLe runs with the standard resolution of 10 kpc (in orange) and with the higher resolution of 2 kpc (in blue). Right panel: stellar SubDLe completeness as a function of stellar mass in cluster CL-13 at $z = 1$, for 2 kpc (in blue) and 10 kpc (in orange) resolution. 155

Figure A.15 Completeness (blue diamonds) and purity (red diamonds) of stellar SubDLe identifications, referred to SubFind catalogs, for galaxies in $z = 0$ clusters. The y-axis refers to metrics calculated on the catalog given by the model trained on two clusters (CL-3 and CL-8), while the x-axis refers to results of the original stellar SubDLe trained on one cluster. The dashed line represents cases where the metrics are identical in the catalogs produced by the two models. 156

Figure B.1 Hubble diagram, obtained with 42 high-redshift type Ia Supernovae from the Supernova Cosmology Project and 18 low-redshift type Ia Supernovae from the Calán/Tololo Supernova Survey. Dashed and dotted lines correspond to different cosmological models, with the corresponding parameters reported at the top-right corner. Reproduced from Perlmutter et al. 1999. 162
Figure D.1 A schematic illustration of the implementation of an oct-tree in two dimensions; the root node (top left) with side l is divided in sub-squares of sides $l/2$ (top right), then the same is done with each of the *children* nodes and with their children as well, until each square of the final partitioned domain contains only one particle (bottom right). Reproduced from Springel et al. (2001a). 179

LIST OF TABLES

Table 3.1 IDs and masses of the DIANOGA cluster simulations at $z = 0$. The last column specifies if the cluster belongs to the set of the 24 most massive clusters (set 1) originally selected in the parent box or to the set of 5 less massive ones (set 2). Clusters D15 and D16 were simulated down to $z = 1$, so their masses at $z = 0$ come from the parent simulation. 51

Table 4.1 Properties of the DIANOGA PCs. Reported below are the ID, the mass of the $z = 0$ cluster, the mass of the main progenitor halo at $z = 2.2$, the radius of the PC region and the number of PC cores (halos with $M_{200} > 1.5 \times 10^{13}\ M_{\odot}$ within the PC region) for each DIANOGA PC. 67

Table A.1 Main properties of the clusters analyzed in this study. For each cluster, listed below are the virial mass, the virial radius, the number of galaxies considered resolved enough for this study (with at least 100 star particles), and the phase of the analysis in which it was used (i.e., training, referred to as "train," validation, referred to as "val," or test). 131

Table A.2 Results of identification of galaxies in 11 simulated galaxy clusters analyzed with stellar SubDLe. The first column reports the cluster ID, the second and third ones show the completeness and purity of the stellar SubDLe catalogs, in comparison with SubFind, and the fourth one reports the wall clock time (in seconds) taken by SubDLe for the processing of the d grids. 147

1

INTRODUCTION

The cosmological principle assumes that the Universe is homogeneous and isotropic on large scales. This assumption is supported by observations of the Cosmic Microwave Background (CMB; serendipitously discovered by Penzias & Wilson 1965) and the large scale distribution of galaxies. The current leading framework describing the homogeneous and isotropic expanding Universe is the Λ Cold Dark Matter (ΛCDM) model. In this model, the Universe is dominated by a constant energy density component (Λ), which drives the recent acceleration of cosmic expansion, as first observed through type Ia Supernovae (Riess et al. 1998). This component – commonly interpreted as "dark energy" – is associated with a density parameter $\Omega_\Lambda \sim 0.7$.

The matter content of the Universe ($\Omega_m \sim 0.3$) is dominated by a cold, collisionless component known as Dark Matter (DM), while baryons contribute only a small fraction of the total matter budget ($\Omega_b \sim 0.05$). Along with the Hubble constant H_0, these parameters define the cosmological background, which is constrained by several independent probes: primordial abundaces from the Big Bang Nucleosynthesis (e.g., Wagoner et al. 1967; Burles et al. 2001), CMB anisotropies (e.g., de Bernardis et al. 2000; Spergel et al. 2003; Planck Collaboration VI 2020), Baryonic Acoustic Oscillations (BAOs) in the large-scale clustering of galaxies (e.g., Eisenstein et al. 2005; Anderson et al. 2014; Adame et al. 2025), cosmic shear (e.g., Troxel et al. 2018; Anbajagane et al. 2025), the Lyman-α forest (e.g., Abdul Karim et al. 2025) and galaxy cluster counts (e.g., Borgani et al. 2001; Vikhlinin et al. 2009; Planck Collaboration XX 2014; DES Collaboration 2025).

On top of the homogeneous ΛCDM background, small-scale inhomogeneities must have been present at early cosmic times. Otherwise, the Universe could not host the observed hierarchy of structures – ranging from galaxies to filaments, and groups and clusters of galaxies (see Fig. 1.1 for a simulated slice of the Universe).

The large-scale structure observed in the so-called "cosmic web" originates from small density fluctuations in the highly homogeneous primordial Universe. In order to describe the statistical properties of these fluctuations, two further cosmological parameters are introduced: the spectral index of the primordial power spectrum, n_s (~ 1), and the amplitude of matter density fluctuations at the scale of 8 h^{-1}Mpc, σ_8 (~ 0.8).

To date, the ΛCDM paradigm remains a successful description of the Universe, requiring only a small set of parameters to explain a broad range of observations across multiple scales – albeit with some tensions between different cosmological probes (e.g., Di Valentino et al. 2021a,b).

Primordial fluctuations left an imprint on the CMB, that provides a snapshot

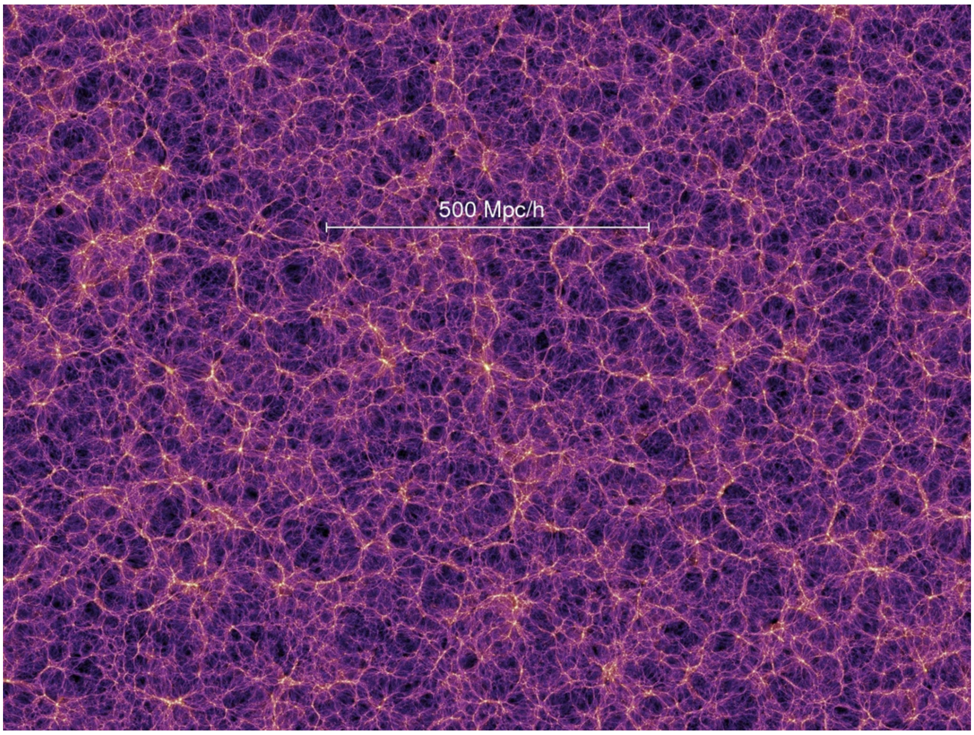


Figure 1.1: Dark matter field in the Millennium Run dark matter-only simulation (Springel et al. 2005b) at $z = 0$, in a ΛCDM cosmology.

of the Universe at the so-called "last scattering surface", around $z \sim 1100$, shortly after recombination occurred at $z \sim 1300$, when cosmic baryons became neutral for the first time. These fluctuations are thought to originate from physical processes (e.g., phase transitions or quantum fluctuations) in the early Universe, during epochs not directly accessible through electromagnetic probes.

As the Universe expanded, these density perturbations grew under the effect of gravitational instability. Their nonlinear growth ultimately led to the formation of DM halos, whose gravitational wells host galaxies. Within the ΛCDM framework, structures form hierarchically: small objects collapse first and merge over time to build up progressively larger systems, such as groups and clusters of galaxies.

Galaxy clusters (e.g., Kravtsov & Borgani 2012) are the most massive gravitationally bound systems in the Universe, residing at the nodes of the cosmic web. Their number counts were initially predicted to serve as a cosmological probe, and they have since been used to constrain cosmological parameters. Clusters are also rich astrophysical laboratories: their baryonic components are widely investigated, spanning from the diffuse, hot intracluster medium (ICM), observed in X-ray emission and through the Sunyaev-Zel'dovich effect (e.g. Sunyaev & Zeldovich 1970, 1972, 1980), to their populations of largely quiescent galaxies.

Studying the progenitors of clusters – known as "protoclusters" (e.g., Overzier 2016) – at early cosmic times (particularly during "cosmic noon", at $2 \lesssim z \lesssim 4$, the epoch of peak star formation and black hole accretion) offers valuable insights into the early phases of cluster assembly. These regions also provide critical information about galaxy evolution under extreme environmental conditions.

From a theoretical point of view, studying these early structures requires modeling the highly nonlinear regime of structure formation. The most accurate approach to this problem involves numerical simulations. N-body simulations solve for the gravitational interaction among a large number of collisionless particles that trace the underlying matter distribution. To account for baryons, with their complex collisional and radiative physics, one must also model hydrodynamics and relevant radiative processes. Semi-analytic models (SAMs) offer approximate recipes to populate dark matter halos with galaxies and gas, bypassing the explicit solutions of the equations of hydrodynamics, while cosmological hydrodynamical simulations solve fluid dynamics equations numerically. However, the latter still rely on simplified effective models to describe processes occurring on unresolved scales such as star formation, feedback, and cooling, which cannot be directly modeled self-consistently.
Hydrodynamical cosmological simulations have become an essential tool for studying structure formation across cosmic time and over a wide range of spatial scales. By comparing their predictions with observations, we can test models of baryonic physics and gain insights into the interpretation of observational data.

In this Thesis, I investigate the early phases of cluster assembly – in the protocluster regime – using state-of-art cosmological hydrodynamical simulations of galaxy clusters. The goal of this work is to address open questions regarding the evolution of cluster galaxies and the emergence of the ICM. First, I focus on characterizing the properties of the galaxy population in protocluster environments at cosmic noon, aiming to provide theoretical predictions on if and how environmental processes influence galaxy evolution since the protocluster regime. Then, I analyze the early phases of ICM assembly in emerging massive halos in protocluster regions and across the transition to the first virialized clusters, in order to shed light on the timing of thermalization of ICM and on the physical processes that are contributing to it – an area where observations are only beginning to offer insight. By comparing simulated properties with recent observations of galaxies and hot gas in protoclusters, I also assess the reliability of current simulations in reproducing the conditions found in these extreme environments.

The structure of this Thesis is as follows.

Chapter 2 reviews the main properties of galaxy clusters, with a focus on their galaxy populations and the ICM, both from an observational and a simulation-based perspective. It also introduces protoclusters, including recent observational efforts to trace galaxy and (proto-)ICM evolution in these environments.

Chapter 3 introduces the cosmological code `OpenGADGET-3`, which was used to produce the simulations analyzed in this Thesis: the `DIANOGA` suite of galaxy cluster simulations. These are then presented, along with a companion run of an average cosmological region ("cosmological box" hereafter). I present key predictions of the simulations regarding galaxy stellar mass functions (GSMFs) and radial profiles of thermodynamic properties of the ICM (electron density, temperature, pressure, entropy) at $z = 0$, and compare them with observations. This Chapter concludes with a description of two tools used for generating synthetic observations from simulation outputs: `SKIRT-9` (Camps & Baes 2020), for radiative transfer of stellar light through modeled dust distributions, and `SMAC` (Dolag et al. 2005), used here to create synthetic X-ray maps of the diffuse hot gas.

Chapter 4 presents a detailed analysis of the galaxy population in protoclusters and in the cosmological box. I explore environmental effects on galaxy properties, as predicted by the galaxy formation model implemented in the `DIANOGA` simulations. I then compare the simulation predictions for protoclusters at $z = 2.2$ with observations of protocluster galaxies in the redshift range $2 \leqslant z \leqslant 2.5$. This analysis includes the post-processing of a sample of simulated protocluster galaxies with `SKIRT-9` to investigate the reliability and biases of an observational diagnostic - the UVJ diagram. The results presented in this Chapter are published in Esposito et al. (2025).

Chapter 5 focuses on the gas component of massive halos during the transition from the protocluster stage to the epoch of fully formed, virialized clusters, across the redshift range $1.5 < z < 3$. In the first part, I characterize the phases of gas within massive halos and trace the build-up of the ICM across this transition. In the second part, I compare simulation predictions with observations of ICM thermodynamic properties in high-redshift clusters and protocluster environments. The results of these analyses will be published in an upcoming article (Esposito et al., 2025, in preparation). This Chapter includes a comparison between synthetic X-ray surface brightness profiles of simulated halos with observations of the first confirmed system at $z > 3$, showing evidence of a primordial ICM, as presented in Travascio et al. (2025b, including M. Esposito).

Chapter 6 summarizes the main findings of this Thesis and outlines future directions for related projects.

Appendix A presents a complementary project based on the application of deep learning techniques for efficient post-processing of cosmological simulations. Here I introduce `SubDLe`: a deep learning-based algorithm for subhalo identification. I show its performance on the total mass distribution in a simulated cluster and then focus on testing on stellar distribution alone. The results of this work are presented in Esposito et al. (2024).

Appendices B and C provide a brief theoretical background on cosmology and on the linear growth of cosmic perturbations, respectively.

Appendix D offers a technical overview of numerical techniques in modern N-body codes, and on Smoothed Particle Hydrodynamics (SPH), the hydrodynamical scheme used in the simulations analyzed throughout this Thesis.

2

GALAXY CLUSTERS AND PROTOCLUSTERS

This chapter provides a general overview of galaxy clusters and their high-redshift progenitors, commonly referred to as protoclusters. The first part (Sect. 2.1) introduces low-redshift galaxy clusters; the observational methods used to identify these systems are discussed in Sect. 2.1.1. Section 2.1.2 reviews key observational findings on cluster galaxy populations, and briefly summarizes predictions from cosmological hydrodynamical simulations. Section 2.1.3 introduces the intracluster medium (ICM), discussing its observed properties and the dichotomy between cool-core and non-cool-core clusters. This section also highlights the achievements and limitations of current simulations in reproducing observed ICM properties. Section 2.2 introduces protoclusters, regions in the high-redshift Universe expected to collapse into clusters by $z = 0$. Section 2.2.1 reviews the main techniques currently employed to identify high-redshift clusters and protoclusters. Finally, Sects. 2.2.2 and 2.2.3 discuss the observed properties of galaxies in protoclusters and high-redshift clusters, and the early phases of ICM assembly in these systems, respectively. Both sections include a discussion on theoretical predictions from cosmological hydrodynamical simulations.

2.1 GALAXY CLUSTERS IN THE NEARBY UNIVERSE

Galaxy clusters are the most massive nodes in the cosmic web, representing the final stage of hierarchical structure formation in a ΛCDM Universe. They are conventionally associated with virialized dark matter (DM) halos of mass $M_{200} \geqslant 10^{14}\ M_\odot$, where M_{200} is the mass enclosed within a spherical region of radius R_{200}, centered on the halo, such that the mean enclosed density is 200 times the critical density of the Universe. Galaxy clusters are populated preferentially by galaxies whose star formation has been mostly quenched. A significant fraction ($85-90\%$) of a cluster's baryonic content ($15-17\%$ of their total mass) resides in the ICM (e.g., Donahue 1993; Kravtsov & Borgani 2012), a hot, pressurized and diffuse plasma that emits in the X-ray band[1]. At the center of most clusters reside Brightest Cluster Galaxies (BCGs; e.g., Collins et al. 2003; Chu et al. 2021, 2022), the most massive and luminous galaxies in the Universe whose growth is shaped by both the cooling of the ICM and feedback from active galactic nuclei (AGN). Surrounding these galaxies is the intracluster light (ICL; e.g., Contini 2021; Montes 2022), composed of stars that are not gravitationally

1 Despite the generic term "intracluster medium", this component typically refers to the hot, diffuse phase of the gas permeating galaxy clusters; however, cooler gas phases have also been observed, traced by Hα and CO emission (e.g., Salomé et al. 2006; McDonald et al. 2010).

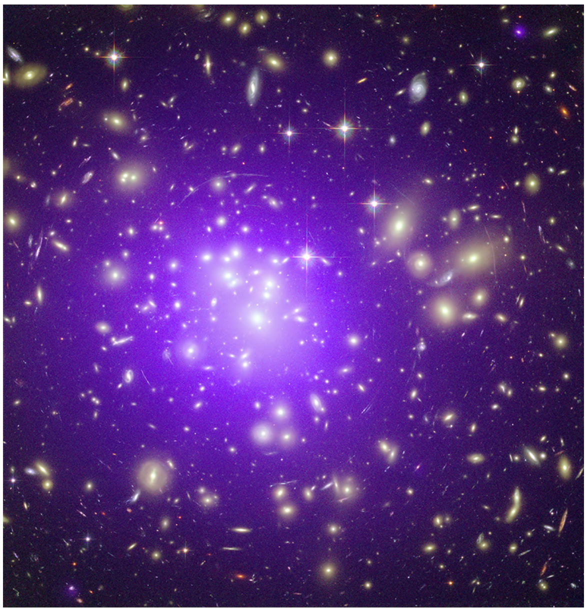


Figure 2.1: Composite image of X-ray and optical views of Abell 1689 at z=0.18. The purple halo shows the distribution of the ICM observed by Chandra X-ray Observatory. The underlying optical image was taken with Hubble Space Telescope. Image credits: X-ray: NASA/CXC/MIT/E.-H Peng et al; Optical: NASA/STScI.

bound to any individual galaxy, most of which have been likely stripped during galaxy–galaxy interactions and mergers throughout the cluster's hierarchical assembly. The dynamics of galaxy clusters provided the first observational evidence for the existence of DM, through velocity dispersion measurements of galaxies in the Coma Cluster by Zwicky (1933). Clusters thus represent both laboratories for studying baryonic physics in extreme environments and sensitive cosmological probes of the matter content and structure formation in the Universe. Figure 2.1 shows a composite image, in optical and X-ray, of the local cluster Abell 1689. The X-ray emission traces the distribution of the ICM, overimposed on the optical image of galaxies, including the central BCG, and of the surrounding ICL.

2.1.1 *Search for low-redshift clusters*

Galaxy clusters can be identified across a wide range of wavelengths, using different observational techniques, which are sensitive to distinct physical components of the cluster. In optical and near-infrared (NIR) surveys, one of the most widely used methods relies on the presence of the tight "red sequence" of early-type galaxies in color–magnitude space (see Sect. 2.1.2). These galaxies, which dominate the cluster galaxy population, exhibit old stellar populations and hence red optical colors, thus forming a distinct locus in color-magnitude diagrams that enables efficient identification of overdensities even in photometric data sets. Large cluster samples have been identified with this technique, including the Red-Sequence Cluster

Survey (RCS; Gladders & Yee 2005) and RCS-2 (Gilbank et al. 2011), and the red-sequence Matched-filter Probabilistic Percolation (redMaPPer) catalogs (Rykoff et al. 2014).

Alternatively, clusters can be identified via their diffuse X-ray emission, originating from the ICM, which traces the deep gravitational potential wells of massive halos. X-ray detection provides valuable information about the thermodynamic state and mass of the cluster, although it is biased toward relaxed, centrally concentrated systems. Cluster samples have been assembled through the detection of the diffuse ICM emission from X-ray satellites, such as ROSAT, Chandra X-ray Observatory, XMM-Newton, and, most recently, eROSITA (e.g., Böhringer et al. 2000, 2004; Ebeling et al. 2001; Burenin et al. 2007; Mehrtens et al. 2012; Pierre et al. 2016; Kluge et al. 2024; Bulbul et al. 2024).

A complementary approach exploits the Sunyaev-Zel'dovich (SZ) effect (Sunyaev & Zeldovich 1970, 1972, 1980), a spectral distortion of the CMB caused by inverse Compton scattering off free electrons in the ICM (see Carlstrom et al. 2002, for a review). In strong contrast with the ICM detection via X-ray emission, the ICM detection via SZ effect is nearly redshift-independent, making it especially powerful for detecting massive clusters out to high redshifts. This technique has enabled the compilation of galaxy cluster catalogs through SZ sky surveys with the Atacama Cosmology Telescope (ACT; Hilton et al. 2021; Klein et al. 2024a; ACT/DES/HSC Collaboration 2025), the South Pole Telescope (SPT; Bleem et al. 2015, 2020, 2024; Klein et al. 2024b), and the Planck satellite (Planck Collaboration XXVII 2016).

The combination of these different, independent techniques improves completeness, purity and characterization of the selected cluster samples, and mitigates selection biases inherent in any single method.

2.1.2 *The population of cluster galaxies*

Observations up to $z \sim 1$ have shown that cluster environments are ubiquitously characterized by a population of massive galaxies with a higher fraction of quiescent sources compared to average field regions (e.g., Haines et al. 2015; van der Burg et al. 2018, 2020). It is indeed established that in the local universe a star formation rate (SFR) - density relation (Dressler 1984; Lewis et al. 2002; Peng et al. 2010) is in place, with local clusters hosting a larger fraction of galaxies whose star formation activity was quenched (quenched galaxies[2]) in comparison with the general field. Quenched galaxies populate the red sequence in color-magnitude (or color-color) diagrams, as shown in Fig. 2.2, for galaxies in the Coma Supercluster (Gavazzi et al. 2010). Early-type galaxies, characterized by a bulge-dominated morphology (Hubble 1926) – typically associated with quenched galaxies, marked in red – form a tight sequence at red g–i colors, while late-type galaxies, with disk-dominated structure (generally star-forming, marked in blue and green)

2 Quenched and star-forming galaxies are typically classified on the basis of their specific star formation rates (sSFR; e.g., Franx et al. 2008) or through their photometric properties, for example using the color-color UVJ diagram (e.g., Williams et al. 2009).

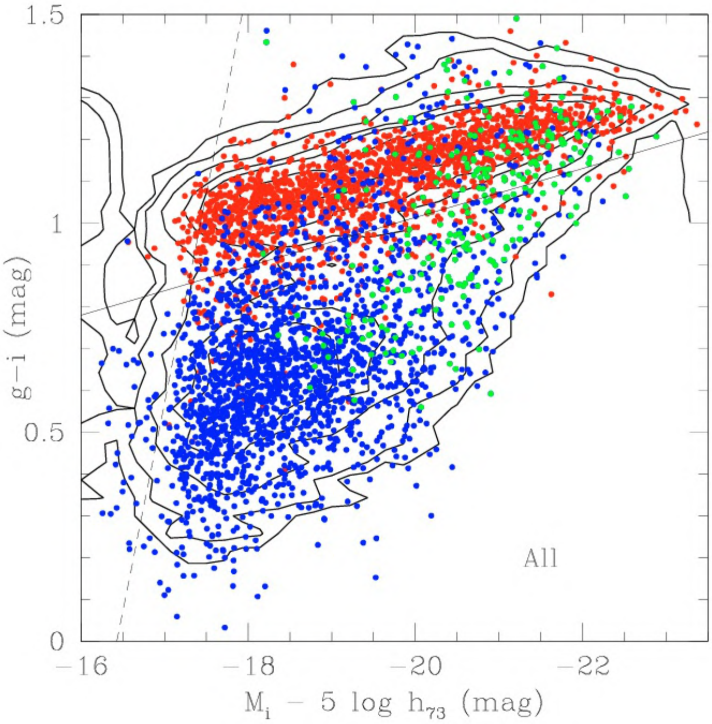


Figure 2.2: Color-magnitude diagram of galaxies in the Coma Supercluster, from the Sloan Digital Sky Survey (SDSS). Different colors mark the morphological type according to Hubble's classification: red for early-type galaxies, green for bulge-dominated spirals and blue for disk galaxies, classified though inspection of SDSS images and spectra. Reproduced from Gavazzi et al. (2010).

populate the so-called "blue cloud" at bluer colors. Quenched galaxies account for the majority of the population in a typical low-redshift cluster, making the emergence of the red sequence a key aspect of cluster galaxy evolution.

The high fraction of quenched galaxies in clusters points to the crucial role of environment in driving galaxy evolution. In particular, galaxy clusters – which originate from the highest peaks in the primordial density field – represent biased and extreme regions where evolutionary processes are accelerated. As such, they serve as natural laboratories for studying the most extreme environmental mechanisms that shape the transformation of galaxies.

Observations have revealed a complex interplay of mechanisms responsible for quenching star formation in galaxies (e.g., Fraser-McKelvie et al. 2018), driven by both internal and external processes, referred to as "mass quenching" and "environmental quenching", respectively (Peng et al. 2010).

The main processes affecting galaxy evolution in cluster environments include:

- Hydrodynamical processes: the gas reservoirs of galaxies are depleted due to interactions with the ICM. These include ram-pressure stripping (Gunn & Gott 1972), starvation (or strangulation; Larson et al. 1980), viscous stripping (Nulsen 1982), and thermal evaporation (Cowie & Songaila 1977). These mechanisms proceed outside-in, first suppressing star formation in the outer regions of galaxies.

- Interactions: these include galaxy-galaxy interactions or perturbations induced by the halo, called tidal interactions. Galaxy interactions can take place in the form of mergers, less disruptive interactions, or simple fly-bys, whose cumulative effect, at high speed, is called "harassment" (Moore et al. 1996). These processes are capable of inducing instabilities in the galactic disks, which may drive gas inflows to the center of the galaxy and trigger starburst events or AGN activity.

- Internal processes: these operate also in isolated galaxies, but may be enhanced or happen at earlier times in overdense environments. These processes include galactic outflows driven by Supernovae (SN) and AGN feedback, that can heat and/or expel gas. AGN feedback is thought to be particularly effective in quenching massive galaxies ($\log_{10}(M_*/M_\odot) > 10.5$; e.g., Boselli & Gavazzi 2006; Wylezalek & Zakamska 2016). Quenching due to these processes is expected to proceed inside-out.

Observations carried out at high resolution (e.g., Schaefer et al. 2017; Fossati et al. 2018; Poggianti et al. 2019; Bluck et al. 2020) can probe spatially-resolved signatures of quenching across galaxy disks, which allows to distinguish between inside-out and outside-in processes, as well as to identify disturbed morphologies and tidal features. On the other hand, statistical studies of galaxy populations (e.g., Kauffmann et al. 2004; Peng et al. 2010; Darvish et al. 2016; Davies et al. 2019) can reveal trends in galaxy properties

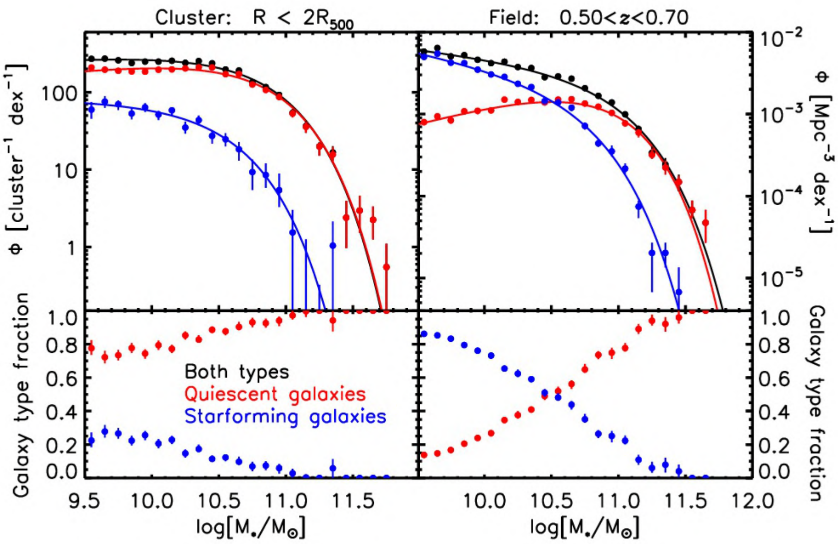


Figure 2.3: GSMFs of cluster galaxies, from a sample of 21 clusters at 0.5<z<0.7 (top-left panel, van der Burg et al. 2018), compared with field galaxies at the same redshift (top-right panel). The GSMFs are shown for all galaxies and for quiescent (quenched) and star-forming populations separately, classified according to U-V and V-J colors. The bottom panels show the fractions of quenched and star-forming galaxies as a function of stellar mass in the two coeval environments. Reproduced from van der Burg et al. (2018).

with respect to environmental proxies, leveraging much larger samples. The definition of galaxy environment is, to some extent, arbitrary, as its optimal choice depends on the physical environmental drivers of galaxy formation and evolution – precisely the processes we aim to constrain. Commonly used definitions of environment therefore include the projected cluster-centric radius, the local galaxy density, and the halo mass.

Fundamental properties of galaxy populations that are investigated, also in connection with the environment, include the fraction of quenched galaxies (quenched fraction), the galaxy stellar mass function (GSMF) and the star formation rates (SFRs). In the following, I review some key results concerning these properties in cluster galaxies.

2.1.2.1 *Galaxy stellar mass functions*

The GSMF is the number density of galaxies in stellar mass bins. It encodes information on all the processes that have contributed to shape the star formation history and the stellar mass assembly of a galaxy population and is thus a fundamental observable in galaxy evolution studies.

In cluster environments, the GSMF shows significant differences compared to the field up to $z \sim 1.5$ (Annunziatella et al. 2014, 2016; van der Burg et al. 2013, 2018, 2020). An excess of massive galaxies is generally found in clusters relative to the field. However, this difference vanishes when considering only star-forming galaxies, a result that holds up to $z \gtrsim 1$ (e.g., van der Burg

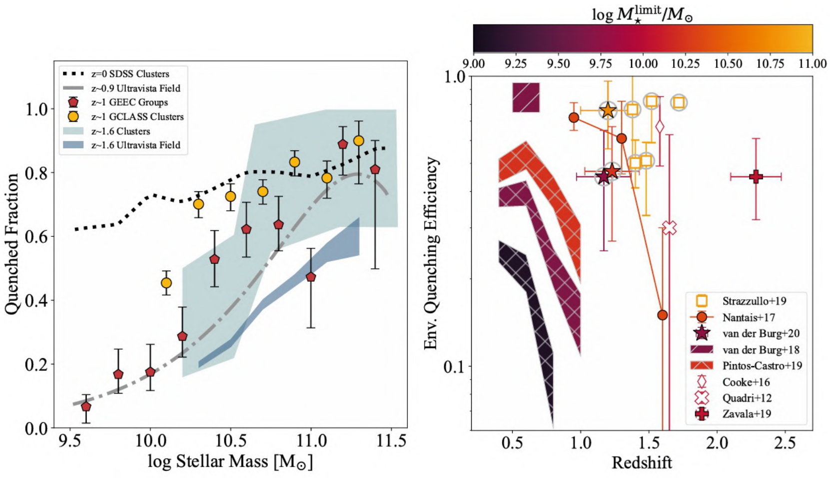


Figure 2.4: Left panel: quenched fraction as a function of stellar mass in clusters and groups at $z \sim 1$ (Balogh et al. 2016, yellow and red points, respectively), and clusters at $z \sim 1.6$ (green shaded region; Lee-Brown et al. 2017), compared to coeval field samples from UltraVISTA (dashed-dotted line at $z \sim 1$ and blue shaded region at $z \sim 1.6$) and SDSS local clusters (dotted line). Right panel: the environmental quenching efficiency in massive clusters (and two protoclusters) is shown as a function of redshift, color coded according to the stellar mass limit of each sample. Observations are taken from Quadri et al. (2012), Cooke et al. (2016), Nantais et al. (2017), van der Burg et al. (2018, 2020), Pintos-Castro et al. (2019), Zavala et al. (2019), and Strazzullo et al. (2019). Filled symbols and regions contain multiple (proto)clusters, while open symbols are single clusters. Gray hatches and circles mark $\epsilon_{\rm env}$ measured within $0.5 \times R_{200}$; all the others are within R_{200}. Reproduced from Alberts & Noble (2022).

et al. 2020). In contrast, the GSMF of quenched galaxies in clusters shows a steeper low-mass end, indicating a higher abundance of low-mass quenched galaxies relative to the field. This is illustrated in Fig. 2.3, showing the GSMFs derived from 21 massive clusters at 0.5<z<0.7, compared to a coeval field sample (van der Burg et al. 2018).

2.1.2.2 *Quenched fractions*

The quenching mechanisms listed in Sect. 2.1.2 are all expected to play a role in shaping galaxy evolution, with their relative importance depending on environment, cosmic epoch, and galaxy mass. Both mass and environmental quenching evolve with redshift, becoming more efficient at later epochs (e.g., Pintos-Castro et al. 2019).

Observations reveal that more massive galaxies begin and complete their quenching process earlier, and that, at fixed stellar mass, galaxies in clusters quench earlier than their counterparts in groups and in the field.

The quenched fraction and the environmental quenching efficiency are typically employed to quantify the level of star formation quenching in a galaxy population. Given the number of quenched galaxies N_Q and of star-forming galaxies N_{SF} in a galaxy sample, the quenched fraction is simply defined as the fraction of quenched galaxies: $f_Q = \frac{N_Q}{N_Q+N_{SF}}$. The environmental quenching efficiency (ϵ_{env}; Peng et al. 2010) quantifies the relative difference in quenched fraction between a given environment (e.g., a cluster; $f_{Q,cl}$) and a reference field ($f_{Q,field}$):

$$\epsilon_{env} = \frac{f_{Q,cl} - f_{Q,field}}{1 - f_{Q,field}}. \tag{1}$$

Observations up to at least $z \sim 1$ show a dependence of ϵ_{env} on halo mass, with massive clusters having higher quenching efficiency than coeval groups (e.g., Balogh et al. 2016). Furthermore, ϵ_{env} decreases with increasing halocentric radius in clusters (e.g., Balogh et al. 2016; Kawinwanichakij et al. 2017; van der Burg et al. 2018, 2020; Papovich et al. 2018).

Early observational studies indicated that environmental and mass quenching could be treated as independent processes in the local Universe (e.g., Peng et al. 2010; Muzzin et al. 2012; Quadri et al. 2012). More recent studies, however, revealed that environmental quenching exhibits a dependence on stellar mass as well, and mass quenching is more efficient in the inner regions of clusters, thereby challenging the assumption of their separability (Balogh et al. 2016; Kawinwanichakij et al. 2017; Pintos-Castro et al. 2019). At $z \gtrsim 1$, this could be explained by the "over-consumption" model (McGee et al. 2014): as a galaxy enters the virial region of a massive halo, it undergoes starvation; since gas accretion is suppressed, star formation and outflows, that are more intense at higher redshifts, deplete gas on a mass-dependent timescale.

Figure 2.4 shows a compilation of observational estimates of quenched fractions in clusters, groups and in the field at $z \sim 0-1.6$ (left panel, Balogh et al. 2016; Lee-Brown et al. 2017) and the environmental quenching efficiency in clusters as a function of redshift, up to $z \sim 2.5$ (right panel, Quadri et al. 2012; Cooke et al. 2016; Nantais et al. 2017; van der Burg et al. 2018, 2020; Pintos-Castro et al. 2019; Zavala et al. 2019; Strazzullo et al. 2019). The left panel shows that a SFR-density relation is already in place in clusters at $z \sim 1$. The comparison between coeval massive clusters and groups highlights that the quenched fraction depends on halo mass in the probed redshift range. The figure also highlights an increase of the quenched fraction with redshift between $z \sim 1.6$ and $z \sim 1$, both in clusters and in the field, across all the probed stellar masses in the field, and at low masses ($\log(M_*/M_\odot) \lesssim 10.5$) in clusters. Figure 2.4 also shows the typical increasing trend of quenched fractions with stellar mass, in all environments and at all redshifts.

The right panel shows the redshift evolution of environmental quenching efficiency up to $z \sim 2.5$, including a heterogeneous collection of data – in terms of stellar mass limits, halo mass range and apertures around the cluster center. Despite the complication of comparing such diverse data, clear trends emerge: the environmental quenching efficiency is a strong function of redshift and is higher at higher stellar masses; more massive clusters (van der Burg et al. 2018, compared to Pintos-Castro et al. 2019) have higher quenching efficiencies at fixed stellar mass limit and redshift. Furthermore, on a more qualitative note, quenched galaxy populations can be found in some clusters out to $z \sim 1.7$ at high stellar mass (e.g., Cooke et al. 2016; Strazzullo et al. 2019), especially in the core region, and even in protoclusters at $z \sim$ 2-2.5 (Zavala et al. 2019, see Sect. 2.2.2 for further discussion on this). The general picture that emerges is that quenched fractions and environmental quenching efficiency strongly depend on stellar mass, halo mass, location within the halo and redshift, highlighting the high complexity and multidimensional dependency of the mechanisms driving star formation quenching.

2.1.2.3 *Star formation rates*

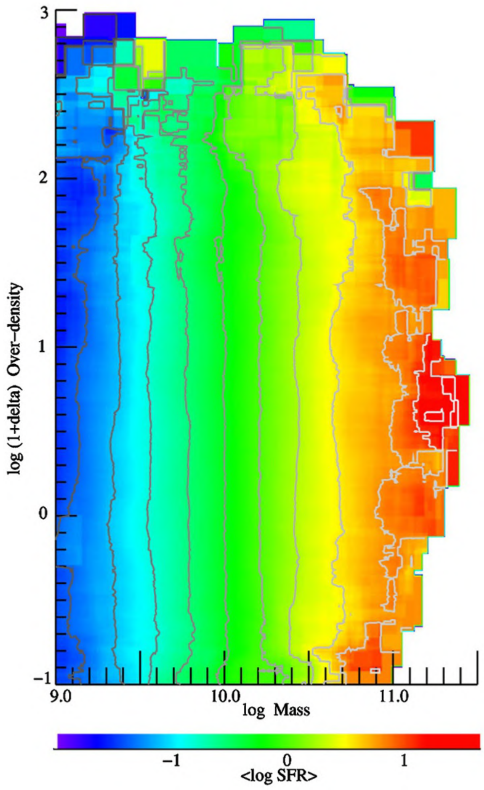


Figure 2.5: Mean SFR of star-forming galaxies at 0.02<z<0.085, as a function of stellar mass and environment – represented by galaxy overdensity – from the SDSS. Reproduced from Peng et al. (2010).

Several studies have shown that the star formation activity of star-forming galaxies is largely independent of environment, at least up to $z \sim 1$ (e.g., Peng et al. 2010; Muzzin et al. 2012). This is illustrated in Fig. 2.5, which shows the mean SFR of local star-forming galaxies from SDSS as a function of stellar mass and local galaxy overdensity, used here as a proxy for environment. While the galaxy SFRs show a clear trend with stellar mass, with more massive galaxies exhibiting higher SFRs, no clear dependence on environment is observed at fixed stellar mass. The lack of environmental dependence of the star-forming "main sequence" – the relation between SFR

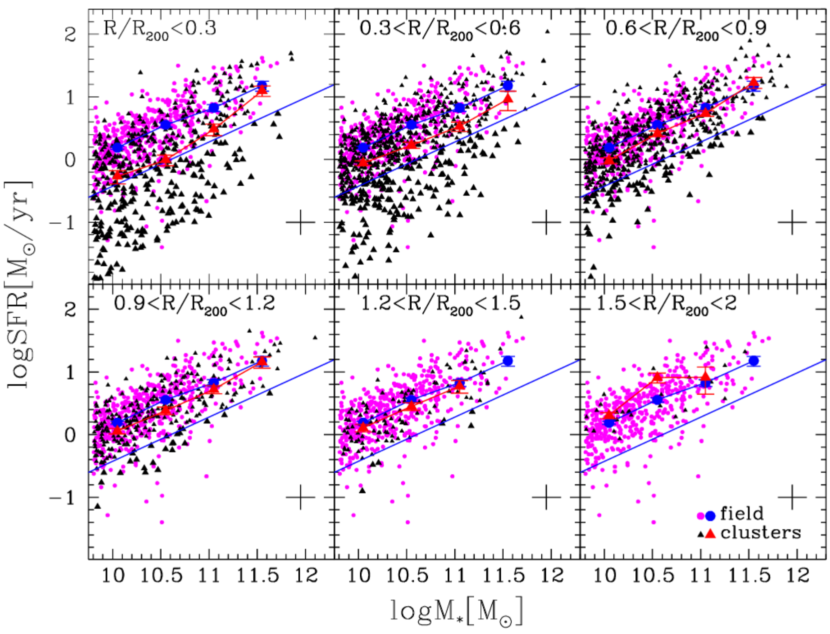


Figure 2.6: Star-forming main sequence of cluster galaxies at $0.04 < z < 0.07$ (black triangles), in 6 different cluster-centric radial bins (from top-left to bottom-right), and of field galaxies at $0.02 < z < 0.09$ (magenta circles). Blue and red symbols show the median SFRs in field and clusters, respectively. Reproduced from Paccagnella et al. (2016).

and stellar mass for star-forming galaxies (e.g., Noeske et al. 2007; Elbaz et al. 2007) – implies that the quenching process in dense environments is relatively fast, rather than a gradual suppression that would allow the detection of transition galaxies (i.e., with partially suppressed star formation). However, other studies have reported a suppression of SFRs in clusters up to $z \sim 1$, compared to the field (e.g., Vulcani et al. 2010; Paccagnella et al. 2016; Wang et al. 2018a; Old et al. 2020), pointing toward a more gradual quenching process in these environments.
For instance, Paccagnella et al. (2016) analyzed a sample of $z \sim 0$ cluster galaxies extracted from the WIde-field Nearby Galaxy-cluster Survey (WINGS; Fasano et al. 2006; Moretti et al. 2014) and its extension OMEGAWINGS (Gullieuszik et al. 2015), finding a population of galaxies with suppressed SFRs, that are rarely found in the field, as shown in Fig. 2.6.
Discrepancies in the reported environmental signatures on star formation activity may partly arise from inconsistent definitions of star-forming galaxies and differences in sensitivity across observational studies.

2.1.2.4 *Predictions of cosmological hydrodynamical simulations*

To complement observations and gain a more complete understanding of cluster galaxy formation and evolution, cosmological hydrodynamical simulations have become an essential tool. In particular, zoom-in simulations of galaxy clusters achieve high resolution while zooming on these rare systems. These simulations have made substantial progress in reproducing some ob-

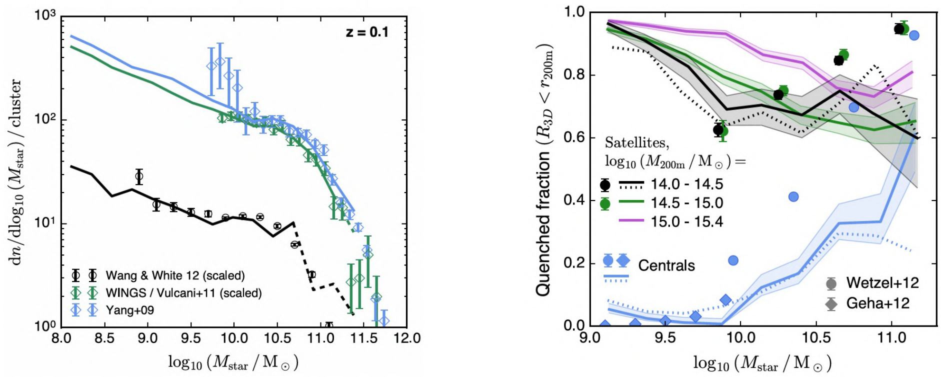


Figure 2.7: Galaxy stellar mass function (left panel) and quenched fractions as a function of stellar mass (right panel) at $z = 0$ from Hydrangea simulations (Bahé et al. 2017). The GSMF is shown in three differently colored lines, representing galaxy samples matching the selections in Vulcani et al. (2011, green) and Wang & White (2012, black), as well as the full population of satellites (blue). Quenched fractions are shown for satellite galaxies in three different halo mass bins, from Hydrangea (solid black, green and purple lines) and, in the lowest halo mass bin and in the field, from EAGLE simulations (black dotted line and blue lines, respectively; see Bahé et al. 2017 for details). Simulations are compared to observations from Wetzel et al. (2012) and Geha et al. (2012). Reproduced from Bahé et al. (2017).

served properties of cluster galaxies, although significant discrepancies with observations persist. In this section, I briefly summarize the main successes and limitations of such simulations in reproducing the observed properties of cluster galaxy populations, focusing on four zoom-in simulation suites: `DIANOGA` (e.g., Bassini et al. 2020), Hydrangea (Bahé et al. 2017), TNG-Cluster (Nelson et al. 2024), and The Three Hundred (Cui et al. 2018). For details on numerical resolution and galaxy formation models, I refer to the respective presentation papers.

The GSMF of satellite galaxies in clusters is generally well reproduced at the high-mass end by simulations, as is the overall stellar mass content in clusters (Bahé et al. 2017; Bassini et al. 2020; Cui et al. 2018, see Fig. 2.7, left panel). However, significant discrepancies remain at the low-mass end ($\log_{10}(M_*/M_\odot) \lesssim 10$–$10.5$), where simulations tend to either overproduce or underproduce satellite galaxies. This mismatch suggests that the modeling of feedback processes — particularly SN-driven winds, which are expected to be more efficient at low masses — still requires further refinement to realistically reproduce the cluster galaxy populations across all the observed stellar mass range.

A particularly challenging observable for simulations is the trend of quenched fraction of satellite galaxies as a function of stellar mass (see Fig. 2.7, right panel). Simulations of galaxy clusters often predict quenched fractions that are too high at low stellar masses, also reversing the observed trend with stellar mass (Bahé et al. 2017; but see also results from the large-volume simulations IllustrisTNG in Donnari et al. 2021). These tensions with observational results may arise from an oversimplified treatment of the interstellar medium (ISM), which could be insufficiently resilient to enhanced feedback processes and hydrodynamical interactions with the pressurized ICM in cluster environments. This clearly defines a limitation on the exploitation of cosmological hydrodynamical simulations in understanding the complexity of star formation quenching.

Another major discrepancy concerns the BCGs. Virtually all current simulations predict BCGs that are overmassive relative to observations (Bahé et al. 2017; Cui et al. 2018; Bassini et al. 2020; Nelson et al. 2024) and generally too star forming. This is likely linked to inefficient AGN feedback, which fails to quench star formation effectively in these extreme galaxies, allowing them to continue their star formation activity at too high levels at low redshift.

In summary, state-of-the-art cosmological hydrodynamical simulations are still affected by some significant limitations concerning the predictions of galaxy population properties in clusters. While some global observed properties are well reproduced, the details of galaxy evolution in the extreme environments of massive clusters are still at odds with observations. One of the most important challenges remains a realistic modeling of quenching of star formation at the opposite extremes of the stellar mass distribution – in low mass satellites and in massive BCGs.

2.1.3 *Intracluster Medium*

The ICM is a hot, fully ionized and diffuse plasma that fills the virialized region of galaxy clusters, with typical temperatures in the range 10^7–10^8 K and electron number densities between 10^{-2} and 10^{-5} cm^{-3}. It was first discovered through its thermal bremsstrahlung emission in the X-ray band in the Coma cluster (e.g., Gursky et al. 1971; Meekins et al. 1971). Subsequently, it was also detected in the microwave band via the SZ effect (e.g., Birkinshaw et al. 1984).

The synergy between X-ray and SZ observations has proven crucial in constraining the radial profiles of the thermodynamic properties of the ICM, including density, temperature, pressure, and entropy[3], as first proposed by Ameglio et al. (2007). This is due to the fact the X-ray emissivity scales as the square of electron density, while the SZ signal is proportional to the gas pressure integrated along the line of sight.

The ICM forms through the gravitational collapse of gas from the filamentary intergalactic medium (IGM) into the potential well of a massive DM halo. As the gas is accreted, its kinetic energy is converted into thermal energy via adiabatic compression and shocks, heating it to temperatures of 10^7–10^8 K, where it settles into hydrostatic equilibrium.

The simplest theoretical framework for describing ICM assembly is the self-similar model proposed by Kaiser (1986). In this model, gravity is the

3 Entropy is defined as $k_B T_e n_e^{-2/3}$, where T_e and n_e are the electron temperature and number density, respectively, and k_B is the Boltzmann constant.

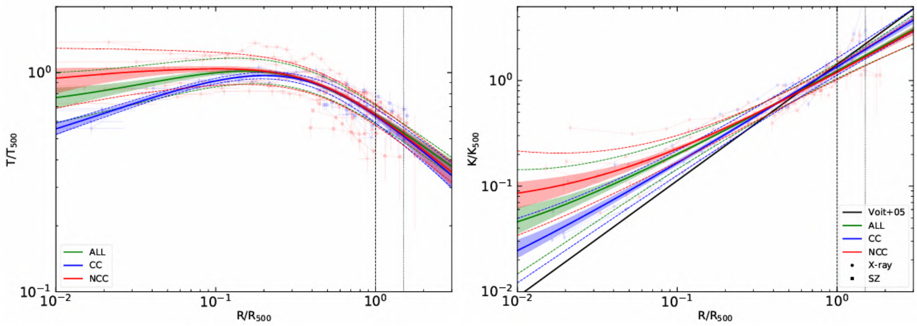


Figure 2.8: Temperature (left panel) and entropy (right panel) profiles of 12 clusters at $0.05 \lesssim z \lesssim 0.09$ from the X-COP sample (Eckert et al. 2017), from Ghirardini et al. (2019). Solid lines show analytic fits to the data for the entire sample ("ALL"), as well as for the cool-core ("CC") and non-cool-core ("NCC") populations separately, overplotted on individual data points. Shaded regions indicate the 1σ uncertainty on the best-fit relations, while the dashed lines show the intrinsic scatter. The dashed and dotted vertical lines mark the positions of R_{500} and R_{200}, respectively. Temperature and entropy profiles are normalized to the self-similar predictions at R_{500}. The black solid line in the right panel shows the theoretical entropy profile from Voit et al. (2005). Adapted from Ghirardini et al. (2019).

main driver of gas thermodynamics during hierarchical structure formation, leading to simple scaling relations between global cluster properties, such as temperature, X-ray luminosity and SZ signal, and total cluster mass. In the absence of non-gravitational processes, such as radiative cooling, star formation, and feedback, low-entropy gas is expected to sink to the inner region and high-entropy gas to rise buoyantly, expanding to large radii. These conditions were reproduced in non-radiative hydrodynamical simulations of galaxy clusters and produced self-similar thermodynamic profiles (e.g., Tozzi & Norman 2001; Voit et al. 2005). Observations generally support these predictions at large radii and up to redshifts of $z \sim 1.9$ (e.g., Leccardi & Molendi 2008; McDonald et al. 2017; Sanders et al. 2018; Ghirardini et al. 2019, 2021).

However, significant deviations from self-similar predictions are observed, particularly in cluster cores. These deviations highlight the importance of non-gravitational processes in shaping the ICM properties on small scales. As a result, the thermodynamic structure of the ICM encodes not only information about the underlying DM potential, but also about the formation and evolutionary history of clusters, including the baryonic processes involved. The profiles of thermodynamic properties of the ICM are therefore fundamental for constraining the role of baryonic physics in the evolution of cluster cores, particularly when compared to theoretical models where these processes can be isolated to test their relative importance.

2.1.3.1 *Cool-core and non-cool-core clusters*

X-ray observations have revealed efficient cooling in the cores of galaxy clusters (e.g., Fabian 1994); the associated short radiative cooling times led to the so-called "cooling flow" model, in which the gas in the cluster core was thought to cool radiatively and flow inward at high rates, potentially fueling star formation and AGN activity in the BCG. However, the measured SFRs in BCGs remain surprisingly low (e.g., McDonald et al. 2011). These observations suggest that only a small fraction of the gas cools to very low temperatures, leading to the so-called "cooling flow problem". A solution to this problem is the presence of heating mechanisms that counterbalance radiative cooling. X-ray data point to feedback from AGNs as the primary process regulating thermal losses (e.g., McNamara & Nulsen 2007), although other mechanisms — such as turbulence and gas bulk motions in the cluster core — have been proposed as alternatives or complementary channels for transporting heat from outer to inner regions (Fujita et al. 2004; ZuHone et al. 2010; XRISM Collaboration 2025).

Closely connected to this problem is the observed dichotomy between "cool-core" (CC) and "non-cool-core" (NCC) clusters. CC clusters are characterized by high central X-ray emissivity, positive entropy and temperature gradients in the core region (see Fig. 2.8), and short cooling times, while NCC clusters lack these features (e.g., Molendi & Pizzolato 2001; Pratt et al. 2010). CC clusters are the systems that were initially associated with cooling flows.

Recent observations, however, indicate that this distinction does not correspond to two separate classes, but instead reflects different stages in cluster evolution. In particular, some NCC clusters show evidence of having undergone a previous CC phase (e.g., Rossetti & Molendi 2010).

The mechanisms responsible for disrupting cool cores are still under debate. The leading hypothesis is that merger events can destroy cool cores via mixing of the convectively stable gas (Allen et al. 2001; Sanderson et al. 2006; Pratt et al. 2010; Riva et al. 2024). Such transitions from CC to NCC are expected to occur on relatively short timescales (e.g., Rossetti et al. 2011). However, the existence of relaxed NCC clusters implies that non-gravitational heating may also play a role, with AGN feedback being the most likely candidate (e.g., McCarthy et al. 2004).

At the same time, NCC clusters may potentially evolve back into CC systems through radiative cooling, although the physical mechanisms and timescales involved in this transition remain uncertain.

2.1.3.2 *Predictions of cosmological hydrodynamical simulations*

Cosmological hydrodynamical simulations are successful in reproducing key observational scaling relations of the ICM, both in X-ray and SZ measurements (e.g., Cui et al. 2018; Pellissier et al. 2023).

Simulations confirm the observational finding that, at large radii, the ICM exhibits remarkable regularity and follows self-similar predictions up to

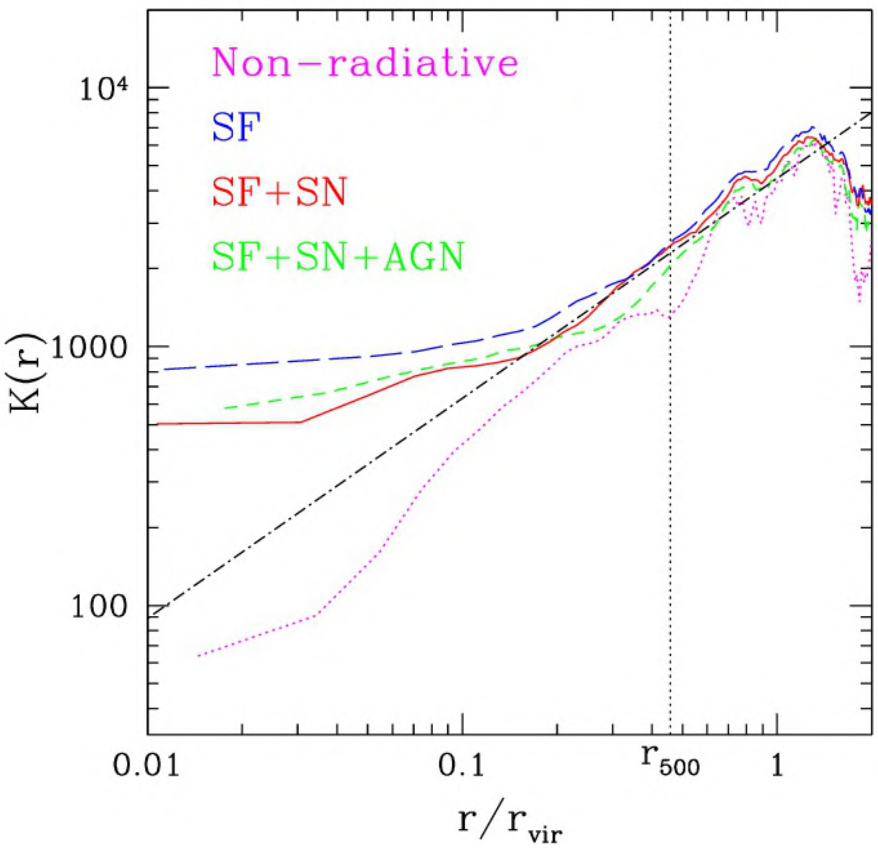


Figure 2.9: Radial profiles of entropy (in units of keV cm^2) of a galaxy cluster of mass $M_{500} = 10^{15}\ M_{\odot}h^{-1}$ at $z = 0$, simulated with GADGET-3 (Springel 2005). Different lines refer to different simulations of the same cluster, as indicated: a non-radiative version and three radiative simulations including *i*) only cooling and star formation, *ii*) also the effect of SN feedback, and *iii*) also the effect of AGN feedback. A power-law entropy profile, as predicted by non-radiative models, is also shown. The vertical dotted line marks the position of R_{500} of the cluster. Reproduced from Kravtsov & Borgani (2012).

$z \sim 2$ (Li et al. 2020, 2023). At the same time, the inclusion of complex baryonic physics introduces significant deviations from these simple scalings in the core regions of clusters.

Figure 2.9 illustrates the effect of astrophysical processes on the entropy profile of the ICM in a simulated cluster. The profiles correspond to simulations including different physical processes: radiative cooling and star formation alone (blue); the addition of SN feedback in the form of galactic winds with kinetic energy comparable to type-II SN releases (red); the further inclusion of AGN feedback (green); and a purely non-radiative run (purple). These are compared to the power-law profile expected from non-radiative models ($K \propto r^{1.1}$; Voit et al. 2005). The inclusion of non-gravitational processes offsets gas cooling in the central region ($r \lesssim 0.1 \times R_{vir}$) by either removing or redistributing low-entropy gas and/or injecting energy into the ICM, thereby creating a central entropy plateau. These results represent significant progress in addressing the longstanding "overcooling problem" – connected to the cooling flow problem in observations – that manifests as an excess of cold gas in the center of simulated clusters. However, discrepancies persist, particularly regarding the thermal structure of the ICM in the central regions of CC clusters.

In fact, while AGN feedback models have improved consistency with observations (e.g., Sijacki et al. 2007; Puchwein et al. 2008), most cosmological simulations still fail to reproduce the observed declining temperature and entropy profiles toward the center of CC clusters (e.g., Borgani et al. 2005; Borgani & Kravtsov 2011; Altamura et al. 2023). Although some idealized and zoom-in simulations can recover aspects of CC structures (e.g., Gaspari 2015; Li et al. 2015; Rasia et al. 2015; McCarthy et al. 2017; Barnes et al. 2017; Tremmel et al. 2019), most cosmological simulations predict flat entropy cores and a suppressed CC population – known as the "entropy core problem". This is thought to arise either because radiative cooling removes low-entropy gas too efficiently, allowing high-entropy gas to sink into the center of the cluster, or due to excessive heating of the gas.

The origin of the entropy core problem remains uncertain. It may stem from inaccurate assumptions in baryonic physics, the omission of key physical processes, or limitations in numerical resolution. Further complicating the picture is the fact that CC can easily be disrupted by merger events, marking their transition to NCC clusters (e.g. Burns et al. 2008; Hahn et al. 2017; Chadayammuri et al. 2021; González Villalba et al. 2025). Simulations support the idea that CC and NCC systems represent different evolutionary stages (e.g., Poole et al. 2008), with transitions between them driven primarily by mergers (e.g., Guo & Oh 2009). However, some studies suggest that mergers alone cannot fully account for the CC-to-NCC transition (Barnes et al. 2018) and, moreover, simulated CC seem to be resilient to mergers without AGN feedback (e.g., Poole et al. 2008). This suggests that non-gravitational processes play a crucial role in increasing entropy values at the cluster center, facilitating CC disruption.

Recent results from the TNG-Cluster simulations present a scenario where halos are mainly born as CC and evolve toward the NCC regime over time, with much more frequent transformations from CC to NCC than the opposite case (Lehle et al. 2025). The transition is connected to merger events, but at the same time, the cumulative effect of repeated AGN feedback episodes is found to be efficiently enhancing central entropy values. So far, though, no theoretical study has been able to determine whether CCs are more frequently destroyed via mergers and gas displacement, via AGN feedback or a combination thereof.

In summary, cosmological hydrodynamical simulations have successfully reproduced the large-scale thermodynamic properties of the ICM, including scaling relations and the self-similar behavior at large radii. However, significant discrepancies persist in the core regions, where most simulations fail to reproduce CC clusters.

2.2 PROTOCLUSTERS AND HIGH-REDSHIFT CLUSTERS

In the ΛCDM hierarchical framework of structure formation, galaxy clusters assemble over cosmic time through the accretion and merging of lower-mass DM halos. The early stages of this process are traced by protoclusters (e.g., Overzier 2016) – extended, dynamically unrelaxed structures representing the progenitor regions of present-day galaxy clusters. The term typically refers to systems identified at $z \gtrsim 2$, before the typical epoch of formation of virialized clusters with masses $M_{200} \sim 10^{14}\, M_{\odot}$.
Figure 2.10 illustrates the assembly history of a $z = 0$ cluster with $M_{200} \sim 8 \times 10^{14}\, M_{\odot}$, as traced by the Millennium-II simulation (Boylan-Kolchin et al. 2009), from $z \sim 6$ to $z = 0$ (from top to bottom). The panels show the evolution of the DM distribution on three comoving scales (100, 40, and 15 Mpc h^{-1}, from left to right), highlighting how a modest matter overdensity is already present at $z \sim 6$, characterized by a web of low-mass halos interconnected by filaments. By $z \sim 2$, a dominant central halo emerges, surrounded by multiple satellite halos that persist until $z \sim 1$. As structure formation proceeds, the central halo eventually evolves into a nearly spherical, virialized cluster by $z = 0$.
Theoretical models predict that protoclusters at $z \sim 2-6$ occupy comoving scales of ~ 10 megaparsecs, significantly exceeding the typical virial radii (of order ~ 1 Mpc) of local clusters (e.g., Chiang et al. 2013). This extended spatial scale reflects the early stage of gravitational collapse, when significant amount of material is still being accreted from the surrounding cosmic web. Figure 2.11 shows the evolution of the effective radius of (proto)clusters across cosmic time. The effective radius, introduced by Chiang et al. (2013), is defined as the mass-weighted second moment of the positions of dark matter halos that are destined to merge and form the cluster by $z = 0$. It therefore quantifies the spatial extent of the region that constitutes the progenitor of the $z = 0$ cluster, and its evolution traces the gravitational collapse and hierarchical assembly of the cluster. As a consequence, at decreasing redshift, as the overdense region collapses and virializes, the effective radius decreases, converging to ~1 Mpc by $z \lesssim 0.5$.

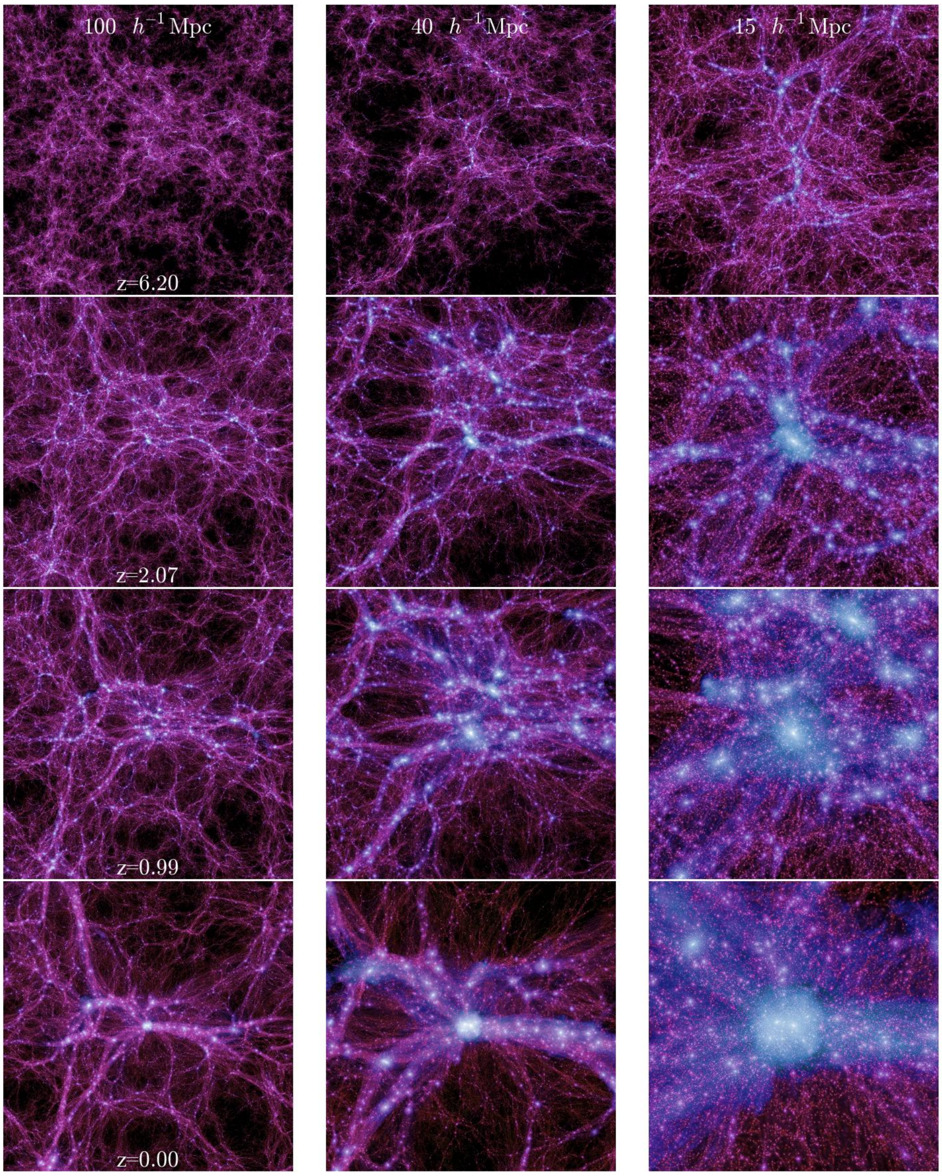


Figure 2.10: Dark matter density maps tracing the evolution of the most massive cluster ($M_{200} = 8 \times 10^{14}$ $M_{\odot}$ at $z = 0$) in the Millennium-II N-body simulation from $z \sim 6$ to $z = 0$ (from top to bottom), shown at three different comoving scales: 100, 45, 15 cMpc h^{-1} (from left to right). Reproduced from Boylan-Kolchin et al. (2009).

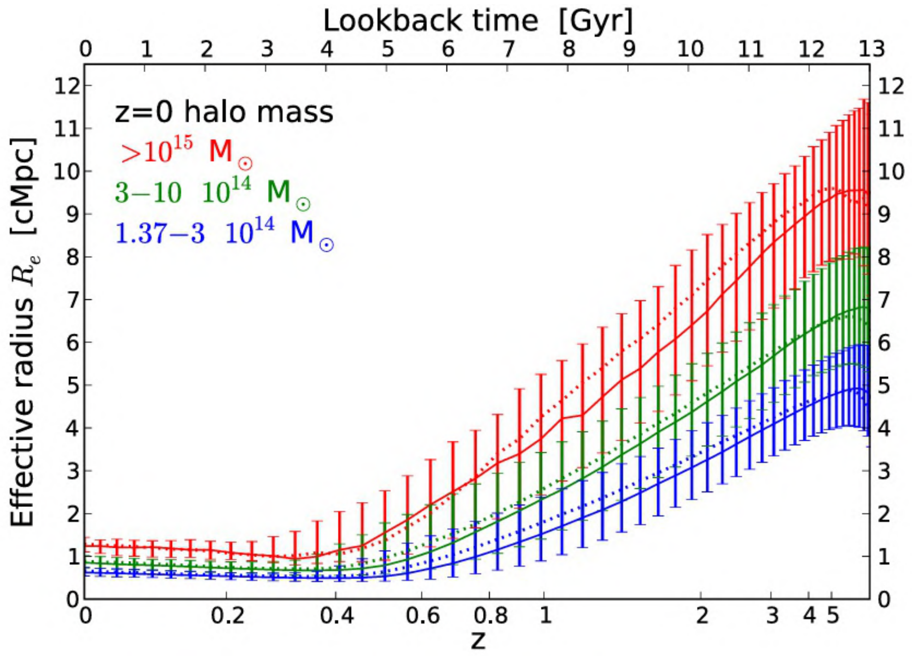


Figure 2.11: Evolution of the effective radius (introduced by Chiang et al. 2013, see text) of (proto)clusters, grouped in bins of $z = 0$ mass. Lines (errorbars) mark the median values (16th-84th percentile ranges). Solid and dotted lines show results for different assumed cosmological parameters. Reproduced from Chiang et al. (2013)

Due to their unvirialized state, protoclusters exhibit substantial substructure and dynamical complexity (e.g., Shimakawa et al. 2014), which complicates their identification in the high-redshift Universe. Nevertheless, understanding how galaxy clusters emerge from the filamentary cosmic web – through the protocluster phase and into the mature systems observed today – is key to constraining the role of dense environments in shaping galaxy evolution. During the transition from large-scale protoclusters to established clusters, galaxy populations undergo a major transformation, the suppression of star formation leading to the emergence of the red sequence observed in low-redshift clusters.

Protoclusters at cosmic noon ($z \sim 2-4$) offer a unique window onto this phase, probing galaxy evolution at a time when the cosmic star formation rate, AGN activity and merger rates reach their peaks. Located at the intersections of gas-rich filaments, these regions provide ideal conditions to probe how local and global environmental processes drive galaxy evolution. Observational evidence suggests that many protoclusters experience phases of enhanced star formation, elevated merger rates, and widespread AGN activity (e.g., Elbaz et al. 2007; Grützbauch et al. 2011; Koyama et al. 2013). On the other hand, significant diversity among observed protoclusters – together with limited statistics of the observed samples – poses challenges for developing a broad, coherent picture of their evolution. Current studies of protoclusters are hampered by small sample size, heterogeneous selection techniques, and the difficulty of confirming their future evolution into

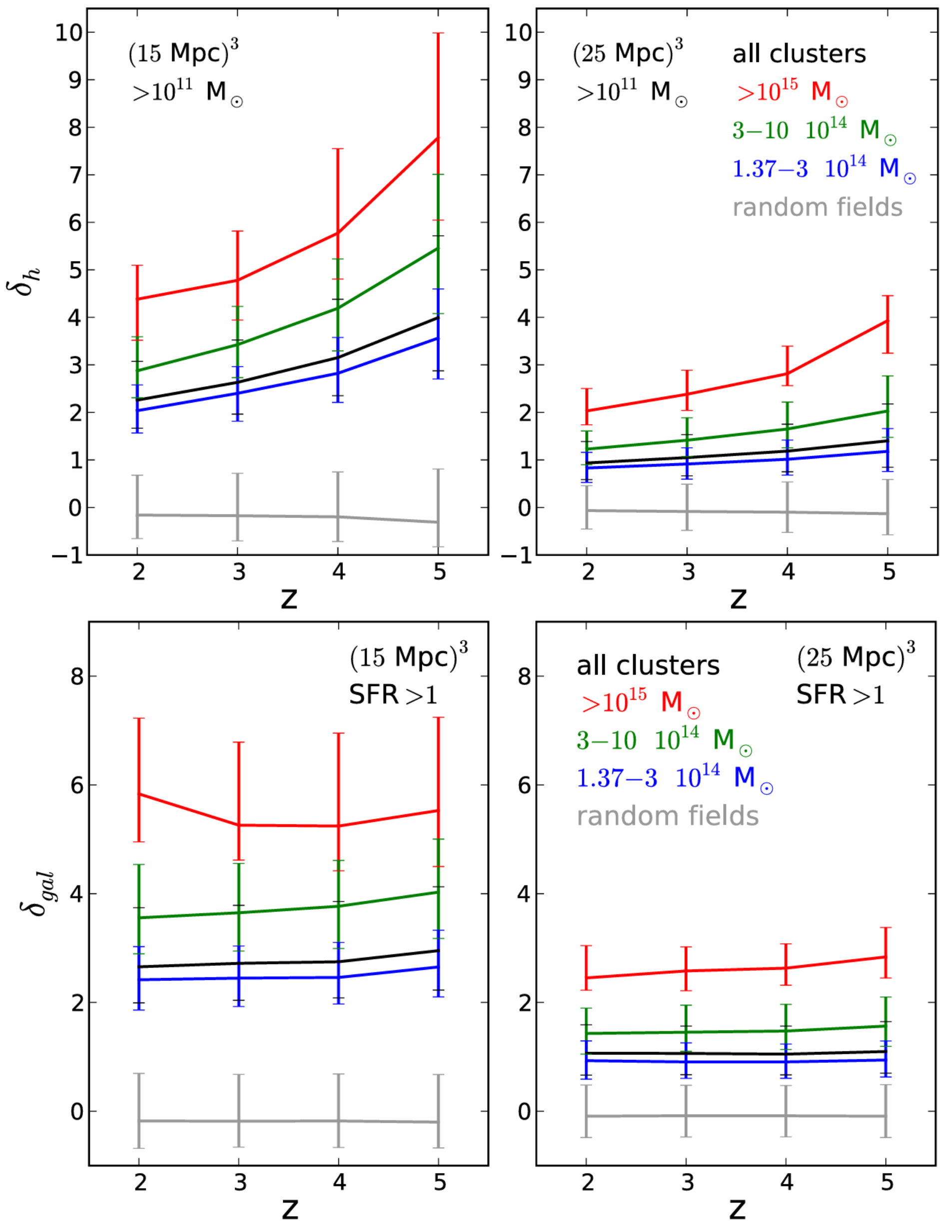


Figure 2.12: Evolution of the overdensity in the number of massive halos ($M_{200} > 10^{11}$ $M_\odot$; top) and star-forming galaxies (SFR>1 $M_\odot yr^{-1}$; bottom) in simulatd protocluster regions, on scales of (15 Mpc)^3 and (25 Mpc)^3, in the left and right panels, respectively. The full protocluster population (black line) is also shown split by halo mass at $z = 0$ (colored lines). For reference, the gray line shows the same measurement in random regions in the simulation. Adapted from Chiang et al. (2013).

clusters. Indeed, detecting a local galaxy overdensity does not guarantee that the structure will evolve into a massive cluster by $z = 0$, as predictions on mass growth and the final destiny of halos are inherently statistical. For example, Angulo et al. (2012) show that the most massive halos at early times do not necessarily evolve into the most massive clusters at late times (see also Remus et al. 2023; Onions et al. 2025). Instead, the large-scale matter overdensity on comoving scales of $\gtrsim$10–25 Mpc provides a more robust predictor of cluster progenitors (Chiang et al. 2013).
This is illustrated in Fig. 2.12, which shows the evolution of overdensities in halo number (top panels; $M_{200} > 10^{11}$ $M_{\odot}$) and in star-forming galaxy number (bottom panels; SFR $> 1\,M_{\odot}\mathrm{yr}^{-1}$) for the progenitors of clusters with different final masses at $z = 0$, relative to average cosmic regions. Both halo and galaxy number overdensities provide statistical leverage to distinguish protocluster regions from the field, with the correlation to final cluster mass improving at larger scales. These results underscore the need for wide-field surveys to efficiently identify high-redshift cluster progenitors.

Investigating protoclusters at cosmic noon also provides a unique opportunity to explore the early phases of ICM assembly. The diffuse hot plasma that dominates the virial regions of mature clusters is expected to assemble around $z \sim 1.5$–2, across the transition from the protocluster to the cluster regime. By characterizing the thermodynamic properties of the high-redshift (proto-)ICM, it is possible to place constraints on the non-gravitational processes that shape the thermal structure of the gas in the core regions of forming massive halos, before secular evolution dilutes their signatures.

2.2.1 *Identification of high-redshift clusters and protoclusters*

Many traditional techniques for identifying galaxy clusters rely on the presence of a well-defined red sequence or a mature ICM. At $z \gtrsim 1$, these methods become increasingly biased toward the most evolved systems, reaching their practical limitations around $z \sim 2$ due to the absence of mature cluster components. Nevertheless, they have provided samples of high-redshift clusters in the range $1 \lesssim z \lesssim 2$. In particular, the red-sequence selection has been successfully applied at $z > 1$ using Near- and Mid-Infrared filters (e.g., Andreon et al. 2009; Muzzin et al. 2009; Wilson et al. 2009).
As we approach $z \sim 2$, the red sequence may not be as prominent in all (proto)cluster environments as in low-redshift clusters. In addition, galaxies become intrinsically fainter, requiring much deeper observations.
Spectroscopic surveys, providing accurate redshifts, offer the most robust confirmation of overdensities associated to high-redshift (proto)clusters (e.g., Cucciati et al. 2018; Lemaux et al. 2018), and several protoclusters have been discovered as by-products of these surveys (e.g., Steidel et al. 1998, 2000, 2005). Given the significant time and resource demands of wide-field spectroscopy, many additional candidates have been identified using photometric surveys, by detecting either 3D galaxy overdensities using pho-

tometric redshifts (e.g., Eisenhardt et al. 2008; Stanford et al. 2012; Scoville et al. 2013; Thongkham et al. 2024) or projected overdensities of specifically selected galaxy types. These include, besides red-sequence galaxies, Lyman Break Galaxies (LBGs), Lyman-α and Hα emitters, passive galaxies, Distant Red Galaxies (DRGs), Dusty Star-Forming Galaxies (DSFGs), and NIR color-selected sources from IRAC imaging (e.g., Steidel et al. 2000; Andreon et al. 2009; Papovich et al. 2010; Spitler et al. 2012; Clements et al. 2014; Strazzullo et al. 2015; Wang et al. 2016; Toshikawa et al. 2016; Daddi et al. 2017; Greenslade et al. 2018; Guaita et al. 2020; Ito et al. 2023). While effective, these methods are intrinsically biased by the choice of the tracer galaxy population, potentially impacting conclusions on galaxy evolution within protoclusters. Examples of high-redshift clusters identified with these methods are CLJ1001 at $z = 2.5$ (identified as an overdensity of DRGs; Wang et al. 2016) and CLJ1449 at $z = 2$ (identified through IRAC color selections; Gobat et al. 2011).

An alternative strategy relies on "biased tracers" – rare, extreme objects expected to reside in massive, overdense environments – to pre-select likely (proto)cluster regions. These include high-redshift radio galaxies, quasi stellar objects (QSOs), and submillimeter galaxies, all thought to be preferentially located in forming clusters (e.g., Pentericci et al. 2001; Hatch et al. 2011b; Wylezalek et al. 2013a,b; Miller et al. 2018; Koyama et al. 2021). In particular, high-redshift radio galaxies are thought to be progenitors of local BCGs (e.g., Best et al. 1998; Zirm et al. 2003; Miley & De Breuck 2008; Collet et al. 2015). Many protoclusters have been identified as overdensities of Lyman-α or Hα emitters around such radio sources (e.g., Pascarelle et al. 1996; Le Fevre et al. 1996; Pentericci et al. 2000; Kurk et al. 2000, 2004; Venemans et al. 2002, 2004, 2005, 2007; Chiaberge et al. 2010; Kuiper et al. 2011; Hatch et al. 2011a,b; Mawatari et al. 2012; Hayashi et al. 2012; Wylezalek et al. 2013a; Cooke et al. 2014), with follow-up studies confirming the presence of diverse galaxy populations, including red galaxies, LBGs and DSFGs (e.g., Overzier et al. 2008; Hatch et al. 2011b; Dannerbauer et al. 2014; Zhang et al. 2024, 2025). Other potential tracers of overdensities include Lyman-α blobs (LABs) and nebulae. LABs trace extended reservoirs of cold gas ionized by intense UV radiation, typically from AGNs or star-forming galaxies (e.g., Prescott et al. 2008; Matsuda et al. 2012).

Only a few high-redshift clusters have been detected through their diffuse hot gas, either via extended X-ray emission or the SZ effect (e.g., Stanford et al. 2001; Mullis et al. 2005; Santos et al. 2011; Willis et al. 2020; Mantz et al. 2020). However, the faintness of the signal due to the immature state of the (proto-)ICM at these epochs, together with observational limitations, poses challenges beyond $z \sim 1.5 - 2$. A notable example is XLSSC 122 at $z \sim 2$ (Willis et al. 2020), detected in the X-ray band by the XMM-LSS and XMM-XXL surveys (Pierre et al. 2004; Pierre et al. 2016) and later spectroscopically confirmed. Figure 2.13 shows the X-ray view of XLSSC 122, as seen by XMM-Newton (Mantz et al. 2018). Subsequent SZ observations yielded a

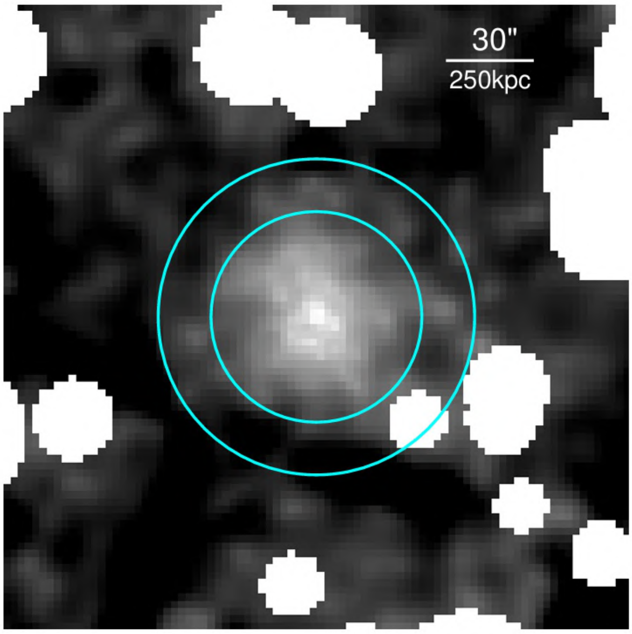


Figure 2.13: Smoothed image of the X-ray emission in the 0.5 - 3 keV band of XLSSC 122 at $z = 2$, measured from XMM-Newton. Nearby point sources are masked out. The cyan circles mark the R_{500} and R_{200} radii. Reproduced from Mantz et al. (2018).

mass of $M_{500} = 6.3 \times 10^{13}\,M_{\odot}$ (Mantz et al. 2018), placing it among the most massive (proto)clusters known at $z \sim 2$. Although often classified as a cluster, recent studies on its ICM have revealed that this system is still undergoing active assembly (van Marrewijk et al. 2024).
ICM-based selection is, to first order, independent of the properties of cluster galaxies, and thus ideally suited for studies of galaxy evolution in cluster environments. Nevertheless, X-ray bright cool-core clusters host higher density ICM in their centers, which may also influence galaxy properties. Another promising approach for identifying overdense environments involves detecting absorption features imprinted by dense gas on the spectrum of a background UV-bright source. This technique has evolved into the Lyman-α forest tomography, which enables 3D mapping of the IGM at Mpc resolution (Lee et al. 2014; Cai et al. 2016, 2017; Zheng et al. 2021).

One of the most extensively studied systems is the Spiderweb protocluster, identified around the radio galaxy MRC 1138-262 at $z = 2.16$ (Miley et al. 2006, see Fig. 2.14 for a multiwavelength view of this system). This structure presents a population of actively star-forming galaxies (e.g., Pérez-Martínez et al. 2023; Pérez-Martínez et al. 2024b), as well as an emerging red sequence (Naufal et al. 2024), a rich AGN population (Tozzi et al. 2022a; Shimakawa et al. 2024), and signatures of a forming ICM (Tozzi et al. 2022b; Di Mascolo et al. 2023; Lepore et al. 2024, see Sect. 2.2.3).
Another example is the protocluster identified in the Multi Unit Spectroscopic Explorer (MUSE) Quasar Nebula 01 (MQN01) field, around one of the brightest QSOs in the Universe, at $z = 3.25$. This region contains one of

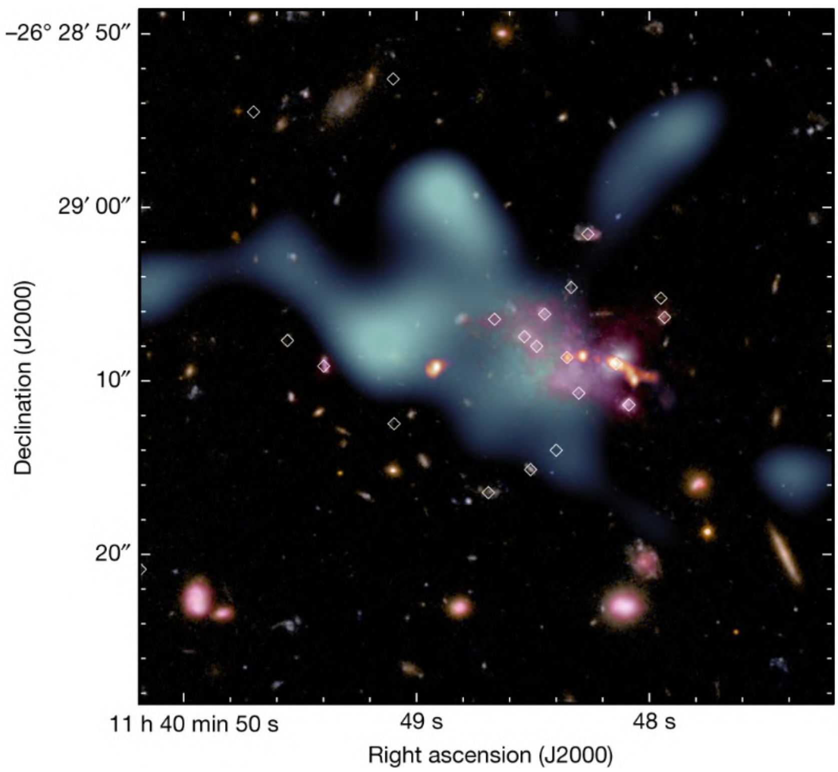


Figure 2.14: Composite image of the Spiderweb protocluster in multiple wavelengths. The optical image is from Hubble Space Telescope and overlaid are: the emission from the Spiderweb galaxy radio jet, measured with the Karl G. Jansky Very Large Array (VLA, in red); the extended Lyman-α nebula observed with FORS1 at Very Large Telescope (VLT, in pink); and the SZ signal from the Atacama Large Millimeter/Submillimeter Array (ALMA, light blue). The white diamonds denote all the spectroscopically confirmed member galaxies from Tozzi et al. (2022a). Reproduced from Di Mascolo et al. (2023).

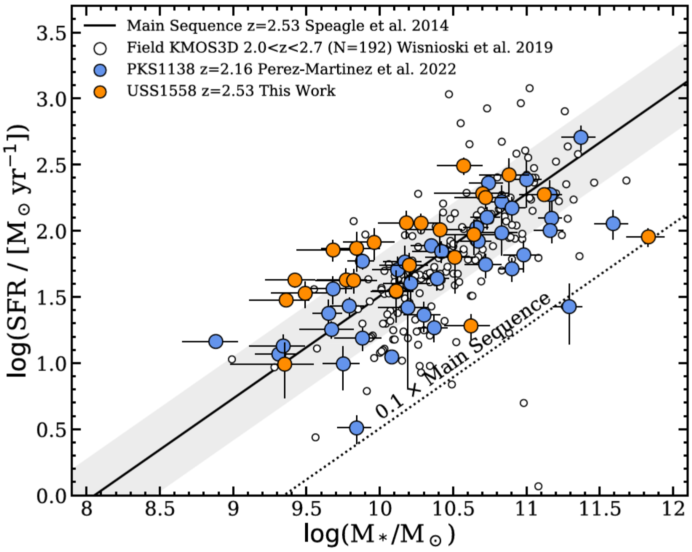


Figure 2.15: Star-forming main sequence of galaxies in the protoclusters USS1558 at $z = 2.53$ (orange circles) and Spiderweb at $z = 2.16$ (blue circles), from Pérez-Martínez et al. (2024a, 2023), compared with coeval field galaxies from Wisnioski et al. (2019, empty circles) and the main sequence fit from Speagle et al. (2014, black solid line) at z=2.53, with its 3σ scatter (shaded area). Reproduced from Pérez-Martínez et al. (2024a).

the brightest Lyman-α nebulae and the largest galaxy overdensity known at $z > 3$, and exhibits a significant overdensity of AGNs (Pensabene et al. 2024; Galbiati et al. 2025; Travascio et al. 2025a; Travascio et al. 2025b). As discussed in Sect. 2.2.3, MQN01 is the first $z > 3$ system with observed diffuse thermal emission possibly associated with a forming ICM.

2.2.2 *Galaxy populations in protoclusters and high-redshift clusters*

Over the past three decades since the detection of the first protocluster, the diverse methods for their identification have painted a heterogeneous picture of the early stages of galaxy cluster formation, revealing a vast diversity in cluster progenitor environments at high redshift. These structures allow the evolution of cluster galaxy populations to be traced back to epochs before the collapse of the first massive clusters at $z \lesssim 2$ and thus to investigate the nature and impact at early times of different kinds of environmental effects that result in environmental signatures in cluster galaxy populations at later epochs. Observations of cluster galaxies up to $z \sim 1$, as discussed in Sect. 2.1.2, show an excess population of quiescent galaxies, as well as of massive galaxies, compared to the coeval field. This implies that galaxy clusters have been sites of accelerated evolution of their galaxy populations. As a consequence, protoclusters must have been sites of enhanced star formation at some point in their evolution, reversing the trend observed at $z \lesssim 1$ (e.g., Oteo et al. 2018; Miller et al. 2018; Smail 2024). In this context,

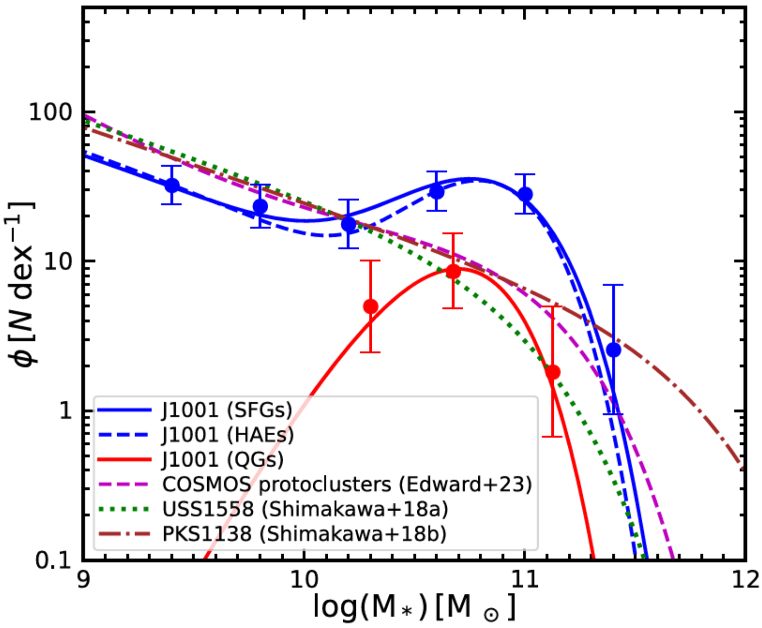


Figure 2.16: Galaxy stellar mass function in the CLJ1001 (proto)cluster at z=2.5, compared with other protoclusters at $z \sim 2-2.5$. Blue (red) solid line shows the GSMF for star-forming (quenched) galaxies in CLJ1001. The blue dashed line shows the GSMF of Hα emitters only, for a fair comparison with the GSMFs in Spiderweb and in USS1558 from Shimakawa et al. (2018b,a, brown dashed and green dotted lines, respectively). All the GSMFs shown are re-scaled to match the normalization of CLJ1001. Reproduced from Sun et al. (2024).

several studies have investigated the SFRs of individual protocluster galaxies compared to field counterparts, looking for environmental signatures on the star-forming main sequence. Some of these studies have reported enhanced star formation in galaxies within protoclusters at cosmic noon (e.g., Wang et al. 2016; Shimakawa et al. 2018a; Monson et al. 2021; Pérez-Martínez et al. 2024a; Staab et al. 2024; Daikuhara et al. 2024). However, other studies have found that SFRs in protoclusters are consistent with the field main sequence at the same redshift (e.g., Cucciati et al. 2014; Koyama et al. 2021; Polletta et al. 2021; Shi et al. 2021; Pérez-Martínez et al. 2023, 2025). This is shown in Fig. 2.15, where the SFRs of galaxies in the Spiderweb protocluster and in USS1558 at $z \simeq 2.5$ (e.g., Shimakawa et al. 2014) are shown as a function of stellar mass and compared to field galaxies. Despite the two protoclusters being almost coeval, the SFRs of their galaxy populations suggest that the two structures are in different evolutionary stages. USS1558 seems to be in a phase of enhanced star formation with respect to the field, while star-forming galaxies in the Spiderweb protocluster have SFRs consistent with the field main sequence (Pérez-Martínez et al. 2023, 2024a, but see also Daikuhara et al. 2024).

Irrespective of an SFR enhancement at the time of observation, the observed protoclusters may have been sites of more intense star formation at earlier times. The GSMF provides key insight into this, as it encapsulates the

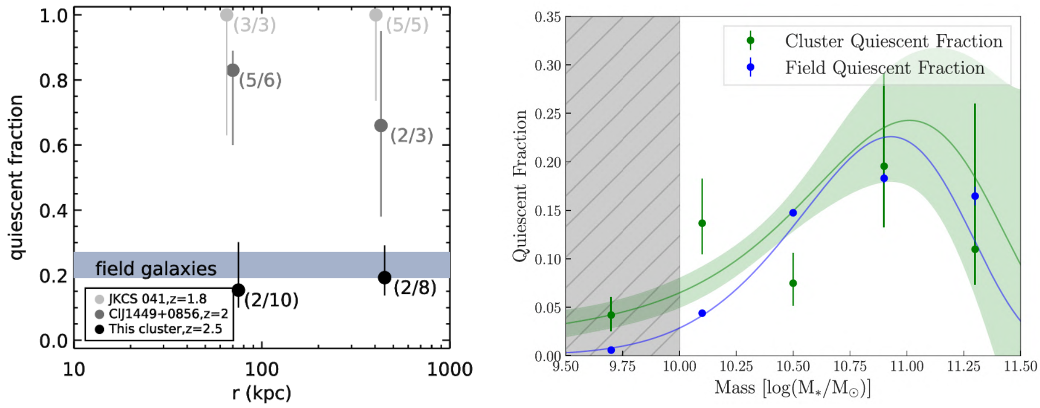


Figure 2.17: Left panel: quenched galaxy fraction (for $M_* > 10^{11}$ $M_\odot$) as a function of cluster-centric radius in the CLJ1001 protocluster ($z = 2.5$), compared with the clusters ClJ1449 ($z = 2$) and JKCS 041 ($z = 1.8$), as well as the field at $z = 2.5$. Reproduced from Wang et al. (2016). Right panel: quenched fraction as a function of stellar mass in the 14 protocluster candidates from Edward et al. (2024, in green), compared with field galaxies (in blue). Reproduced from Edward et al. (2024).

integrated star formation history, as well as information on the stellar mass assembly via mergers, of galaxies over time. Recent studies of some protoclusters at $z \sim 2-3$ have identified an excess of high-mass galaxies in dense environments compared to the coeval field (Sun et al. 2024; Forrest et al. 2024; Galbiati et al. 2025), suggesting that these environments may have undergone an enhanced star formation phase in the past. In contrast, other studies reported no significant deviation from the field GSMF in other protoclusters at $z \sim 2-2.5$ (e.g., Edward et al. 2024), indicating that their star formation history might not differ greatly from that of the field. Figure 2.16 presents a collection of GSMFs of protoclusters at $z \sim 2-2.5$, including CLJ1001 (Sun et al. 2024), Spiderweb (Shimakawa et al. 2018b), USS1558 (Shimakawa et al. 2018a) and 14 protocluster candidates identified in the COSMOS field at $z = 2-2.5$ (Edward et al. 2024). The sample of COSMOS protoclusters yielded a GSMF that is consistent with field galaxies, while the other systems present a relative excess of massive galaxies. These results, in line with those on the main sequence, highlight different evolutionary paths for different protoclusters observed in the same redshift range.

Another key area of investigation is the fraction of quenched galaxies in protoclusters, which provides insight into the efficiency of quenching mechanisms in these early stages of cluster assembly. While some studies have found that the fraction of quenched galaxies in high-redshift protoclusters is similar to field levels (e.g., Wang et al. 2016; Forrest et al. 2024; Edward et al. 2024), overdensities of passive galaxies have been observed up to $z \sim 4$ (e.g., Zavala et al. 2019; Willis et al. 2020; McConachie et al. 2022; Ito et al. 2023; Tanaka et al. 2024). Figure 2.17 (left panel) shows the quenched fraction (above the limiting stellar mass of 10^{11} $M_\odot$) as a function of halo-centric distance for the CLJ1001 protocluster ($z = 2.5$), compared to the high-redshift clusters CLJ1449 ($z = 2$, Strazzullo et al. 2013) and JKCS 041 ($z = 1.8$; Newman et al. 2014; Andreon et al. 2014). The quenched fraction in CLJ1001 remains consistent with the field level down to the inner cluster core, evidence of no environmental quenching effects on massive galaxies in this system. Conversely, CLJ1449 and JKCS 041 (at lower redshift $z \sim 2$, where the field level is ~40%) have much higher quenched fractions testifying an already efficient environmental quenching. The right panel shows the stellar mass dependence of the quenched fraction in the sample of $z = 2 - 2.5$ protoclusters from Edward et al. (2024). At $M_* > 10^{10.5}$ $M_\odot$, no significant deviation from the trend observed in the field is present, suggesting that also in these protoclusters environmental quenching is not significant. While a general trend of the emergence of environmental quenching with redshift can be identified, observations so far cannot precisely define when it starts to significantly affect the evolution of galaxies in cluster progenitors, and what drives the potentially non-negligible cluster-to-cluster variation seemingly suggested by the available measurements.

In conclusion, limited statistics and observational limitations, combined with the complex selection function of currently known protoclusters, complicate the understanding of the observed diversity in galaxy population properties in cluster progenitors.

2.2.2.1 *Predictions of cosmological hydrodynamical simulations*

Theoretical predictions from cosmological hydrodynamical simulations can contribute to improving our picture of galaxy evolution, by accounting for gas dynamics and astrophysical processes self-consistently in a cosmological context. However, these models need to be validated through a detailed comparison with observations before they can be used as predictive tools. In the very first attempts at comparing results on protoclusters from cosmological hydrodynamical simulations to observations (e.g., Saro et al. 2009), simulated protoclusters were predicted to be strongly star forming, in general agreement with the observational constraints available at the time. On the other hand, reproducing observations of galaxy populations in high-redshift protoclusters in detail has been found to be challenging for simulations (e.g., Granato et al. 2015; Bassini et al. 2020; Lim et al. 2021; Remus et al. 2023). One of the main inconsistencies lies in the underestimation of SFRs in simulated star-forming galaxies at cosmic noon, irrespective of their environment (e.g., Bassini et al. 2020; Akins et al. 2022; Andrews et al. 2024; Ragone-Figueroa et al. 2024), which also results in lower integrated SFRs than those observed in protocluster environments. Recently, Lim et al. (2024) showed that the FLAMINGO simulations can partially reconcile this discrepancy, with the most massive protocluster cores in the 1 cGpc and 2.8 cGpc boxes of the FLAMINGO flagship runs matching the observed integrated SFRs at $z \sim 2 - 3$.

The underestimation of the normalization of the star-forming main sequence in simulations is not limited to protocluster environments, it also affects cosmological boxes representing "average" regions of the Universe, equivalent to the "field" in observational studies. Thus, not only do simulations struggle to match the extreme SFRs reported for a possibly biased population of the most star-forming protoclusters, but they also underestimate field-level star formation at $z \sim 2-4$. This issue indicates a deeper inadequacy of the star formation and feedback models in capturing the details of galaxy formation and evolution, which reflects also in the limitations of simulations in reproducing the properties of galaxies at $z = 0$ (see Sect. 2.1.2.4). Despite these limitations, simulations can still provide valuable insights into the trends of galaxy population properties with their environment, even though the detailed star formation history is not yet well reproduced.

2.2.3 *The first observations of proto-ICM*

Constraining the early phases of assembly of the hot diffuse ICM is crucial for reconstructing the formation history of galaxy clusters and investigating feedback processes and their role in shaping the thermodynamics of this hot gas. Observations at high redshift are progressively pushing the detection of this component toward earlier cosmic epochs, revealing systems in which the ICM is not yet fully thermalized or centrally concentrated. However, observing the (proto-)ICM at high redshift poses significant technical challenges. Mainly, the low density and temperature of the gas in these early stages result in extremely faint signals, both in X-ray and in SZ signal, which are often close to or below the sensitivity limits of current instrumentation. Contamination from AGNs in X-ray, and radio galaxies in SZ also constitutes a significant limitation. A further complication is the angular resolution, which needs to be very high in order to carry out spatially resolved studies of gas at such high redshifts.
In this section, I review results on the very few high-redshift systems where diffuse thermal emission associated with a forming ICM has been detected.

XLSSC 122

Mantz et al. (2018) have presented the analysis of the ICM of XLSSC 122 at $z \simeq 2$, based on 100 ks observations with XMM-Newton and interferometric measurements at 30 GHz with the Combined Array for Research in Millimeter-wave Astronomy (CARMA). The X-ray data provided the bulk thermodynamic and chemical properties of this gas component: the gas temperature was measured to be $k_B T = 5.0 \pm 0.7$ keV, with an emission-weighted metallicity of $Z/Z_\odot = 0.33^{+0.19}_{-0.17}$, indicating significant metal enrichment already in place at $z \sim 2$ (Mantz et al. 2018; Willis et al. 2020). The global properties of XLSSC 122 are consistent with observations of local clusters when assuming self-similar evolution. This suggests a "simple" (self-similar) evolution of the ICM up to $z \sim 2$, at least for massive systems such as XLSSC 122.
However, a spatial offset (~ 80 kpc; van Marrewijk et al. 2024) between the peak of the X-ray emission and the peak of the SZ signal, as shown in Fig. 2.18, suggests a merging configuration. Recent high-resolution follow-up

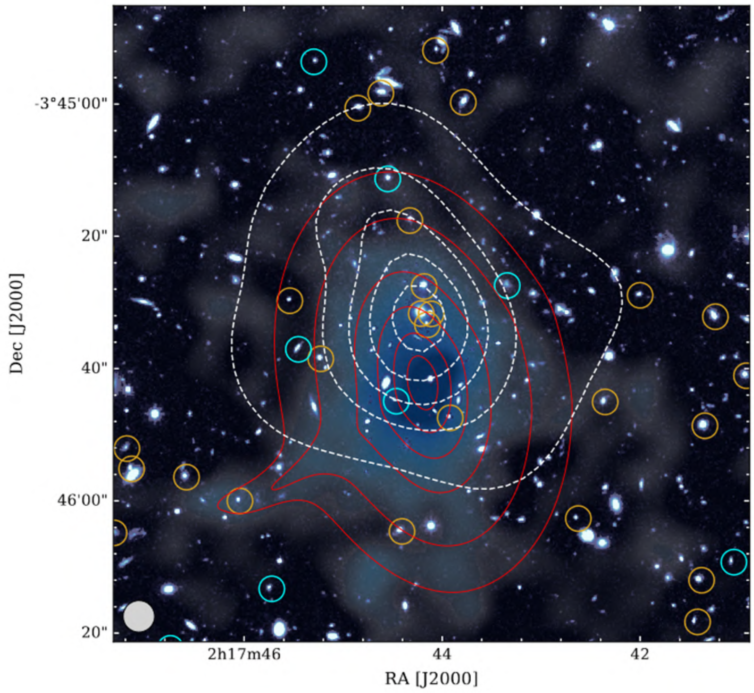


Figure 2.18: Multiwavelength view of XLSSC 122: Hubble Space Telescope image in the background; XMM-Newton smoothed contours in white; SZ flux measured with ALMA in blue. The red contours represent the best model reconstruction of the SZ signal and the circles mark cluster member galaxies (in cyan for star-forming and dusty galaxies). Reproduced from van Marrewijk et al. (2024).

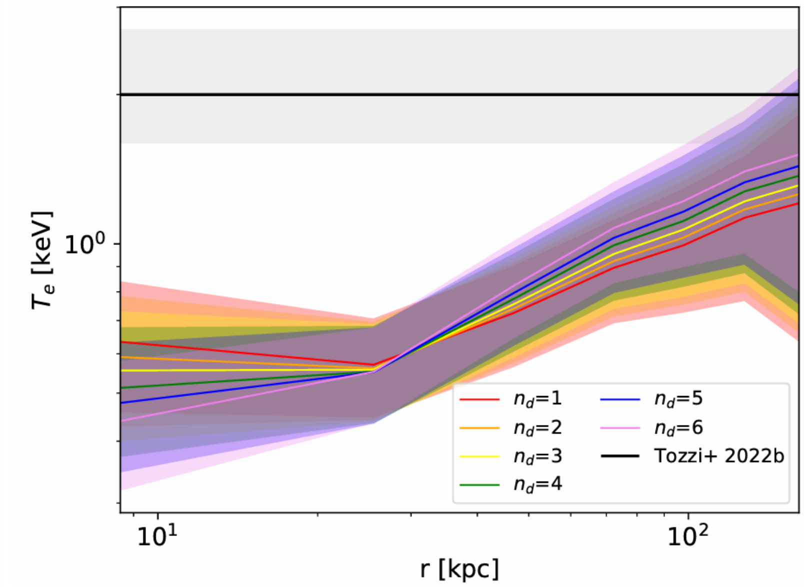


Figure 2.19: Radial temperature profiles of the proto-ICM in the Spiderweb protocluster. Different colors show the profiles for different assumptions of the n_d parameter that accounts for the unknown surface brightness in the innermost region contaminated by the radio galaxy. The solid black line shows the temperature value constrained by Tozzi et al. (2022b) by assuming an isothermal profile, with 1σ error shown by the shaded region. Reproduced from Lepore et al. (2024).

analyses in SZ have confirmed a disturbed dynamical state in XLSSC 122, pointing toward an ongoing active assembly process (van Marrewijk et al. 2024).

All together, these observations suggest that XLSSC 122 is caught in an active phase of cluster formation: the hot gas is present and enriched, but still distributed in a non-symmetric configuration.

Spiderweb

In the Spiderweb protocluster, observations from deep 700 ks Chandra exposures have revealed extended X-ray emission (Tozzi et al. 2022b), with a diffuse component interpreted as thermal bremsstrahlung from hot gas. The best-fit thermal model yielded a temperature of $k_B T \approx 2.0^{+0.7}_{-0.4}$ keV and a metallicity $Z < 1.6\,Z_\odot$. However, these estimates are subject to significant uncertainties due to the complex dynamical state of the system and the presence of non-thermal components. Independent support for the presence of a forming ICM comes from ALMA observations of the SZ effect, reported by Di Mascolo et al. (2023). The detection of a $\sim 6\sigma$ thermal SZ signal confirms the presence of a hot gas component with a pressure profile compatible with expectations for a young ICM.

More recently, Lepore et al. (2024) have combined X-ray and SZ data of Spiderweb to reconstruct, for the first time, density and temperature profiles of a forming ICM in a protocluster at $z > 2$. These profiles reveal a central

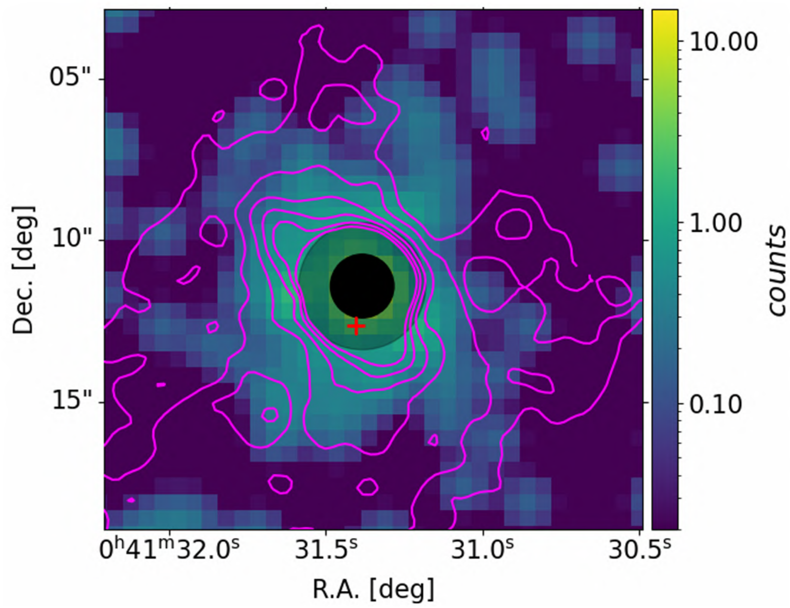


Figure 2.20: Smoothed Chandra X-ray image (map of counts) in the 0.5-2.0 keV rest-frame band of the central halo of MQN01, after subtraction of the contribution of the central QSO. The solid (shaded) black dot marks the 1 (2) arcsec radius around the QSO center. The magenta contours are Lyman-α isophotes and the red cross marks the position of a companion QSO detected with ALMA. Reproduced from Travascio et al. (2025b)

core in the proto-ICM, with relatively high density and a negative temperature gradient (see Fig. 2.19), indicative of a strong cool core, comparable to local clusters. The inferred classical mass deposition rate[4] is $250-1000$ $M_{\odot}yr^{-1}$, consistent with the FIR estimates of the SFR of the central Spiderweb galaxy (Seymour et al. 2012; Drouart et al. 2014). This suggests that the Spiderweb protocluster core might be hosting a cooling flow that feeds both the star formation and the AGN activity in the central galaxy. This phase is expected to be followed by the onset of AGN feedback, whose effect is already observable in morphological asymmetries and cavities inflated in the proto-ICM. Overall, the Spiderweb system provides a detailed snapshot of a proto-ICM in a dynamically complex environment, where gas accretion, cooling, and feedback from the central AGN are all actively shaping the thermodynamics of the diffuse gas.

MQN01

As the most distant system in which a diffuse thermal component has been detected, MQN01 pushes the study of the ICM to very early cosmic epochs, at $z = 3.25$ (Travascio et al. 2025b). The 634 ks X-ray observations with Chandra in the 0.5-2 keV rest-frame energy band (see Fig. 2.20) have highlighted a $\sim 8\sigma$ detection of diffuse emission up to 30 kpc from the central QSO,

4 The classical mass deposition rate is the rate at which gas condenses and is deposited into the central galaxy in a steady-state, isobaric cooling flow model (Fabian 1994).

consistent with thermal emission from a plasma in collisional ionization equilibrium. The combined spatial and spectral statistical analysis yielded a very steep density profile and an average temperature of the gas $k_B T = (1.8 \pm 0.4)$ keV. This system shows a very steep X-ray surface brightness profile, with higher central values than the proto-ICM in the Spiderweb protocluster. The exceptionally high X-ray luminosity of $L_X = 2.25 \times 10^{45}$ erg s^{-1} makes it a clear outlier in the L_X-T_X plane in comparison with local clusters, highlighting that the thermodynamic state of the observed gas component is significantly different from the low-redshift ICM.

2.2.3.1 *Prediction of cosmological hydrodynamical simulations*

The investigation of the early stages of ICM assembly in cosmological hydrodynamical simulations is still in its infancy, with virtually no study focusing on this specific subject. Some studies have nonetheless yielded predictions on this mostly unexplored topic. Saro et al. (2009) have presented the predictions of zoom-in cosmological hydrodynamical simulations of two galaxy clusters, focusing on their progenitors at $z = 2.1$. This analysis has highlighted that simulations predict that a hot diffuse gas component is in place in the potential well of cluster progenitors already at $z \sim 2$, with an average temperature of $k_B T = 2$ (5) keV for a poor (rich) cluster of $z = 0$ mass $M_{200} \sim 10^{14}$ (10^{15}) $M_\odot$ and with a chemical enrichment already comparable to local clusters. These predictions have been confirmed by observations in XLSSC 122 and (to some extent) in the Spiderweb protocluster, as reported above.

A hot gas component (selected to have temperatures $T > 10^7$ K) has been found to dominate the gas budget in cluster progenitor halos since $z \sim 3$ also in The Three Hundred simulations (Li et al. 2023).

In the context of comparison with the available observations, Di Mascolo et al. (2023) have shown that simulated halos produce idealized SZ signals in agreement with ALMA observations in the Spiderweb protoclusters.

These initial results mark an important first step toward bridging simulations and observations of the early phases of ICM evolution. Nevertheless, a systematic investigation of the physical origin, thermodynamic history, and eventual connection to the mature ICM at later epochs remains an open challenge for current and future cosmological hydrodynamical simulations.

3 THE DIANOGA SIMULATIONS OF GALAXY CLUSTERS

In this Chapter, I introduce the cosmological code OpenGADGET-3 (section 3.1), its gravity and hydrodynamics solver and the sub-resolution models implemented to describe astrophysical processes relevant for the simulations presented in this Thesis. Then, I describe the latest version of the DIANOGA suite of cosmological hydrodynamical simulations of galaxy clusters (Sect. 3.2), produced with OpenGADGET-3, which I have analyzed throughout this Thesis. Finally, I present some predictions of the DIANOGA simulations at $z = 0$, on the properties of the galaxy population and of the ICM.

3.1 GALAXIES WITH DARK MATTER AND GAS INTERACT (GADGET)

GADGET is a massively parallel TreeParticle-Mesh (TreePM) Smoothed Particle Hydrodyanamics (SPH) code; the simulations I have analyzed in this Thesis were produced with a developer version of OpenGADGET-3 (Groth et al. 2023; Damiano et al. 2024), successor of GADGET-3, which is an improved version of the publicly available GADGET-2 code (Springel 2005). In the following, I briefly describe how the N-body and SPH schemes presented in Appendix D are implemented in GADGET-2, followed by a summary of the improvements introduced in GADGET-3 and in OpenGADGET-3.

3.1.1 *Gravity*

Gravity is described through a hierarchical multipole expansion, which can optionally be applied in the form of a hybrid TreePM method (see Sect. D.1.4). The "tree" part of the scheme performs a grouping of particles with an oct-tree structure as described in Sect. D.1.3. The expansion of the gravitational potential is limited to monopole moments, which gives a lower complexity per evaluation at the price of more iterations compared to higher order versions. This means that, at fixed accuracy, the walk along the tree (introduced in Sect. D.1.3) takes longer, with more partial force evaluations, each of them being however easier to compute.
The opening criterion (introduced in Sect. D.1.3) adopted for a node of dimension l and total mass M at distance r is:

$$\frac{GM}{r^2}\left(\frac{l}{r}\right)^2 \geqslant \alpha|\vec{a}|, \qquad (2)$$

where α is a tolerance parameter, $\vec{a}$ is the total acceleration of the target particle obtained in the previous time step and G is the universal gravitational constant. Since $\vec{a}$ is not available for the first estimation of forces,

the geometrical opening criterion described in Sect. D.1.3 is used instead. To avoid cases where the particle being evaluated lies within a node that does not satisfy the opening criterion, an additional condition is imposed: $|r_k - c_k| \leqslant 0.6\, l$, where r_k and c_k denote the k-th components of the particle position and the node's geometric center, respectively.
The tree is also used to find neighbors in the SPH implementation (see Sect. D.2.1); for the i-th SPH particle, the walk along the tree is done with an opening criterion based on the existence of an overlap between the search region (a sphere centered at the particle position $\vec{r}_i$ with radius equal to the SPH smoothing length of the particle h_i) and the node volume.

One can opt to replace the tree with a TreePM scheme for gravity, where the walk in the tree needs to be done only locally around the target particle, while long-range forces are calculated by the PM code. This approach can result in improved performance and better accuracy in the computation of long-range forces.
The PM part of the code employs a Cloud in Cell (CIC, see Sect. D.1.2) scheme for mass assignment and the four-point differencing rule to approximate the forces by finite differences of the potential ϕ:

$$\left.\frac{\partial\phi}{\partial x}\right|_{ijk} = \frac{1}{\delta}\left[\frac{2}{3}(\phi_{i+1,j,k} - \phi_{i-1,j,k}) - \frac{1}{12}(\phi_{i+2,j,k} - \phi_{i-2,j,k})\right], \tag{3}$$

where δ is the grid spacing and the presence of the indexes ijk indicates that potential is evaluated at position (r_i, r_j, r_k) on the grid.
In GADGET-2 there is also the possibility of applying a two-level hierarchy of meshes, useful for simulations with a high dynamic range, for which the TreePM is not very efficient (see Springel 2005 for details).
Since the dynamical range of timescales is large in cosmological simulations and evolving all particles with the smallest time step would be a waste of computational resources, the time integration in the TreePM algorithm is performed with individual and adaptive time steps, using a generalization of the leap-frog integration schemes. The time step criterion for DM particles adopted by GADGET-2 is:

$$\Delta t = \min\left(\Delta t_{max}, \sqrt{\frac{2\eta\epsilon}{|\vec{a}|}}\right), \tag{4}$$

where η is an accuracy parameter, ϵ is the gravitational softening (introduced in Sect. D.1.1) and Δt_{max} is the maximum allowed time step, usually a small fraction of the dynamical time of the simulated system. Further details on the gravity solver of GADGET-2 can be found in Springel (2005). This implementation is largely unchanged in OpenGADGET-3.

3.1.2 *Hydrodynamics*

The gas dynamics is followed with an SPH formulation (described in Sect. D.2.1) which manifestly conserves energy and entropy, and makes use of an artificial viscosity in order to resolve shocks.

GADGET-2 uses a density estimate in the form of Eq. 119 and the following interpolation kernel:

$$W(r,h) = \frac{8}{\pi h^3} \begin{cases} 1 - 6R^2 + 6R^3 & 0 \leqslant R \leqslant 0.5 \\ 2(1-R)^3 & 0.5 \leqslant R \leqslant 1 \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $R = r/h$.
An adaptive smoothing length is used, so that mass is constant in the smoothing volumes, i.e.:

$$\frac{4\pi}{3} h_i^3 \rho_i = N_{sph} \bar{m}, \tag{6}$$

where N_{sph} is the number of smoothing neighbors and $\bar{m}$ the average particle mass; this implicit equation is solved with the bisection method.

Shocks are treated by inserting an artificial viscous term in the equation of motions, as in Eq. 128, where the employed parametrization of artificial viscosity is the one derived by Monaghan (1997):

$$\Pi_{ij} = \begin{cases} -\frac{\alpha}{2} \frac{v_{ij}^{sig} w_{ij}}{\rho_{ij}} & \vec{v}_{ij} \cdot \vec{r}_{ij} < 0 \\ 0 & otherwise, \end{cases} \tag{7}$$

where $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$, $\vec{v}_{ij} = \vec{v}_i - \vec{v}_j$, $\rho_{ij} = \frac{\rho_i + \rho_j}{2}$, $w_{ij} = \vec{v}_{ij} \cdot \vec{r}_{ij}/|\vec{r}_{ij}|$ and $v_{ij}^{sig} = c_i + c_j - 3w_{ij}$ is the signal velocity between particles i and j (c_i is the sound speed at the position of particle i). Viscosity is thus only active when particles are approaching. The dimensionless factor α ($\approx 0.5 - 1.0$) defines the strength of viscosity. To avoid viscosity in case of shear motions (for which $\vec{v}_{ij} \cdot \vec{r}_{ij} < 0$ is also verified), the viscous term is multiplied by a limiter, following Balsara (1995), in the form

$$f = \frac{|\nabla \times \vec{v}|}{|\nabla \cdot \vec{v}| + |\nabla \times \vec{v}|}, \tag{8}$$

which distinguishes shocks from shear motions, suppressing viscosity in the latter case.

The hydrodynamic scheme implemented in `GADGET-3` and in `OpenGADGET-3` carries some differences from the one adopted by `GADGET-2`, which I briefly mention below.

- Velocity gradients in the limiter are computed from the full velocity gradient matrix rather than low-order classic kernel derivatives. This yields improved results especially in the estimate of the curl of the velocity field, which gives more accuracy in the suppression of artificial viscosity outside shocks.
- Artificial conductivity is applied at contact discontinuities, where spurious surface tension arises, to promote the transport of heat between particles and avoid such tension.

- A higher order interpolating kernel can be chosen to achieve better numerical convergence; the one adopted in the simulations described below is the Wendland C4 kernel:

$$W(r,h) = \frac{495}{32\pi h^3} \begin{cases} (1-R)^6 \left(1+6R+\frac{35}{3}R^2\right) & 0 \leqslant R \leqslant 1 \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

  where $R = r/h$.

- Since `GADGET` uses individual time steps, particles at a given step are split into "active"[1] and "inactive". Where these are overlapping spatially, some issues might arise. For example, if an active particle gains a lot of momentum and enters an inactive region, particles there are not affected by its presence, leading to unphysical results. To solve this, a particle wake-up limiting scheme is used to shorten time steps and "activate" particles earlier when necessary.

- The artificial viscosity coefficient α is calculated for each particle, following the approach developed by Cullen & Dehnen (2010).

For details, I refer to the paper by Beck et al. (2016).

3.1.3 *Additional physics*

In addition to gravity and hydrodynamics, state-of-the-art cosmological hydrodynamical codes incorporate a variety of astrophysical processes. Due to the limited resolution and the incomplete understanding of many of these processes, they are modeled through sub-resolution prescriptions – effective, phenomenological models that approximate small-scale phenomena such as star formation, SN explosions, and gas accretion onto supermassive black holes (SMBHs). These processes span a wide range of spatial and temporal scales, often far below the typical resolution and time steps of cosmological simulations. For reviews on sub-resolution models in cosmological simulations I refer to Somerville & Davé (2015), Vogelsberger et al. (2020), and Valentini & Dolag (2025).
In this Section, I describe the main sub-resolution models implemented in `OpenGADGET-3`.

3.1.3.1 *Star formation and stellar feedback*

The model for star formation included in the `DIANOGA` simulations described below is based on the original "effective model" presented in Springel & Hernquist (2003, SH03 hereafter), where a detailed description can be found. The formation of stars from the collapse and fragmentation of clouds of molecular gas in the interstellar medium (ISM), besides being still theoretically uncertain, involves much smaller scales than those probed by cosmological simulations. For this reason, a sub-resolution modeling

1 Active particles are those for which hydrodynamical equations are integrated in the current time step.

is required to describe the complex structure of the ISM in an effective way. This is done by performing a coarse-graining procedure, where a fluid element – an SPH gas particle (see Sect. D.2.1) – is described by spatially-averaged properties.

When gas particles reach a density threshold ρ_{th}, corresponding to a hydrogen number density of $n_H = 0.1\ \mathrm{cm}^{-3}$ (and optionally a temperature threshold), they are promoted to "multiphase" particles. In such particles, a cold and a hot sub-resolution phases –representing cold molecular clouds and a hot ambient gas, respectively – are assumed to be in pressure equilibrium. The density of stars (ρ_*), on average, evolves proportionally to the density of the cold phase (ρ_c), on a timescale t_* and also decreases as a consequence of SN explosions:

$$\frac{d\rho_*}{dt} = (1-\beta)\frac{\rho_c}{t_*}, \tag{10}$$

where β is the mass fraction of stars dying as SN (dependent on the initial mass function), assumed to explode instantaneously. The timescale t_* is assumed to be proportional to the local dynamical time of the gas (which results in a Schmidt-like scaling of star formation rates with densities), thus:

$$t_* = t_{0,*}\,(\rho/\rho_{th})^{-0.5}\,, \tag{11}$$

where $t_{0,*}$ is a free parameter, tuned to fit the Kennicutt relation (Kennicutt 1989). Star formation is then treated stochastically: multiphase gas particles may spawn collisionless star particles and only the expectation value of the density of star formation evolves consistently with Eq. 10.
The evolution of cold and hot phases is followed by taking into account, besides star formation, radiative cooling and molecular cloud evaporation:

$$\frac{d\rho_c}{dt} = -\frac{\rho_c}{t_*} - A\beta\frac{\rho_c}{t_*} + \frac{1-f}{u_h - u_c}\Lambda_{net}(\rho_h, u_h), \tag{12}$$

$$\frac{d\rho_h}{dt} = \beta\frac{\rho_c}{t_*} + A\beta\frac{\rho_c}{t_*} - \frac{1-f}{u_h - u_c}\Lambda_{net}(\rho_h, u_h), \tag{13}$$

where A is the efficiency of evaporation of cold molecular clouds caused by SN explosions, and u_c (u_h) is the internal specific energy for the cold (hot) component of the ISM. Λ_{net} is the cooling function describing the rate at which radiative processes cause the cooling (or heating) of the hot phase. The hot phase of the multiphase particles is indeed the only component assumed to be subject to radiative processes. The parameter f acts as a flag distinguishing between two regimes: $f = 0$ when thermal instability and cloud condensation from the hot gas occur (i.e., when $\rho \geqslant \rho_{th}$); conversely, $f = 1$ corresponds to ordinary cooling in the absence of condensation (i.e., when $\rho < \rho_{th}$).
The cold clouds are assumed to have a fixed temperature; therefore, their

thermal energy u_c remains constant. On the other hand, the evolution of the hot component is given by:

$$\rho_h \frac{du_h}{dt} = \beta \frac{\rho_c}{t_*}(u_{SN} + u_c - u_h) - A\beta \frac{\rho_c}{t_*}(u_h - u_c) - f\Lambda_{net}, \tag{14}$$

where $u_{SN} = (1-\beta)\beta^{-1}\epsilon_{SN}$ and ϵ_{SN} is the average SN energy expected per unit stellar mass formed, dependent on the initial mass function (IMF). The SN evaporation parameter A is assumed to be a power-law of density, following McKee & Ostriker (1977):

$$A(\rho) = A_0 \left(\frac{\rho}{\rho_{th}}\right)^{-4/5}, \tag{15}$$

where A_0 is a parameter of the model. This can be constrained by requiring that, when the density threshold is reached, the hot ambient gas is at an equilibrium temperature that allows thermal instability to take place, meaning that the cooling function has to drop, which happens at T$\sim 10^5$ K. Combining this with the fact that Eq. 14, in case of thermal instability, gives a thermal energy $u_h \sim u_{SN}/A$, yields $A_0 \sim 1000$. The remaining parameter is ρ_{th}, which can be constrained if we consider that, in this model, stellar feedback and star formation enter a self-regulated cycle, where the effective pressure of the medium should be constant with time. By imposing the continuity of the effective pressure with density at the onset of self-regulated star formation, the parameter ρ_{th} can be linked to t_0^*, thus leaving the latter as the only actual free parameter, constrained by observations as discussed above.

To account for galactic outflows, this model is complemented by a phenomenological description of winds driven by SN. The rate of mass loss in galactic winds ($\dot{M}_w$) is assumed to be proportional to the star formation rate ($\dot{M}_*$):

$$\dot{M}_w = \eta \dot{M}_*, \tag{16}$$

where η, the mass loading factor, is of the order of unity. The other parameter of the model defines the fraction of SN energy that goes into kinetic energy of the wind, χ. Defining v_w as the velocity of the wind, then the kinetic energy rate is:

$$\frac{1}{2}\dot{M}_w v_w^2 = \chi \epsilon_{SN} \dot{M}_*. \tag{17}$$

In practice, multiphase star-forming gas particles can be stochastically converted into "wind particles" by receiving a velocity kick of magnitude v_w.

3.1.3.2 *Black hole growth and AGN feedback*

`OpenGADGET-3` includes a sub-resolution description of the effects of active galactic nucleus (AGN) feedback driven by gas accretion onto SMBHs, following the original model presented in Springel et al. (2005a), with some variations from Hirschmann et al. (2014). Black holes (BHs) are treated as collisionless sink particles [2], seeded in DM halos, that satisfy specific criteria in terms of mass and baryon content, to guarantee that BHs are seeded only in halos representing clearly resolved galaxies:

- the DM mass of the halo is larger than 5×10^{10} $M_\odot$ h^{-1};
- the stellar mass is a fraction larger than 0.02 of the total mass of the halo and larger than 0.05 times the DM mass;
- the gas mass is at least 0.1 times the stellar mass;
- the halo does not host a BH particle already.

In halos that satisfy the conditions above, the star particle with the highest binding energy is converted into a BH particle, with initial mass:

$$M_{BH}^{seed} = M_0 \frac{M_{*,h}}{f_* M_{DM}^{seed}}, \tag{18}$$

where $M_0 = 2 \times 10^5$ $M_\odot$ h^{-1}, $f_* = 0.02$ and $M_{DM}^{seed} = 5 \times 10^{10}$ $M_\odot$ h^{-1}. Gas accretion onto BHs is assumed to be spherically symmetric, for simplicity, following a Bondi-Hoyle-Littleton parametrization (Hoyle & Lyttleton 1939; Bondi & Hoyle 1944; Bondi 1952):

$$\dot{M}_{Bondi} = \frac{4\pi G^2 M_{BH}^2 \alpha \rho}{(c_s^2 + v^2)^{3/2}}, \tag{19}$$

where M_{BH} is the BH mass, ρ and c_s are the density and the sound speed at the position of the BH and v is the BH's relative velocity with respect to the surrounding gas. α is a dimensionless boost parameter, introduced to account for the unresolved high gas density in the surroundings of the BH. In the simulations analyzed in this Thesis, the accretion was assumed to be limited to the Eddington rate:

$$\dot{M}_{Edd} = \frac{4\pi G M_{BH} m_p}{\epsilon_r \sigma_T c}, \tag{20}$$

where m_p is the proton mass, σ_T is Thomson cross-section and ϵ_r is the radiative efficiency, which defines the fraction of accreted mass that is converted to radiated luminosity. In the simulations used in this Thesis, the value for ϵ_r is assumed to be 0.1 as for radiatively efficient Shakura & Sunyaev (1973) accretion onto a Schwarzschild BH. Different values of α are assumed for the accretion of the cold ($T < 10^5$ K, $\alpha_c = 100$) and hot ($T > 10^5$ K, $\alpha_h = 10$) components of the gas (see Steinborn et al. 2015). In this model, as the BH accretes gas, it injects feedback energy into the surrounding gas. The energy that is thermally coupled to the gas is:

$$\dot{E}_{thermal} = \epsilon_r \epsilon_f \dot{M}_{BH} c^2, \tag{21}$$

2 Sink particles (Bate et al. 1995) are effective particles used to treat regions of gravitational collapse, such as an accreting BH, in simulations. They accumulate mass accreted from the surrounding gas through an immediate collapse onto a point mass, artificially stopping the gravitational collapse at a scale that can still be resolved by the simulation.

where ϵ_f is the feedback efficiency. Following Sijacki & Springel (2006) and Fabjan et al. (2010), this model assumes different values for ϵ_f according to the Eddington ratio η_{Edd} of the BH, defined as the ratio between the accretion rate onto the BH and its Eddington limit. For high Eddington ratios ($\eta_{Edd} > 0.01$), in the so called "quasar mode", the feedback efficiency ϵ_f is 0.05, while at lower accretion rates the "radio mode" is switched on, with a feedback efficiency four times higher than the one assumed in quasar mode, to account for massive BHs that accrete at low rates but are very efficient in inflating bubbles in the surrounding gas.

The finite resolution of simulations causes artificial displacements of BH particles from the center of their host galaxy, due to spurious scattering interactions with their surrounding particles, leading to the formation of artificially "wandering" BHs (e.g., Bahé et al. 2022; Damiano et al. 2024). In the `DIANOGA` simulations introduced below, this numerical artifact is approached by "boosting" the dynamical mass (e.g., Debuhr et al. 2011; Bassini et al. 2020), which enters gravity calculations. At the time of seeding, this is set to the mass of DM particles, until it exceeds the real mass of the BH (which grows following Eq. 19), then it starts growing through gas accretion. This enhances the numerically resolved dynamical friction, preventing BHs from escaping their host galaxy.

Another consequence of the finite resolution of simulations is that the mergers of BHs cannot be resolved. Therefore, to model mergers in `OpenGADGET-3`, it is assumed that two black holes merge, with the one in the deepest potential well swallowing the other, if the following conditions are satisfied (Di Matteo et al. 2008):

- their separation is below 5 times the gravitational softening scale of BH particles;
- their relative velocity v_{rel} is less than half the local sound speed (c_s);
- the values of the gravitational potential at their positions (Φ_1 and Φ_2) need to satify:

$$\frac{|\Phi_1 - \Phi_2|}{a} < 0.5c_s^2 - v_{rel}^2, \tag{22}$$

  where a is the adimensional scale factor.

The sound speed is used as a simple local velocity scale of the galaxy. The condition on the relative velocity, compared to the local sound speed, is introduced to prevent early mergers during a fly-by in a merging event (see Di Matteo et al. 2008).

3.1.3.3 *Chemical enrichment*

The model describing chemical evolution in `OpenGADGET-3` is presented in Tornatore et al. (2007). In this model, each star particle is described as a simple stellar population (SSP), i.e., an ensemble of coeval stars with uniform metallicity. Besides the time of birth and the metallicity, each SSP is characterized by its initial mass and an IMF. The evolution of the SSPs is traced by considering the mass loss and metal production from Type II and Ia SN, as well as from asymptotic giant branch (AGB) stars, based on the choice for an IMF, stellar life-times, and stellar yields. At each time step, given the IMF, the number of stars ending their life as SN (type II or Ia) or entering the AGB phase is computed for each SSP, along with their contribution to chemical enrichment. In particular, the rate of SN Ia, assumed to be binary stars in the mass range $0.8 - 8\ M_\odot$ (Greggio & Renzini 1983), at time t is:

$$R_{SNIa}(t) = -\frac{dm(t)}{dt} \times 24Am(t)^2 \int_{M_{Bm}}^{M_{BM}} \phi(m_B)\, \frac{1}{m_B^3}\, dm_B, \tag{23}$$

where $\frac{dm(t)}{dt}$ defines the rate of mass of stars dying at time t, m is the mass of stars dying at time t, ϕ is the IMF and A is the fraction of stars in binary systems that are progenitors to SN Ia systems. The integral is over the total mass of the binary system, between M_{Bm} and M_{BM}. In `OpenGADGET-3`, the parameter values follow those adopted in the theoretical model by Matteucci & Greggio (1986): $A = 0.1$, $M_{Bm} = 3\ M_\odot$ and $M_{BM} = 16\ M_\odot$.
The rate of stars that end their lives as type II SN or enter the AGB phase is given by:

$$R_{SNII|AGB}(t) = \phi[m(t)] \times \left(-\frac{dm(t)}{dt}\right). \tag{24}$$

This rate needs to be multiplied by a factor $1 - A$ for AGB stars that fall in the same mass range of a secondary binary star of a SN Ia progenitor.

The evolution of the mass density for the i-th chemical element is obtained by solving (see Matteucci 2003):

$$\begin{aligned}
\rho_i(t) = & -\psi(t)Z_i(t) \\
& + A\int_{M_{Bm}}^{M_{BM}} \phi(m) \left[\int_{0.5}^{\mu_{min}} f(\mu)\, \psi(t-\tau_{m_2})\, p_{Z_i}(m,Z)\, d\mu\right] dm \\
& + (1-A)\int_{M_{Bm}}^{M_{BM}} \psi(t-\tau(m))\, p_{Z_i}(m,Z)\, \phi(m)\, dm \\
& + \int_{M_L}^{M_{Bm}} \psi(t-\tau(m))\, p_{Z_i}(m,Z)\, \phi(m)\, dm \\
& + \int_{M_{BM}}^{M_U} \psi(t-\tau(m))\, p_{Z_i}(m,Z)\, \phi(m)\, dm
\end{aligned} \tag{25}$$

where the first term describes the amount of the element that is locked in newly forming stars, with $\psi(t)$ the rate of star formation and $Z_i(t)$ the abundance of the i-th element in the ISM at time t. The second term describes the contribution of SN Ia, with total mass of the system between M_{Bm} and M_{BM}. The internal integral accounts for all possible combinations of masses in the binary system, with $\mu = \frac{m_2}{m_1+m_2}$, where m_1 (m_2) is the mass of the primary (secondary) star in the binary system and $f(\mu)$ is the distribution of μ. Here, τ_m refers to the lifetime of a star of mass m and

$p_{Z_i}(m, Z)$ is the yield for the i-th element. The third and fourth terms describe, respectively, the contribution of low- and intermediate-mass stars, while the last term is the contribution of type II SN. The mass ranges are defined by the following values (Greggio & Renzini 1983; Matteucci & Greggio 1986): $M_L = 0.1\, M_\odot$, $M_{Bm} = 3\, M_\odot$, $M_{BM} = 16\, M_\odot$ and $M_U = 100\, M_\odot$. The distribution of μ in the SN Ia term is taken from Matteucci & Recchi (2001) and μ_{min} for a binary system of total mass m_B is $\mu_{min} = max\left(\frac{m_2}{m_B}, \frac{m_B - 0.5 M_{BM}}{m_B}\right)$. The details of the numerical implementation of this model can be found in Tornatore et al. (2007).

3.1.3.4 *Radiative cooling of gas*

OpenGADGET-3 also includes a model for radiative cooling of the gas component, depending on its temperature, metallicity, and redshift. Cooling and heating rates are computed using the photoionization code CLOUDY (Ferland et al. 1998), which accounts for the most relevant atomic processes over the temperature range $10^4 - 10^8$ K , as covered by the simulations. The gas is assumed to be optically thin, as expected for gas in the temperature and the density ranges typically probed by the simulations, dust-free, in collisional ionization equilibrium, and exposed to both the CMB and a spatially uniform UV/X-ray background. Further details can be found in Wiersma et al. (2009).

3.2 DIANOGA SIMULATIONS

The analyses presented in this Thesis are based on a set of 14 simulated galaxy clusters, based on the DIANOGA initial conditions of zoom-in simulations of galaxy clusters, presented in Bonafede et al. (2011). Zoom-in simulations are an invaluable tool for investigating rare systems, such as massive galaxy clusters. They are high-resolution simulations focused on small regions of interest, which are embedded within a larger, lower-resolution simulation that captures the surrounding large-scale environment. This approach enables the detailed study of specific structures at a resolution that would be computationally prohibitive for large volumes.
Various versions of the DIANOGA simulations have been presented in the literature, implementing different models of sub-resolution physics and with different resolutions (e.g., Rasia et al. 2015; Ragone-Figueroa et al. 2018; Bassini et al. 2020). The adopted cosmology is a flat ΛCDM, with $h = 0.72$, $\Omega_m = 0.24$, $\Omega_b = 0.04$ for the density parameters[3] associated with total and baryonic matter, $n_s = 0.96$ for the primordial spectral index and $\sigma_8 = 0.8$ for the normalization of the linear power spectrum. The DIANOGA initial conditions have been generated with the zoomed-in initial conditions

3 We note that the adopted cosmology differs from the fiducial flat ΛCDM and the standard Planck 2018 (Planck Collaboration VI 2020) cosmologies adopted in recent observational studies. However, we verified that the impact of cosmology on halo masses, stellar masses and radii in the epoch around cosmic noon, which is the focus of this Thesis, is well within typical statistical and systematic uncertainties in current observations.

technique (Tormen 1997): cluster-size halos of DM particles are identified at $z = 0$ in a 1024^3 DM particles box of $1\,h^{-1}$cGpc per side, then traced back to their positions in the initial conditions at $z \sim 180$; the regions defined in this way are re-simulated at higher resolution than the one adopted in the parent simulation, adding gas particles with an initial mass chosen so as to reproduce the cosmic baryon fraction of the assumed cosmology. Outside of these regions, low-resolution DM particles are included to describe the tidal field. The original `DIANOGA` sample includes the 24 most massive halos with a virial mass $M_{vir} > 10^{15}\ h^{-1}M_\odot$ in the parent box (note that, when re-simulated at higher resolutions, the halo masses change slightly) and 5 more halos in the mass range $M_{vir} \simeq (1-7) \times 10^{14}\ h^{-1}M_\odot$. The initial setup was chosen so as to guarantee a spherical volume around each cluster with a radius of 5-6 virial radii simulated at high resolution and free of low-resolution contaminant particles at $z = 0$. These regions are large enough to include satellite clusters and groups other than the target systems of the re-simulations. The masses of the `DIANOGA` clusters analyzed in this Thesis are listed in Tab 3.1. Two out of the 14 clusters in the set were simulated down to $z = 1$, whereas the remaining clusters were simulated to $z = 0$.

Figure 3.1 shows a rendering of the distributions of DM, gas and stars

Table 3.1: IDs and masses of the `DIANOGA` cluster simulations at $z = 0$. The last column specifies if the cluster belongs to the set of the 24 most massive clusters (set 1) originally selected in the parent box or to the set of 5 less massive ones (set 2). Clusters D15 and D16 were simulated down to $z = 1$, so their masses at $z = 0$ come from the parent simulation.

ID	M_{200c} [$10^{15}\ h^{-1}M_\odot$]	set
D2	0.5	2
D3	0.8	2
D4	0.6	2
D5	0.2	2
D6	1.6	1
D7	1.6	1
D9	0.2	2
D10	1.5	1
D13	1.8	1
D15	1.8	1
D16	3.2	1
D17	1.8	1
D21	1.8	1
D25	1.1	1

(from left to right) at $z = 2$, 1 and 0 (from top to bottom) of a cluster extracted from the `DIANOGA` set. This figure captures the evolution of the three components as the cluster assembles, highlighting the transition from a

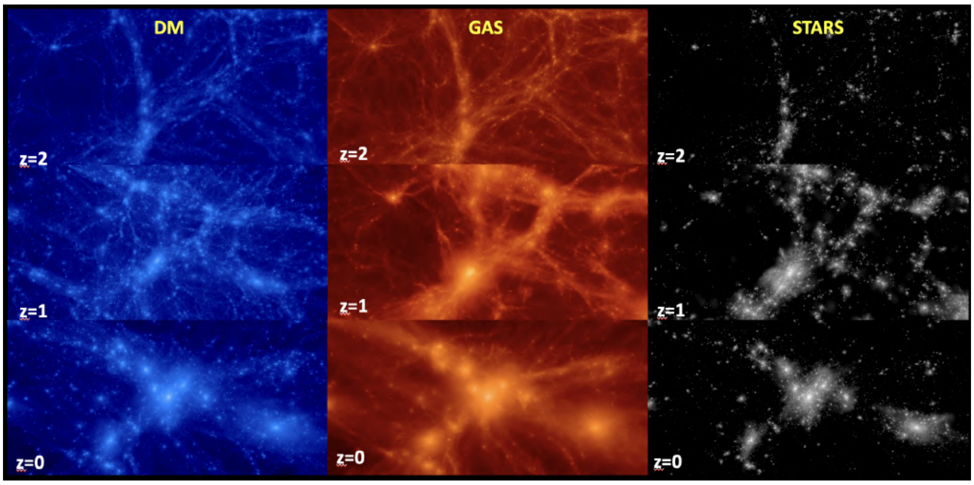


Figure 3.1: Evolution of a DIANOGA cluster from $z = 2$ (top panels), to $z = 1$ (middle panels) and then to $z = 0$ (bottom panels). From left to right, the panels show the rendering of the DM, gas and stars distributions.

filamentary diffuse protocluster system to a fully formed cluster by $z = 0$, that is interacting with a less massive group of galaxies.

Hereafter, I refer to the high-resolution version of the simulations described above as the DIANOGA set of simulations of galaxy clusters. The re-simulations were performed with a state-of-art developer version of the TreePM SPH code OpenGADGET-3 introduced in Sect. 3.1. The mass resolution for DM particles and the initial mass of gas particles are $m_{DM} = 3.4 \times 10^7\ h^{-1}M_\odot$ and $m_{gas} = 6.2 \times 10^6\ h^{-1}M_\odot$, respectively. The Plummer-equivalent gravitational softening lengths in the high-resolution regions at $z = 0$ are $\epsilon_{DM} = 1\ h^{-1}$kpc, $\epsilon_{gas} = 1\ h^{-1}$kpc, and $\epsilon_{*,BH} = 0.25\ h^{-1}$kpc, for DM, gas, and star and BH particles, respectively. All softening lengths are kept fixed in comoving units, except ϵ_{DM}, held constant in physical units below $z = 2$, and fixed in comoving units for $z > 2$.
In these simulations, all the sub-resolution models described in Sect. 3.1.3 were included. In the star formation model, each gas particle can spawn four generations of star particles so that the typical mass of the latter is $m_* = 1.5 \times 10^6\ h^{-1}M_\odot$. The model of stellar evolution described in Sect. 3.1.3 assumes a Chabrier (2003) IMF and follows the evolution of 14 chemical elements: He, C, O, N, Ne, Mg, Si, S, Ca, Fe, Na, Al, Ar, and Ni. The mass-dependent life-times of stars are taken from Padovani & Matteucci (1993) and the stellar yields are taken from Thielemann et al. (2003) for type-Ia SN, from Karakas (2010) for AGB stars, and from Nomoto et al. (2013) for type-II SN. The parameters regulating accretion onto BHs and the associated feedback (reported in Sect. 3.1.3.2) have been tuned to reproduce the observed relation between BH mass and stellar mass of galaxies in the local Universe (Magorrian et al. 1998), using the relation from McConnell & Ma (2013) and, for BCGs, data from Gaspari et al. (2019).

In this Thesis, I analyzed the properties of simulated protoclusters and high-redshift clusters and, in order to study the environmental effects on

the galaxy population properties, I compared them with coeval analogs in a reference, average density region. To this purpose, I used the population of galaxies identified within a cosmological box with size 49 cMpc/h per side, simulated with the same version of `OpenGADGET-3`, assuming the same cosmology and sub-resolution models for baryonic physics. To have the same mass resolution as the `DIANOGA` zoom-in regions, initial conditions were generated for 576^3 DM particles and as many gas particles.
In these simulations, galaxies have been identified with the `SubFind` algorithm (Springel et al. 2001b; Dolag et al. 2009). This subhalo finder searches for DM substructures hosting galaxies as bound systems within a larger parent halo. The first step is thus the identification of DM halos, which is performed with a Friends-of-Friends (FoF) algorithm (Davis et al. 1985). Here the particles are grouped such that the interparticle distance is always below a parameter, the linking length, usually given in units of mean interparticle distance. The linking length is usually assumed to be about 0.2 (the exact value being dependent on cosmology), since this was proved to identify halos with a density threshold similar to what is predicted for virialized systems by the spherical collapse model. Within FoF DM halos, substructure candidates are then identified by looking for regions enclosed by density contours crossing a saddle point of the density field. After that, particles belonging to each of the subhalo candidates undergo an unbinding procedure to exclude those particles that are not gravitationally bound to the assigned subhalo.

To mitigate possible resolution-related effects, when investigating galaxy properties I adopt a lower stellar-mass threshold of $M_* = 10^9\ M_\odot$, corresponding to approximately 500 particles.

3.2.1 *Properties of `DIANOGA` simulations at z=0*

In this Section, I present some fundamental predictions from the `DIANOGA` simulations and the cosmological box at $z = 0$, focusing on galaxy properties and on the thermodynamics of the ICM. This is a preliminary assessment of the reliability of the simulations in reproducing observations, which does not represent an exhaustive investigation of their predictions at $z = 0$.
By comparing simulations with key observables from the main fields of investigations of this Thesis – galaxies and the ICM –, I aim to highlight possible limitations of the implemented galaxy formation model. At $z = 0$ observations are far more abundant and affected by smaller uncertainties and biases than at cosmic noon, where most of the results presented in this Thesis are obtained. Thus, this comparison represents a significant constraint on the predictive power of the simulations and its conclusions should be considered while discussing the predictions of simulations at higher redshifts in Chapters 4 and 5.
The `DIANOGA` clusters considered here are the 12 out of 14 that were simulated down to $z = 0$.

3.2.1.1 *The galaxy population*

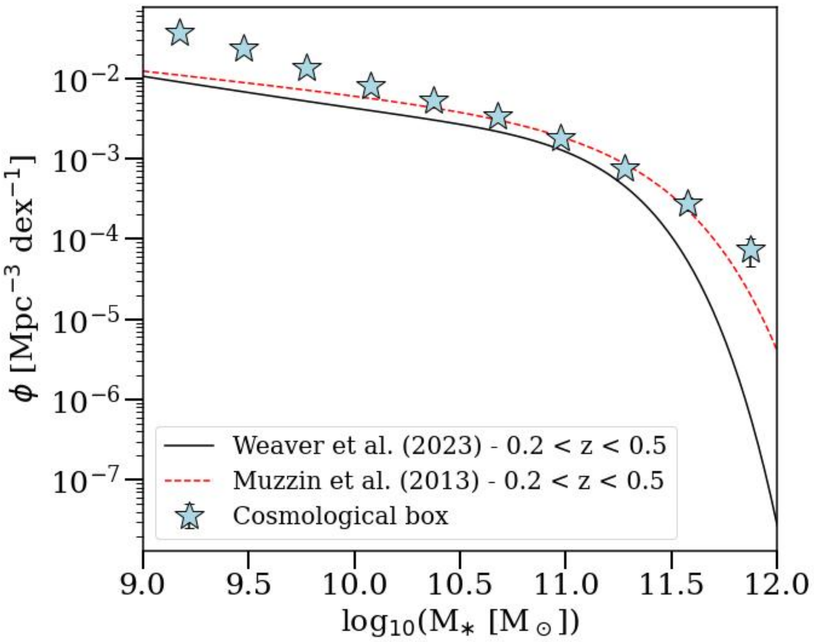
One of the most fundamental observables concerning the galaxy population, especially for the purpose of validation of theoretical models, is the galaxy stellar mass function (GSMF). Being the distribution of stellar masses of galaxies, this observable encapsulates information on the (integrated) star formation history and on the stellar mass assembly. The GSMF is also a sensitive probe of feedback processes. To reconcile the observed GSMF with cosmological simulations, feedback processes are indeed fundamental, both from SN and AGN. Theoretical models have shown that the GSMF in the low- and intermediate-mass regime is mostly affected by galactic winds driven by SN, while the slope of the high-mass end is more sensitive to the inclusion of the AGN feedback (see, e.g., Cattaneo et al. 2007; Somerville et al. 2008; Oppenheimer et al. 2010; Scharré et al. 2024). In fact, the high-mass end of the GSMF can be reconciled with observations only if AGN feedback is included in the modeling.
The simulated GSMF at $z = 0$ is shown in Fig. 3.2, for galaxies in the cosmological box (top panel) and in the `DIANOGA` clusters (bottom panel), compared with observational data in the COSMOS field (Muzzin et al. 2013; Weaver et al. 2023) and in 21 clusters at $z \sim 0$ from the WIde-field Nearby Galaxy-cluster Survey (WINGS; Vulcani et al. 2011), respectively. Stellar masses in these observational studies were obtained either with the same IMF as in simulations (Chabrier 2003) or with the Kroupa (2001) IMF; no correction was applied to convert the latter to estimates based on a Chabrier (2003) IMF, as the correction is negligible compared to the measurement uncertainties. In the simulations, galaxy stellar masses were obtained by considering only the star particles associated to each galaxy by `SubFind`, within a sphere of radius 30 kpc around the galaxy center. The chosen radius, although arbitrary to some extent, is large enough to encompass approximately 3 times the half-light radius of observed massive galaxies ($M_* \sim 10^{11}$ $M_\odot$; e.g., van der Wel et al. 2014). Further increasing the radius to 40 kpc affects the number density of galaxies by less than $\sim 1\%$ at $M_* \lesssim 10^{10.5}$ $M_\odot$ and by $\sim 3\%$ at $M_* \sim 10^{11}$ $M_\odot$, well within observational uncertainties ($\sim 0.1 - 0.15$ dex for $\log_{10}(M_*\ [M_\odot]) \sim 10 - 11$ at $z \sim 0$; Muzzin et al. 2013). This level of uncertainty is not relevant for what concerns the purpose of this comparison in the context of results presented in this Thesis.
The GSMF from Vulcani et al. (2011) is a Schechter function (Schechter 1976) fit to the data, including all galaxies of the 21 clusters. In the figure, I show the fit, divided by the number of clusters, to compare it with the mean GSMF in the `DIANOGA` simulations. To have an approximate quantitative comparison between the normalizations of the simulated and observed GSMFs, one should consider the different mass ranges spanned by observed and simulated clusters. The median mass of WINGS clusters is indeed ~ 3 times lower than the `DIANOGA` median mass and the former are thus expected to have lower richness. To approximately account for this, the observed GSMF is rescaled by a factor of 2 (see, e.g., Lin et al. 2004). To replicate the selection of galaxies in the WINGS clusters, galaxies in the `DIANOGA` clusters

are selected within a sphere of radius $0.6\times$ R_{200} from the cluster center. In the simulations, the GSMFs are convolved with a Gaussian distribution of errors on the stellar masses, with dispersion equal to 0.3 dex, corresponding to the typical mass uncertainty reported in Vulcani et al. (2011).

The comparison shows a good agreement between the GSMF in the cosmological box and the field observations around the knee of the mass function. At the low-mass end, the GSMF is too steep, which is a known limitation of models for kinetic SN feedback (see Sect. 3.1.3.1) that assume a constant wind velocity (e.g., Choi & Nagamine 2011). This is due to the fact that, for typical values adopted in models using constant wind velocity and mass loading factor, SN-driven winds are inefficient in low-mass galaxies, as shown by the comparison with models that scale such parameters with the local velocity dispersion or the stellar mass of galaxies (e.g., Oppenheimer & Davé 2006; Oppenheimer et al. 2010). At the same time, the high-mass end of the simulated GSMF shows a deviation from observations, with an excess of very massive galaxies in simulations.
As for the `DIANOGA` clusters, the median prediction of the $z = 0$ GSMF is consistent with observational constraints across almost all the probed stellar mass range. Due to the approximate nature of the mass-richness correction applied to the observational data, the normalization should be interpreted with caution, and the comparison should primarily focus on the relative trends. The high-mass end of the GSMF of the `DIANOGA` simulations – from which BCGs have been excluded – predicts an excess of very massive galaxies ($M_* \gtrsim 10^{11.5}$ $M_\odot$) compared to observations, mirroring the inconsistency seen in the cosmological box. This suggests that the interplay of feedback processes, especially the AGN feedback which is expected to be more relevant for massive galaxies, is not efficient enough to suppress star formation at the highest stellar masses.

3.2.1.2 *Profiles of ICM thermodynamic properties*

To characterize the simulated ICM in `DIANOGA` clusters, I analyze the radial profiles of its thermodynamic properties, including electron density (n_e), temperature (T), pressure ($P=n_eT$), and entropy ($K= k_BTn_e^{-2/3}$). These properties are shaped during the assembly of the ICM through diffuse accretion and mergers, and are also influenced, on small scales, by feedback processes.
The characterization of these profiles in simulations, and their comparison with observational data, is essential for several reasons. First, the comparison can reveal limitations in the implementation of feedback, radiative cooling, or in the treatment of viscosity and conductivity within the hydrodynamical scheme. Moreover, it may point to missing physical processes not included in the current modeling.
Additionally, the study of simulated profiles can help identify biases in observational measurements, which may propagate into systematic errors in scaling relations — for example, those relying on cluster mass estimates derived from density and temperature profiles under the assumption of



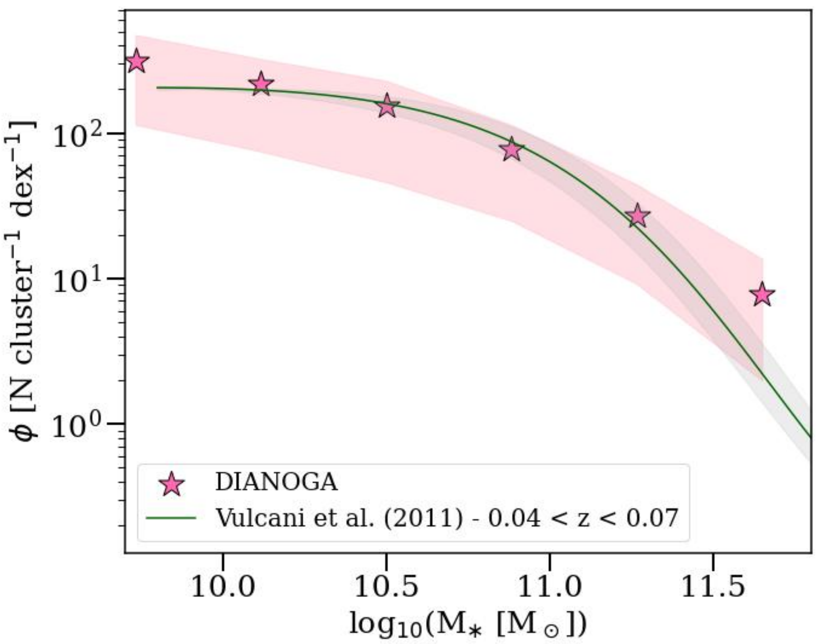


Figure 3.2: Galaxy stellar mass function at $z = 0$ in simulations, compared with observational data. The top panel shows the prediction for galaxies in the cosmological box (with Poissonian uncertainties), compared with the GSMF in the COSMOS field at $0.2 < z < 0.5$, from Muzzin et al. (2013) and Weaver et al. (2023). The bottom panel shows the GSMF in DIANOGA clusters, excluding the BCGs. Star symbols are the total number of galaxies in all clusters divided by the number of clusters and the shaded region shows the 16th-84th percentile range among the 12 clusters. These are compared with the observed GSMF of WINGS clusters ($0.04 \leqslant z \leqslant 0.07$, Vulcani et al. 2011). Galaxies are selected within $0.6 \times R_{200}$ to match the galaxy sample in almost all WINGS clusters. The observed GSMF in clusters is shown above the mass completeness limit of $10^{9.8}$ $M_\odot$.

hydrostatic equilibrium (e.g., Rasia et al. 2006).
Density and temperature are the fundamental properties from which pressure and entropy are derived in simulations. Observationally, they are inferred from X-ray measurements of the ICM, or from their combination with pressure profiles derived from the thermal SZ effect. The pressure profile provides a key diagnostic of the thermodynamic state of the ICM. Compared to density and temperature profiles, it is more robust and nearly universal (e.g., Arnaud et al. 2010), because pressure governs hydrostatic equilibrium and thus closely traces the depth of the gravitational potential well. As a result, it is less sensitive to the cluster's dynamical history and to non-gravitational processes.
Entropy profiles, on the other hand, are particularly useful for investigating the thermal history of the diffuse gas in clusters. Entropy acts as a fossil record of past heating processes – whether due to shock-heating from gravitational collapse or to non-gravitational feedback mechanisms. In the cores of clusters, entropy profiles exhibit a bimodal behavior, distinguishing "cool-core" systems – where entropy decreases sharply toward the center – from "non-cool-core" systems, characterized by an almost flat, isentropic core (see Sect. 2.1.3.1). Cool-core clusters have historically posed a challenge for simulations (see Sect. 2.1.3.2), with only a few studies managing to reproduce them successfully in cosmological hydrodynamical simulations (e.g., Rasia et al. 2015).
Figure 3.3 shows the radial profiles of the thermodynamic properties of the ICM of `DIANOGA` clusters at $z = 0$, compared with profiles extracted from observations of 12 clusters from the X-COP sample (Eckert et al. 2017) in the local universe ($0.05 \lesssim z \lesssim 0.09$), from Ghirardini et al. (2019), with a median mass that is consistent with (~ 1.4 times smaller than) that of the `DIANOGA` clusters. X-COP is a very large program on XMM-Newton, targeting the most significant SZ sources in the Planck survey. The goal is to study the virial regions of galaxy clusters, combining X-ray and SZ data, even beyond their R_{200}. The observed profiles are resolved in two decades in radius, between 0.01 and $2{\times}R_{500}$ and are thus ideal for a comparison with the simulated profiles from the virial radius down to the innermost core of the clusters. I show profiles of free electron density, temperature, pressure and entropy of the ICM. In the simulations, the ICM is made up of the gas particles with temperature exceeding a threshold of 0.3 keV (e.g., Rasia et al. 2025) and having less than 10% of their mass in the cold star-forming phase. Temperature, pressure and entropy are normalized to the values predicted by the self-similar scaling at R_{500} (Voit et al. 2005):

$$T_{500} = 8.85 \frac{M_{500}}{h_{70}^{-1} 10^{15}\ M_\odot} E(z)^{2/3} \frac{\mu}{0.6} \mathrm{keV}, \tag{26}$$

$$P_{500} = 3.426 \times 10^{-3}\, E(z)^{8/3} \left(\frac{M_{500}}{10^{15}\ h_{70}^{-1}\ M_\odot} \right)^{2/3} \frac{f_b}{0.16} \frac{\mu}{0.6} \frac{\mu_e}{1.14}\ \mathrm{keV\,cm^{-3}}, \tag{27}$$

$$K_{500} = 1667 \left(\frac{M_{500}}{10^{15}\,h^{-1}\,M_\odot}\right)^{2/3} E(z)^{-2/3} \left(\frac{f_b}{0.16}\right)^{-2/3} \frac{\mu}{0.6} \times \left(\frac{\mu_e}{1.14}\right)^{2/3} \text{keV cm}^2, \quad (28)$$

where μ, μ_e and f_b are the mean molecular weight, the mean molecular mass per electron and the baryon fraction, $h_{70} = H_0/70$ km s^{-1} Mpc^{-1} and $E(z) = \frac{H(z)}{H_0}$.

The number of free electrons is computed by assuming complete ionization of hydrogen and helium in the gas and neglecting the contribution of metals. The profile of the electron density (top left) agrees with the observational fit quite well over almost all the virial volume, although it shows an excessive flattening in the central region ($r \lesssim 0.1\,R_{500}$) compared to observations.
The temperature profiles shown for simulations follow two different definitions of temperature that can be computed from simulations:

- mass-weighted temperature T_{MW}, defined as:

$$T_{MW} = \frac{\sum_i m_i T_i}{\sum_i m_i}, \quad (29)$$

  where m_i and T_i are the mass and the temperature of the i-th particle and the summation is over all particles in a spherical shell.

- The spectroscopic-like temperature T_{SL} (Mazzotta et al. 2004), defined as:

$$T_{SL} = \frac{\sum_i m_i \rho_i^{SPH} T_i^{0.25}}{\sum_i m_i \rho_i^{SPH} T_i^{-0.75}}, \quad (30)$$

  where ρ_i^{SPH} is the SPH density of the i-th particle. This definition of temperature was shown to be a good approximation to spectroscopic temperatures inferred from X-ray observations.

The temperature profile from Ghirardini et al. (2019) was obtained from spectroscopic temperatures within R_{500} and from gas-mass-weighted temperatures, between 0.7 R_{500} and 2 R_{500}. The latter comes from the combination of the pressure profile from SZ measurements and the density from the X-ray data. So, for comparison with this profile I show the simulated profiles obtained with both the listed definitions of temperature from simulations. At $0.1 \lesssim r/R_{500} \lesssim 1$, both the simulated profiles agree with the observed temperature profile, while in the inner region the DIANOGA clusters do not show the decrease in temperature that is observed in the X-COP clusters. This is an example of the limitations of most simulations in reproducing cool-core systems. Radiative cooling in the center of clusters is expected to remove low-entropy gas particles, with low cooling times, from the ICM phase (e.g., Li & Bryan 2012), producing a flat temperature profile, in absence of a source of heat. Thus, the AGNs at the centers of DIANOGA clusters may be inefficient at counteracting this effect, or their thermal feedback may

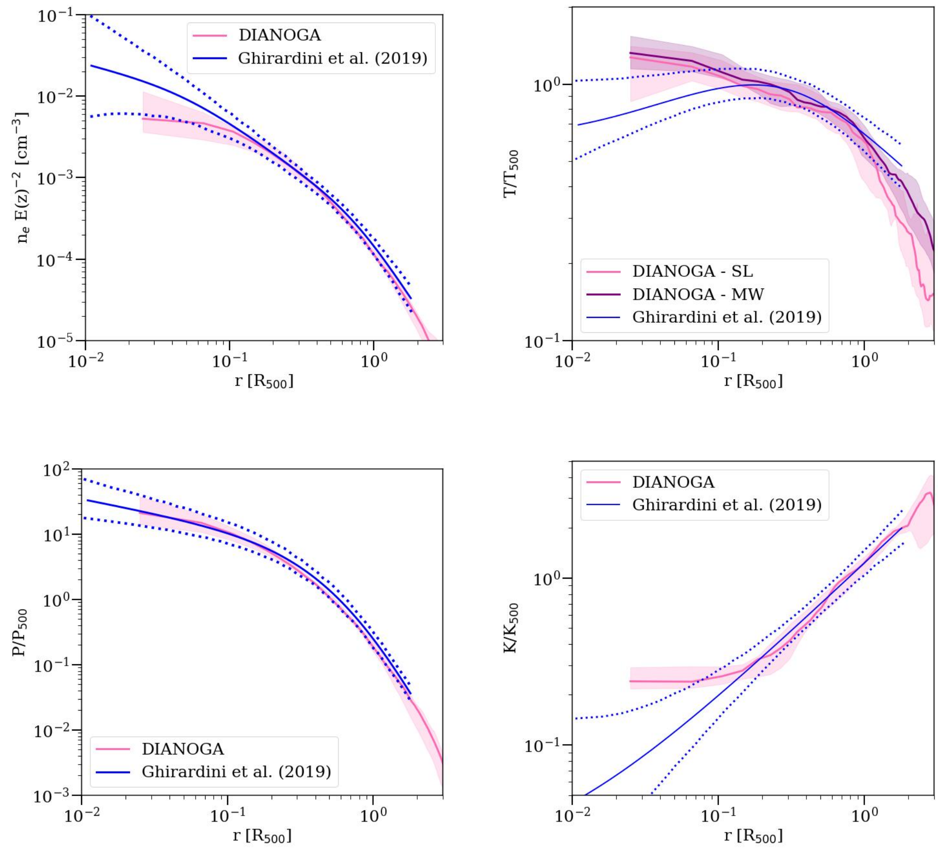


Figure 3.3: Radial profiles of thermodynamic properties of the ICM in DIANOGA clusters, compared with observations from Ghirardini et al. (2019). For DIANOGA simulations, the solid line represents the median profile among the clusters and the shaded area is the 16th-84th percentile range. The solid lines corresponding to the data are the global functional forms fitted to the observational data, while dotted lines mark the intrinsic scatter on the profiles. The top-left panel shows the comparison between the radial profiles of number density of free electrons. The top-right panel shows the temperature profiles, where two possible definitions are used in simulations, mass-weighted (MW) and spectroscopic-like (SL) temperature. The bottom-left and right panels show the radial profiles of pressure and entropy of free electrons, respectively.

overheat the central gas, preventing the turn-around seen in the temperature profile at $r \sim 0.2\,R_{500}$.
At $r \sim R_{500}$, the simulated temperature declines more steeply than observed. The mass-weighted temperature profile, which corresponds most closely to the observational definition in this regime, agrees better with the observational fit. In contrast, the spectroscopic-like temperature drops more sharply, being more sensitive to cold clumps and substructures in the cluster outskirts.
The simulated pressure profiles are perfectly consistent with the observations, as expected, given their low sensitivity to non-gravitational processes. Indeed, the simulated pressure profiles are tracing the structure of the gravitational potential of the cluster, largely unaffected by energy injections by the central AGN.
As for the entropy profile, at $0.1 \lesssim r/R_{500} \lesssim 1$, the median simulated profile follows the power-law fit to the data, while in the innermost regions it shows a plateau at higher entropy values compared to observations. This is consistent with the too high central temperatures and too low central density values that the DIANOGA clusters exhibit in comparison with observations.

Overall, this suggests that the modeling of sub-resolution processes – primarily AGN feedback – which are expected to influence the thermodynamics of the ICM in the very central regions of clusters, is still not well constrained in this regime. The possible explanations would be that either the AGN thermal feedback is heating the gas too efficiently, preventing the formation of cool core-like systems, or it is failing at preventing the cooling of low-entropy particles.

In conclusion, DIANOGA simulations were shown to give an overall realistic representation of galaxies and the ICM, however with quantitative differences with observational data: the simulated GSMF shows an excess of low-mass galaxies in the "field", and an excess of high-mass galaxies also in cluster environments, in comparison with observations; the ICM in the inner regions of clusters has too shallow density profiles and too high temperatures, reflecting also in flat entropy profiles.

3.2.2 *Post-processing of* DIANOGA *simulations*

To enable a proper comparison between simulations and observations in Chapters 4 and 5, I post-processed the simulations to produce observables, for both the galaxy populations and the diffuse gas component, that are more directly comparable to actual observations. For galaxies, I produced dust-attenuated spectral energy distributions (SEDs) of galaxies to extract NIR-to-UV magnitudes. For the ICM, I produced X-ray luminosity maps, modeling the Bremsstrahlung emission of this component.

3.2.2.1 *Radiation transfer of stellar emission*

In this Section, I describe the radiation transfer simulations performed to produce dust-attenuated SEDs for the simulated protocluster galaxies, used in Chapter 4. These are exploited to analyze rest-frame UVJ color-color diagrams, and in particular compare the results of this widely used photometric classification of star-forming and quenched galaxies in observational studies (e.g., Labbé et al. 2005; Williams et al. 2009), with the classification based on their intrinsic properties in the simulations. A proper comparison with the UVJ photometric classification approach requires accounting for the effect of the extinction produced by dust on the UV/optical/NIR emission associated with the stellar populations in the simulated galaxies. I resorted to the Monte Carlo radiation transfer code `SKIRT-9`[4] (Camps & Baes 2020) for this purpose. The `DIANOGA` simulations do not include a self-consistent treatment for the generation and evolution of dust grains. However, this can be implemented and indeed has been done in a lower-resolution version of these simulations (see Gjergo et al. 2018, for details). Therefore, I modeled the dust distribution through the spatial distribution of metals in the gas component of the simulated galaxies by adopting a fixed dust-to-metal ratio, for which I used the value $f_{DTM} = 0.2$ (e.g., Trčka et al. 2022). To each SPH gas particle, with mass m_{gas} and metallicity Z (fraction of mass in metals), is then attributed a dust mass equal to $m_{dust} = f_{DTM}\, Z\, m_{gas}$.

I adopted the THEMIS model (Jones et al. 2017) to assign a composition and grain size distribution to dust in the ISM, with associated optical properties that are based on laboratory measurements of materials physically similar to interstellar dust. To discretize the simulation physical domain, in terms of properties of the absorbing medium, I adopted an oct-tree structure (e.g., Saftly et al. 2013). In this kind of grid, cubic meshes are recursively split to half their linear size, thus generating eight "child nodes" each, until they meet a convergence criterion in the last "leaves" of the tree. In the radiation transfer simulations, I considered boxes of 64 kpc per side encompassing each galaxy and chose, for the construction of the oct-tree, a maximum fraction of the total dust mass in a leaf equal to $\sim 10^{-6}$, and a maximum number of refinement levels equal to seven, corresponding to a minimum size of leaf nodes equal to the gravitational softening of the gas component in the simulations. In each cell of the final grid, the properties of the medium (dust) and of the radiation field originating from the sources, represented by star particles of the hydrodynamical simulations, are considered uniform. The stellar emission is modeled by approximating each star particle as an SSP, consistent with the stellar evolution model included in the `DIANOGA` simulations. For each SSP, I adopted a Bruzual & Charlot (2003, BC03) SED, which is specified by the age, the initial mass and the metallicity of the SSP, as predicted by the simulations, once a Chabrier (2003) IMF is assumed.

The self-consistent treatment described above does not account for age-dependent extinction (e.g., Silva et al. 1998; Granato et al. 2001), with young

4 `https://skirt.ugent.be/root/_home.html`

stars highly attenuated by the remnants of their high-opacity birth clouds on spatial scales that are not resolved by cosmological simulations. To gauge the impact of age-dependent dust extinction on the results presented in this Thesis, I also ran simulations with SKIRT-9, including both a model for the effect molecular clouds (MCs) on young stars and the self-consistent treatment of the diffuse ISM described above. I followed Granato et al. (2021) but adopted a higher age threshold of 10 Myr to select stars still affected by their birth cloud, as in Baes et al. (2024). To model the extinction of young stars, I placed them all, for a give galaxy, at the center of an idealized macro-MC with a fixed M/r^2 ratio, where M is the molecular gas mass, approximated with the total mass of the "cold" gas phase of the ISM (in the SH03 model) of each galaxy, and r is the radius of the idealized cloud. To assign an opacity to the macro-MC, I refer to Silva et al. (1998), who showed that the M/r^2 ratio, and not the single values of M and r, determines the opacity of MCs. Then, I constrained the radius of the idealized MC by adopting the same M/r^2 ratio of a giant MC with a mass of molecular gas equal to 10^6 $M_\odot$ and a radius of 15 pc, as done in Granato et al. (2021). Finally, the dust mass in this cloud is given by $M_{dust,MC} = f_c \, f_{DTM} \, Z_c \, M_{gas}$, where f_c is the fraction of cold gas in each galaxy, as predicted by the simulations, Z_c is the metallicity of the cold component of the ISM and M_{gas} is the total gas mass. For consistency, I removed the cold gas assigned to the MCs from the diffuse medium that produces the attenuation of the emission of older star particles, as described above, by multiplying the mass of each gas particle (therefore of each dust particle) by $1 - f_c$. SEDs of individual galaxies are then computed by summing the SEDs of the old and young star populations.

In both the runs with and without a model for MCs, I ran SKIRT-9 in extinction-only mode, where calculations on the dust thermal re-emission at IR wavelengths are omitted to speed up the computations. The full exploration of the parameter space concerning the sub-resolution modeling of MCs is beyond the purpose of this Thesis, but will be object of future works aiming at refining the modeling of realistic photometric properties of simulated galaxies. Nevertheless, in Sect. 4.5.1, I show that the conclusions of the analysis based on SKIRT-9 SEDs are not affected by the modeling of MCs.

3.2.2.2 *X-ray luminosity of ICM*

Here, I describe the methods adopted to produce X-ray luminosity maps of the simulated ICM. The goal of this analysis is to produce X-ray surface brightness profiles that can be directly compared with observed profiles, as done in Chapter 5.

I used the SMAC[5] visualization tool, introduced in Dolag et al. (2005). This code produces maps and light-cones of multiple quantities from SPH simulations, for idealized observations. I used SMAC to produce X-ray luminosity maps of the diffuse gas in simulated halos. The X-ray emissivity of each gas

5 https://www.mpa-garching.mpg.de/~kdolag/Smac/

particle is calculated based on the tabulated cooling tables produced with the APEC model[6]. In the maps analyzed in Chapter 5, I used a CIC scheme (see Sect. D.1.2) to distribute particles to grid points on the map, and their luminosity distribution is smoothed with a cubic spline kernel.

6 https://heasarc.gsfc.nasa.gov/docs/software/xspec/manual/node134.html

4 GALAXY PROPERTIES IN SIMULATED PROTOCLUSTERS

This Chapter focuses on the characterization of cluster progenitors, or protoclusters, at cosmic noon, specifically looking into the predictions of DIANOGA simulations at $z = 2.2$. The epoch around $z \sim 2$ marks the emergence of the first virialized massive clusters ($M_{200} \sim 10^{14}\ M_\odot$) from the cosmic web, offering a view into the very early stages of cluster evolution. In addition, at $z \sim 2$ observational studies of protoclusters are more abundant, compared to higher redshifts, thus this epoch is ideal for a detailed comparison between simulations and observations. First, I provide an operational definition of protocluster regions in simulations (Sect. 4.1); then, I introduce the observed systems used in the comparison with simulations and compare their masses with those of halos populating the DIANOGA protoclusters (Sect. 4.2). In Sect. 4.3, I present the analysis of the galaxy stellar mass function (GSMF); I then discuss the star-forming main sequence (MS; Sect. 4.4) and the fraction of quenched galaxies (Sect. 4.5), also addressing their identification with observational techniques and on possible related biases, through the photometric properties of simulated galaxies modeled with radiative transfer simulations (Sect. 4.5.1). Finally, I analyze possible systematics affecting the adopted definitions of stellar mass and of star formation rate (SFR) in simulations (Sects. 4.6 and 4.7).

4.1 IDENTIFICATION OF PROTOCLUSTERS

It is worth reminding that no unique and objective definition exists for protoclusters (PCs) in observations, since the fate of observed galaxy overdensities can only be predicted statistically. In the following, I adopt an operational definition of a PC as the volume occupied by the DM that collapses into the final R_{200} radius of the $z = 0$ cluster. To do this, for each cluster I trace the DM particles that fall within R_{200} at $z = 0$ back to $z = 2.2$ and I identify the main progenitor of the cluster as the halo to which the majority of them are associated by the Friends-of-Friends halo finder. For the two DIANOGA clusters that were not simulated to $z = 0$, I assume that the main progenitor at $z = 2.2$ is the most massive halo within the high-resolution region. I verified that this is indeed the case for nine out of the twelve regions that were simulated to $z = 0$. As for the remaining three regions, in two cases the main progenitor is the second most massive halo (with mass ratio to the most massive of ~ 1.1), while in the third case it is the fifth most massive halo (mass ratio ~ 1.8).
After identifying the main progenitor halo, I define the PC region as the spherical region around its center, encompassing 80% of the traced DM

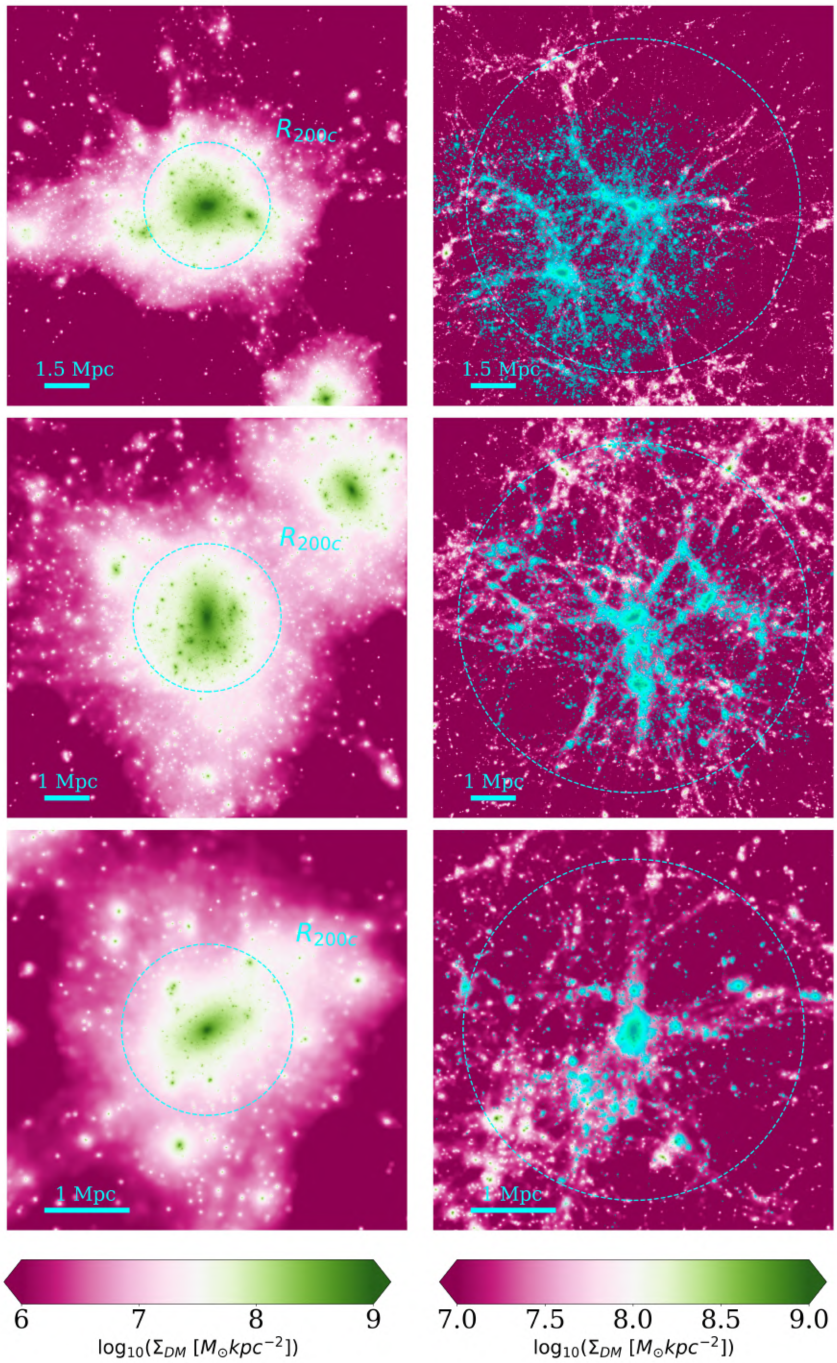


Figure 4.1: Projected DM density maps of three simulated clusters at $z = 0$ (left column), with total masses $M_{200} = 2 \times 10^{15}\ M_{\odot}$ (top), $9 \times 10^{14}\ M_{\odot}$ (middle), and $2 \times 10^{14}\ M_{\odot}$ (bottom), and their $z = 2.2$ progenitors (right column). The cyan circle at $z = 0$ traces the R_{200} radius of each cluster, the one at $z = 2.2$ traces the PC region as defined in the text. DM particles at $z = 2.2$ that collapse within the R_{200} radius by $z = 0$ are plotted in cyan. Densities are computed in pixels of size 15 kpc in the top panels and 10 kpc in the middle and bottom panels.

particles. This specific value of the percentage of DM particles is clearly arbitrary, but represents a sensible operational definition in order to encompass a region that is large enough to have a complete view of the cluster progenitor environment, while avoiding regions that are dominated to a very large extent by structures that are not part of the cluster by $z = 0$. Indeed, I verified that, when considering the region between the spheres that contain 80 and 100% of the traced particles, 94%, on average, of the particles in this region do not fall within the descendant cluster by $z = 0$, compared to approximately 40% within the chosen boundary.
For the two simulations that do not reach $z = 0$, I know from the parent DM-only simulation that the cluster masses M_{200} at $z = 0$ are in the range $(2-3) \times 10^{15}\ M_{\odot}$. I then assume the radial extent of these PC regions to be equal to the mean radius (expressed in units of R_{200} of the main progenitor halo at $z = 2.2$) of the progenitors of the six most massive clusters, whose masses are in the range $(1-2) \times 10^{15}\ M_{\odot}$. The size of such PCs at $z = 2.2$ is approximately $15\times\ R_{200}$. I summarize the properties of the PC regions in Table 4.1.

Table 4.1: Properties of the DIANOGA PCs. Reported below are the ID, the mass of the $z = 0$ cluster, the mass of the main progenitor halo at $z = 2.2$, the radius of the PC region and the number of PC cores (halos with $M_{200} > 1.5 \times 10^{13}\ M_{\odot}$ within the PC region) for each DIANOGA PC.

ID	$M_{200,z=0}$ [$10^{14}\ M_{\odot}$]	$M_{200,MP}$ [$10^{13}\ M_{\odot}$]	radius [Mpc]	N_c
D2	5.2	3.4	3.9	2
D3	7.6	4.8	4.4	3
D4	6.0	1.8	4.0	1
D5	1.8	4.0	2.1	1
D6	16	8.9	6.2	2
D7	16	12	6.0	4
D9	1.5	2.8	2.1	1
D10	15	7.0	5.3	6
D13	18	3.2	7.9	8
D15	18	11	7.1	4
D16	32	9.1	6.7	7
D17	18	8.3	6.5	5
D21	18	11	6.2	2
D25	11	3.4	5.5	3

Figure 4.1 shows the DM density maps of three DIANOGA clusters at $z = 0$ (left column), with $z = 0$ masses $M_{200} = 2 \times 10^{15}\ M_{\odot}$ (top), $9 \times 10^{14}\ M_{\odot}$ (middle) and $2 \times 10^{14}\ M_{\odot}$ (bottom) and of their $z = 2.2$ progenitors (right column). At $z = 2.2$, the figure also shows particles ending up within R_{200} at $z = 0$. The choice of setting the boundary of the PC at the sphere

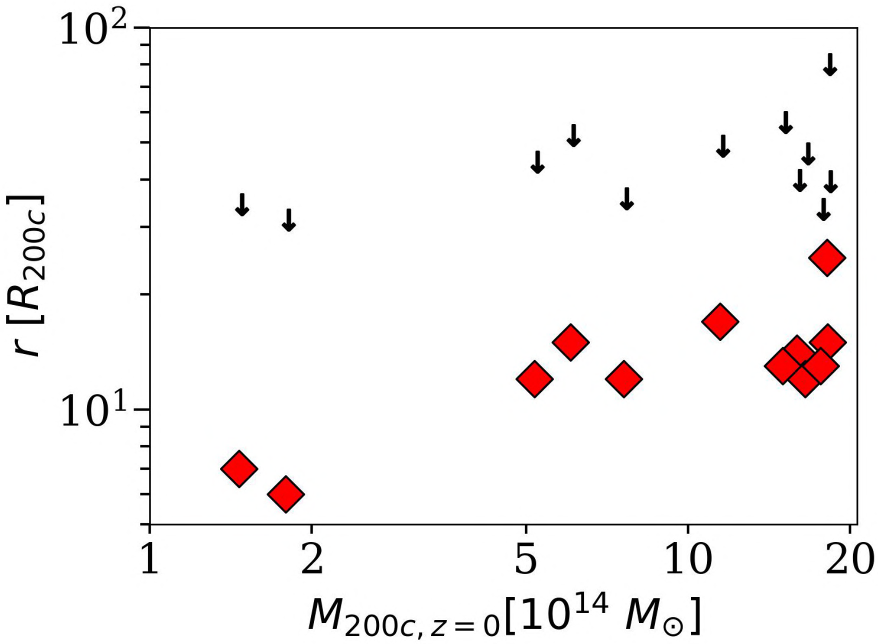


Figure 4.2: Extension of the PC regions (in units of R_{200} of the main progenitor halos) at $z = 2.2$ as a function of the mass, M_{200}, of the descendant clusters at $z = 0$ (red diamonds). The downward arrows show, for each progenitor, the distance from the boundary of the high-resolution region (see main text).

containing 80% of the total DM particles that are found within R_{200} by $z = 0$ is sufficient for encompassing all relevant DM structures that are merging into the final cluster. However, the assumption of spherical symmetry is not always optimal, as shown in the top-right panel of Fig. 4.1. This PC region exhibits two halos of comparable size, and the center of the main halo here does not approximate the center of mass of the distribution of the traced DM particles. Nevertheless, I maintained this definition by analogy with PCs defined in observations, often identified around the most massive overdensity observed in a PC region.

I show the radius of the PC regions as defined above in Fig. 4.2, in units of R_{200} of the main progenitor halo, as a function of the final ($z = 0$) mass of the cluster. I also show, in Fig. 4.2, the closest boundary of the high-resolution region, defined as the distance of the closest low-resolution DM particle from the center of the main progenitor halo. I point out that this is a conservative definition of the boundary of the high-resolution zoom-in region. In fact, few low-resolution DM particles could spuriously be present within the regions where high-resolution particles are largely dominant. I verified that no such particles are present within the defined PCs, thus guaranteeing that the chaotic diffusion of low-resolution particles does not contaminate these regions.

As expected, there is a correlation between the extension of the PC radius

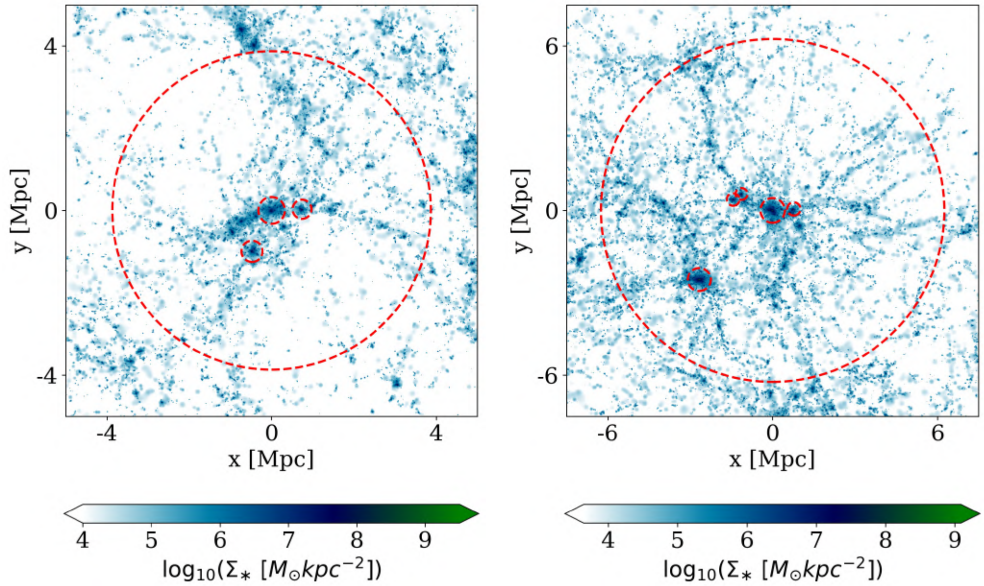


Figure 4.3: Projected stellar mass density maps of two DIANOGA PC regions, which are the progenitors of clusters of total mass M_{200} equal to 5×10^{14} $M_\odot$ and 2×10^{15} $M_\odot$ at $z = 0$ (left and right panel, respectively). In each panel, the larger red dashed circle corresponds to the PC region as defined in the text, while the smaller circles show the R_{200} of halos in these regions with $M_{200} > 10^{13}$ $M_\odot$. Densities are computed in pixels of size 15 (10) kpc in the right (left) panel.

and the final mass of the cluster it will collapse into, reflecting the hierarchical build-up of structures. On the other hand, the arbitrariness of the PC center for large systems, such as the one shown in the top-right panel of Fig. 4.1, produces an artificial increase of the sphere's radius for more massive systems.

Figure 4.3 shows the projected maps of stellar mass densities in the progenitors of two clusters, with $M_{200} = 5 \times 10^{14}$ $M_\odot$ and 2×10^{15} $M_\odot$ at $z = 0$. Both regions are characterized by filaments of galaxies intersecting each other at the locations where massive halos are emerging.

In the following, I compare the properties of PC galaxies with those in the cosmological box to highlight the environmental signatures predicted by the DIANOGA simulations. In this box, two cluster-size halos of masses $M_{200} = 1.8$ and 1.7×10^{14} $M_\odot$ are present at $z = 0$; however, I did not exclude their progenitors from the analysis of the galaxy population, following the "field" definition adopted in observational studies, which refers to a large, representative volume, with all the possible overdense structures it contains.

4.1.1 *Definition of galaxy properties*

Here I provide the definitions of the stellar and gas masses, as well as of SFRs, adopted in this Chapter. I use SubFind catalogs to define the centers of

galaxies and the stars or gas particles bound to them. To compare to galaxy properties as measured from observations, I impose a fixed 2D aperture of 1 arcsec (approximately 8.5 kpc at $z = 2.2$) in radius, along a random line of sight, for consistency with observations typically adopting similar apertures. I compare stellar masses computed within these apertures to those directly provided by `SubFind` in Sect. 4.6.
After identifying the star particles that belong to each galaxy, I compute the corresponding SFR as follows. Each star particle is characterized by an age (time elapsed since it has been generated, see Sect. 3.1.3.3). I then compute SFRs averaged over a typical timescale (which was chosen based on the observational data used for the comparison, see Sect. 4.4), by considering the total initial mass of star particles younger than the given timescale and dividing by it, within the adopted aperture. This allows for a more proper comparison with the observations rather than relying on noisier instantaneous SFRs that could be calculated by summing over the SFRs of all the star-forming gas particles that `SubFind` associates with the galaxy. I show the comparison between the instantaneous and averaged SFRs in Sect. 4.7.
When comparing with observations in this Chapter, observed stellar masses and SFR estimates are converted to Chabrier (2003) IMF, adopted in the simulations, where necessary.

4.2 PROPERTIES OF PROTOCLUSTER HALOS

In this Section, I present the properties of the halo population in `DIANOGA` protoclusters. I compare their mass range with that of observed halos associated with protoclusters or high-redshift clusters, which serve as a reference for the comparison of simulated galaxy properties in the following sections. Additionally, I examine the velocity dispersion of their member galaxies to provide a more complete comparison with dynamically inferred masses and to identify potential sources of discrepancy in such estimates. Indeed, velocity dispersions are used in dynamical studies of observed halos as proxies for their total masses (e.g., Shimakawa et al. 2014), assuming spherical symmetry and local relaxation. In the simulated PCs, I considered all halos with $M_{200} > 10^{13}$ $M_{\odot}$, and calculated the 1D velocity dispersions of galaxies identified within the 3D R_{200} of each halo, along three orthogonal lines of sight.
I show these quantities in Fig. 4.4, where the masses M_{200} are plotted as a function of the galaxy velocity dispersions along the three lines of sight. These results are compared with observational estimates for (proto)clusters at $z \sim 2 - 2.5$ from Gobat et al. (2013), Shimakawa et al. (2014), Wang et al. (2016, 2018b), Mantz et al. (2018) and Di Mascolo et al. (2023). The estimates from Shimakawa et al. (2014) for the Spiderweb PC at $z = 2.2$ and the two subgroups of USS1558 at $z = 2.53$, within their estimated R_{200}, come from a dynamical analysis based on velocity dispersions. The remaining data include estimates based on observations of the ICM: X-ray data for CLJ1001 at $z = 2.5$ (Wang et al. 2016, in agreement with estimates based on stellar

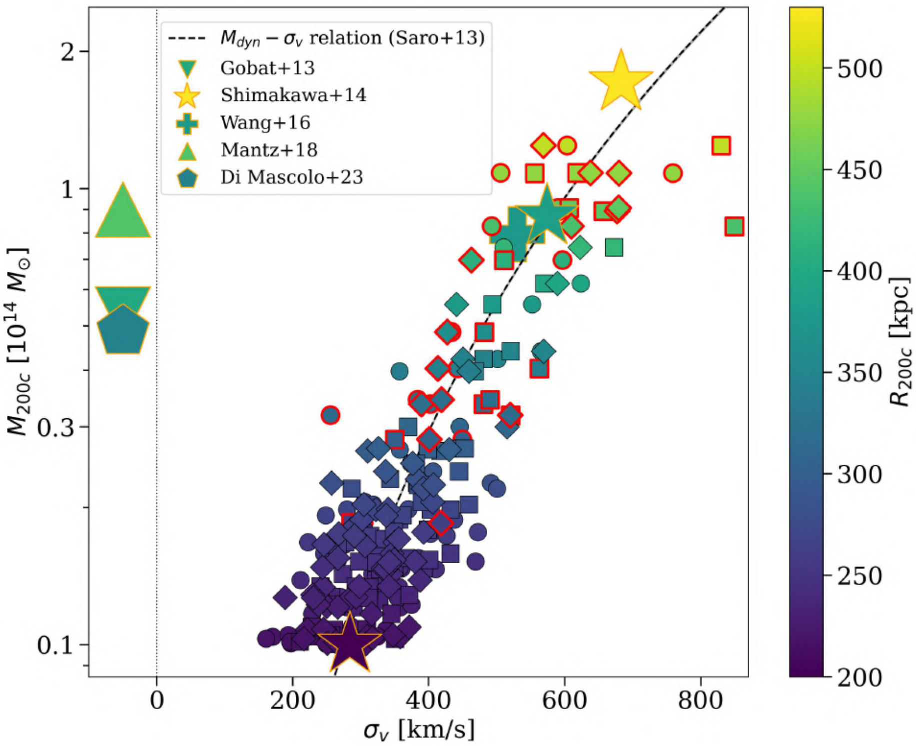


Figure 4.4: Mass M_{200} of massive halos in PC regions as a function of velocity dispersion σ_v along three orthogonal lines of sight (represented by circles, squares and diamonds) as traced by galaxies with stellar masses larger than 10^9 $M_\odot$ and lying within the R_{200} of each halo. Each point is color coded according to the halo R_{200}. The symbols with red contours mark the central halos of DIANOGA PCs. The dashed curve shows the relation between dynamical masses and galaxy velocity dispersions calibrated in simulations by Saro et al. (2013) up to $z = 1.2$, and extrapolated to $z = 2.2$. Star symbols represent velocity dispersions and dynamical mass estimates within the core of the Spiderweb PC and for the two subgroups of USS1558 from Shimakawa et al. (2014). The cross symbol corresponds to CLJ1001 (Wang et al. 2016). The pentagon symbol shows M_{200} for the Spiderweb PC, from Sunyaev-Zeldovich effect measurements with ALMA (Di Mascolo et al. 2023), the triangle represents the cluster XLSSC 122 (Mantz et al. 2018; Willis et al. 2020; Noordeh et al. 2021) and the downward triangle represents CLJ1449 (Gobat et al. 2013). The latter three points are shown at $\sigma_v < 0$ for illustrative purposes, since velocity dispersions measured consistently within the estimated radii are not available.

mass and velocity dispersion), XLSSC 122 at $z = 1.98$ (Mantz et al. 2018; Willis et al. 2020; Noordeh et al. 2021), and CLJ1449 at $z = 2.00$ (Gobat et al. 2013; Valentino et al. 2016; Strazzullo et al. 2018), as well as estimates from the SZ signal for the Spiderweb PC (Di Mascolo et al. 2023, consistent with X-ray measurements from Tozzi et al. 2022b). Wherever needed, I convert M_{500} to M_{200} (and R_{500} to R_{200}) using the relations from Ragagnin et al. (2021).

I also show an analytic redshift-dependent relation between dynamical mass and velocity dispersion, calibrated by Saro et al. (2013) on simulations up to $z = 1.2$, and extrapolated to $z = 2.2$, in Fig. 4.4. In general, the simulated halos agree well with this relation. For the few large halos with $M_{200} \sim 10^{14}\ M_{\odot}$, the figure suggests an increased scatter in velocity dispersions, though with very limited statistics. This might be due to the earlier stage of assembly of these objects. In such cases, the assumption of spherical symmetry underlying the definition of R_{200} may not be valid. The comparison with Shimakawa et al. (2014) shows good overlap with the mass and velocity dispersion ranges of the simulated halos studied here, with the Spiderweb PC core, lying at the high-mass end of the range probed by simulations, a factor $\lesssim 2$ more massive than the most massive simulated halos. On the other hand, this measurement of the Spiderweb PC core mass is a factor ~ 3 larger than the ICM-based estimates (Tozzi et al. 2022b; Di Mascolo et al. 2023), which are well within the mass range of simulated halos, as well as of the other observed systems mentioned above.

The inconsistency between different observations of the same system is likely due to the complications of a dynamical analysis in highly disturbed systems, especially at these redshifts. Such analyses rely on the assumptions of local virialization, spherical symmetry, and the reliable removal of line-of-sight interlopers, all of which may bias the measurement of the velocity dispersion.

With these caveats, I conclude that the DIANOGA PC regions are populated with halos similar in mass and velocity dispersion (when available) to clusters and PC cores identified in observations at $z \sim 2 - 2.5$.

4.3 GALAXY STELLAR MASS FUNCTION

The galaxy stellar mass function is an important observable in constraining galaxy formation models since it traces the integrated star formation history in a given region. In cosmological simulations, the GSMF at various redshifts is often used to calibrate sub-resolution models related to star formation and feedback processes. However, in the DIANOGA simulations, no calibration was performed to reproduce the GSMF, making it a genuine prediction.

In this section I present the simulated GSMF and investigate its environmental dependence by selecting galaxies from different regions within the PCs and in the cosmological box. I also compare the simulated GSMF with observed GSMFs in PCs at $z \sim 2 - 2.5$, in order to assess whether DIANOGA simulations can reproduce the integrated star formation history observed in these systems.

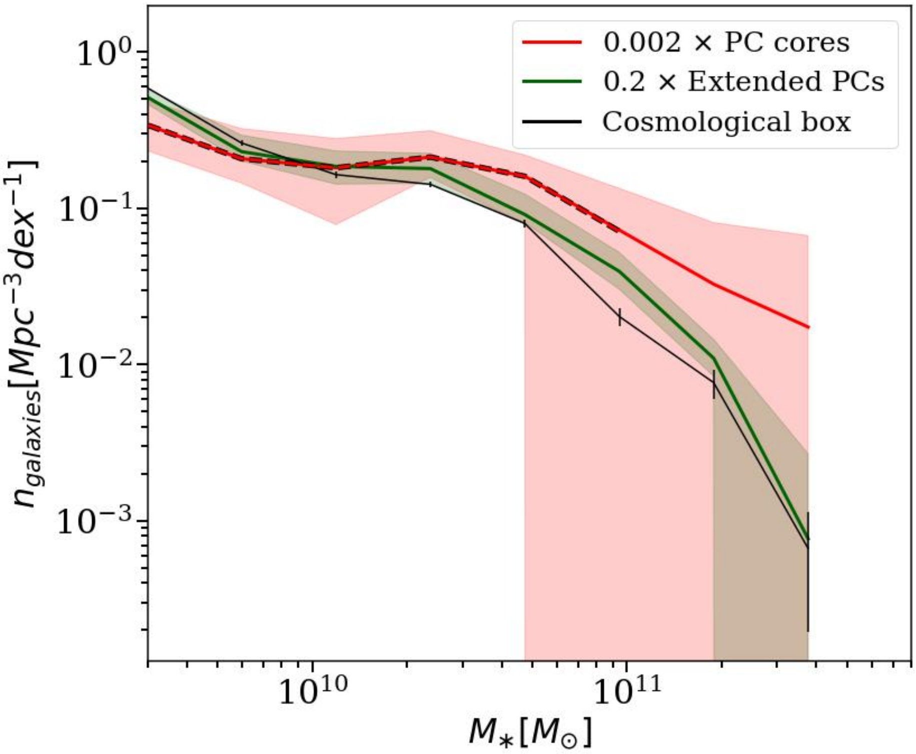


Figure 4.5: Galaxy stellar mass functions in the PC environment compared with the cosmological box. The red line is the median of the GSMFs in PC cores, with and without central galaxies (solid and dashed lines, respectively; I note that the dashed line stops at $M_* \sim 10^{11} M_\odot$ because the median GSMF in PC cores in the higher mass bins is zero when central galaxies are excluded). The green line shows the median GSMF in the less dense, diffuse PC regions, masking out all PC cores. The normalizations have been rescaled by the factors reported in the legend. The shaded regions define the 16th-84th percentile range spanned by the different PCs (or PC cores). The black line is the GSMF in the cosmological box, with Poissonian uncertainties.

The simulated GSMF for `DIANOGA` PCs is shown in Fig. 4.5. To highlight environmental signatures associated with the most massive halos in the PC regions, I separately considered galaxies hosted by halos with $M_{200} > 1.5 \times 10^{13}\ M_\odot$ in PCs. I refer to such halos as PC cores, whose extension is defined by their R_{200} radius. This separation is intended to investigate differences between galaxy populations in the most massive halos within PCs and those in the extended PC structure. The specific threshold of $M_{200} = 1.5 \times 10^{13}\ M_\odot$ corresponds to the mean mass of the third most massive progenitor halo, with the median mass ratio between first and third most massive progenitors at this redshift being 3.5. I show the predicted GSMF for PC cores, for the extended PC regions[1] and for the cosmological

1 To calculate the GSMFs of PC cores I used the volume of the sphere with the radius R_{200} of each halo, while for the extended PCs I used the radius of the PC region as defined in Sect. 4.1, from which I subtract the cumulative volume occupied by the PC cores.

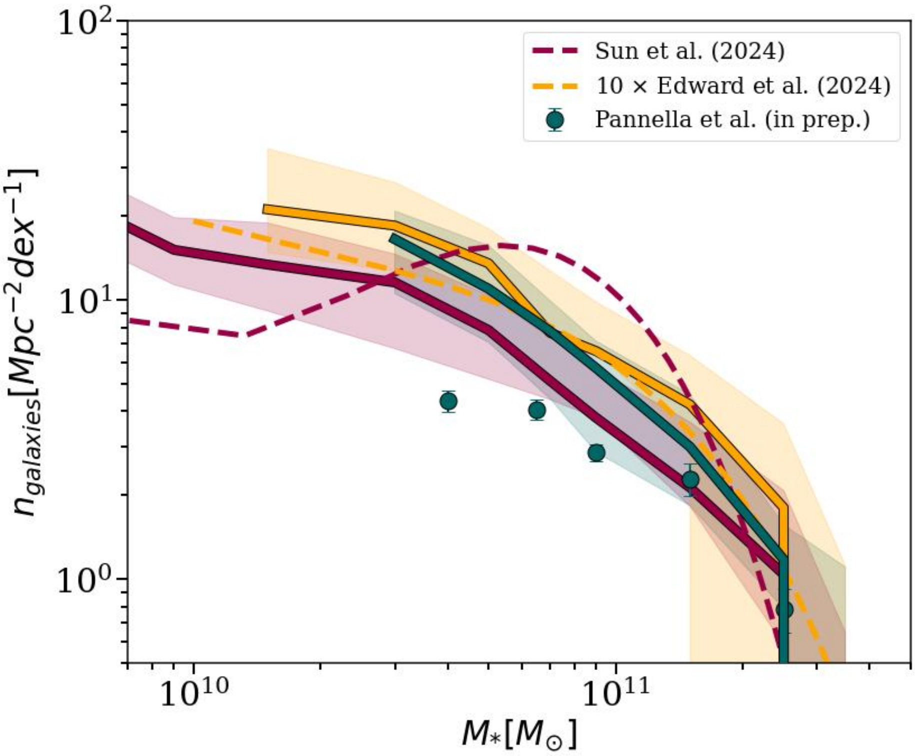


Figure 4.6: Comparison between the GSMF in DIANOGA PCs and in observations of PCs at $z \sim 2-2.5$ (Spiderweb, CLJ1001, 14 PCs in the COSMOS field; Pannella et al. in prep; Sun et al. 2024; Edward et al. 2024, respectively). For each set of observations, a simulated GSMF has been produced to match the selection of galaxies as closely as possible. For each selection, GSMFs have been extracted along three orthogonal lines of sight, convolved with Gaussian errors with scatter 0.3 dex (see text), and the solid lines (shaded regions) show the median GSMFs (16th-84th percentile ranges, respectively) across all realizations, in the same color as the corresponding data. The GSMF from Edward et al. (2024) has been arbitrarily re-scaled by a factor of 10 to match the normalization in simulations, since the extension of these PCs along the line of sight in unconstrained and, thus, it cannot be replicated in the simulations.

box. The GSMF normalization is obviously higher in denser regions; the GSMFs shown in Fig. 4.5 are thus rescaled to highlight differences in their shape.

The comparison between the GSMFs in the cosmological box and in the extended PC regions reveals a slight relative excess of massive galaxies, with $M_* > 10^{10}\ M_\odot$, in the latter. This indicates that even when excluding galaxies in large halos, PCs exhibit a faint signature of accelerated evolution at earlier times. As expected, the excess of massive galaxies is more pronounced for the median GSMF in PC cores. This signal is completely dominated by central galaxies at $M_* \gtrsim 10^{11}\ M_\odot$. This result is qualitatively consistent with recent observational findings, such as the top-heavy feature in the GSMF of CLJ1001 at $z = 2.5$ (Sun et al. 2024) and the large ratio of high-mass to low-mass galaxies relative to the field in the Elentári proto-supercluster at $z \sim 3.3$ (Forrest et al. 2024). I note that the GSMF of PC cores shows significant variability at high stellar masses among different cores, with a considerable scatter toward low number densities at $M_* \gtrsim 5 \times 10^{10}\ M_\odot$, which seems to be primarily due to the lower-mass PC cores not having formed a significant fraction of massive galaxies yet.

I note that the presence of massive halos in the cosmological box does not impact the global properties of the galaxy population. I identified the three most massive halos, with M_{200} in the range $(1-3) \times 10^{13}\ M_\odot$, in the cosmological box and defined a PC region around each of them, assuming an extent of 6.5 times their R_{200} (the average radius of the `DIANOGA` PCs with descendant mass $M_{200} \sim 10^{14}\ M_\odot$). The galaxies within such regions comprise only 4% of the total galaxy population within the box, and their presence has only a negligible effect on the GSMF.

In Fig. 4.6 I compare the predicted GSMF with observations in PCs at similar redshift, including the Spiderweb PC (Pannella et al., in prep.), CLJ1001 (Sun et al. 2024) and 14 PCs identified in the COSMOS field at $2 < z < 2.5$ (Edward et al. 2024). I try to replicate the selection of the observational data as closely as possible for an accurate comparison between simulations and observations. For the comparison with Pannella et al. (in prep.) and Sun et al. (2024), I compute the GSMF within the same projected area covered by the observations and apply a redshift cut (including the contribution from peculiar motions) equivalent to the one imposed by the narrow-band filters in the data. Edward et al. (2024) use photometric redshifts with statistical subtraction of the background, meaning that the actual extension of the probed PC regions along the line of sight is unconstrained. As a result, for comparison to this work I only match the projected area of these observations and rescale their GSMF fit by a factor of 10 to approximately match the normalization of the simulated GSMF. For each simulation, I repeat the selection along three orthogonal lines of sight, extracting the median across all 3×14 realizations of the GSMF and computed the 16th-84th percentiles to define the scatter among simulated PCs. I apply a convolution with a Gaussian distribution of errors in the logarithm of stellar masses, with a dispersion of 0.3 dex, to approximate observational uncertainties (e.g., Conroy 2013, Mobasher et al. 2015, Lower et al. 2020). For Sun et al. (2024), I show the sum of the observed GSMFs of star-forming and quenched galaxies, although the latter is affected by incompleteness in the shown stellar mass range.

This comparison highlights an overall agreement in the normalization of the GSMF with observations (excluding Edward et al. 2024, where the normalization has been arbitrarily rescaled as mentioned above). The shape of the simulated GSMF is in very good agreement with Edward et al. (2024) and consistent with Pannella et al. (in prep.) at high stellar masses (note that these observations may suffer from significant incompleteness at lower stellar masses; see Tozzi et al. 2022a). On the other hand, the simulations do not reproduce the pronounced top-heavy shape observed in the GSMF of CLJ1001 (Sun et al. 2024), which is dominated by an excess of massive star-forming galaxies. The authors propose two possible scenarios to reconcile this feature with observations of lower-redshift clusters. First, galaxies in the central region of CLJ1001 may be in the process of merging to form a BCG, which is currently absent in this system. Second, massive star-forming galaxies, with gas depletion timescales shorter than 0.5 Gyr, may undergo rapid quenching. While these scenarios were introduced to address the tension with lower-redshift cluster GSMFs, if they capture the physical origin of the peculiar GSMF of CLJ1001, they might also be considered to explain the discrepancy with other observed protoclusters at $z \sim 2$, as well as with simulations. However, in this context, concerning the quenching scenario: even if galaxies in simulations (and in other protoclusters) were characterized by different quenching timescales, the excess of massive galaxies observed in CLJ1001 could not be reproduced, since such an excess is not present in the simulations even when both star-forming and quiescent populations are included. On the other hand, concerning the merging scenario: if CLJ1001 is caught immediately before the merger-driven assembly of a BCG, its GSMF might at least in principle be reconciled with simulations (and observations of other coeval protoclusters), though a dedicated analysis is necessary to quantify the actual occurence of this scenario in DIANOGA simulations.

4.4 STAR-FORMING MAIN SEQUENCE

The star-forming MS is a key observable for constraining the mode of star formation as a function of stellar mass at different epochs. It is therefore useful to assess the effect of environment on the star formation activity of galaxies. However, theoretical models struggle to match the normalization of the observed MS at cosmic noon, with most semi-analytic models and cosmological simulation suites underestimating galaxy SFRs at high redshift (see, e.g., Granato et al. 2015; Hirschmann et al. 2016; Donnari et al. 2019; Akins et al. 2022; Andrews et al. 2024; Fontanot et al. 2024), despite the fact that simulations can be tuned to reproduce the observed GSMFs quite well. Star formation has been investigated in lower-resolution versions of the DIANOGA simulations, showing indeed lower SFRs of galaxies in PCs with respect to the observed MS (e.g., Bassini et al. 2020) and a star formation

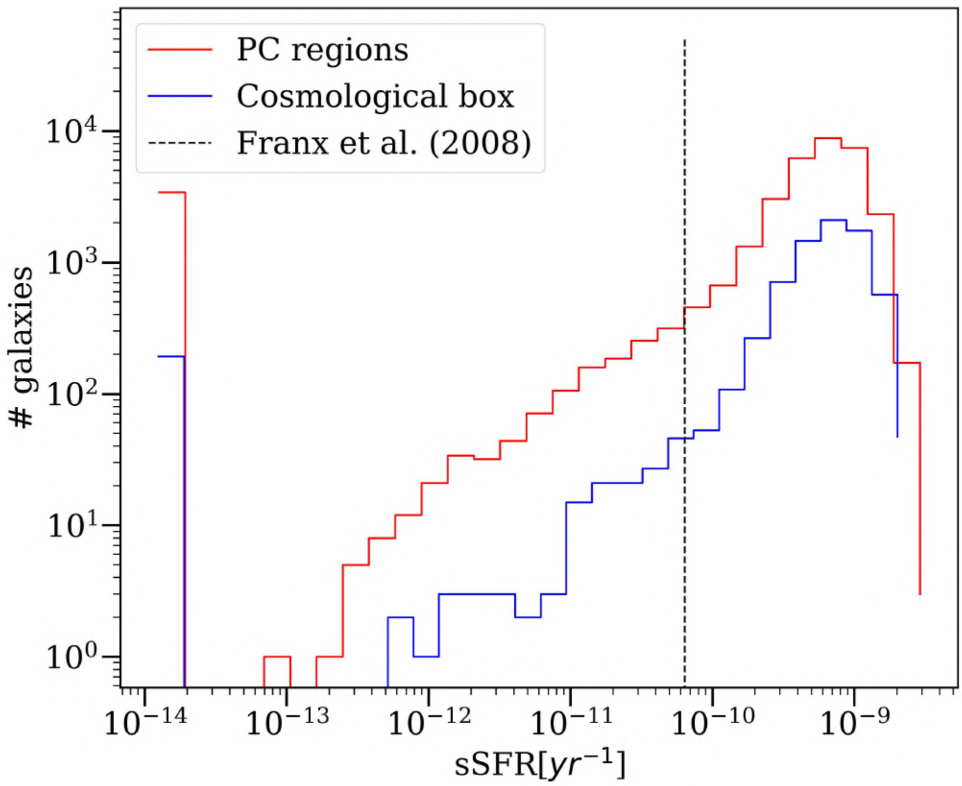


Figure 4.7: Distributions of sSFRs of galaxies in the DIANOGA PC regions (red line) and in the cosmological box (blue line) at $z = 2.2$ compared with the threshold separating star-forming from quenched galaxies, at this redshift, from Franx et al. (2008). Galaxies with SFR=0 have been artificially assigned the minimum non-zero sSFR, corresponding to 10^{-14} yr^{-1}.

history that is not sufficiently bursty and thus is unable to reproduce far-infrared (FIR) observations in high-redshift PCs (Granato et al. 2015).

In the following, I compare the star-forming MS in the DIANOGA PCs and in the cosmological box at $z = 2.2$ to investigate the impact of environment on the population of star-forming galaxies, and I compare these results with observations in the field and in the Spiderweb PC.

I classify galaxies as star forming if their specific SFR (sSFR) exceeds $0.3/t_H$ (where t_H is the Hubble time at $z = 2.2$; Franx et al. 2008). Figure 4.7 shows that this threshold appropriately separates star-forming and quenched galaxies in DIANOGA simulations, falling approximately 1 dex below the peak of the simulated MS and intersecting the sSFR distribution at the point where quenched galaxies lead to a change in the slope of the distribution, both in the PCs and the cosmological box.

In Fig. 4.8, I compare the simulated MS with observational data for PC and field galaxies at the same redshift. Specifically, I compare the median SFRs as a function of stellar mass across all the simulated PCs and in the cosmological box with observations in the Spiderweb PC from Pannella et al. (in prep., with SFR measurements based on dust-corrected rest-frame

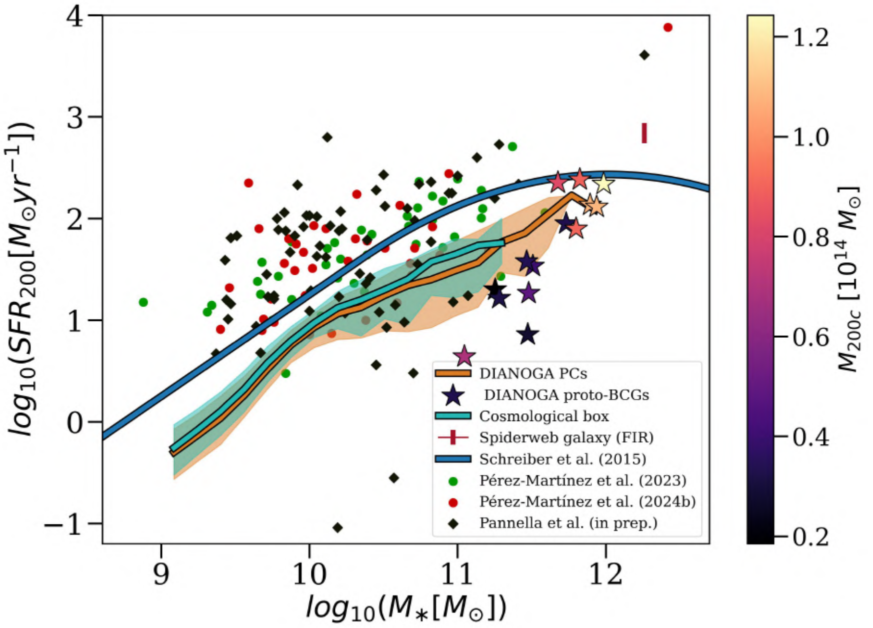


Figure 4.8: Star-forming MS of galaxies in the DIANOGA PCs (orange line) and in the cosmological box (green line). The solid orange and green lines show the median SFRs of all star-forming galaxies within each stellar mass bin while the shaded areas mark the 16th-84th percentile intervals. The star symbols show the proto-BCGs identified in the simulations, color coded according to the mass M_{200} of the main progenitor halo of each PC. Black diamonds, green and red dots are data for the Spiderweb PC from Pannella et al. (in prep.) and from Pérez-Martínez et al. (2023) and Pérez-Martínez et al. (2024b), respectively. The blue line is the best fit for the observed field star-forming MS by Schreiber et al. (2015) at $z = 2.2$. The red bar shows the interval of estimated SFRs for the Spiderweb galaxy from FIR measurements by Seymour et al. (2012) and Drouart et al. (2014). Simulated SFRs are averaged over 200 Myr (see text). For readability, error bars on observational data have been omitted.

UV fluxes), Pérez-Martínez et al. (2023, with Hα-based SFRs) and Pérez-Martínez et al. (2024b, with Paβ-based SFRs), and with field measurements from Schreiber et al. (2015, based on Herschel observations). To ensure a more proper comparison with these observations, I compute the SFRs of simulated galaxies averaged over 200 Myr, which is approximately the timescale of the SFR tracers used in Pannella et al. (in prep.) and in Schreiber et al. (2015), while Hα and Paβ observations used in Pérez-Martínez et al. (2023) and Pérez-Martínez et al. (2024b) trace an almost instantaneous SFR, with the relevant timescales being 5-10 Myr. I show in Fig. 4.17 that the SFRs in the cosmological box and in the PCs at $z = 2.2$ decline over the past 200 Myr in DIANOGA simulations, with instantaneous SFRs systematically lower than averaged ones across all probed stellar masses. Nevertheless, the offset between these two SFRs estimates (~ 0.1 dex on average) is much smaller than typical observational uncertainties. Therefore, considering also the limited sample sizes in MS investigations in observed high-redshift PCs, I expect the comparison between simulated and observed SFRs in Fig. 4.8 not to be significantly affected by the different timescales of the mentioned SFR tracers. Thus, I use SFRs averaged over 200 Myr in Fig. 4.8 and in the following.

In this analysis, I compare SFRs from the simulations with observational SFR estimates derived from the tracers mentioned above, which are affected by different complications. In particular, dust attenuation affects UV/optical SFR tracers: although the observational data used in this comparison are corrected where needed for this effect, such correction remains affected by limitations. In this analysis, I use all SFR estimates as they are provided in the original studies, and compare them with the intrinsic SFRs from the simulations, without accounting for possible residual differences among tracers of obscured and unobscured star formation.

The simulated MS is systematically lower than the observed one, both in the field and in the Spiderweb PC. The 11 star-forming proto-BCGs[2] also have lower SFRs compared to the Spiderweb galaxy. A similar underestimation of SFRs at high redshift is present also in lower-resolution simulations of the DIANOGA PCs (Granato et al. 2015; Ragone-Figueroa et al. 2018; Bassini et al. 2020), which also adopt slightly different prescriptions for the feedback processes. This suggests that the primary cause for the lower MS normalization predicted by simulations lies in the star formation model (SH03), which implements a quiescent mode of star formation that inherently prevents any bursty star formation episodes, which are instead expected at cosmic noon. Ragone-Figueroa et al. (2024) show that by adopting a different model for star formation and stellar feedback (Murante et al. 2010, 2015; Valentini et al. 2020, 2023), coupled with H_2 formation on dust grains (Granato et al. 2021), the discrepancy between the simulated and observed MS can be reduced, also resulting in a cosmic star formation history that better matches the observed one.

Within the simulation, I observe that less massive PC main progenitors host

2 I define as proto-BCG the central galaxy of the main progenitor halo in each PC. 3 out of 14 proto-BCGs, also shown in Fig. 4.8, are quenched, according to the adopted definition.

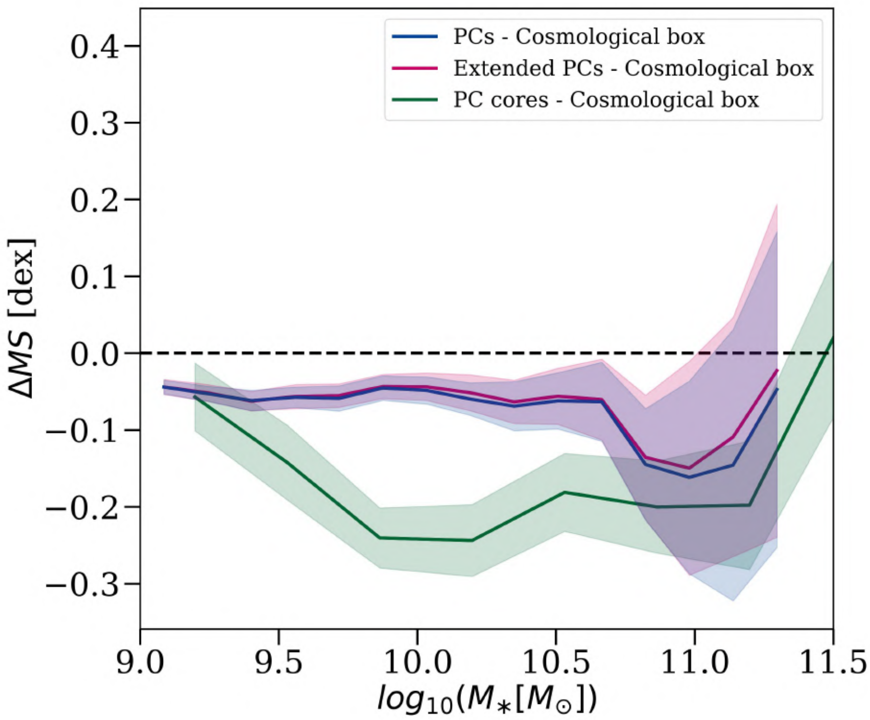


Figure 4.9: Environmental dependence of the simulated star-forming MS at $z = 2.2$ as a function of stellar mass. Results are shown for the whole sample of PC galaxies, as well as separating PC galaxies located within/outside R_{200} of halos with $M_{200} > 1.5 \times 10^{13}$ $M_{\odot}$ (PC cores/extended PCs), compared to galaxies in the cosmological box. The solid lines and shaded regions show the MS offset, and its intrinsic scatter, between pairs of environments as indicated in the legend (see text for details).

proto-BCGs with suppressed star formation compared to the (simulated) MS. The simulated MS in Fig. 4.8 also shows a rather minor suppression of star formation in PCs compared to the cosmological box[3]. This is further explored in Fig. 4.9 where, for a better understanding of the effect of the environment, I separate star-forming galaxies hosted by PC cores ($M_{200} > 1.5 \times 10^{13}$ $M_{\odot}$) from those in the lower-density extended regions of PCs, comparing them also with galaxies in the cosmological box. Figure 4.9 shows the environmental dependence of the star-forming MS by comparing offsets in SFRs between the aforementioned galaxy populations. To compute the offsets, I generate 1000 realizations of the MS of each galaxy population, assuming an (asymmetric) Gaussian distribution of SFRs as a function of stellar mass based on the median SFR and 16th-84th percentile range for the given population of simulated galaxies. For each population (and thus environment) pair, I then compute the mean of the SFR offsets as a function of stellar mass over these 1000 realizations. The error on the offset is estimated as the scatter of the SFR offsets across the realizations.

3 I verified that masking the "PC" regions in the cosmological box, as defined in Sect. 4.3, does not impact this result.

The simulations predict a very small though significant suppression of approximately 0.05 dex in the SFRs of PC galaxies compared to the cosmological box (note that this is insensitive to including or excluding the most massive halos, as their galaxy populations are subdominant compared to the full PC structure). This suppression is independent of stellar mass, though for $\log_{10}(M_*/M_\odot) > 11$ statistical uncertainties are too large to draw any conclusions. The observational uncertainties in the current MS determinations do not allow the detection of a 0.05 dex offset. Therefore, this prediction from simulations is compatible with studies that found either a low-significance or no suppression in the SFRs of PC galaxies compared to their field counterparts (e.g., Cucciati et al. 2014; Koyama et al. 2021; Polletta et al. 2021; Shi et al. 2021; Pérez-Martínez et al. 2023). On the other hand, the predicted offset contrasts with other studies finding an enhanced star formation in PC galaxies (e.g., Pannella et al. in prep.; Shimakawa et al. 2018a; Monson et al. 2021; Pérez-Martínez et al. 2024a).
Figure 4.9 further shows that the star formation suppression is more significant in PC cores, with an MS offset up to ~ 0.25 dex compared to the cosmological box. This offset decreases to ~ 0.05 dex for galaxies with $M_* \sim 10^9$ $M_\odot$. For these galaxies, whose star formation history has experienced a more marked recent decline (see Fig. 4.17), the SFR averaged over 200 Myr may be less sensitive to their relatively rapid evolution, especially in PC cores, as discussed in the following.

Given that star formation in the DIANOGA simulations depends on the gas reservoir and on the efficiency of gas conversion into stars, I further explore how these quantities compare with observational constraints. The reservoir of star-forming gas is represented by the cold component of the multiphase gas particles that describe the ISM in the SH03 star formation model (see Sect. 3.1.3.1). I compute the cold gas fraction, $\mu_{\text{cold}} = M_{\text{cold}}/M_*$, where M_{cold} is the total cold gas mass in a galaxy. I compare this fraction with the molecular gas fraction, M_{mol}/M_*, which is estimated in observational investigations, noting that M_{cold} represents an upper limit to the (sub-resolution) molecular gas content. I also compute the depletion time $t_{\text{dep}} = M_{\text{cold}}/\text{SFR}$, using the SFR averaged over 200 Myr.
In Fig. 4.10 I show μ_{cold} and t_{dep} as a function of stellar mass for star-forming galaxies, both in PCs and in the cosmological box, comparing them with observational results from Magdis et al. (2012), Liu et al. (2019), Sanders et al. (2023), Mérida et al. (2024), Leroy et al. (2024) and Pérez-Martínez et al. (2025) at $z \sim 1.5 - 3$. In both PCs and the cosmological box, the simulations predict cold gas fractions that are systematically lower than the observed values, to different extents, with offsets ranging from ~ 0.1 to ~ 1 dex, also likely depending on the kind of measurement and on the galaxy sample selection adopted in different observational studies. On the other hand, depletion times (thus the star formation efficiencies) appear consistent with observations, with the exception of the lower-mass data point in the Spiderweb protocluster from Pérez-Martínez et al. (2025).
Considering the realistic GSMF in the simulated PCs discussed in Sect. 4.3,

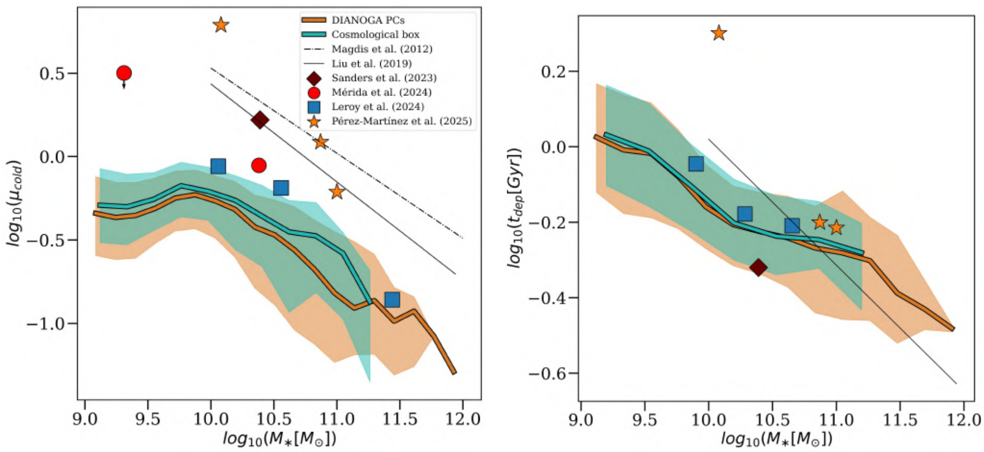


Figure 4.10: Cold gas fractions (left panel) and depletion times (right panel) in star-forming galaxies in the DIANOGA PCs (orange) and in the cosmological box (green) as a function of stellar mass. Solid lines (shaded regions) show median values (16th-84th percentile ranges, respectively). For comparison, I show observations at $z \sim 1.5-3$ from Magdis et al. (2012), Liu et al. (2019), Sanders et al. (2023), Mérida et al. (2024), Leroy et al. (2024) and Pérez-Martínez et al. (2025). In the latter, the two higher-mass values correspond to massive galaxies ($\log_{10}(M_*\ [M_\odot]) > 10.5$) that do and do not host AGNs (8 and 12 galaxies, respectively), at median masses $M_* \sim 10^{10.9}$ and $M_* \sim 10^{11}\ M_\odot$, respectively.

the cold gas deficit suggests a possible inefficiency in gas cooling, excessive heating due to feedback processes, or highly efficient interactions with the hot ambient gas, which begins to permeate the simulated halos at these redshifts (see Chapter 5), rather than an excessive early consumption of gas. As shown in Fig. 4.10, the fraction of cold gas is slightly lower in PCs than in the cosmological box. To investigate this further, I derive offsets in the median cold gas fraction and depletion times of galaxies in PCs in comparison with galaxies in the cosmological box, as a function of stellar mass, using the same methodology applied for the MS offsets in Fig. 4.9. The resulting offsets are shown in Fig. 4.11. For the cold gas fraction, I observe a small (~ 0.05 dex) though significant suppression in the cold gas reservoir in PC galaxies across all stellar masses, similar to the trends seen for the MS. For star-forming galaxies in PC cores, the offset in μ_{cold} relative to the cosmological box decreases at lower masses (increasing in absolute value), suggesting a more efficient depletion or less efficient replenishment of gas in low-mass galaxies. Since this feature is specific to the most massive halos in the PC regions, it is possible that interactions with the hot ICM emerging in these massive halos are either stripping cold gas from galaxies or halting cosmological gas accretion, leading to more rapid gas depletion, especially in lower-mass galaxies. In particular, the offset in μ_{cold} shows a sharp drop at $\log(M_*/M_\odot) \lesssim 9.5$. This is because approximately 40% of galaxies in this mass range within PC cores, classified as star forming based on their averaged SFRs, have completely lost their cold gas reservoir, effectively shutting off star formation over the past 200 Myr. While these

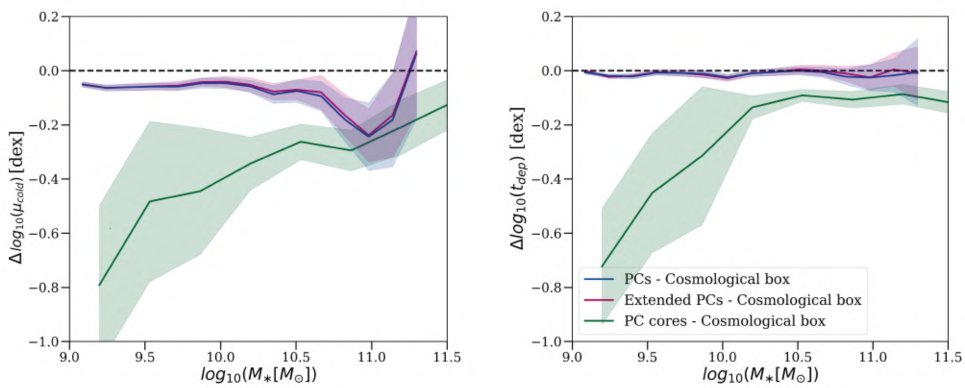


Figure 4.11: Environmental dependence of the fraction of cold gas (left panel) and of the cold gas depletion time (right panel) at $z = 2.2$ as a function of stellar mass. Here I consider four environments: the cosmological box, the whole PCs, the low-density extended PCs and the PC cores. The solid lines define the offsets between the cold fractions and the depletion times in the PC environments with respect to the cosmological box, as defined in the legend, resulting from 1000 Monte Carlo realizations (see text). The shaded regions represent the intrinsic scatter on the offsets.

galaxies are in a regime where the results could be influenced by the finite numerical resolution of the simulations, galaxies of similar mass in the cosmological box and in smaller halos within the PC regions do not show such an abrupt decrease in cold gas fractions. Therefore, I conclude that this is not purely a numerical artifact, although the extreme environment in the largest halos within the simulated PC regions, coupled with the relatively low number of gas particles in galaxies with $\log(M_*/M_\odot) < 9.5$, may amplify a genuine physical process occurring in massive halos at this redshift. Figure 4.11 also shows that the environmental dependence of depletion times qualitatively mirrors the trends observed for the cold gas fraction. The offset in the median depletion time of galaxies in the full PC region in comparison with the cosmological box (with no sensitivity to the inclusion or exclusion of the most massive halos) is marginal and statistically insignificant. However, low-mass PC galaxies within the most massive halos exhibit a dramatic decrease in their median depletion time, corresponding to the low median cold gas fractions discussed above.

4.5 QUENCHED GALAXY FRACTIONS

In this Section I analyze the population of quenched galaxies in the DIANOGA PCs and in the cosmological box at $z = 2.2$, to quantify how the environment of massive halos in PCs suppresses star formation in galaxies. This represents the first necessary step to exploit the simulations in probing the nature of early quenching processes in cluster progenitor environments, in the context of understanding the physical mechanisms resulting in the

predominance of quenched galaxies in mature clusters at later epochs.
I investigate the effects of the local environment, by isolating galaxies hosted by PC cores of different halo masses and galaxies populating the extended PC structures. I also separate galaxies in two stellar mass bins ($10^9 < M_*/M_\odot < 10^{10}$ and $M_* > 10^{10}$ $M_\odot$) to examine how galaxies of different masses are affected by the environment. The median quenched fractions as a function of halo mass, and the fractions in the wide PC structures and in the cosmological box, are shown in Fig. 4.12.
Across all stellar masses, the quenched galaxy fraction increases significantly in more massive halos, reaching up to 90% (70%) for the $10^9 < M_*/M_\odot < 10^{10}$ ($M_* > 10^{10}$ $M_\odot$) galaxy population at $M_{200} \gtrsim 10^{14}$ $M_\odot$. The scatter around the median is representative of the halo-to-halo variation, though higher halo mass bins might be affected by limited statistics.
The DIANOGA simulations predict an increased efficiency of quenching mechanisms compared to the cosmological box, across all the probed stellar mass and halo mass ranges, already at $z = 2.2$. Noticeably, the fractions of quenched galaxies in the extended PC regions exceed slightly those in the cosmological box, suggesting that star formation in galaxies might be suppressed before their infall into the virial region of a massive halo.
While these simulations predict an increasing quenched fraction as a function of stellar mass in the cosmological box and in the extended PC regions, in qualitative agreement with field observations (e.g., Muzzin et al. 2013; Kawinwanichakij et al. 2017; Sherman et al. 2020; Park et al. 2024), Fig. 4.12 shows instead higher quenched fractions for $10^9 < M_*/M_\odot < 10^{10}$ than for $M_* > 10^{10}$ $M_\odot$ galaxies in PC cores, at odds with observations in the field and in observed clusters and PCs (e.g., van der Burg et al. 2013; Cooke et al. 2016; Lee-Brown et al. 2017; van der Burg et al. 2020; Edward et al. 2024). This issue is common in various cosmological simulation suites across different redshifts and environments (e.g., Bahé et al. 2017; Donnari et al. 2021). This discrepancy between DIANOGA simulations and observations is likely a direct consequence of the low cold gas fractions in low-mass star-forming galaxies, as is discussed in Sect. 4.4.

In Fig. 4.13, I compare the quenched galaxy fraction in the most massive PC cores, with $M_{200} > 5 \times 10^{13}$ $M_\odot$, with observations in similar environments at $z \sim 2 - 2.5$, including XLSSC 122 (Willis et al. 2020; Noordeh et al. 2021), CLJ1001 (Wang et al. 2016) and CLJ1449 (Gobat et al. 2011; Strazzullo et al. 2018). Note that, while the quoted quenched fractions for XLSSC 122 and CLJ1001 are computed in very similar apertures (approximately 600 and 700 kpc, respectively) and the two (proto)clusters have similar halo masses ($M_{200} \sim 8 - 9 \times 10^{13}$ $M_\odot$), the quoted quenched fraction in CLJ1449 (with $M_{200} \sim 5 \times 10^{13}$ $M_\odot$) is computed within the much smaller aperture of 200 kpc.
To compare the quenched fractions in simulated PCs with these observations, I select galaxies in the same apertures and with stellar masses larger than the completeness limits in the referenced observational studies. I thus stress that measurements in different stellar mass bins shown in Fig. 4.13 are

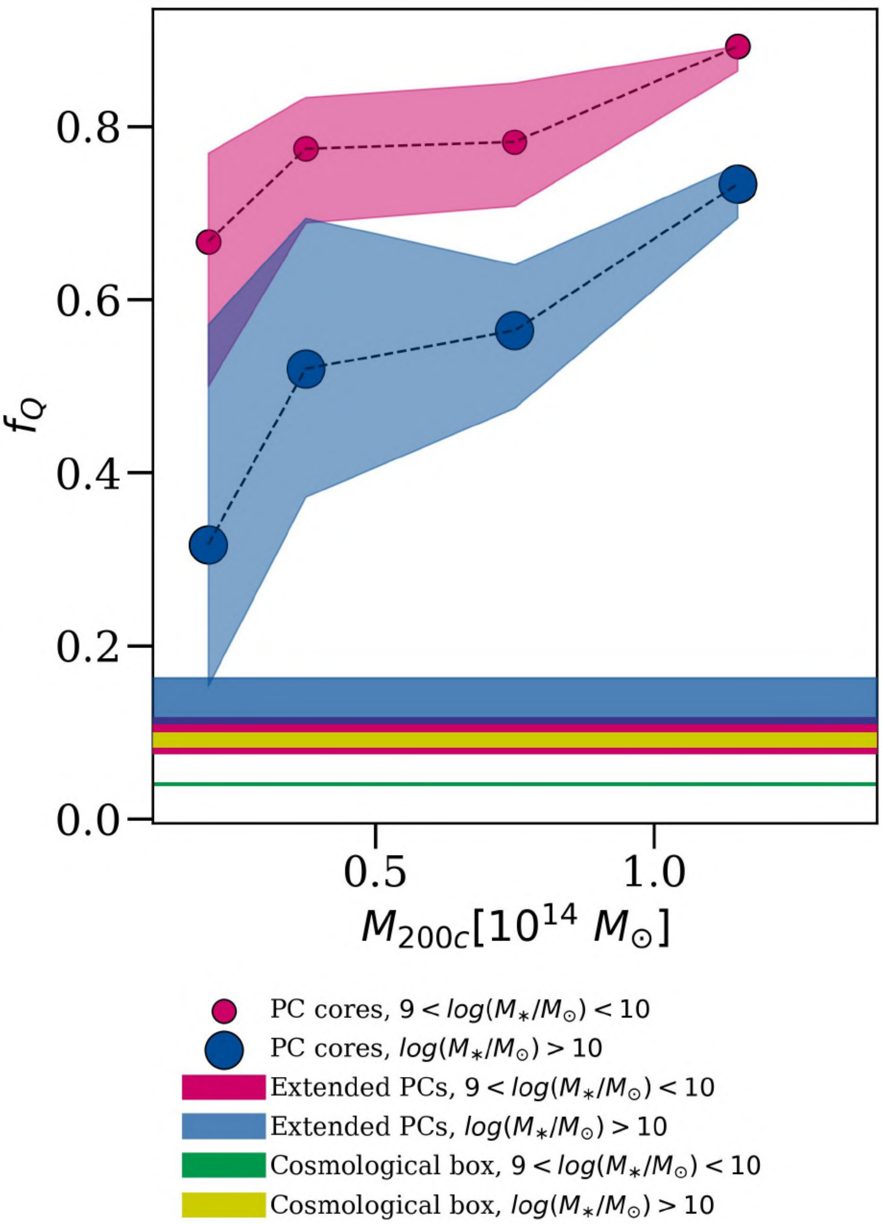


Figure 4.12: Fraction of quenched galaxies in PC cores as a function of the core mass M_{200} in the `DIANOGA` PCs at $z = 2.2$. The dots are median quenched fractions across different cores in each mass bin and the shaded areas around them are the 16th-84th percentile ranges. The horizontal bands show the fraction of all quenched galaxies in the coeval cosmological box (its width is the 1σ confidence interval calculated following Cameron 2011), and the 16th-84th percentile interval of quenched fractions in the PC extended regions (masking out the PC cores), across the 14 different simulations, independently of their host halo mass. Results are shown for galaxies with stellar masses in the range $10^9 < M_*/M_\odot < 10^{10}$ and $M_* > 10^{10}\ M_\odot$ separately.

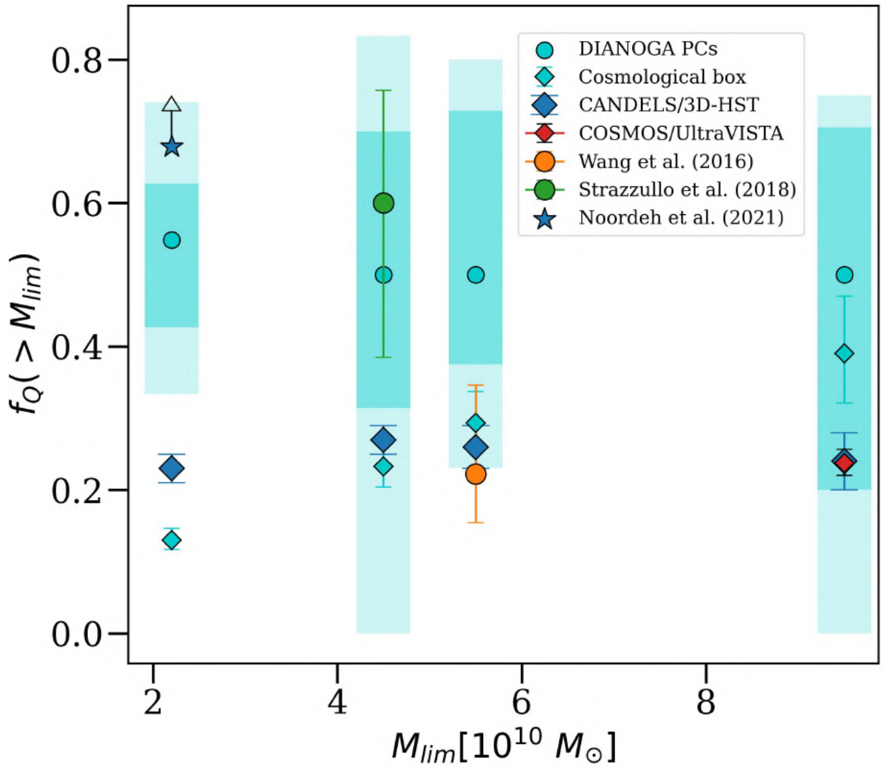


Figure 4.13: Comparison of quenched fractions in the most massive simulated PC cores (with $M_{200} \geqslant 5 \times 10^{13}\ M_{\odot}$) and observations at $z \sim 2-2.5$, including all galaxies with stellar masses larger than the completeness limit (M_{lim}) of each observational study. I note that measurements in this figure are not homogeneous in terms of the size of the considered region around the PC center (see text). The cyan dots show the median quenched fractions among selected halos, while shaded dark (light) regions are the 16th-84th percentile (minimum-maximum) ranges. The cyan diamonds, with binomial population uncertainties (Cameron 2011), are quenched fractions extracted from the cosmological box. For comparison with the observations in CLJ1001, XLSSC 122 and ClJ1449, galaxies in PCs were selected above a limiting stellar mass and within a physical aperture around halo centers, matching those adopted in Wang et al. (2016), Noordeh et al. (2021) and Strazzullo et al. (2018), respectively (see text for details). For comparison with the cosmological box, I also show the quenched fractions in the field from COSMOS/UltraVISTA catalog (Muzzin et al. 2013) and from the CANDELS/3D-HST survey (Skelton et al. 2014) at $2 < z < 2.5$.

not homogeneous in terms of the size of the considered region around the PC centers. I also compare the quenched fractions in the cosmological box with field observations in COSMOS/UltraVISTA (Muzzin et al. 2013) and in CANDELS/3D-HST (Skelton et al. 2014) at $z \sim 2-2.5$. For completeness, I also show the quenched fractions in the PC cores in the highest stellar mass bin, even though there is no direct comparison with mass-complete galaxy samples in observed PCs. I calculate these quenched fractions, as a reference, within the average R_{200} (440 kpc) of the considered halos. The quenched fractions at $M_* > 10^{11}$ $M_\odot$ are consistent with the value of $60 \pm 20\%$ reported for the Spiderweb protocluster by Naufal et al. (2024) in the same stellar mass range. However, the underlying galaxy sample in that study is not mass-complete, which limits the robustness of a direct quantitative comparison.

The massive halos in the simulated PCs exhibit a broad range of quenched galaxy fractions, reflecting the diversity in the timing of evolutionary processes affecting galaxies in overdense environments at cosmic noon. In particular, the most star forming among massive PC cores show fractions of quenched galaxies comparable to those in the cosmological box as do observed PCs with negligible environmental quenching (e.g., CLJ1001, Wang et al. 2016), while the most quiescent PC cores resemble the most mature observed systems (e.g., XLSSC 122, Noordeh et al. 2021)[4]. Quenched fractions in the cosmological box in the stellar mass range probed in Fig. 4.13 are close to observations in the field at similar redshift, especially at $M_* \sim 4-5 \times 10^{10}$ $M_\odot$, whereas at higher (lower) masses they are slightly overestimated (underestimated).

Based on results from `DIANOGA` simulations, the large scatter observed in the quenched fraction of PC cores is primarily driven by intrinsic variability in the evolutionary phase of PC cores of similar mass at $z \sim 2$. I also note that the quenched fractions of galaxies in the innermost regions of massive PC cores, as shown in comparison with CLJ1449, have the largest scatter. While the intrinsic variability might indeed be enhanced in the selected innermost region of massive halos, the increased scatter might also be due to lower statistics in such small volumes. A more detailed investigation of the processes affecting galaxy evolution in halos of different masses and at different halo-centric radii will be the focus of future work.

4.5.1 *UVJ diagrams*

In observational studies, UVJ diagrams are widely used to classify galaxies as star forming or quenched (e.g., Labbé et al. 2005; Williams et al. 2009; Whitaker et al. 2012). Donnari et al. (2019) found that the fractions of quenched galaxies inferred from UVJ diagrams in the IllustrisTNG simulations were consistent with those estimated from the galaxy distance from the MS at redshifts $0 \leqslant z \leqslant 2$. This was achieved by adopting a separation line between star-forming and quenched galaxies in the UVJ diagram that

4 The estimate of the mass completeness limit is based on the spectroscopic completeness limit from Willis et al. (2020), assuming a BC03 SSP with formation redshift of $z = 5$ and no dust.

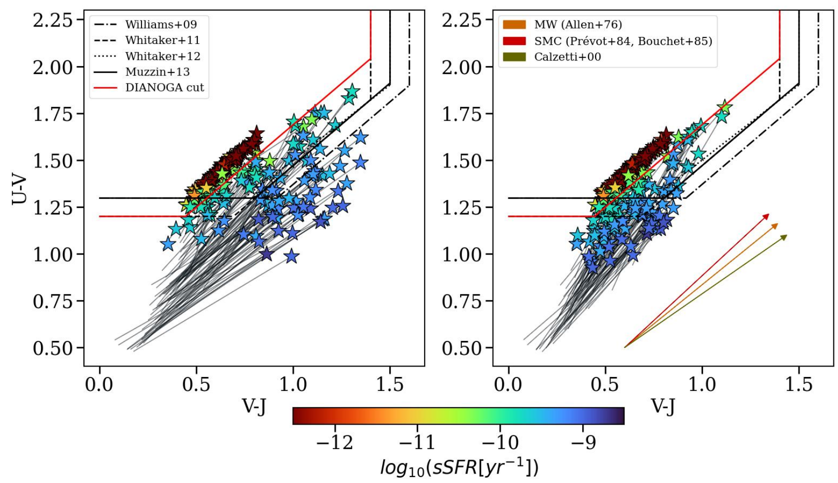


Figure 4.14: Simulated UVJ diagrams for galaxies within seven massive halos in DIANOGA PC regions at $z = 2.2$, within a 700 kpc aperture, produced with SKIRT-9. Intrinsic colors are connected to reddened colors by gray arrows that end with star symbols color coded according to the sSFRs of galaxies. The left panel shows the diagram obtained by tracing dust in the diffuse ISM through a fixed dust-to-metal ratio. The right panel shows the diagram produced with the model for MCs described in Sect. 3.2.2.1. Typical cuts adopted in UVJ diagrams to separate quenched and star-forming galaxies (Williams et al. 2009; Whitaker et al. 2011, 2012; Muzzin et al. 2013) are shown as black lines, while the red line defines a custom-defined boundary to separate these populations in the DIANOGA simulations. The colored arrows show the reddening vectors for typical attenuation laws (Allen 1976; Prevot et al. 1984; Bouchet et al. 1985; Calzetti et al. 2000).

preserves the slope of the commonly used classifications (e.g., Whitaker et al. 2011), but is shifted to account for the slightly different color distribution compared to observations. I test this approach in the DIANOGA simulations, coupled with the SKIRT-9 radiative transfer code (see Sect. 3.2.2.1), to investigate potential complications in the comparison between quenched galaxy populations in simulations and observations, particularly in the context of high-redshift massive halos.

I produced UVJ diagrams for a random sample of seven massive halos, with $M_{200} = (0.6 - 1.4) \times 10^{14}\ M_\odot$, in the simulations. I select galaxies with $M_* > 2 \times 10^{10}\ M_\odot$, which is approximately the mass range that can be reasonably probed with mass-complete samples in deep observational investigations of $z \sim 2$ PC galaxy populations thus far, and located within a 3D aperture radius of 700 kpc from the center of each halo, corresponding to $\sim 1.5 \times R_{200}$. The resulting UVJ diagrams are shown in Fig. 4.14. The two panels refer to results based on two models, corresponding to including or not a prescription for molecular clouds (MCs, see Sect. 3.2.2.1). In the following, I refer to these two models as "diffuse model" (left panel) and "MC model" (right panel), respectively. I note that the diffuse model relies

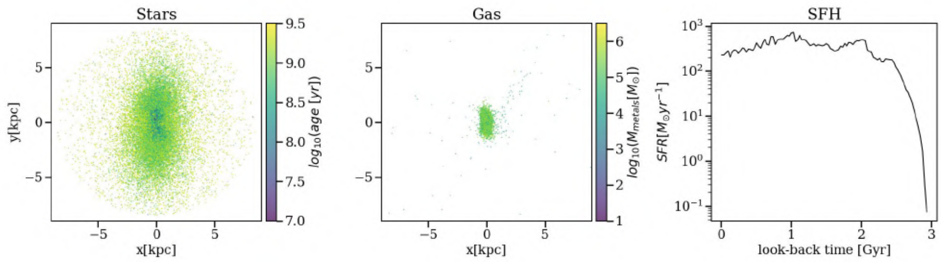


Figure 4.15: Properties of a galaxy characterized by a steep reddening vector. The left panel shows the distribution of stars, color coded according to their age, while the middle panel shows the distribution of gas particles color coded according to their mass in metals, both within 0.25 kpc from the center of the galaxy along the line of sight. The right panel shows the galaxy star formation history.

on a smaller number of assumptions and represents a more direct prediction of the simulations, though it does not resolve the scales of MCs.

The diagrams for both models reveal that galaxy populations in the simulated halos exhibit a red sequence (e.g., Bell et al. 2004) that is already in place at $z = 2.2$, reflecting the high quenched fractions presented in Sect. 4.5. These red galaxies form a tight sequence in the quenched region of the diagram and are largely unaffected by dust attenuation, due to their limited gas reservoir, which translates into little dust to absorb the stellar emission in this modeling. Qualitatively, the distribution of both quenched and star-forming galaxies is in agreement with observed UVJ diagrams. However, the actual color distribution of the simulated star-forming galaxies is not the same as for observed counterparts. Therefore, to quantitatively match the separation between the two galaxy populations based on their sSFR (see Fig. 4.7), I need to shift the separation line by approximately 0.2 mag toward redder U-V colors compared to standard separations used in the literature (e.g., Williams et al. 2009; Whitaker et al. 2011, 2012; Muzzin et al. 2013). I note that this is in quantitative agreement with the UVJ separation found by Donnari et al. (2019) in the IllustrisTNG simulations, as well as with the color distributions in SIMBA simulations at $z = 2$ from Akins et al. (2022). For star-forming galaxies, the diffuse model produces overall stronger attenuation (the average A_V for the studied sample is 1, compared to 0.5 in the MC model), which is dominated by the attenuation of the diffuse dust on the more abundant older stellar population. In this respect, I recall that the MC model subtracts the cold gas reservoir from the total gas contributing to the attenuation of older stars in the ISM. Since star-forming galaxies in PCs reach very high fractions of cold to total gas, this leaves a rather small amount of diffuse gas, and consequently dust, in the MC model, resulting in lower dust attenuation compared to the diffuse model. The details of this effect are related to the assumptions adopted in the MC model implemented here, however, as discussed in Sect. 3.2.2.1, tuning this model is not the focus of this analysis; my primary interest in modeling the age-dependent attenuation of MCs is understanding its qualitative influence on the reddening of galaxy colors in the UVJ diagrams.

Figure 4.14 shows that, in my modeling of dust attenuation applied to DIANOGA simulations (diffuse model), the majority of the galaxy population has reddening vectors in agreement with typical attenuation laws (e.g., Allen 1976; Prevot et al. 1984; Bouchet et al. 1985; Calzetti et al. 2000). However, a subset of star-forming galaxies exhibit steeper reddening vectors. This is consistent with the results from Akins et al. (2022) on SIMBA simulations coupled with dust radiative transfer. In fact, they found that the assumption of a universal Calzetti et al. (2000) attenuation law might impact the efficacy of the UVJ classification, especially for star-forming galaxies near the quenched region of the UVJ diagram. This might lead to the incorrect classification of these galaxies as quenched when applying standard UVJ cuts (Williams et al. 2009; Whitaker et al. 2011, 2012; Muzzin et al. 2013). This behavior appears in both models. Thus, although the MC model seems to extend this feature to a larger fraction of the galaxy population compared to the diffuse model, this is not a numerical artifact introduced by the MC modeling. I note that galaxies that are characterized by such "atypical" steep reddening vectors tend to be central galaxies or located close to the halo center and have high stellar masses ($M_* \gtrsim 10^{11}\ M_\odot$). The comparison between the two models suggests that the limited resolution of cosmological simulations, which cannot capture the small-scale high-density structure of MCs, does not explain this inconsistency in the simulated UVJ diagram. Indeed, if anything, the MC model produces a more pronounced tilt of the reddening vector in the UVJ plane.

I investigated the "atypical" galaxies further by analyzing the spatial distributions of their stars and metals (whose mass is mapped into dust mass in my modeling) and their star formation histories. As an example, I show in Fig. 4.15 these quantities for one of these galaxies. The star formation history (right panel) shows approximately constant SFR for the past ~ 2 Gyr at $\sim 400\ M_\odot yr^{-1}$. The distribution of stars (left panel) shows a positive gradient of stellar ages, while the distribution of metals (central panel) is concentrated in a dense central core. The steepness of the reddening vector is thus produced by the presence of highly attenuated young stars in the metal-rich central region of the galaxy. This is a genuine signal of age-dependent extinction produced by the simulations (though at larger scales than the unresolved MCs), which causes a stronger reddening of the U-V color than of the V-J color (see Fig. 4.14).

4.6 IMPACT OF APERTURES ON STELLAR MASSES OF SIMULATED GALAXIES

Here I examine the impact of using stellar masses calculated within fixed apertures (2D, with an 8.5 kpc radius, corresponding to 1 arcsecond at $z = 2.2$) as opposed to using total SubFind masses. Figure 4.16 shows the ratio of these two mass definitions as a function of aperture stellar

mass. This ratio is ~ 1 on average for stellar masses up to $M_* \sim 10^{11}\ M_\odot$, with negligible scatter (~ 0.005), for galaxies both in the PCs and in the cosmological box. For $M_* > 10^{11}\ M_\odot$, however, the deviation of the mass ratio from unity becomes significant, with the `SubFind` masses exceeding the aperture masses by a factor ~ 1.3 (1.2) on average in the PCs (cosmological box, respectively).
Most galaxies (72%) with $M_* > 10^{11}\ M_\odot$ are hosted by the central subhalo in massive halos, where `SubFind` does not separate the actual central galaxy from the diffuse stellar component (see, e.g., Marini et al. 2021). For $M_* > 10^{11}\ M_\odot$ satellites (which are thus not affected by this feature of `SubFind`) the `SubFind` mass estimate still exceeds the aperture mass by a factor 1.1 on average, reaching up to a factor $\gtrsim$2 for 1% of these most massive galaxies.

4.7 EFFECTS OF AVERAGING SFRS

Here I assess differences between galaxy SFRs averaged over 200 Myr (SFR_{200}), which more properly compare to some observational estimates used in this work (see Sect. 4.4), and the instantaneous SFRs associated with gas particles in the SH03 star formation model (SFR_{gas}). Figure 4.17 shows the ratio between these two SFR estimates as a function of stellar mass. SFR_{gas} is lower than SFR_{200} across all the probed stellar mass range, with a weak increasing trend with increasing mass. Proto-BCGs behave similarly to other galaxies, with instantaneous SFRs tendentially lower than those averaged over the past 200 Myr. The same trends are observed in PCs and in the cosmological box, reflecting a declining star formation history of galaxies in both environments at $z = 2.2$, possibly more pronounced at lower masses.
Figure 4.18 shows the distributions of specific SFR_{gas} (SFR_{200}) for galaxies in which the SFR_{200} (SFR_{gas}, respectively) is zero, which are not accounted for in Fig. 4.17. There are very few galaxies with $SFR_{200} = 0$ and $SFR_{gas} > 0$, and their SFR_{gas} corresponds anyway to a negligible level of residual star formation. Galaxies with $SFR_{gas} = 0$ but $SFR_{200} > 0$ are more common; these galaxies have relatively low (averaged) sSFRs, with the bulk of them falling below the level of the adopted threshold for quenched galaxies. Galaxies for which both SFR estimates are zero also do not appear in Fig. 4.17. However, they would in principle add to the population of galaxies for which the estimates are the same and, since they are a minority across all stellar masses, they would not significantly change the trend shown in Fig. 4.17.

4.8 SUMMARY AND CONCLUSIONS

In this Chapter, I have analyzed the properties of galaxy populations in PCs at $z = 2.2$ in the `DIANOGA` set of zoom-in cosmological hydrodynamical simulations of galaxy clusters and in the cosmological box, introduced in Sect. 3.2.1.

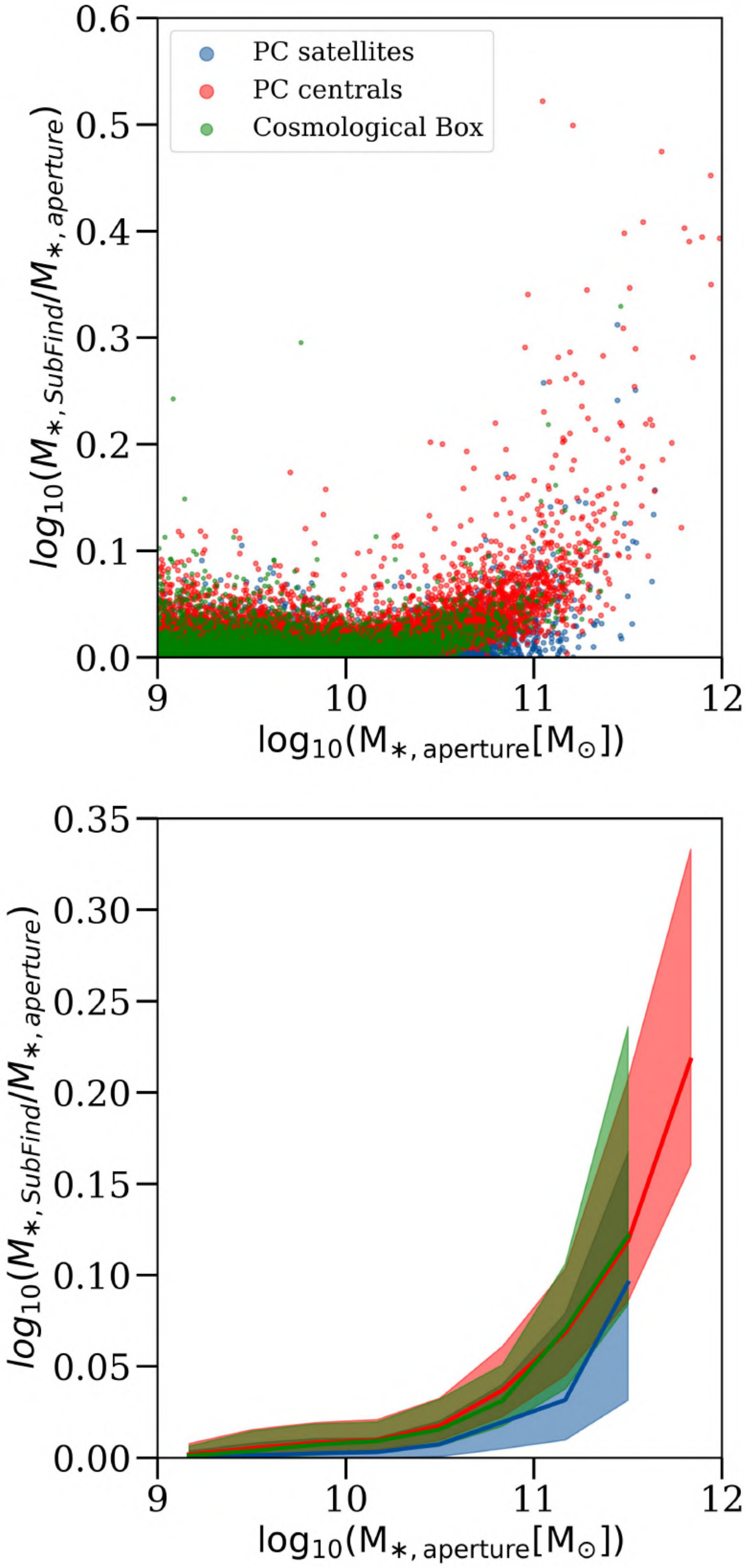


Figure 4.16: Ratio of `SubFind` and aperture stellar masses, as a function of aperture mass for galaxies in `DIANOGA` PCs, separated in centrals and satellites, and in the cosmological box at $z = 2.2$. Aperture masses are estimated within a two-dimensional projected radius equal to 8.5 kpc. The bottom panel shows the median (16th-84th percentiles) of the mass ratios of galaxies plotted in the top panel, in solid lines (shaded areas, respectively).

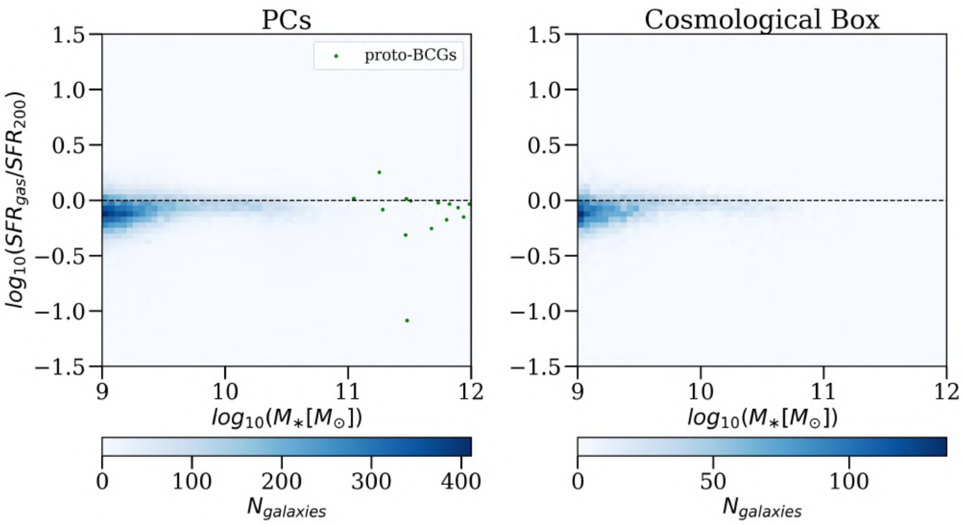


Figure 4.17: Two-dimensional histograms of the ratios between instantaneous SFR (SFR_{gas}) and the SFR averaged over 200 Myr (SFR_{200}) of galaxies in DIANOGA PCs (left panel) and in the cosmological box (right panel) at $z = 2.2$, as a function of stellar mass. Proto-BCGs are plotted separately.

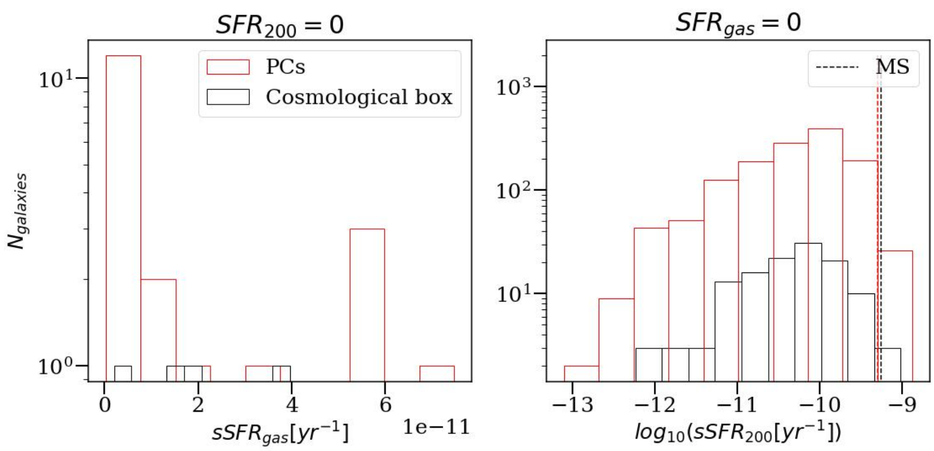


Figure 4.18: Distribution of the instantaneous (averaged) sSFRs for galaxies with $\mathrm{SFR}_{200} = 0$ ($\mathrm{SFR}_{gas} = 0$), in the left (right, respectively) panel. The dashed lines show the average sSFR of main-sequence galaxies in the cosmological box (black) and in the PCs (red).

To define PCs in simulations, I traced back the DM that falls within the R_{200} of the $z = 0$ cluster and identified, around the most massive progenitor halo, a region that encompasses 80% of all these DM particles.
I investigated the properties of simulated galaxy populations and I compared them to observations to test the predictions of simulations, focusing on halos with properties similar to the observed (proto)cluster cores. I compared galaxy properties in the simulated PCs and in the cosmological box to highlight the impact of environment on galaxy evolution, as predicted by the analyzed simulations. I also performed radiative transfer simulations using the `SKIRT-9` code (Camps & Baes 2020) to estimate dust-attenuated rest-frame UV to NIR colors of simulated galaxies, by consistently tracing dust through metals in the gas component. From these, I tested the impact of adopting a UVJ photometric classification of star-forming and quenched galaxies, as routinely applied in observational studies. I summarize the main conclusions as follows.

- The GSMF in the `DIANOGA` PCs is broadly consistent with observations in PCs at $2 \leqslant z \leqslant 2.5$, both in shape and normalization (see Fig. 4.6). This demonstrates that the integrated star formation history in the simulated PCs realistically reproduces the observational data. The median PC GSMF reveals an excess of galaxies with $M_* > 10^{10}$ $M_\odot$ compared to the cosmological box (see Fig. 4.5). This excess is detected both in PC cores ($M_{200} > 1.5 \times 10^{13}$ $M_\odot$) and, albeit with lower significance, in the extended PC regions (i.e., outside the cores). Since this excess is indicative of early stellar mass assembly, its presence in the extended PC regions suggests the onset of an overall accelerated galaxy evolution in the broader environments hosting PCs.

- The SFRs of star-forming galaxies in both the PCs and the cosmological box are lower than expected from observations (see Fig. 4.8). This is consistent with previous results from lower-resolution versions of the `DIANOGA` simulations (Granato et al. 2015; Bassini et al. 2020) and other suites of cosmological hydrodynamical simulations (e.g., Akins et al. 2022; Edward et al. 2024). The SFRs in PC galaxies show a small (~ 0.05 dex) yet significant suppression compared to galaxies in the cosmological box (see Fig. 4.9). This signal is dominated by galaxies in the extended structures of the PCs, while galaxies in the PC cores exhibit a more pronounced suppression (up to ~ 0.25 dex compared to SFRs in the cosmological box).

- The reservoirs of cold star-forming gas are systematically lower than observational estimates at $z \sim 1.5-3$, while depletion times are broadly consistent with those observed in star-forming galaxies, both in the PCs and in the cosmological box (see Fig. 4.10). This suggests that the underestimation of SFRs in the simulations is related to galaxies not having sufficient gas available to convert into stars, possibly due to the effect of a too efficient feedback inhibiting cooling or expelling gas.
  Protocluster galaxies in these simulations have marginally lower cold gas content than their counterparts in the cosmological box on average,

while depletion times do not show significant differences (see Fig. 4.11). However, galaxies in PC cores exhibit more marked differences in both cold gas fractions and depletion times compared to galaxies in the extended PC structure and in the cosmological box. In massive cores, both cold gas fractions and depletion times show a significant drop for galaxies with $\log_{10}(M_*/M_\odot) < 9.5$, and they remain significantly lower than the cosmological box levels at higher masses. The sharp drop is due to a large fraction ($\sim$ 40%) of galaxies with $\log_{10}(M_*/M_\odot) < 9.5$ that are classified as star forming based on their SFR averaged over the past 200 Myr but have actually completely lost their cold gas over this time.
I note that these galaxies may be in a regime influenced by the finite numerical resolution. However, since galaxies of similar mass in other simulated regions do not show such a sudden loss of gas, this feature could reflect a genuine physical process emerging in simulated massive halos at high redshift, possibly amplified by the combined effect of relatively limited numerical resolution and the extreme environment of the most massive high-redshift halos.

- The fraction of quenched galaxies in the simulations increases significantly with the mass of the host halo in the PC cores, reaching approximately 0.9 (0.7) at $M_{200} \geqslant 10^{14}\ M_\odot$ for galaxies with $10^9 < M_*/M_\odot < 10^{10}$ ($M_* > 10^{10}\ M_\odot$, respectively; see Fig. 4.12). DIANOGA simulations predict an increasing fraction of quenched galaxies toward lower stellar masses in PC cores, which is at odds with indications from observations. The fraction of quenched galaxies in PCs, even beyond massive cores, is higher than in the cosmological box, implying accelerated quenching also in lower-density PC regions, prior to galaxy infall into massive halos.

- The DIANOGA PCs include halos in which quenching has progressed in very different ways (see Figs. 4.12,4.13), similar to the observed PCs. In some halos, the galaxy quenched fraction is as low as in the cosmological box (comparable to observed starbursting PC cores, e.g., CLJ1001 at $z = 2.51$, Wang et al. 2016), while in others it is as high as in observed mature clusters (e.g., XLSSC 122 at $z = 1.98$, Willis et al. 2020; Noordeh et al. 2021). The quenched fractions in the cosmological box also qualitatively agree with observations in the field.

- The distribution of galaxies in the simulated UVJ diagram in a sample of 7 massive simulated halos is qualitatively consistent with observed distributions of star-forming and quenched galaxies (see Fig. 4.14). However, from a more quantitative point of view, I needed to shift the UVJ selection boundary by 0.2 mag toward redder U-V colors, compared to typical classification criteria used in observations, to optimally match the sSFR-based classification of simulated galaxies. Despite the overall qualitative agreement, I note indeed the presence of a population of star-forming galaxies (mainly massive galaxies in the central regions of PC cores) with a steep reddening vector, leading to significant contamination of the quenched region of the diagram when adopting standard selections.

In this Chapter, I have shown that the `DIANOGA` cosmological hydrodynamical simulations can broadly reproduce the properties of galaxy populations in observed PCs at $z \sim 2-2.5$, which is a particularly significant time for the evolution of clusters and their host galaxies since it corresponds to the transition between PC structures and the first established clusters. This redshift range also corresponds to a cosmic epoch where the cosmic SFR density is at its peak and, at the same time, the first signatures of environmental quenching appear in the densest environments. I have shown that the environment of cluster progenitors has an impact on galaxy populations, particularly through the early onset and subsequent suppression of star formation, which is most significant in the massive PC cores. These results provide an interpretative framework, that consistently accounts for the full cosmological context, for understanding the evolution of galaxies in PCs.
The results of this analysis highlight that the current generation of cosmological hydrodynamical simulations are invaluable instruments to shed light on the complex interplay between environment and the physical processes that shape galaxy evolution. In this respect, cluster progenitor environments at cosmic noon present a significant challenge to models of star formation and feedback processes, as they couple extreme regions of the Universe and a cosmic time where transformational processes are at their peak. In this context, and while keeping in mind the current limitations that still affect the comparison of observations and simulations at high redshift, the results presented in this Chapter highlight the need for improved modeling of galaxy formation and evolution to better capture observed galaxy population properties (the level of star formation in particular) in the densest high-redshift regions. This is a crucial part of the broader effort to consistently reproduce the evolution of galaxy populations across cosmic time and across different environments.

5 DIFFUSE GAS ACROSS THE PROTOCLUSTER TO CLUSTER TRANSITION

In this Chapter, I present the evolution of gas in massive halos in the environment of cluster progenitors at cosmic noon. The goal of this analysis is to address the open questions on the epoch and timing of ICM assembly, and to assess the capability of simulations to reproduce the available observations of the (proto-)ICM. To this end, I analyze the budget of different gas phases in halos of different masses, across the transition from the protocluster to the cluster regime (Sect. 5.1) and then focus on the diffuse hot component, progenitor of the ICM that permeates mature clusters at lower redshifts (Sect. 5.1.1). I analyze the build-up of this component by tracing it across the protocluster to cluster transition to pinpoint its origin. Then, I present the comparison between the predictions of simulations and observations of the thermodynamic properties of high-redshift clusters and protoclusters, up to $z = 3$ (Sect. 5.2). The results presented here will be published in Esposito et al. (2025, in preparation).

5.1 GAS COMPOSITION IN HIGH-Z HALOS

In order to investigate the properties of gas permeating massive halos at high redshift, I analyze all halos within the high-resolution regions of the `DIANOGA` simulations with masses above 10^{13} $M_{\odot}$, regardless of their final fate at $z = 0$. The gas within their R_{500} is classified into four phases, based on its position in the temperature-density phase diagram, as indicated by the separation lines in Fig. 5.1. At low densities, at progressively decreasing temperatures I identify the ICM, the warm-hot intergalactic medium (WHIM) and the IGM. A high-density component is referred to as the "condensed" phase. Technically, the latter could be separated into a cooling phase and a star-forming phase, at higher temperatures, made up of multiphase particles as described in the SH03 model. Nevertheless, for the purpose of this analysis the separation between the two is not relevant, since I am only interested in identifying the cold gas that is mainly confined to galactic halos. It is worth noting that for halos in the mass range $10^{13} \leqslant M_{200}/M_{\odot} < 10^{14}$, the hot, low-density gas phase should be formally classified as the intra-group medium (IGrM). Nonetheless, for simplicity, I refer to this phase as "ICM" throughout, regardless of halo mass. The classification of ICM, WHIM, and IGM phases follows the typical temperature ranges of their observational counterparts. However, the specific boundaries adopted between phases (as shown in Fig. 5.1) are somewhat arbitrary.

Figure 5.1 shows the phase diagrams of gas within and around a halo of

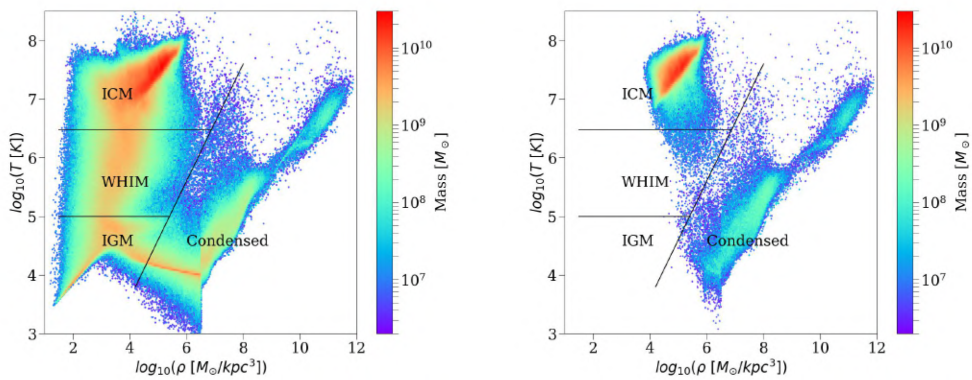


Figure 5.1: Phase diagram of the gas in the region within $5{\times}R_{500}$ (left panel) and R_{500} (right panel) of a cluster of mass $M_{200} = 1.2 \times\ 10^{14}\ M_{\odot}$ at $z = 2.2$. The solid black lines define the separation between different phases of the gas. The discontinuity at $\rho \sim 10^{6.5}\ M_{\odot}kpc^{-3}$ corresponds to the transition to the star-forming phase in the SH03 model for star formation.

mass $M_{200} = 1.2 \times 10^{14}\ M_{\odot}$ at $z = 2.2$, which already qualifies as a cluster. The left panel includes gas within $5{\times}R_{500}$, while the right panel focuses on gas within R_{500}. The comparison highlights that the ICM is predominantly confined within R_{500}, where it constitutes the dominant gas component. Furthermore, only the hottest and densest portion of the ICM – reaching temperatures up to $T \sim 10^8$ K – is found within R_{500}. Besides minor WHIM and IGM components, the remaining gas inside R_{500} is in the condensed phase, fueling intense star formation activity (the integrated SFR within R_{500} is $\sim 600\ M_{\odot}\ yr^{-1}$).

On the other hand, most of the WHIM and of the IGM components are distributed on large scales. Figure 5.2 shows the spatial distribution of the gas components, as well as of DM and stars, in a region of 4 Mpc per side, centered on the same halo shown in Fig. 5.1. The condensed phase has a clumpy distribution, that closely traces the galactic (sub)halos and overlaps with the densest gas filaments of the IGM. The latter does not reach inside the R_{500} of the central halo, where mainly ICM can be found. Similarly, the WHIM is largely absent within the massive central halo, but it is distributed in the smaller halos, that can be identified in the DM map, and along the most prominent filaments of the IGM. Overall, the gas budget within the central massive halo is already dominated by the ICM, much like a mature $z = 0$ cluster, as already shown in Fig. 5.1.

For a quantitative analysis of the gas budget across the full range of halo mass probed here, I compute the fraction of gas mass in the different phases within the R_{500} of halos at $z = 3$, 2.2 and 1.5. This redshift interval captures the transition from the regime of sparse protoclusters to the emergence of the first massive virialized clusters ($M_{200} \sim 10^{14}\ M_{\odot}$).
These fractions are shown, as a function of halo mass, in Fig. 5.3, at $z = 3$

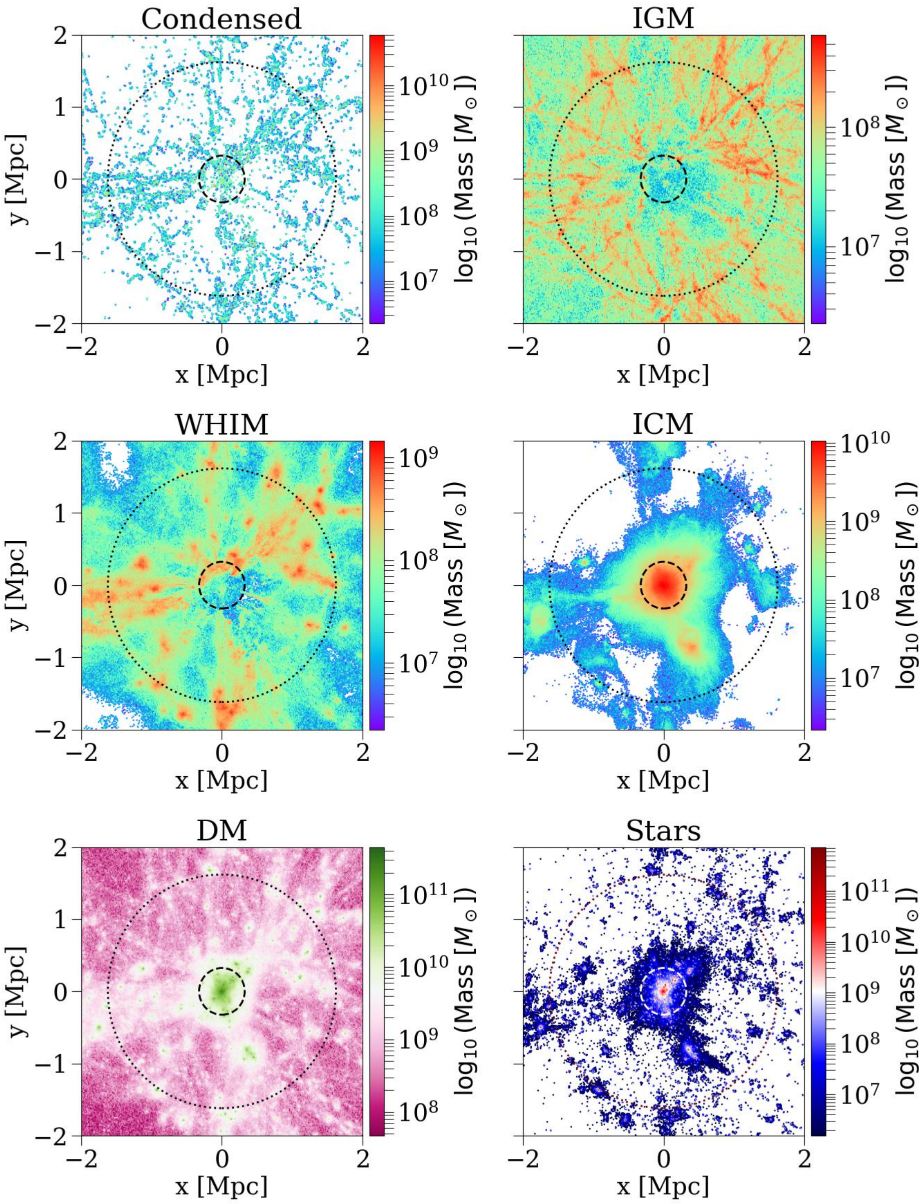


Figure 5.2: Projected mass map of the gas components, as well as of DM and stars, in the region spanning 4 Mpc per side around a simulated cluster of mass $M_{200} = 1.2 \times 10^{14}\ M_\odot$ at $z = 2.2$. The different gas components (condensed phase, IGM, WHIM, ICM) are shown in the upper- and middle-row panels. The bottom-row panels show the distribution of DM mass (left panel) and of stellar mass (right panel). In all panels, the dashed (dotted) circle defines R_{500} ($5{\times}R_{500}$) of the central halo.

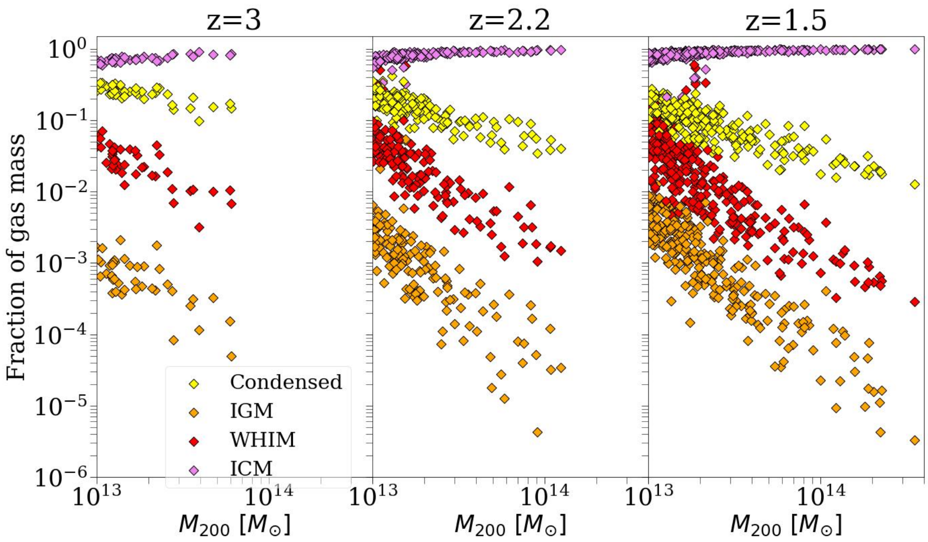


Figure 5.3: Fraction of gas mass in four different phases (condensed, IGM, WHIM and ICM), within R_{500} of each halo, as a function of the halo mass. From left to right the halos are selected at $z = 3$, 2.2 and 1.5.

(left panel), 2.2 (middle panel) and 1.5 (right panel). Given the remaining degree of arbitrariness of the exact boundaries in the phase diagram, the specific values of these gas mass fractions are to some extent arbitrary as well. The interesting information comes rather from their overall relative values and their trend with halo mass. In this respect, Fig. 5.3 shows that the ICM consistently dominates the gas budget across the probed halo mass and redshift range ($\log_{10}(M_{200}/M_\odot) \geqslant 13$, $1.5 \leqslant z \leqslant 3$), showing that hot ambient gas is already present in all massive halos as early as $z = 3$. At all probed redshifts, the fraction of ICM increases with halo mass across the range from $M_{200} \sim 10^{13}$ $M_\odot$ to $\sim 10^{14}$ $M_\odot$, although the relative change remains within a factor of $\lesssim 2$.

In contrast, the fractions of all other gas phases (IGM, WHIM, condensed phase) decrease with increasing halo mass. These gas phases follow a consistent hierarchy in terms of abundance: condensed gas is the most prevalent, followed by WHIM and finally the IGM, which contributes the smallest fraction to the gas budget within the massive high-redshift halos probed here. These trends with halo mass suggest that in more massive halos, whose evolution is expected to be in a more advanced stage, the cold gas that fuels star formation has been depleted more and the accreted IGM has been converted more efficiently into the hot phases, possibly passing through the WHIM and then to the ICM. I note that Fig. 5.3 shows the gas fractions on a logarithmic scale, thus the visual comparison of the relative contributions of the different gas components may be somewhat misleading: the IGM component is in practice negligible within R_{500}.

To better illustrate the redshift evolution of these trends, Fig. 5.4 shows the median fraction for each gas phase as a function of halo mass, comparing the three probed redshifts. For the IGM and WHIM components there

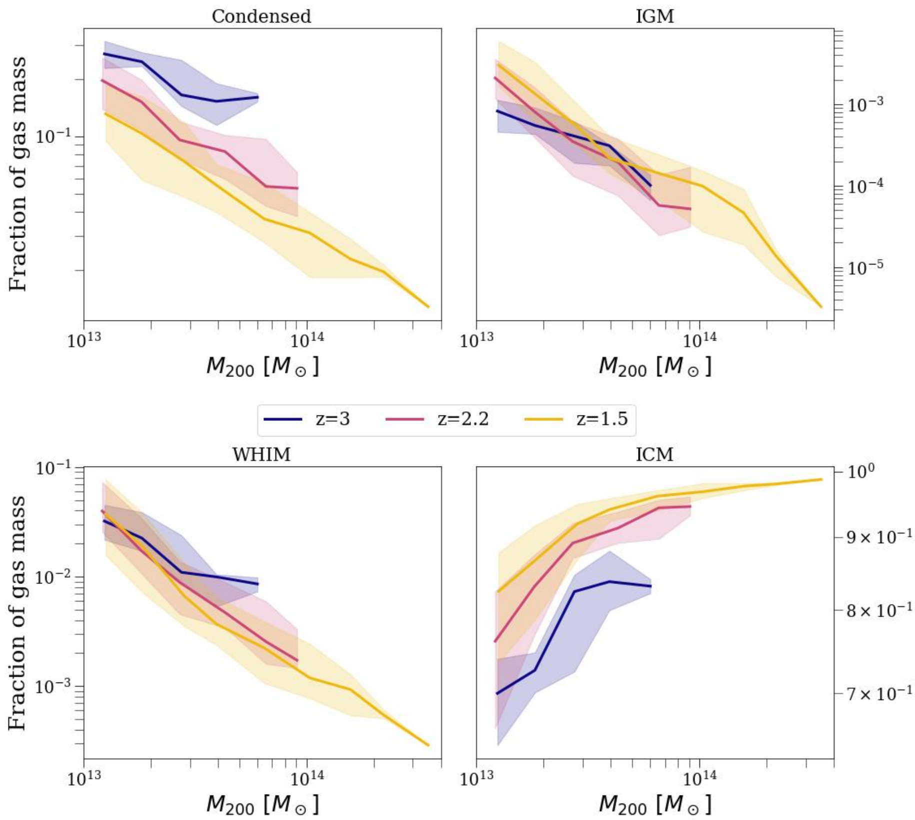


Figure 5.4: Median mass fractions of different gas components (condensed, IGM, WHIM and ICM) lying within R_{500} of massive halos, as a function of halo mass. For each component, the different lines show the median fraction at different redshifts (z = 1.5, 2.2 and 3), with shaded areas showing the corresponding 16th-84th percentile range.

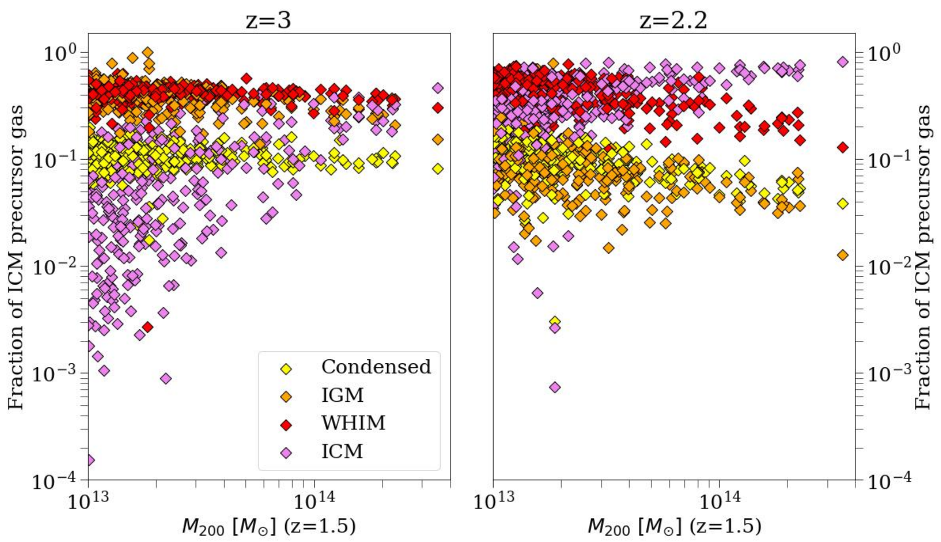


Figure 5.5: Gas components across the ICM build-up: the left (right) panel shows, as a function of the halo mass at $z = 1.5$, what mass fraction of gas belonging to the ICM at $z = 1.5$ is in different phases – condensed, IGM, WHIM and ICM – at $z = 3$ (2.2, respectively).

is rather marginal evolution of the gas fraction - halo mass relation in the probed redshift range. There is instead significant redshift evolution between $z = 3$ and 1.5 in the fraction of both the condensed phase and ICM components at fixed halo mass. For all the halo masses analyzed here, the fraction of gas in the condensed phase decreases towards lower redshifts, while the relevance of the ICM component increases. This is in line with expectations for massive halos across the transition from highly star forming protoclusters to massive clusters characterized by a strong ICM component and low levels of star formation.

5.1.1 *The build-up of ICM*

The study of the build-up of ICM in protoclusters is a relatively young field of investigation (see Sect. 2.2.3). In order to exploit the `DIANOGA` simulations to gain insight on this important process, I analyze the ICM of simulated massive halos at $z = 1.5$.

For this purpose, I mark the gas particles that populate the ICM component in halos identified at $z = 1.5$, and trace them back to the previous simulation outputs, at $z = 2.2$ and 3. This approach allows me to investigate the recently formed hot gas component by linking it to its properties at higher redshifts, where its thermalization took place. Focusing on recently formed ICM offers a clear view of the processes that contribute to shaping its thermal structure, without the cumulative effects and mixing that may dilute their signatures by $z = 0$.

Figure 5.5 shows the fraction of this traced gas – hereafter referred to as ICM precursor – that resides in different gas phases (condensed, IGM, WHIM

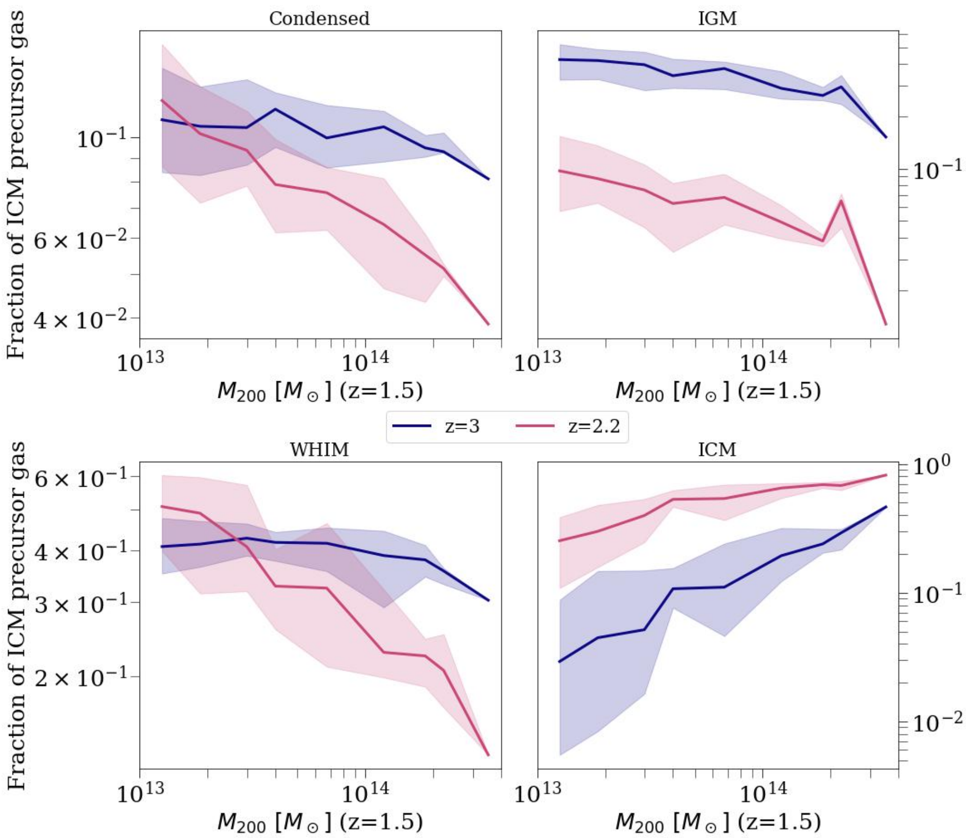


Figure 5.6: Median values (solid lines; and 16th-84th percentile ranges, shaded areas) of the fractions (shown in Fig. 5.5) of the different components of the $z = 1.5$ ICM precursor, as a function of the $z = 1.5$ halo mass. The figure shows the fractions of condensed (top-left panel), IGM (top-right panel), WHIM (bottom-left panel) and ICM (bottom-right panel) phases, at $z = 3$ and 2.2.

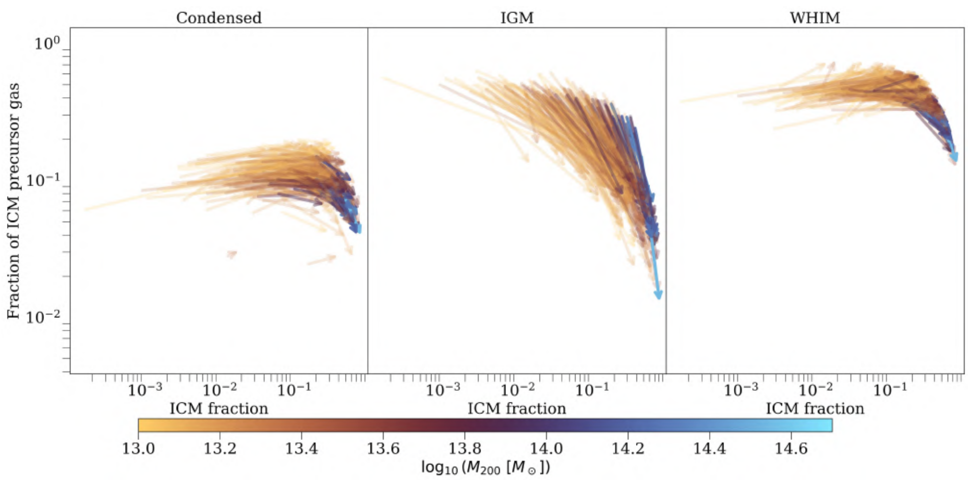


Figure 5.7: Variation of the mass fraction of the condensed, IGM and WHIM components (from left to right, as indicated) of the $z = 1.5$ ICM precursor gas, shown against the variation of fraction of gas that is already in the ICM phase, between $z = 3$ and 2.2. The arrows, color coded according to the halo mass at $z = 1.5$, start at fractions at $z = 3$ and point to fractions at $z = 2.2$.

and ICM) at $z = 2.2$ and 3, as a function of the halo mass at $z = 1.5$. At $z = 3$, most of the ICM precursor is in the IGM and WHIM phases, and about 10% of it is in the condensed phase, largely irrespective of the $z = 1.5$ halo mass. Indeed, at $z = 3$ only the fraction of ICM precursor that is already in the ICM phase shows a dependence on the $z = 1.5$ halo mass. The most massive ($M_{200} \gtrsim 10^{14}$ $M_\odot$) $z = 1.5$ halos have an appreciable fraction ($\gtrsim$ 20%) of their ICM already in place by $z = 3$.
On the other hand, at $z = 2.2$ the ICM component is almost fully assembled at high halo masses, with a fraction $\gtrsim$ 50 % of the precursor gas already in ICM phase for halos with $M_{200} \gtrsim 4 \times 10^{13}$ $M_\odot$. At lower halo masses, WHIM remains the main progenitor phase of the $z = 1.5$ ICM. The IGM and condensed phases contribute comparable amounts of mass, both being subdominant at the $\lesssim$ 10% level. In this context, recent observations have revealed large, extended molecular gas reservoirs around star-forming galaxies in protoclusters. Although the constraints on their total mass budget remain uncertain, these reservoirs could potentially contribute to the build-up of the ICM at later times (e.g., Chen et al. 2024).

Figure 5.6 shows the median values (with 16th-84th percentile ranges) of the fractions presented in Fig. 5.5, as a function of the $z = 1.5$ halo mass, now emphasizing the redshift evolution of the different gas phases.
As mentioned above, the fraction of ICM precursor in the condensed phase (top-left panel) at $z = 3$ is largely independent of the $z = 1.5$ halo mass, while at $z = 2.2$ it decreases significantly with increasing mass. This reflects the earlier evolution of higher-mass halos, where cold gas is more efficiently shock-heated as it falls into the gravitational potential well of the halo. As a result, the contribution of condensed gas to the total ICM precursor drops to just a few percent for $M_{200} \gtrsim 10^{14}$ $M_\odot$.
The fraction of ICM precursor in IGM phase (top-right panel) shows a weak decreasing trend with halo mass at $z = 3$. At $z = 2.2$, it remains below 10% throughout the entire mass range, dropping to $\lesssim$ 5% at $M_{200} \gtrsim 10^{14}$ $M_\odot$, where most ($\gtrsim$ 70%) of the ICM is already in place.
The WHIM contribution (bottom-left panel) shows a qualitative trend similar to that of the condensed phase. Interestingly, at the lowest halo masses ($\sim 10^{13}$ $M_\odot$), the WHIM fraction actually increases with time, as infalling gas is shock-heated and transitions through this phase before thermalizing as ICM.
Finally, the fraction of $z = 1.5$ ICM gas that is already in place in the ICM phase (bottom-right panel) at $z = 2.2$ or 3 effectively traces the formation rate of this component, which is faster for progenitors of more massive halos — consistent with their overall accelerated evolution.

To better visualize the evolution of the different gas phases within the ICM precursor between $z = 3$ and 2.2, Fig. 5.7 shows the variation in the mass fraction contributed by each phase versus the variation in the fraction of traced gas already in the ICM phase.
For each halo, an arrow connects the values of these fractions at $z = 3$ and at $z = 2.2$, and is color coded according to the halo mass at $z = 1.5$.

This visualization highlights that the ICM assembly at $3 \lesssim z \lesssim 1.5$ follows distinct evolutionary paths in halos of different mass.
Low-mass halos ($M_{200} \sim 1 - 3 \times 10^{13}$ $M_\odot$) at $z = 1.5$ experience significant accretion and shock-heating of the IGM to the ICM temperatures between $z = 3$ and 2.2. At the same time, the condensed phase in the ICM precursor increases – likely due to compression of the infalling gas – along with the WHIM component.
Instead, the ICM of massive clusters ($M_{200} \gtrsim 10^{14}$ $M_\odot$) at $z = 1.5$ is caught in a different evolutionary phase between $z = 3$ and 2.2. The contributions of condensed and WHIM phases to the ICM precursor are generally decreasing, with a corresponding increase of the ICM component. At the same time, active accretion from the IGM is still ongoing, with the contribution of the IGM phase dropping from $20 - 40\%$ of the overall precursor budget at $z = 3$ to a few percent at $z = 2.2$.

In conclusion, $20 - 50\%$ of the ICM at $z = 1.5$ comes from gas that was still in the filamentary IGM at $z = 3$. By $z = 2.2$ further accretion only contributes to less than 10% of the final mass, a fraction that drops to a few percent at the highest halo masses. The contribution from initially cold gas that is shock-heated to ICM temperatures is found to be between 4 and 10% between $z = 3$ and 2.2. The WHIM component dominates the ICM precursor, along with the IGM phase, at $z = 3$. At $z = 2.2$, WHIM is still the main ICM precursor at the lowest halos masses probed, while it becomes subdominant in massive ($z = 1.5$) clusters, where most of the ICM is already in place. The WHIM component appears to be a transition phase for gas that is thermalizing as it undergoes gravitational collapse. This is supported by the increase of its contribution at $M_{200} \sim 10^{13}$ $M_\odot$ between $z = 3$ and 2.2, as well as by its deficiency in halos that have already built most of their ICM.

5.2 comparison with observations

In this Section, I compare the simulated properties of the hot diffuse gas with available observations, to assess how well they reproduce the observed thermodynamic state of the (proto-)ICM. This comparison is performed with the caveat that observational constraints at high redshift are severely affected by small-number statistics (see Sect. 2.2.3) and by technical limitations.
First, I compare the radial profiles of thermodynamic properties in massive halos at $z = 1.5$ and 2.2 with profiles inferred from X-ray and SZ observations of high-redshift clusters and protoclusters (Sect. 5.2.1). Then, I model the X-ray emission of the gas component in simulated halos at $z = 3$ to produce X-ray surface brightness maps to compare with the first detection of diffuse emission from hot gas at $z \gtrsim 3$ (Sect. 5.2.2). In order to select only a reliable counterpart to the observed (proto-)ICM component, in the simulated ICM (defined as in Sect. 5.1) I subtract the fraction of gas mass that is treated as reservoir for star formation in the SH03 model, which is anyway negligible in the ICM region of the phase diagram ($10^{-5} - 10^{-2}$ in the probed halos).

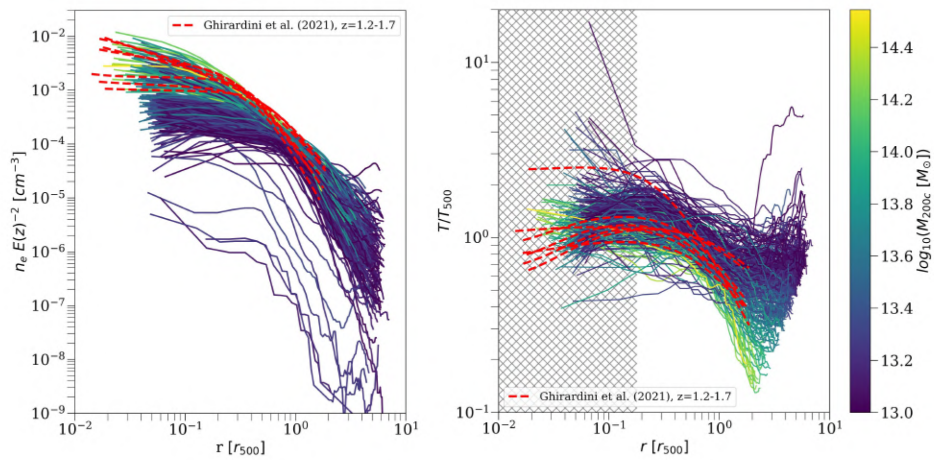


Figure 5.8: Free electron number density (left panel) and temperature (right panel) profiles of simulated halos with $M_{200} \geqslant 10^{13}$ $M_\odot$, color coded by their mass, in comparison with observations from Ghirardini et al. (2021) of seven clusters at $z = 1.2 - 1.7$. All the temperature profiles are normalized by the self-similar T_{500}. The hatched area marks the region where the observed profiles are not resolved.

5.2.1 *Thermodynamic profiles*

As in Sect. 5.1, I consider all halos with $M_{200} \geqslant 10^{13}$ $M_\odot$, and compute their 3D radial profiles of electron number density, temperature, pressure and entropy, as in Sect. 3.2.1.2. All simulated profiles shown below are smoothed using a median filter[1] to reduce stochastic fluctuations.

5.2.1.1 *Profiles at* $z = 1.5$

The observational sample used as a reference for comparison with the simulations at $z = 1.5$ consists of seven clusters at $1.2 < z \leqslant 1.7$, detected by SPT via their SZ signal, with signal-to-noise ratio $S/N > 6$ and mass $M_{500} > 3 \times 10^{14}$ $M_\odot$ (Bleem et al. 2015). The radial profiles of their thermodynamic properties were obtained from X-ray observations with the XMM-Newton and Chandra satellites, with a total exposure time of ~800 ks with Chandra and ~1 Ms with XMM-Newton, and are presented by Ghirardini et al. (2021). The median mass of these clusters is $M_{200} \sim 4 \times 10^{14}$ $M_\odot$, making them slightly more massive than the most massive clusters in the DIANOGA simulations at $z = 1.5$.
Figure 5.8 (left panel) shows the comparison between simulated and ob-

1 The median filter is applied with a window size of 3: for each radial bin, the value shown is the median of the values in the bin and in the two adiacent bins.

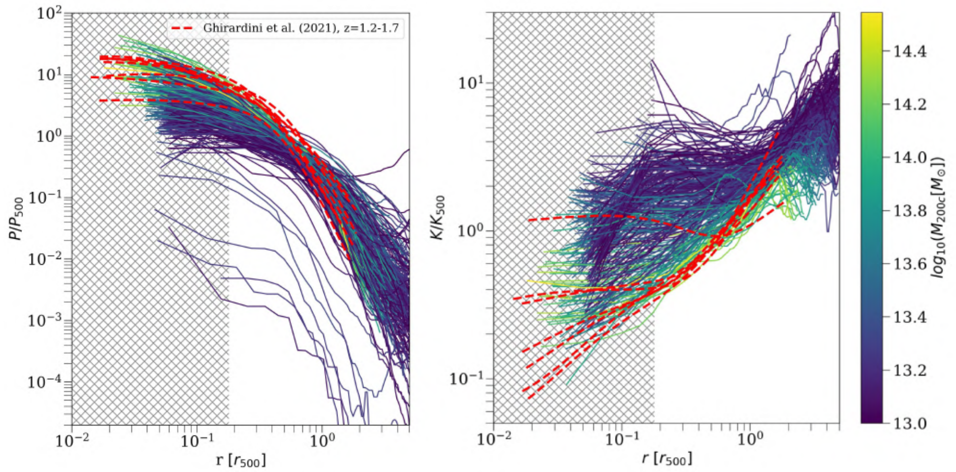


Figure 5.9: Pressure (left panel) and entropy (right panel) profiles of simulated halos with $M_{200} \geqslant 10^{13}$ $M_\odot$, color coded with their mass, in comparison with observations from Ghirardini et al. (2021) of seven clusters at $z = 1.2 - 1.7$. All the pressure and entropy profiles are normalized by the self-similar P_{500} and K_{500}, respectively. The hatched areas mark the regions where the observed profiles are not resolved.

served radial profiles of electron number density. The fit to the observed profiles was obtained using the model proposed by Vikhlinin et al. (2006):

$$n_e(r)^2 = n_0^2 \frac{(r/r_c)^{-\alpha}}{[1+(r/r_c)^2]^{3\beta/2-\alpha/2}} \frac{1}{[1+(r/r_s)^3]^{\epsilon/3}}, \tag{31}$$

with n_0, r_c, α, β, r_s and ϵ as free parameters. The best-fit values for these parameters are taken from Tab. 2 of Ghirardini et al. (2021). The simulated profiles for clusters with $M_{200} > 10^{14}$ $M_\odot$ are consistent with the observed ones across all radii. This agreement also extends to the innermost core regions of the simulated halos, where the `DIANOGA` simulations instead struggle to match observations at $z = 0$ (see Sect. 3.2.1.2).
The observed temperature profiles shown in Fig. 5.8 (right panel) correspond to the best-fit 3D profiles following the Vikhlinin et al. (2006) model:

$$T(r) = T_0 \frac{\frac{T_{min}}{T_0} + \left(\frac{r}{r_{cool}}\right)^{a_{cool}}}{1 + \left(\frac{r}{r_{cool}}\right)^{a_{cool}}} \frac{1}{[1+(r/r_t)^2]^{c/2}}, \tag{32}$$

with T_0, T_{min}, r_{cool}, a_{cool}, r_t, and c as free parameters. Their best-fit values are listed in Tab. 3 of Ghirardini et al. (2021). The observed temperature profiles are only resolved down to $\sim 0.2{\times}R_{500}$. The shape of the simulated profiles, showing spectroscopic-like temperatures (see Sect. 3.2.1.2), reproduces the observed ones reasonably well across the resolved radial range. At the smallest resolved radii, the observed profiles tend to flatten into isothermal cores. This is in contrast with the simulated profiles, which maintain a negative gradient down to small radii. Although this difference arises in

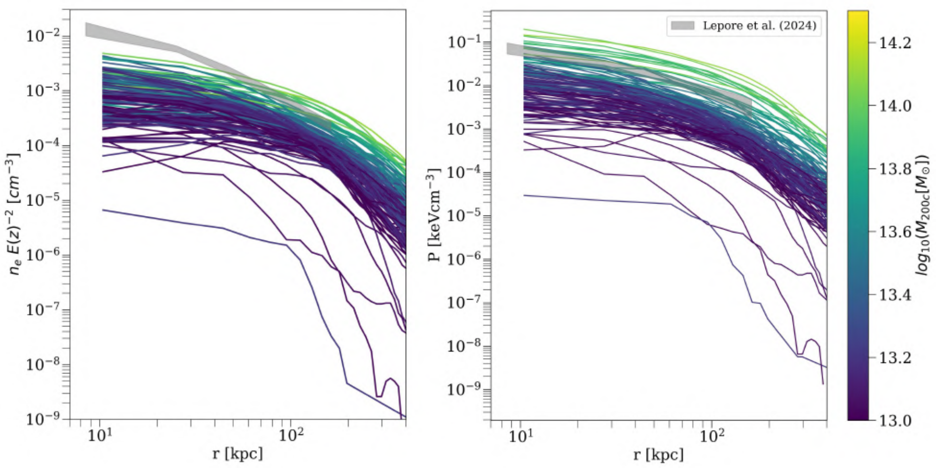


Figure 5.10: Free electron density (left panel) and pressure (right panel) profiles of simulated halos with $M_{200} \geqslant 10^{13}$ $M_\odot$ at $z = 2.2$, color coded with their mass, and compared with observations from Lepore et al. (2024) of the Spiderweb protocluster at $z = 2.2$ (shaded grey region). The range shown for the observed density profile represents its variation when assuming different values of the n_d parameter that accounts for the unknown surface brightness in the innermost region contaminated by the radio galaxy. The range shown for the pressure profile of the Spiderweb protocluster represents the statistical uncertainty around the best-fit pressure profile.

a radial range that is only marginally resolved in the observations, the discrepancy is qualitatively similar to what is seen in the cores of simulated clusters at $z = 0$ (see Sect. 3.2.1.2).

Figure 5.9 shows the comparison between the pressure (left) and entropy (right) profiles in simulations and observations. The observed profiles are obtained by combining the density and temperature profiles, and are therefore also resolved only beyond ~ 0.2 R_{500}. In the resolved radial range, the simulated profiles qualitatively reproduce the observed trends, with the observed pressure profiles being slightly shallower and the entropy profiles (at least the bulk of them) somewhat steeper than the simulated profiles.

5.2.1.2 *Profiles at $z = 2.2$*

At $z \gtrsim 2$ the detection of (proto-)ICM is more challenging, both in the SZ signal and in the X-ray emission. The Spiderweb protocluster is the first (and only to date) system at $z > 2$ for which the nascent ICM has been detected through the SZ signal measured with ALMA observations (Di Mascolo et al. 2023). In addition, Lepore et al. (2024) have derived the thermodynamic profiles of the proto-ICM of Spiderweb, by combining the SZ measurements with X-ray data from Chandra, that I compare to simulated profiles in this Section. In the following Figures, I do not rescale the radius to R_{500} nor the pressure, temperature and entropy values to the self-similar normalizations,

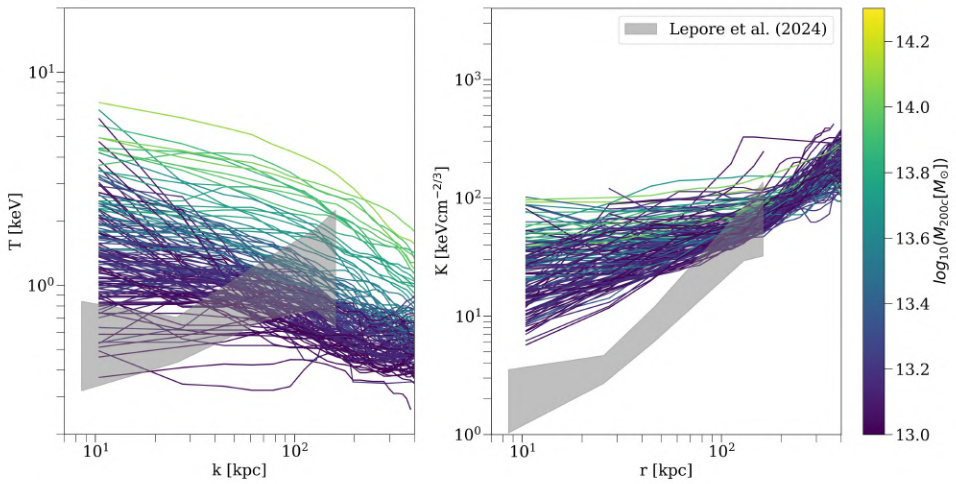


Figure 5.11: Temperature (left panel) and entropy (right panel) profiles of simulated halos with $M_{200} \geqslant 10^{13}$ $M_{\odot}$ at $z = 2.2$, color coded with their mass, compared with observations from Lepore et al. (2024) of the Spiderweb protocluster. The range shown for the observations represents variations associated with the assumptions of different values of the n_d parameter that accounts for the unknown surface brightness in the innermost region contaminated by the radio galaxy.

given the large uncertainties in the mass estimates of the central halo in Spiderweb (see Extended Data Table 2 in Di Mascolo et al. 2023).

Figure 5.10 (left panel) shows the comparison between the simulated profiles of electron density and the one derived by Lepore et al. (2024). The observed electron density profile is obtained by fitting a β-model (Cavaliere & Fusco-Femiano 1976) to the X-ray surface brightness profile.

The comparison highlights that the observed density profile is steeper than that of any `DIANOGA` halo, reaching central values that are $2-4$ times higher than the highest densities in simulations at r∼10 kpc.

The pressure profile of Spiderweb is derived by forward modeling the ALMA SZ data. This is shown in Fig. 5.10 (right panel) and compared to the simulated profiles of `DIANOGA` halos at $z = 2.2$. The observed pressure profile is consistent with simulations for halo masses $M_{200} \sim 5-7 \times 10^{13}$ $M_{\odot}$.

Temperature (mass-weighted) and entropy profiles of the $z = 2.2$ simulated halos are compared with the observational profiles of the Spiderweb protocluster in Fig. 5.11. The latter are obtained by combining the X-ray density and the SZ pressure profiles. As a consequence of the discrepancy in the central densities, the temperature profiles in simulations reach higher central values than in the Spiderweb system. Moreover, the observed profile shows a decreasing temperature toward the center, typical of a cool-core system (see Sect. 2.1.3.1), in contrast with increasing – or constant – temperatures toward the center of simulated halos. As a consequence, also the entropy profiles of simulated objects reach higher (by more than one order of magnitude) values than in observations. Since there are no other observations to compare with in terms of thermodynamic profiles at $z \gtrsim 2$, it is hard to draw general conclusions. The Spiderweb protocluster might be a rare system that does not find a counterpart in the limited sample of the DIANOGA simulations. Therefore, more observations and a larger set of simulated protoclusters are necessary to understand possible marked discrepancies in the properties of simulated and observed diffuse hot gas at cosmic noon.

5.2.2 *X-ray surface brightness at z=3*

The first detection of diffuse X-ray emission at $z \gtrsim 3$, in the MQN01 structure ($z = 3.25$), has been recently presented in Travascio et al. (2025b, including M. Esposito). In this work, we use X-ray luminosity maps in the 0.5-2 keV rest-frame energy range, produced with SMAC (see Sect. 3.2.2.2), of the diffuse gas component of halos extracted from the DIANOGA simulations at $z = 3$. We use these to extract surface brightness profiles to compare with the observational data. Here the temperature threshold to select the hot diffuse gas component is decreased to 0.1 keV, to account for the early phases of thermalization of the gas at such high redshift. To match the estimated mass of the halo in MQN01 ($\sim 3 \times 10^{13}$ $M_{\odot}$), a sample including the 12 most massive simulated halos at $z = 3$ is selected, spanning the mass range $M_{200} \sim 2 - 6 \times 10^{13}$ $M_{\odot}$, with median mass 3×10^{13} $M_{\odot}$.
Each X-ray map encompasses a cubic region of 500 kpc around each halo center, thus going far beyond the region where the observed X-ray emission is traced, which extends to ~ 30 kpc. Zoomed-in maps of the modeled X-ray emission of the 12 most massive halos in the DIANOGA simulations at $z = 3$ are shown in Fig. 5.12. The maps show a large diversity in the state of the gas, both in the intensity of the X-ray emission – which approximately traces the density-squared distribution – and in its highly disturbed morphology. From these maps, surface brightness profiles were extracted and compared with observations in MQN01 in the 0.5-2 keV rest-frame band, obtained with Chandra. The comparison is shown in Fig. 5.13. The median prediction of the surface brightness from DIANOGA halos lies a factor $30 - 60$ below the observations, at $15\,\mathrm{kpc} < r < 30\,\mathrm{kpc}$ (where detections and not upper limits or extrapolations are available), with a much shallower profile compared to MQN01. On the other hand, the highest surface brightness profile in simulations is consistent with the observed surface brightness at radii larger than 20 kpc.
Figure 5.14 shows the comparison of the simulated profiles of free electron densities, including all 41 DIANOGA halos with $M_{200} > 10^{13}$ $M_{\odot}$ at $z = 3$, and the β-model fit to the data of MQN01. The comparison shows a discrepancy in normalization and slope, with the DIANOGA halos having lower and shallower density profiles, which explains the mismatch in X-ray surface brightness profiles. Similarly to what I discussed concerning the comparison of simulated halos with the Spiderweb protocluster, MQN01 is also likely to be a one-of-a-kind system, representing the highest overdensity of galaxies at $z \sim 3$ and hosting one of the largest Lyman-α nebula. This is

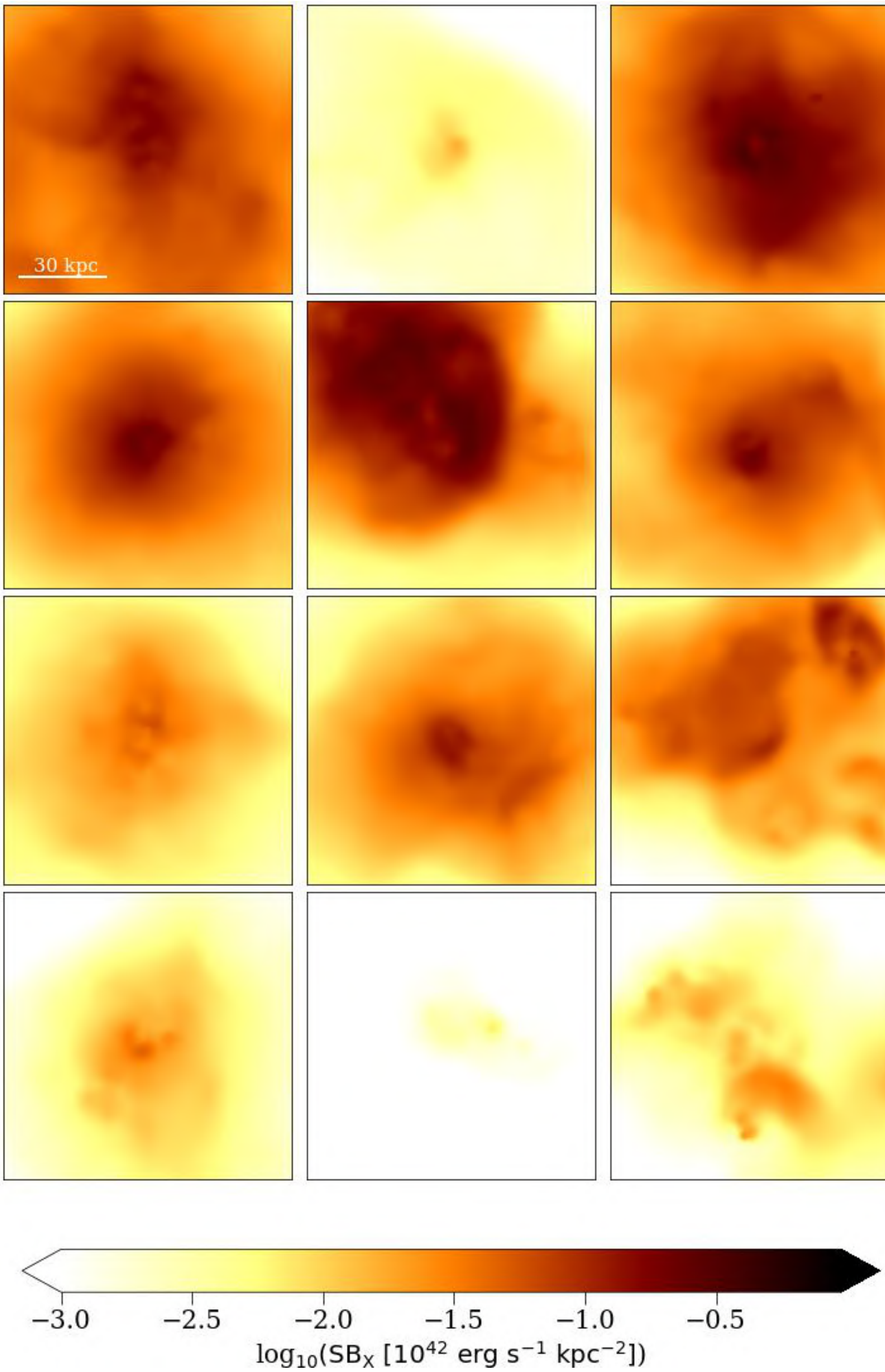


Figure 5.12: X-ray surface brightness maps of the 12 most massive halos in the DIANOGA simulations at $z = 3$, modeled with SMAC. From the top-left to the bottom-right panel, the halo mass is decreasing, from $M_{200} = 6 \times 10^{13}$ to 2×10^{13} $M_{\odot}$. All panels are on the same spatial scale as indicated in the top-left one.

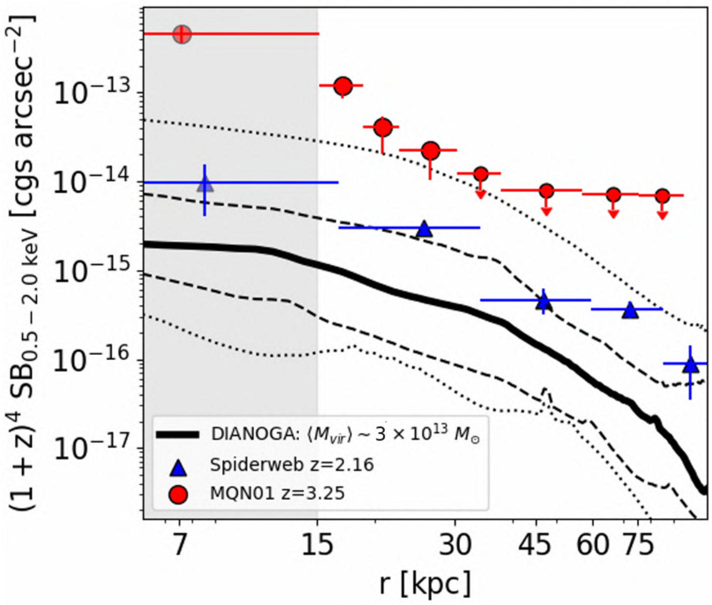


Figure 5.13: X-ray surface brightness profiles in the 0.5–2 keV band for halos at $z \sim 3$ (and Spiderweb at $z = 2.2$). The solid black line shows the median profile of the 12 selected DIANOGA halos at $z = 3$ (see text), derived from SMAC X-ray luminosity maps. Dashed and dotted lines indicate the 16th–84th and minimum–maximum ranges, respectively. Red dots mark Chandra X-ray data for MQN01 at $z = 3.25$. The data point below 15 kpc (gray shaded area) is extrapolated from the best-fit model of the data, while beyond 30 kpc only upper limits are available. For comparison, the X-ray surface brightness profile of the Spiderweb system from Lepore et al. (2024), corrected for differences in the redshift-dependent intrinsic energy band, is shown in blue. Reproduced from Travascio et al. (2025b).

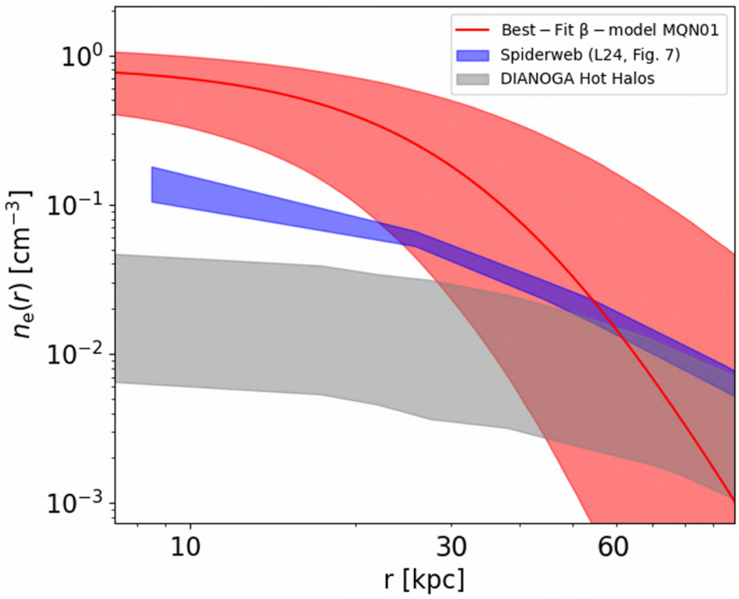


Figure 5.14: Electron density profiles of $M_{200} > 10^{13}$ $M_\odot$ DIANOGA halos at $z = 3$, compared with the best-fit model to observational data in MQN01 at $z = 3.25$. The grey shaded area represents the full range covered by simulated halos. The red solid line is the best fit to the data, with the shaded area defining the 68% confidence interval. For comparison, the density profile of Spiderweb from Lepore et al. (2024) is shown in blue.

thus an extreme structure that does not necessarily represent the general protocluster population at $z \sim 3$. Also in this case, more observations of protoclusters at this redshift and a larger set of simulated structures will be necessary to assess the reliability of the galaxy formation model implemented in these simulations in reproducing the properties of diffuse gas in cluster progenitors at cosmic noon.

5.3 SUMMARY AND CONCLUSIONS

In this Chapter, I analyzed the gas components in massive ($M_{200} > 10^{13}$ $M_\odot$) halos in DIANOGA simulations $z = 3$, 2.2 and 1.5. In this way, I could probe the transition between the regime of sparse protoclusters and the emergence of the first population of massive ($M_{200} \sim 10^{14}$ $M_\odot$) clusters. The goal of this analysis is to pinpoint the origin and timing of the ICM assembly, and to compare with available obervations of the (proto-)ICM to assess the capability of simulations to reproduce the properties of recently established (or forming) ICM. The main results can be summarized as follows.

- I identified different phases in the gas populating the region within R_{500} of the simulated halos according to their location in the phase diagram (see Fig. 5.1). In this way, I defined the simulated counterparts to the observed ICM, WHIM and IGM, as well as a condensed phase made up of gas that is either cold and dense or forming stars. The ICM refers to the hottest low-density phase of the gas within the

halo[2].
The ICM component dominates - in terms of mass fraction over the total gas mass - the gas budget within all the probed halos and redshifts, followed by the condensed phase, the WHIM and then, subdominant to all the others, the IGM component (see Fig. 5.3). The ICM accounts for a larger fraction of the total gas mass in more massive halos, while the mass fraction of the other gas phases decreases with halo mass. This is indicative of an earlier evolution of higher-mass halos: they have converted most of the accreted IGM in ICM, possibly transitioning through the WHIM phase, while depleting more of the multiphase star-forming gas that makes up the ISM. Marginal redshift evolution takes place for the WHIM and IGM components in halos at fixed mass; on the other hand, ICM shows a significant increase from $z = 3$ to $z = 1.5$, while the condensed component decreases over the same redshift range at fixed halo mass (see Fig. 5.4).

- In order to track the ICM assembly, I selected gas particles that are found in this phase within halos identified at $z = 1.5$, and traced them back to higher redshifts to investigate the origin of this component. I identified large, comparable fractions of ICM precursor in the WHIM and IGM phases at $z = 3$ (see Fig. 5.5). Quite interestingly, at this redshift $\sim 10\%$ of the ICM precursor is instead in the condensed phase. This component is converted to the hot phase by $z = 1.5$ possibly through feedback processes. The fraction of ICM that has already thermalized is higher for systems that are more massive by $z = 1.5$. At $z = 2.2$, the most massive halos have already built-up most of their $z = 1.5$ ICM, while the ICM precursor in WHIM phase is still dominating less massive halos. The IGM contribution drops down from $20 - 40\%$ at $z = 3$ to less than 10% at $z = 2.2$ (see Fig. 5.6).
  Low-mass halos ($M_{200} \sim 10^{13}$ $M_\odot$) at $z = 1.5$ have undergone a very active phase of ICM build-up between $z = 3$ and 2.2: most of the gas that ends up in the ICM (technically IGrM) by $z = 1.5$ has been accreted from the large-scale IGM in this epoch. Then, part of this gas has entered the condensed phase - possibly due to compression during accretion - and part has been heated to WHIM or ICM temperatures. As a consequence, the budget of all phases (except the IGM) that the ICM precursor transitions through in this epoch has increased from $z = 3$ to 2.2 (see Fig. 5.7). On the other hand, more than about 50% of the ICM of massive clusters at $z = 1.5$ is already in place by $z = 2.2$. While significant accretion ($20 - 30\%$) from the IGM has occurred between $z = 3$ and 2.2 also in this mass range, the more efficent shock-heating of gas to ICM temperatures as it falls in their deep potential wells overrides any other phase transition in the gas. As a result, all the fractions of IGM, WHIM and condensed phase through which the

2 For simplicity, I use the "ICM" label irrespective of the mass of the host halo, thus also including the IGrM for $M_{200} < 10^{14}$ $M_\odot$

ICM precursor transitions are decreasing with redshift, in favor of the emergence of the ICM phase.

- The comparison with the profiles of electron number density, temperature, pressure and entropy in a sample of seven massive clusters at $1.2 < z \leqslant 1.7$ from Ghirardini et al. (2021) has shown broad agreement with the most massive cluster-sized simulated halos at $z = 1.5$ (see Figs. 5.8, 5.9). At $z = 2.2$, however, the comparison with thermodynamic profiles in the Spiderweb protocluster from Lepore et al. (2024) shows some significant discrepancies. While the pressure profile is in good agreement with predictions of simulated halos with $M_{200} \sim 5 - 7 \times 10^{13}\ M_{\odot}$, the electron density profile is steeper than any simulated profile (see Fig. 5.10). This translates also in substantially different temperature and entropy profiles: while Spiderweb seems to host a very prominent cool-core, all the simulated halos present a negative or flat temperature gradient in their central regions. However, to account for the possibility that the Spiderweb protocluster is a rare system that does not find a counterpart in the limited sample of simulated halos, more observations and larger simulation sets will be necessary to robustly assess the reliability of the simulations in reproducing the thermodynamic state of the proto-ICM.
- At $z = 3$, I have presented the comparison between X-ray surface brightness profiles of simulated halos, modeled with `SMAC` (see Sect. 3.2.2.2), and the profile of the central halo of MQN01 at $z = 3.25$, observed with Chandra (Travascio et al. 2025b). The observed profile is a factor 30-60 above the median profile of simulated halos of comparable mass and overall much steeper (see Fig. 5.13). As for the comparison with the Spiderweb protocluster, larger samples of simulated and observed systems will be necessary for a robust assessment of the predictions of simulations at such high redshifts.

Overall, these results provide a coherent picture of the formation of the ICM at cosmic noon, highlighting the mass- and redshift-dependent paths through which hot gas emerges in massive halos. At the same time, the comparison with observations at $z > 2$ reveals significant discrepancies in the gas profiles, emphasizing the importance of extending both simulation efforts and observational campaigns to better constrain the early phases of assembly of the ICM.

6 CONCLUSIONS AND FUTURE PERSPECTIVES

In this Thesis, I have analyzed the DIANOGA set of cosmological hydrodynamical simulations of galaxy clusters, with the aim of characterizing the environment of protoclusters at cosmic noon. The focus has been placed on the analysis of galaxy populations and on the early phases of intracluster medium (ICM) build-up. The simulations used throughout this Thesis are the latest version of the DIANOGA simulations, targeting 14 $z = 0$ clusters, simulated with OpenGADGET-3, with a galaxy formation model that includes radiative cooling of gas, star formation, accretion onto supermassive black holes (SMBHs), feedback from Supernovae (SN) and Active Galactic Nuclei (AGN), stellar evolution and chemical enrichment (see Chapter 3). To complement this analysis, in particular in the comparison with observations in protocluster environments, I employed post-processing tools, such as SKIRT-9 to model the effect of dust on galaxy SEDs, and SMAC to model the X-ray thermal emission of diffuse hot gas in massive halos.
In this concluding Chapter, I summarize the main results presented throughout this Thesis, draw the main conclusions of this study and outline future lines of investigations. Section 6.1 reviews the key findings on the properties of protocluster galaxies discussed in Chapter 4, while Sect. 6.2 focuses on the evolution of the gas component presented in Chapter 5, with particular attention to the build-up of the (proto-)ICM. General conclusions about this work are drawn in Sect. 6.3. Finally, Sect. 6.4 outlines future observational perspectives, and ongoing and future research directions in simulations that stem from the results of this work.

6.1 PROPERTIES OF PROTOCLUSTER GALAXIES

Chapter 4 presented an analysis of the properties of galaxy populations in protocluster environments in the DIANOGA simulations at $z = 2.2$. This analysis was developed in the context of exploring the impact of the environment on the evolution of galaxies, in the cosmic epoch around $z \sim 2$, where most of the activity – in terms of star formation, black hole accretion and mergers – is ongoing. Moreover, this is the epoch when the first $M_{200} \sim 10^{14}$ $M_\odot$ clusters assemble.
At the same time, this work provided a validation for the galaxy formation model implemented in the DIANOGA simulations, through a comparison with (proto)cluster galaxy observations, which are most abundant at $z \sim 2 - 2.5$. The analysis is based on 14 DIANOGA cluster progenitors and on a cosmological box representative of an average region of the Universe, simulated at the same resolution and with the same galaxy formation model as for the

DIANOGA clusters. I compared predictions and observations of fundamental properties of galaxy populations: the galaxy stellar mass function (GSMF), the star-forming main sequence, the fraction of cold star-forming gas, gas depletion times, and the fraction of quenched galaxies.
The DIANOGA simulations produce a GSMF in broad agreement with observations of protoclusters at $z \sim 2-2.5$ (see Fig. 4.6). The simulated GSMF shows a higher fraction of high-mass galaxies ($M_* > 10^{10}$ $M_\odot$) within massive protocluster cores ($M_{200} > 1.5 \times 10^{13}$ $M_\odot$) compared to the cosmological box (see Fig. 4.5). The same signal, albeit with a lower significance, is also observed in the wide-field protocluster structures, indicating an accelerated evolution of galaxies before their infall into massive halos.
On the other hand, DIANOGA simulations underestimate star formation rates of star-forming galaxies both in protoclusters and in the cosmological box compared to thir observed counterparts (see Fig. 4.8), due to an underestimation of the star-forming gas reservoirs (see Fig. 4.10). The gas depletion timescales – that quantify the efficiency of star formation – are instead in agreement with observational results. Star formation rates in DIANOGA protocluster galaxies are weakly suppressed (by ~ 0.05 dex on average) in comparison with counterparts in the cosmological box (see Fig. 4.9). This suppression increases up to ~ 0.25 dex in massive protocluster cores, reflecting suppressed cold gas reservoirs.
The fraction of quenched galaxies varies significantly across different protocluster halos, which is consistent with observations at $z \sim 2-2.5$ (see Fig. 4.13). The simulations show a strong dependence of the quenched fraction on host halo mass, as well as an excess of quenched galaxies in the wide-field protocluster region outside the massive halos, compared to the cosmological box (see Fig. 4.12).
To enable a more proper comparison with observational studies, which commonly use color–color UVJ diagrams to classify star-forming and quenched galaxies, I computed the rest-frame UV to near-IR colors of galaxies using the SKIRT-9 radiative transfer code. This analysis showed qualitative agreement with observed color distributions of star-forming and quenched galaxies, except for few massive galaxies ($M_* \gtrsim 10^{11}$ $M_\odot$) in the cores of massive halos where a form of age-dependent extinction results in steeper reddening vectors than typically assumed in observational studies (see Fig. 4.14). This discrepancy might highlight a potential limitation of the UVJ diagnostic in the regime of massive star-forming galaxies.

6.2 GAS PROPERTIES ACROSS THE PROTOCLUSTER TO CLUSTER TRANSITION

Chapter 5 presented the characterization of the gas component in massive halos ($M_{200} \geqslant 10^{13}$ $M_\odot$) in the DIANOGA simulations across the redshift range $1.5 \leqslant z \leqslant 3$, bridging the protocluster regime and the epoch of formation of the first $M_{200} \sim 10^{14}$ $M_\odot$ clusters. This analysis aimed to shed light on the early phases of ICM assembly, which X-ray and SZ observations have only recently begun to probe at $z \sim 2$–3. To address the evolution of gas in

high-redshift halos, I first focused on the relative contribution of different gas phases to the overall gas budget, and then on the thermal history of the newly assembled ICM at $z = 1.5$.
Simulated halos are already dominated by hot, diffuse gas – within the density and temperature ranges characteristic of the ICM – at $z = 3$, across the full halo mass range probed here (see Fig. 5.3). The remaining gas is mainly in a condensed phase, consisting of cold and dense, as well as of actively star-forming gas, followed by a warm, low-density component interpreted as the simulated counterpart of the WHIM. Finally, the IGM contributes the smallest fraction of the total gas mass in massive halos, as it is primarily distributed in large-scale filamentary structures.
The gas mass fractions in these components show a clear dependence on halo mass at fixed redshift: more massive halos host a larger fraction of their total gas mass in the ICM, while showing lower fractions of condensed gas, WHIM, and IGM within R_{500}. At fixed halo mass, the fraction of gas in the ICM phase increases from $z = 3$ to $z = 1.5$, while the contribution of the condensed phase decreases. The other components exhibit no significant redshift evolution at fixed halo mass (see Fig. 5.4).
To investigate ICM assembly, I identified the ICM component in halos at $z = 1.5$ and traced its origin back to $z \sim 2$–3 (see Fig. 5.5). This analysis revealed interesting properties of the ICM precursor: at $z = 3$, the gas that becomes ICM by $z = 1.5$ is distributed mainly in the IGM and WHIM phases, in comparable fractions ($30 - 40\%$), while approximately 10% is in the condensed phase, later heated by feedback processes (see Fig. 5.6).
Thermalization of gas to ICM temperatures proceeds more rapidly in more massive halos: the ICM in $z = 1.5$ cluster-sized halos ($M_{200} \gtrsim 10^{14}\ M_{\odot}$) is mostly ($\gtrsim 60\%$) in place by $z \sim 2$. In contrast, in small groups ($M_{200} \sim 1 - 2 \times 10^{13}\ M_{\odot}$ at $z = 1.5$) only $20 - 30\%$ of the ICM – technically intragroup medium (IGrM) – is in place by $z \sim 2$, while the bulk ($\sim 50\%$) of the remaining part of ICM precursor is transitioning through a WHIM phase, before complete thermalization to ICM temperatures by $z = 1.5$.
Finally, I compared the thermodynamic properties of simulated halos with those observationally inferred at $z \sim 1.5$–3. At $z = 1.5$, the most massive simulated halos broadly reproduce the radial profiles of thermodynamic quantities – electron number density, temperature, pressure and entropy – observed in high-redshift clusters (see Figs. 5.8, 5.9). However, at $z \sim 2$–3, simulations fail to reproduce the strong cool core-like structures observed in some high-redshift protoclusters (see Figs. 5.10, 5.11 and 5.14). These systems are characterized by steep central density profiles and, where observational data are available, by positive central temperature gradients.
At $z = 3$, I generated X-ray luminosity maps of the hot gas in simulated halos to compare with the first detection of diffuse thermal emission from a protocluster at $z \gtrsim 3$ (Travascio et al. 2025b, see Fig. 5.12). This comparison showed that X-ray surface brightness profiles from simulations are too shallow – and generally too faint – compared to observational indications, thus reflecting the insufficient central gas densities in the simulated (proto-)ICM (see Fig. 5.13).

Nevertheless, observational studies of the (proto-)ICM are still in their infancy and are strongly affected by small-number statistics and technical limitations.

6.3 SUMMARY AND CONCLUSIONS

The work carried out in this Thesis has shed light on the properties of galaxies and of the emerging ICM in the progenitor environments of $z = 0$ clusters, at cosmic noon. This has been achieved by analyzing cosmological hydrodynamical simulations that include gas dynamics and the most relevant processes in galaxy formation, in a cosmological context. The DIANOGA simulations predict that galaxy populations hosted by massive cores in protocluster environments exhibit signatures of environmental effects, reflecting an early onset and the subsequent quenching of their star formation activity. Interestingly, similar – though less significant – signatures are also found in galaxies populating the large-scale protocluster regions, suggesting that environmental effects on star formation begin before galaxies become satellites of group- or cluster-sized halos.
At the same time, the analysis of massive ($M_{200} \geqslant 10^{13}\ M_{\odot}$) high-redshift halos in DIANOGA simulations has shown that a hot, diffuse gas component – a (proto-)ICM – dominates the gas budget within such halos as early as $z = 3$. The build-up of the ICM and IGrM follows different, mass-dependent evolutionary paths and timescales.
While highlighting the importance of cosmological hydrodynamical simulations in interpreting the complex environments of cluster progenitors at cosmic noon, this work has also underlined the limitations of simulations in reproducing high-redshift observations. In particular, these limitations include the systematically low galaxy SFRs as compared to observations, and the too shallow density profiles of the proto-ICM compared to the (very limited) observational studies currently available.

6.4 FUTURE PERSPECTIVES FOR PROTOCLUSTER STUDIES

6.4.1 *New and future observational facilities*

On the observational side, current protocluster studies are hampered by small-number statistics and complex selection functions. Ongoing and upcoming wide-field surveys, such as the ESA Euclid mission, will soon enable the selection of statistically robust protocluster samples, thereby reducing selection biases and enabling a first statistical probe of the intrinsic diversity of the protocluster population. A significant step forward in the investigation of galaxy evolution in these environments will then require specific follow-up observations probing galaxy population properties. Such follow-up programs will take advantage of current state-of-art and upcoming facilities both ground- and space-based, spanning multiple wavelengths:

from the near-IR/optical regime – with, e.g., the Multi-Object Optical and Near-infrared Spectrograph (MOONS) at the ESO Very Large Telescope (VLT) and the James Webb Space Telescope, as well as wide surveys from Euclid, Subaru, Roman and Rubin–LSST – to the submillimeter and radio bands, with ALMA, MeerKAT, and the Square Kilometer Array (SKA).
Future and recent facilities in both the X-ray and submillimeter regimes will also advance the observational investigation of ICM assembly, which is currently constrained by technical limitations. SZ studies with ALMA will benefit from the recent introduction of Band 1 (and the forthcoming Band 2), providing improved sensitivity and better sampling of large scales – ideal for extended signals such as the SZ effect. Further progress will come from synergy with the future Atacama Large Aperture Submillimeter Telescope (AtLAST), which will provide deep, high-resolution observations with its 50-meter diameter dish and large field of view.
On the X-ray side, next-generation observatories, such as the Advanced X-ray Imaging Satellite (AXIS, Reynolds et al. 2023) and NewAthena (Advanced Telescope for High ENergy Astrophysics, Barret et al. 2023), are expected to provide significant enhancements thanks to their improved collecting area, and spectral and angular resolutions.

6.4.2 *Ongoing work and future perspectives from simulations*

Building on the results presented in this Thesis, several paths for future investigations of protocluster environments in cosmological hydrodynamical simulations can be identified. In the context of galaxy evolution, the next crucial step is to constrain when, where, and how the evolution of (proto)cluster galaxies is affected by the broad range of environments that they cross throughout their evolutionary history.

In this context, a natural question that arises is how the galaxy populations and the emerging ICM co-evolve in cluster progenitors at cosmic noon. Since some quenching paths are associated with the presence of a hot, pressurized ICM, understanding how this component starts to impact galaxy evolution in high-redshift protoclusters is fundamental in the broader picture of environmental quenching.
Star formation properties of coeval protocluster cores with similar mass are observed to vary significantly from object to object at $z \sim 2$. This heterogeneity is also found in the `DIANOGA` simulations (see Sect. 4.5), as evidenced by the diverse fractions of quenched galaxies in halos of comparable mass. These differences are attributed to variations in the timescales of evolutionary processes that regulate star formation in galaxies within these highly active environments. Part of this diversity may reflect the timing of the ICM build-up in massive halos, as recent observations begin to reveal possible signatures of ram-pressure stripping at work in high-redshift protoclusters (e.g., Xu et al. 2025)
In this context, Fig. 6.1 shows preliminary results on the possible correlation between star formation activity in satellite galaxies and the timing of ICM

build-up in massive ($M_{200} \geqslant 10^{13}$ $M_{\odot}$) halos in DIANOGA simulations. The central galaxy is excluded, as its evolution is possibly sensitive to different processes compared to the overall population of satellite galaxies. The integrated specific star formation rate (sSFR) of the population of satellite galaxies at $z = 1.5$ is shown as a function of the fraction of the ICM that was already thermalized at $z = 3$ and 2.2 (right and left panels, respectively). This fraction is a measure of how fast the hot diffuse phase has built up between $z = 3$ (or 2.2) and $z = 1.5$. The halos are separated in three mass intervals to investigate the scatter in sSFRs among systems of similar mass. The median sSFR in each mass bin generally decreases with the fraction of ICM progenitor already thermalized at $z = 2.2$, though with a large scatter. Therefore, halos whose hot diffuse gas component has thermalized earlier are less star-forming on average, at fixed halo mass. At $z = 3$, this trend is qualitatively the same, though with even larger scatter and with the exception of the $M_{200} \sim 10^{14}$ $M_{\odot}$ halos. In these massive halos, no clear correlation is found between the sSFR and the fraction of already thermalized ICM at $z = 3$. This suggests that some other processes, not related to the presence of a mature ICM component, might influence the star formation activity in satellite galaxies in $z = 1.5$ cluster-sized halos, between $z = 3$ and 2.2. However, the number of such massive systems in the simulated sample is limited. The overall picture highlights the complexity of galaxy evolution in these extreme environments, with multiple (internal and environmental) processes acting to quench – and possibly also rejuvenate – star formation in galaxies.

In this context, future work will be directed also on the investigation of AGN activity in protoclusters, to address the open questions of whether bright AGNs are expected to be more abundant in these environments and what are the conditions that trigger their intense activity (e.g., Tozzi et al. 2022a; Vito et al. 2024; Travascio et al. 2025a).

The preliminary analysis presented here is currently hampered by the limited statistics in the sample of simulated halos, and by the temporal resolution of the saved simulation outputs. In fact, in the very active environment of protoclusters, it is crucial to trace galaxies between subsequent outputs of the simulations with a time cadence that allows to reliably track their evolution, before, during and after their infall in massive halos. In this way, by studying the evolving star formation properties of galaxies as they cross different environments, much better constraints could be placed on the processes that are most effective in quenching their star formation. At the same time, more accurate predictions could be made on the timing of the thermalization of the ICM and on the influence of feedback processes in shaping its properties in the innermost regions of clusters.

The numerical constraints highlighted above – limited sample sizes and sparse temporal sampling in simulation outputs – reflect broader limitations in current simulations relative to the computational cost of their production and post-processing, which are being actively addressed by the community. To this purpose, substantial effort is being devoted to the development of efficient cosmological codes that exploit advanced parallelization strategies

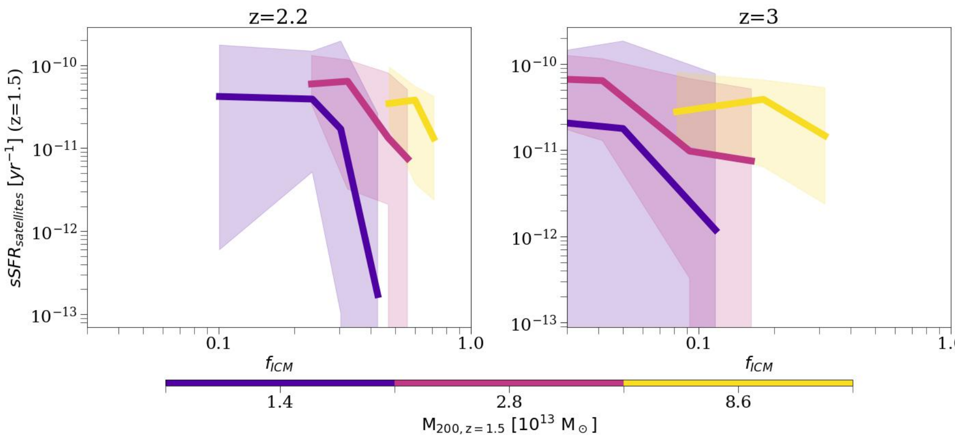


Figure 6.1: Correlation between integrated specific star formation rates of satellite galaxies in massive halos at $z = 1.5$ and timing of build-up of their ICM. The latter is quantified by the fraction of ICM identified at $z = 1.5$, that was already thermalized by $z = 2.2$ (left panel) and $z = 3$ (right panel). Here, the sSFRs are calculated by considering all satellites within R_{500} and with stellar mass $M_* > 10^{10}$ $M_\odot$, so as to exclude galaxies whose evolution might be affected by resolution in the densest environment (see Sect. 4.4). Stellar masses and SFRs are calculated within a 8.5 kpc radius around `SubFind` centers, with the SFR being averaged over the past 200 Myr. Solid lines (shaded areas) show the median (16th-84th percentile range of) sSFR, color coded according to the value of the median mass in three halo mass intervals.

and hardware accelerators, effectively speeding-up the production of simulations. Recent cluster zoom-in simulations are thus expanding in statistical power, with new suites encompassing samples of $\sim 10^2$ clusters at high resolution, as in TNG-Cluster (e.g., Nelson et al. 2024) and The Three Hundred project (e.g., Cui et al. 2018).

Since the link between simulations and the observable Universe is represented by halo and subhalo catalogs, considerable work is also focused on developing efficient post-processing tools for substructure identification (see Appendix A). The application of substructure finders to a high number of simulation outputs – especially at high resolution – can indeed represent a significant bottleneck.

Besides technical limitations, simulations are currently challenged by the difficulties in reproducing in detail the increasingly abundant and accurate observations, at both low and high redshift – as shown throughout this Thesis. Both galaxy and ICM properties are highly sensitive to the details of star formation and feedback models, making their accurate calibration essential for reliably reproducing observations. New suites of cosmological hydrodynamical simulations are currently adopting machine-learning based approaches to calibrate sub-resolution models of galaxy formation, thereby enabling an effective exploration of the parameter space (e.g., Kugel et al. 2023). Accurate calibrations, along with increasingly sophisticated models to capture the complexity of sub-resolution processes driving galaxy formation and evolution, are paving the way for an improved, realistic description of galaxy and gas properties in the Universe.

A SUBSTRUCTURE IDENTIFICATION WITH DEEP LEARNING

The identification of substructures (or subhalos) within halos in cosmological hydrodynamical simulations is a fundamental step for identifying the simulated counterparts of observed systems, namely galaxies. For this reason, substructure finders play a key role in extracting relevant information from the simulation outputs. Traditional algorithms are based on physically motivated definitions of substructures, performing multiple steps of particle-by-particle operations, and for this reason they are computationally expensive.

In this Appendix, I propose a fast algorithm based on deep learning techniques to identify substructures, particularly galaxies, in simulations. The main goal, besides a faster production of substructure catalogs, is to provide an algorithm fast enough to be applied with a fine time cadence during the evolution of the simulations. Accessing galaxy catalogs while the simulation is evolving is indeed necessary for sub-resolution models that depend on the global properties of galaxies. Here, I present `SubDLe` (Substructure identification with Deep Learning), an algorithm that combines a 3D adaptation of the well-known fully convolutional network `U-Net` (Ronneberger et al. 2015) with the Friends-of-Friends algorithm.

The structure of this Appendix is as follows: Sect. A.1 reviews traditional halo and subhalo finders and discusses their limitations, motivating the search for faster alternatives using machine learning techniques. In Sect. A.2, I describe the lower-resolution version of the `DIANOGA` simulations used for training and testing the model, as well as the basics of convolutional neural networks (CNNs) and `U-Net`, along with details on the training strategy for `SubDLe`. Sect. A.3 presents the results of the trained model for substructures including dark matter, gas, and stellar particles at $z = 0$. A focused analysis is then presented for cases where substructures are identified only from the distribution of star particles at $z = 0$. Additional tests were performed at $z = 1$ and increasing the resolution of input grids. Conclusions and future perspectives are reported in Sect. A.4. The results presented here are published in Esposito et al. (2024).

A.1 INTRODUCTION

In order to fully exploit the predictive power of cosmological simulations, advanced analysis tools are required, especially structure finders. The identification problem becomes particularly non-trivial when structures are embedded within larger, virialized systems. This prompts the definition of

two main classes of algorithms for structure finding: halo finders, strictly speaking, are those that search for virialized structures around local maxima of the matter density field (the "knots" of the cosmic web); subhalo (or substructure) finders are faced with the more challenging task of identifying structures orbiting within a larger parent halo. In addition to these, there are also algorithms designed to identify filaments in simulations.
Subhalo finding has evolved and become a discipline itself during the past decades, given its crucial role of mapping the simulated phase-space distribution of matter to observable objects, such as galaxies. This is a necessary step to study galaxy properties in simulations and compare them with observations.
Though a proper classification of the algorithms for halo and subhalo finding cannot be easily defined, as reviewed in Knebe et al. (2013b), they can be broadly divided into two categories stemming from the first two well-known algorithms developed for this purpose:

- Direct collectors of groups of close particles, such as the Friends-of-Friends (FoF) method (Davis et al. 1985). In the classical FoF, particles are grouped such that the interparticle distance is always below a parameter, the linking length, usually given in units of mean interparticle distance. This is used as the starting point of the algorithm `SubFind` (Springel et al. 2001b; Dolag et al. 2009) to identify substructures; the FoF run finds the dark matter parent halos and within them the search for substructures is performed by `SubFind`. The linking length is usually assumed to be about 0.2 (the exact value being dependent on cosmology), since this was proved to identify halos with a density threshold similar to what is predicted for virialized systems by the spherical collapse model. The classical FoF algorithm, when applied to subhalo identification, has the disadvantage of linking close structures together through thin filaments of particles. Nevertheless, extensions of the original FoF have been developed to account for this, for example ROCKSTAR (Behroozi et al. 2013), which does a first separation of large groups in the classical FoF approach, then refines it by hierarchically searching for subgroups in 6D phase-space, while also tracking the temporal coordinate.

- Density peak locators complemented with a neighbor search, such as the Spherical Overdensity (SO) method (Lacey & Cole 1994). In this kind of approach, the first step is the identification of "locations" of candidate halos, through a search of peaks in the density field or local minima in the potential field, among many possibilities. Then, the structures are grown around these locations by grouping neighboring particles, up to a given boundary, whose definition varies across different algorithms. `SubFind` follows a similar approach to identify subhalo candidates, by locating overdense regions, which are enclosed by isodensity contours intersecting saddle points in the density field. Another example is SKID, described in detail in Stadel (2001), where particles are repeatedly "slid" across the local gradient of the density

> field until they are collected around local maxima. Then in both cases, the catalog of subhalo candidates has to undergo an unbinding procedure, where particles that are not dynamically bound to their candidate subhalo are pruned from the catalog.

A detailed comparison of different halo and subhalo finders can be found in Knebe et al. (2011, 2013a) and Onions et al. (2012) for dark matter-only simulations and in Knebe et al. (2013b) for hydrodynamical simulations, where the authors have investigated the source of scatter in halo properties among different codes.
Halo and subhalo finders are now routinely integrated into the simulation codes, allowing for identification of subhalos on-the-fly, during the evolution of the simulation. This allows to save subhalo catalogs rather than the complete raw output of simulations and also access information on galaxy properties on-the-fly. Unfortunately, the overhead in computational cost when using such subhalo finders is usually non negligible, reaching up to the point where their cost over a significant number of outputs becomes comparable to the simulation itself. This is due to the large number of particle-by-particle operations involved in the different steps of structure finding, like the search for saddle points in `SubFind` or for the local maxima in SKID, which might also require several iterations. Furthermore, as the resolution of simulations increases, so do the number of particles and the computational cost of structure finders.
Many algorithms have been updated or designed to work in parallel, like VELOCIraptor (Elahi et al. 2011), which searches for substructures as peaks above the velocity distribution of the background, or `SubFind` itself. However, the cost of running these algorithms while the simulation evolves is still a large limitation. On the other hand, performing frequent on-the-fly identifications of subhalos would be desirable for different applications. For example, one such application is the implementation of models for sub-resolution physics which rely on global properties of galaxies.

In this context, I introduce `SubDLe` (`Sub`structure identification with `Deep Le`arning), an algorithm for substructure identification which is based on methods of deep learning (e.g., Chapter 3 of Aggarwal 2018), a subclass of machine learning (see Baron 2019, for an overview of machine learning techniques applied to astronomy). Given the computational efficiency of deep learning methods, this approach aims to drastically reduce the cost of identifying substructures in large cosmological simulations.
The application of the so-called supervised techniques requires access to a "training set," consisting of input-output pairs, which define the problem to solve and the ideal solution. Then, through the optimization of an objective function, a set of parameters, which define how the algorithm solves a specific task, are iteratively adjusted in order to reproduce the solutions. This is the "training" stage which is the most computationally expensive step in the definition of the machine learning model, but only needs to be performed once if the data domain does not change significantly. A proper training requires the final model to be able to work outside the training set,

which means that it needs to be able to solve problems belonging to the same class of the training examples, but characterized by different data sets.

Artificial neural networks (e.g., Chapter 1 of Aggarwal 2018) are a class of algorithms capable of approximating strongly nonlinear functions, with a flexible structure which makes them suitable for various tasks. Convolutional neural networks (CNNs, e.g., Chapter 8 of Aggarwal 2018) are a network architecture tailored for visual tasks. For example, Teodoro et al. (2023) applied a CNN to clean galaxy catalogs of contaminants by identifying star formation regions or shreds of galaxies in large photometric surveys containing millions of sources. Indeed, object detection with CNNs has been widely implemented in different fields and can be a suitable option to perform frequent on-the-fly identification of galaxies, given the computational advantages of a trained network.

In this work I present an application of a specific CNN, `U-Net` (Ronneberger et al. 2015), to 3D density maps of galaxy clusters obtained from a set of cosmological hydrodynamical simulations carried out with `GADGET-3` (see Sect. 3.1). In this analysis, I performed the training on subhalo catalogs provided by the built-in subhalo finder `SubFind`.

A.2 METHODS

A.2.1 *Simulations*

The analysis presented here is based on a set of 13 simulated galaxy clusters extracted from a lower-resolution version of the `DIANOGA` simulations of galaxy clusters (e.g., Bonafede et al. 2011; Rasia et al. 2015, see Sect. 3.2), compared to the version analyzed in the main body of the Thesi. These were re-simulated with a developer version of `GADGET-3`. The mass resolution for DM and gas particles is $m_{DM} = 8.47 \times 10^8\ h^{-1}M_\odot$ and $m_{gas} = 1.53 \times 10^8\ h^{-1}M_\odot$, respectively, while the Plummer-equivalent softening lengths are $\epsilon = 5.6\ h^{-1}$ kpc for DM and gas, and $\epsilon_* = 3\ h^{-1}$ kpc for stars. The simulations analyzed here include the description of the following processes described in Sect. 3.1.3: metal-dependent radiative cooling, star formation, chemical enrichment from stellar evolution and feedback from Supernovae (SN) and Active Galactic Nuclei (AGN). I refer to Ragone-Figueroa et al. (2013) for a detailed description of this set of simulations.

In the simulated regions, galaxies have been identified with the `SubFind` algorithm (Springel et al. 2001b; Dolag et al. 2009). As described in Sect. A.1, this subhalo finder performs its search within DM halos identified with a FoF algorithm. Within them, substructures are then identified by looking for regions enclosed by isodensity contours crossing a saddle point of the density field. After that, particles belonging to each of the subhalo candidates undergo an unbinding procedure to exclude those particles that are not gravitationally bound to the assigned subhalo. In Section A.3 I refer

Table A.1: Main properties of the clusters analyzed in this study. For each cluster, listed below are the virial mass, the virial radius, the number of galaxies considered resolved enough for this study (with at least 100 star particles), and the phase of the analysis in which it was used (i.e., training, referred to as "train," validation, referred to as "val," or test).

ID	M_{vir} [10^{15} $M_\odot$]	R_{vir} [Mpc]	N_{gal}	phase
CL-1	0.74	2.4	192	train and val
CL-2	0.73	2.4	231	test
CL-3	2.3	3.4	571	test
CL-4	0.92	2.5	236	test
CL-5	0.24	1.6	75	test
CL-6	2.1	3.3	510	test
CL-7	0.18	1.5	40	test
CL-8	1.7	3.1	505	test
CL-9	1.9	3.2	604	test
CL-10	2.4	3.5	602	test
CL-11	2.3	3.5	535	test
CL-12	2.2	3.4	543	test
CL-13	2.3	3.2	542	test

to the clusters I have analyzed as `CL-i`, where i runs from 1 to 13. The main properties of these clusters are reported in Table A.1.

A.2.2 *Convolutional neural networks*

Convolutional neural networks (CNNs) are a class of machine learning algorithms, with a multilayer architecture (see Chapter 8 of Aggarwal 2018), where each computational layer maps local input to local output via linear or nonlinear operations. CNNs are designed to process grid-like inputs, like images, text, time series or data sequences, that exhibit strong variations. In image data, the semantic meaning of an object does not depend on its position within the image. This property makes images particularly suitable for CNNs, which are designed to achieve translation invariance. This can be explained as the core operations of a CNN act locally on the input grid and depend only on relative spatial coordinates. This characteristic is essential for substructure identification, enabling the detection of galaxies at any location within the simulated domain.

CNNs are defined as artificial neural networks that use a convolution operation (described in Sec. A.2.3) in at least one layer. A network consisting only of layers that perform convolution operations is called a fully convolutional network (FCN, Shelhamer et al. 2017), which is the case of `U-Net`, chosen for this work.

In the following, I describe CNNs and their core operations (Sect. A.2.3). Then, I introduce `U-Net` (Sect. A.2.4), implemented in `SubDLe`, in a generalized 3D version to address the identification of substructures in cosmological simulations. The goal is to outline the fundamental principles underlying this subhalo finder and to clarify the sources of its strengths and limitations.

A.2.3 *Basic structure of a CNN*

In the case of applications to image data, each layer of a CNN processes an input grid characterized by height, width, and depth, where height and width correspond to the spatial dimensions, and depth represents the number of color channels. Throughout this Section, I refer to CNN input data as images, following the standard convention in which a pixel corresponds to a location on a two-dimensional grid (*height* × *width*) with multiple associated values along the *depth* dimension (typically three in RGB images, representing the intensities of the different color channels). The extension from two to three spatial dimensions is conceptually straightforward; hence, the 2D case is described here for simplicity of notation.

A.2.3.1 *Convolution*

A convolution is an operation between an input grid and a filter – a small kernel whose values (parameters) are optimized during the training process. Consider the q-th layer of a CNN, where the input grid has dimensions $L_q \times B_q \times d_q$, and a filter with dimensions $F_q \times F_q \times d_q$, with the depth d_q matching that of the input. Typically, $F_q \ll L_q$ and $F_q \ll B_q$, with common filter sizes $F_q = 3$ or 5.
The convolution operation is performed by shifting the filter across each possible location in the grid, where a dot product is computed between the filter parameters and the corresponding overlapping region of the grid, producing a single scalar value in the output feature map. To illustrate this, consider the example shown in Fig. A.1: here $L = B = 5$, $F = 3$ and $d = 1$, for simplicity. The filter is first applied to the top-left 3×3 region of the grid; the element-wise products are then summed, yielding the first output value, $1 \times 1 + 1 \times 1 + 1 \times 1 + 1 \times 1 = 4$, with multiplications by zero omitted. The filter is then translated by one position along the horizontal axis, and the operation is repeated to obtain the second entry of the output map, until the output map is completely filled. The number of filter placements is defined by the number of the alignments along the height $(L - F + 1)$ times the number of alignments along the width $(B - F + 1)$, 3×3 in the example. The convolution yields output values (features) that are not connected indiscriminately to all the elements of the input data, but only to a local region. This enables the network to capture spatial patterns

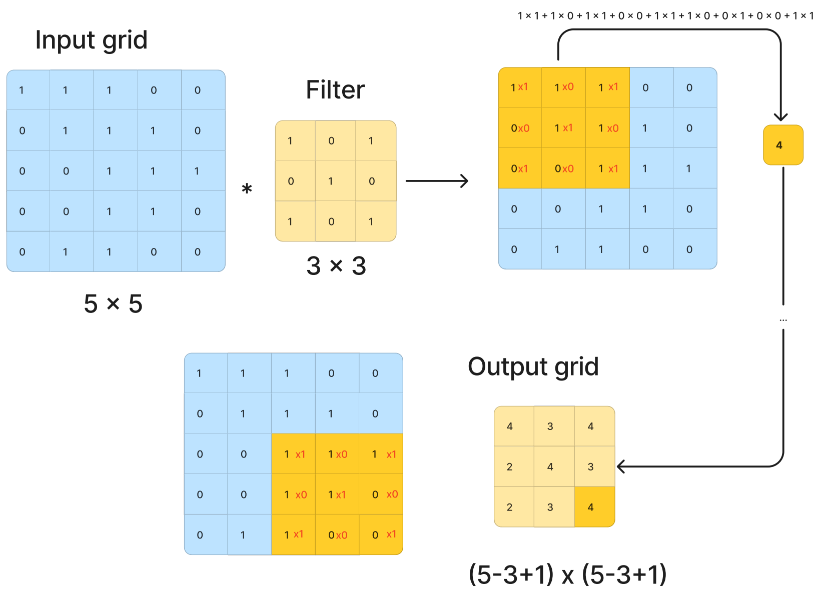


Figure A.1: Schematic representation of a convolution between a 5×5 grid and a 3×3 filter, with depth 1. The filter is slid across the grid and the dot-product is performed at each alignment (represented by the yellow subset of the grid in the figure), until all the grid is covered and the 3×3 output grid is completed.

and shapes. This property makes the application of CNNs particularly well suited to associate local overdensities in the simulated mass distribution with substructures.
Figure A.2 shows a more complex configuration, with $L = 64$, $B = 32$, $d = 3$ and $F = 5$. The convolution works as in the example with $d = 1$, the only difference being that the dot product at each position of the grid is performed across the third dimension as well, thus outputting a $(64 - 5 + 1) \times (32 - 5 + 1)$ grid. Then, the depth of the output grid is built by stacking the output of convolutions with different filters (2 in the example). Here, strides larger than 1 are not considered, meaning that the filter is shifted by only one position at each step. For a more comprehensive treatment, I refer to Chapter 8 of Aggarwal (2018).

In general, for the q-th layer of a CNN having as input a $L_q \times B_q \times d_q$ grid, the depth d_q is given by the number of filters applied in the $(q-1)$-th layer. The total number and the dimensions of filters in all layers of the CNN define the number of parameters, also known as the capacity, of the model.
The underlying idea in applying multiple filters is that each one can capture a particular spatial pattern, so a large number of them are required in order to capture different patterns in complex input images.
Let the parameters of the p-th filter in the q-th layer be defined by the tensor

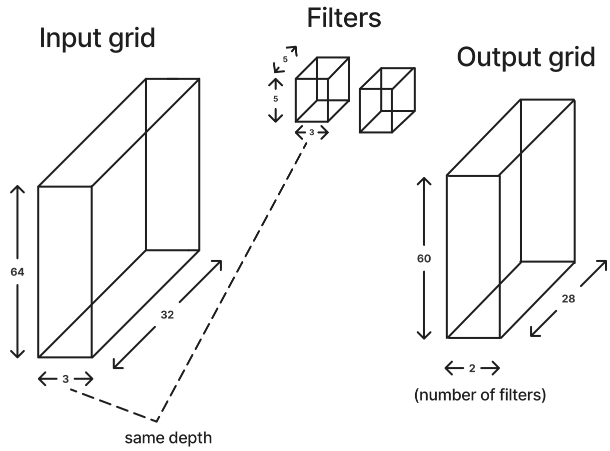


Figure A.2: Convolution between a $64 \times 32 \times 3$ input grid with two $5 \times 5 \times 3$ filters gives a $60 \times 28 \times 2$ grid. Note that the depth of the filter and that of the grid must be the same, while the depth of the output grid is given by the number of filters.

$w_{ijk}^{(p,q)}$, where the indices i, j, k correspond to positions along the height, width, and depth of the filter. Let $h_{ijk}^{(q)}$ represent the k-th feature map (i.e., the output of a convolution) in the q-th layer. The p-th feature map in the $(q+1)$-th layer is then given by:

$$h_{ijp}^{(q+1)} = \sum_{r=1}^{F_q} \sum_{s=1}^{F_q} \sum_{k=1}^{F_q} w_{rsk}^{(p,q)} h_{i+r-1,j+s-1,k}^{(q)} + b^{(p,q)} \qquad (33)$$
$$\forall i \in [1, L_q - F_q + 1],\ \forall j \in [1, B_q - F_q + 1],\ \forall p \in [1, d_{q+1}];$$

where $b^{(p,q)}$ is the bias for the p-th filter of the q-th layer, an extra parameter which is summed to the element-by-element product in the convolution, thus only producing a uniform offset in the feature map. d_{q+1} is equal to the number of filters applied in the q-th layer.

In standard multilayer convolutional architectures, filters in the initial layers encode primitive shapes, while those in deeper layers detect increasingly complex patterns. Successive convolutional operations progressively increase the receptive field, i.e., the local region of the input image that influences each feature.
In the example shown in Fig. A.1, the application of a 3×3 filter produces feature maps in which each feature encodes information from a 3×3 patch of the input grid. If the same filter is applied once more, given that the grid now has dimensions equal to the filter size, only a single feature is generated. This feature now contains information over a larger spatial region, corresponding to all 5×5 pixels from the original input grid.
In addition to convolutional layers, which perform convolution operations,

CNNs typically include two other types of layers: pooling layers and ReLU (Rectified Linear Unit) layers, where the pooling operation and the ReLU function are applied, respectively.

A.2.3.2 *Rectified Linear Unit and pooling*

The Rectified Linear Unit (ReLU) is a function defined as:

$$f_{\mathrm{ReLU}}(x) = \max\{x, 0\}. \tag{34}$$

This is frequently used in artificial neural networks, especially CNNs to introduce nonlinearity into the model, since convolution is a linear operation mapping the input to a feature map. In a CNN, ReLU is applied element-wise to all $L \times B \times d$ values of the input grid, leaving its dimensions unchanged, and typically follows a convolution operation.

Pooling is an operation that acts on small regions of a grid, producing an output with the same depth. In max pooling, the most common variant, the maximum value is taken from each region, adding some nonlinearity in the model, while other variants, such as average pooling, are linear. Pooling also increases the receptive field of each feature.

A.2.4 `U-Net`*, data set, and training*

`U-Net` (Ronneberger et al. 2015) is a FCN developed to perform semantic segmentation in biomedical images. Semantic segmentation is a task involving classification of each pixel of an image to identify object boundaries instead of bounding boxes, as in classical object detection. I refer to Ronneberger et al. (2015) for a detailed description of the architecture of `U-Net`.
For this analysis, I modified the original network to apply it to the identification of subhalos in 3D, in cosmological hydrodynamical simulations. In this application, the pixel values represent the mass densities. I built a 3D version of `U-Net` using the pyTorch[1] library.

The training of the 3D `U-Net`, that performs most of the calculations in `SubDLe`, requires a set of reliable data to "teach" the network how to identify substructures. This was provided by the subhalo catalogs produced by `SubFind` in the simulations described in Section A.2.1. For my aims, using the direct `SubFind` output, which associates each particle with a substructure (or to none, if the particle is not dynamically bound to any subhalo) was not feasible. The reason is that a CNN works with grid-like inputs with a fixed dimension; the particles of a halo instead define a sparse data structure. I thus mapped `SubFind` catalogs to grids having fixed dimensions. A grid is defined as a box having an origin in a given point of space (X_g, Y_g, Z_g), a size L_g, and N_g points per dimension. A particle with coordinates in the ranges $[X_g, X_g + L_g]$, $[Y_g, Y_g + L_g]$, and $[Z_g, Z_g + L_g]$ falls within the grid and will be mapped to a grid pixel. Let us consider the simple situation in which each grid pixel contains at most one particle, as in the upper panel

1 `https://pytorch.org/`

of Fig. A.3. The red particles as those associated, according to previous knowledge (such as SubFind's identification), with the objects I want to identify, whereas the blue particles are just part of a "background." Then, I can assign the label "1" to the substructures and "0" to the background. Thus, in Fig. A.3, grid pixels which contain red particles are assigned the label "1" in the right grid of the top panel, containing labels for all grid pixels, while all the others are assigned the label "0". Typically, more particles fall in the same pixel; then, the pixel is assigned to a substructure or to the background using a majority vote, as in the lower panel of Fig. A.3. In the unlikely case of parity, I assign the pixel to a substructure. I refer to the grid carrying labels as "s grid."

For the problem of identification of substructures in simulations, a twin grid sampling the mass density is also computed, a "d grid," built with a Nearest Grid Point (NGP) assignment scheme, for simplicity. This is a case of binary segmentation, in which a grid pixel is classified as a member of one class (label "1," denoting substructures) or not (label "0"), resulting in s grids containing only 1s and 0s, as in the simplified examples shown in Fig. A.3. This approach is thus based only on the density field, without any information on the kinematics of particles.

The training of U-Net is performed with one NVidia P100 GPU with 16 GB memory, provided by the online data science platform Kaggle[2]. The training, as is standard for neural networks (Aggarwal 2018, Section 1.2.1) is done through the minimization of a "loss function," which defines the error of the model's prediction at each iteration ("epoch"). In simple cases of regression the chosen loss function is often a simple Mean Squared Error. The minimization is performed with respect to the model parameters, with backpropagation (see Aggarwal 2018, Section 1.3). Since it is conceptually wrong to evaluate the best set of parameters on the same data set on which the optimization of the loss function is performed, a set of data is reserved for validation. The loss function is then evaluated on the validation data set and the best model is selected from the set of parameters that yields the minimum loss function. The chosen loss function here is the Binary Cross Entropy, suited for binary segmentation tasks.

Because of memory constraints during training, each simulated cluster is covered with multiple sub-grids. Employing a single grid to encompass the entire cluster would demand an impractical amount of memory to achieve the spatial resolution necessary for resolving the smallest substructures with a few pixels. The dimension of each sub-grid determines the largest scale on which the identification can benefit from the network's architecture, since identifying structures larger than the sub-grid itself is problematic. In such case, indeed, large substructures are split into more than one grid, thereby forcing SubDLe to try to identify "shreds" of a substructure in each of them. However, I emphasize that this represents only a technical limitation, strictly

2 https://www.kaggle.com/

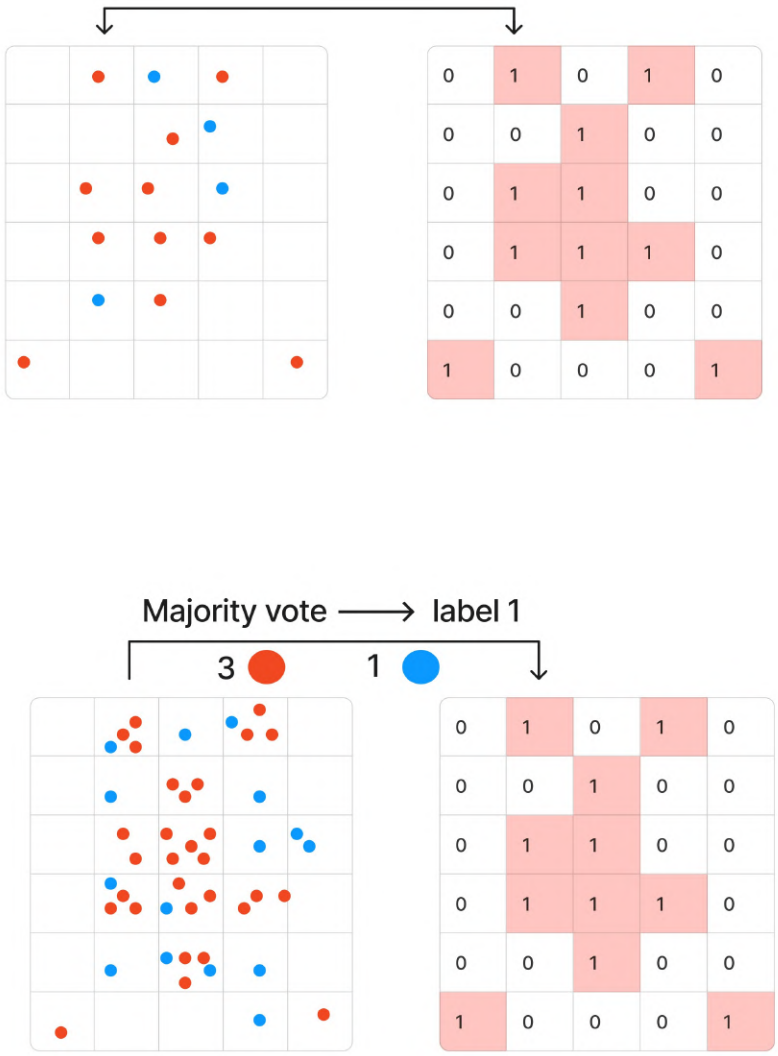


Figure A.3: Example of the assignment scheme of labels adopted in the training set of the 3D `U-Net` implemented in `SubDLe`. Here red particles trace the substructures to be identified, and blue particles are part of a background. In the upper panel, one particle at most is present in each grid pixel, thus the substructures (associated with the label "1" in the right-hand grid, whereas the background is labeled as "0") are simply associated with pixels containing one red particle. In the lower panel, closer to the real identification problem, more than one particle falls in each grid pixel, and the assignment of labels is performed with a majority vote.

determined by the accessible hardware, and not an intrinsic limitation of the method.

I adopt a number of grid meshes per side N_g=64, following the convention for images of having a power of 2 as the number of pixels per side. The physical size of each pixel is chosen to be approximately 10 kpc, such that a full d-grid encompasses a region larger than the typical extent of $z = 0$ galaxies in `SubFind` catalogs. Unless otherwise stated, in Section A.3, I always assume a resolution of about 10 kpc, with small variations (ranging from 9 to 12 kpc), which make little to no difference in the final subhalo catalog. `U-Net` is trained using both s and d grids, learning how to assign d grid pixels to substructures as defined by the s grid.

The `U-Net` sector in `SubDLe` is designed to output a 64^3 grid with pixel values defining the probability p of each pixel to belong to a substructure (the class labeled as "1"). A threshold is then chosen in order to decide which pixels are assigned to substructures. The most common choice is a threshold of $p_{th} = 0.5$, so I adopt it without tuning for this parameter. Then all pixels with a probability $p > 0.5$ are assigned the label "1," whereas the remaining ones are labeled as background (label "0"). This outputs an s grid containing recovered substructures.

Once pixels with a high probability of belonging to a substructure have been selected to produce a binary s-grid, a final step is required to regroup pixels that can be attributed to the same subhalo across all s-grids. This step is necessary because all subhalos are assigned the same label ("1"): binary semantic segmentation enables the identification of regions associated with a given class of objects, but does not provide information on individual object instances.
Figure A.4 shows an example of output of `SubDLe` at this stage of the identification process. The map shows the 2D projection of density of active pixels (i.e., pixels with label "1"), reconstructed from the s grids sampling the whole cluster `CL-5` at $z = 0$ (see Section A.2.1).

`SubDLe` performs the grouping of active pixels in the s grid using a proximity criterion based on a FoF algorithm, implemented through the `pyfof` package[3]. The linking length is set to $\sqrt{3}$ times the pixel size, in order to group pixels that are contiguous along all spatial directions. After this step, a catalog of substructures defined on the grid is obtained. Particles are then assigned to substructures by inverting the NGP assignment. This approach relies on the intrinsic assumption that substructures are spatially disconnected. While traditional subhalo finders may face challenges in identifying close or interacting structures as separate entities, the inclusion of kinematic information can mitigate this issue. In `SubDLe`, the FoF-based analysis explicitly searches for disconnected groups on the grid, thereby preventing the distinction of spatially overlapping subhalos.

3 `https://pypi.org/project/pyfof/`

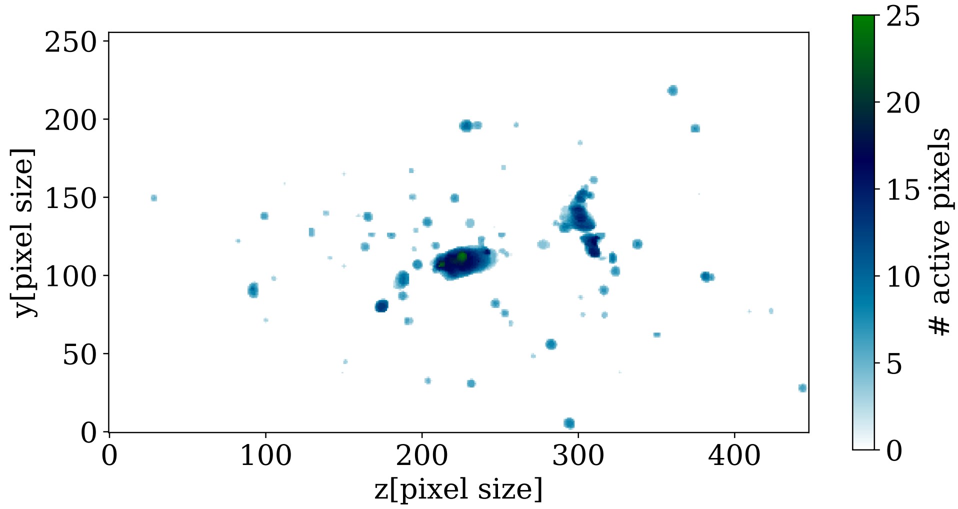


Figure A.4: Projected map of the active pixels associated with the substructures identified by the first step performed by SubDLe (through a 3D U-Net), in a run on the whole cluster CL-5 at $z = 0$. The blobs of active pixels need to be grouped together to identify subhalos, since at this stage all subhalos are associated with the same class label.

A.3 RESULTS

A.3.1 *SubDLe*

In this Section, I present the results from the application of SubDLe to the total mass density 3D maps of simulated galaxy clusters, including the contribution of DM, gas and stellar particles, and compare the output subhalo catalog with the one based on the application of SubFind.

The training of U-Net within SubDLe is performed on approximately 1000 d grids (700 dedicated to the training itself and 300 to the validation) extracted from the cluster CL-1 at $z = 0$, with a total virial mass $M_{vir} \sim 7.4 \times 10^{14}$ $M_{\odot}$. Each 64^3 d grid extracted from it is visually inspected in order to provide highly representative training data, with a high number of substructures. An increase in training data set size is expected to enhance model performance, although most of the results presented in this exploratory study are based on training data set extracted from a single cluster. Nevertheless, preliminary experiments with a larger data set were performed, as detailed in Sect. A.3.5.

The trained SubDLe was tested on the galaxy cluster CL-2 at $z = 0$, with approximately the same virial mass as CL-1 (see Table A.1). Here the d grids were extracted so that they overlap with their neighboring grids by half their sides in all three directions in order to properly sample border regions, for a total of 6851 grids covering a box of about $(10 \times 5 \times 5)$ Mpc^3. Each d grid covers a cubic region of about 640 proper kpc per side. Here the mass density sampled by the d grids includes all the particles contained in SubFind's catalogs, that are DM, stars and gas particles. A visual inspection of the results can be done by comparing the two panels of Fig. A.5. The left panel

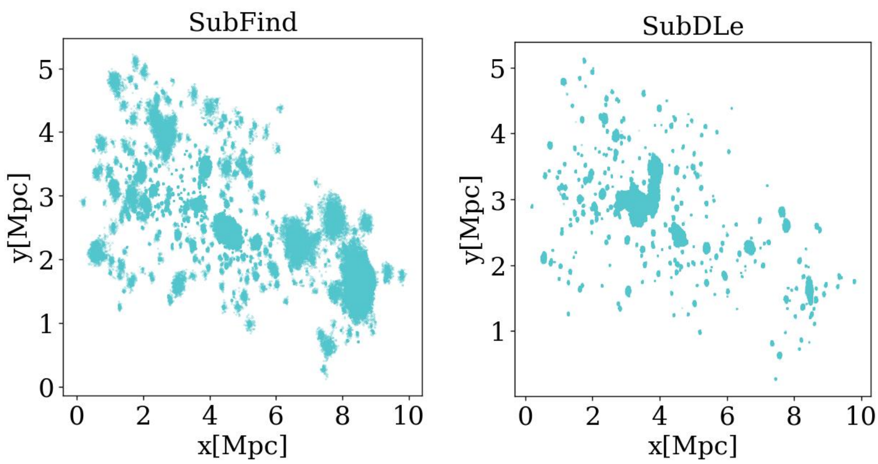


Figure A.5: Projected map of particles belonging to subhalos of the cluster `CL-2` at $z = 0$. The left and right panels show the subhalos identified by `SubFind` and `SubDLe`, respectively.

shows the projected distribution of particles belonging to substructures identified by `SubFind` in the cluster `CL-2`, whereas the right panel shows the substructures identified by `SubDLe` in the same cluster. Qualitatively, `SubDLe` appears to be able to locate most of the major substructures, while missing some of the smaller ones. The latter are inevitably described by a weaker signal in the d grids. On the other hand, the spatial extension of `SubDLe`'s substructures is generally smaller than the size of their `SubFind` counterpart. This behavior can be attributed to the fact that `SubDLe` detects more effectively regions characterized by high density contrast, such as the centers of subhalos. Conversely, the network tends to miss external low-density cells, even when they are dynamically bound to the same subhalo. Nevertheless, the most challenging task in substructure identification is the location of density peaks, which `SubDLe` performs pretty well, as I will discuss more quantitatively later in this Section. Once these peaks are identified, the detailed matter distribution of each subhalo can be recovered with a fast neighbor search algorithm, as mentioned in Sect. A.1 for the SO method.

Another detail that emerges from the comparison of the distribution of `SubDLe` and `SubFind` substructures in Fig. A.5 is the presence of a very large subhalo among `SubDLe` substructures at $(x, y) \approx (3\ \mathrm{Mpc}, 3\ \mathrm{Mpc})$, while the corresponding region in the left panel of Fig. A.5 is devoid of large structures. This corresponds to the largest "subhalo" of the cluster, hosting the Brightest Cluster Galaxy (BCG). This was excluded from `SubFind`'s substructures because this algorithm does not separate the largest subhalo of a FoF group from the diffuse component that is not bound to any specific subhalo (see Springel et al. 2001b). `SubDLe` can thus also be applied to separate the central subhalo from the diffuse component, though further specific tests are required to assess this application. A machine learning algorithm specifically designed to disentangle the BCG from the intracluster light (ICL) in the

spatial distribution of star particles within simulated clusters, by exploiting their dynamical properties, has already been developed and validated by Marini et al. (2022).
In the following, I focus on the ability of `SubDLe` to reproduce the average subhalo population in cluster-sized halos. Tests performed in different environments are not presented here, but better results can reasonably be expected in less dense regions, where galaxies are more isolated and therefore more easily separated from the background by a density-based tool like `SubDLe`.

In order to quantitatively compare catalogs produced by `SubFind` and `SubDLe`, I match each subhalo found by `SubDLe` to the nearest counterpart identified by `SubFind`. Here only `SubFind` subhalos resolved with at least 100 particles are included. `SubDLe` substructures are processed in decreasing order of total mass, so as to avoid matching a `SubFind` subhalo to a close small clump of particles found by `SubDLe` before the actual corresponding `SubDLe` subhalo is processed. The proximity was estimated through the distance between `SubFind` center, identified as the position of the most bound particle in each substructure, and `SubDLe` center, which I define as the center of mass of the subhalo, for simplicity. The search for matching substructures is performed within a radius of 50 kpc, then the matched `SubDLe` substructures that do not share any particle with their closest match are excluded from the catalog, since they are unlikely to be matching any `SubFind` subhalo. Since `SubDLe` tends to miss the external lower-density region of subhalos, while efficiently detecting the central density peaks, as qualitatively shown in Fig. A.5, the criterion of requiring at least one common particle was proved to be sufficient to guarantee a significant overlap and thus a correct match between `SubFind` and `SubDLe` subhalos (i.e., the central most bound particle of a `SubFind` subhalo is typically present in the corresponding `SubDLe` substructure). Since my primary interest lies in assessing `SubDLe`'s performance as a peak locator, I do not apply a stricter matching criterion.

The distribution of distances between the centers of `SubDLe` substructures and their `SubFind` counterparts is shown in Fig. A.6, in blue. I find that 91% of pairs of substructures have distances below the gravitational softening of the DM particles, approximately representing the spatial resolution of the simulation. There is a tail of pairs which expands up to few tens of kiloparsecs, which might be due to a displacement between the center of mass and the center of the gravitational potential well or to an artificial merging of close substructures, as shown below in this Section. Overall, 80% of all the resolved substructures (i.e., with more than 100 total particles) identified by `SubFind` have been identified by `SubDLe` as well.

Fig. A.7 shows the comparison between matched pairs of `SubDLe` and `SubFind` substructures in three representative cases:

- the left column shows an ideal case, with a close to perfect match between the identifications performed by `SubFind` and `SubDLe` (98% of

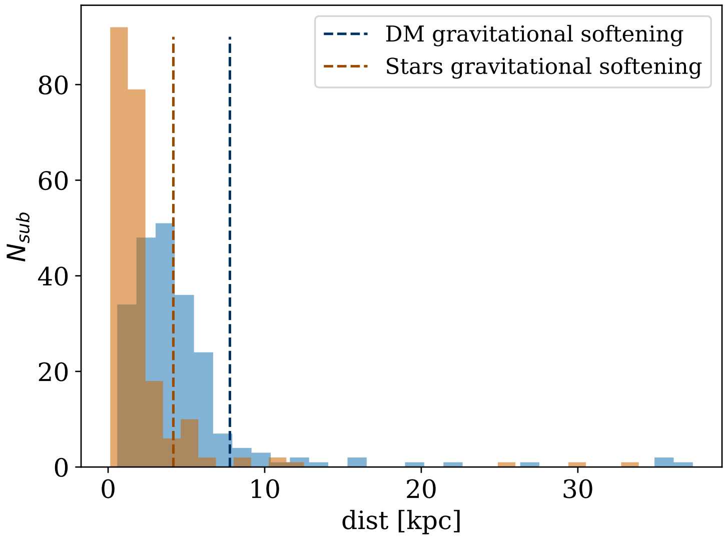


Figure A.6: Distribution of pairs of SubFind and SubDLe subhalos as a function of the distance between their centers, in physical kiloparsecs, for the identification performed by the all-particle SubDLe (in blue) and stellar SubDLe (in orange). The vertical dashed lines mark the gravitational softening used in the gravitational force calculations for the DM particles (in blue) and for star particles (in orange).

the mass of the original subhalo is recovered by SubDLe). Bottom and top panels show SubFind and SubDLe subhalos, respectively. The size of the subhalo, which is few tens of kiloparsecs, is well reproduced, while individual particles have been added or missed by SubDLe, due to the lack of information on their velocities, which prevents a direct assessment of their bound state. I note that this subhalo is well reproduced despite being resolved with few pixels per dimension.

- The central column illustrates the previously discussed trend of SubDLe truncating the most extended subhalos and retaining only their high-density cores. Note that this substructure spans several hundreds of kiloparsecs, thus it was necessarily split in multiple grids, reducing its detectability as a whole structure. This limitation does not represent a major limitation in the application of SubDLe, since the most challenging task of locating the peak of the density field has been accomplished in this instance as well.

- The right column shows a case in which two visibly (and dynamically, as SubFind identifies them as two different subhalos) distinct substructures were merged together by SubDLe because of their proximity. Their centers are indeed few resolution elements away and SubDLe is not able to disentangle them on the basis of their mass distribution alone. The artificial merging of close substructures is one of the causes of missing subhalos in SubDLe's catalogs. I discuss how the increase of

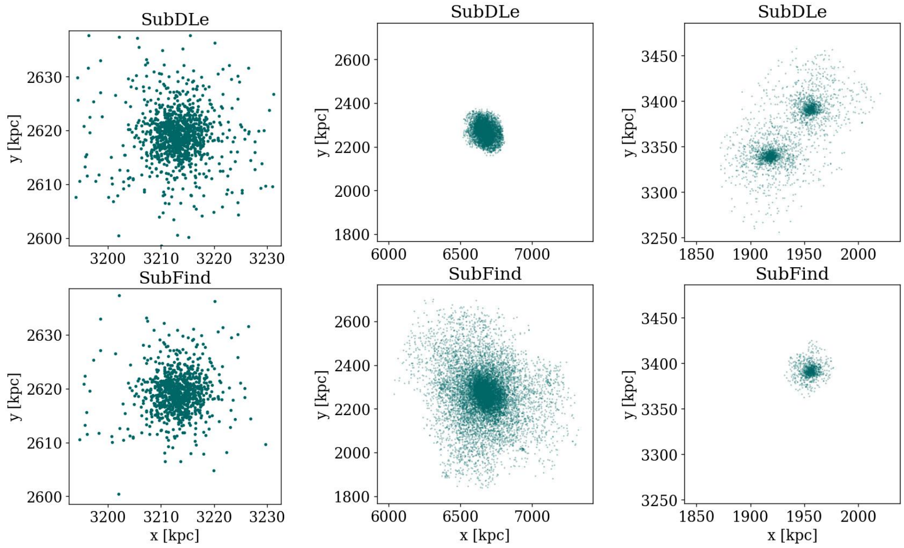


Figure A.7: Scatter plot of all particles (DM, gas, and stars) belonging to a subhalo identified by `SubDLe` (top-row panels) and the matching subhalo identified by `SubFind` (bottom-row panels), projected in 2D. Three typical cases are shown: in the left column, a subhalo with an extension of few tens of kiloparsecs, perfectly identified by `SubDLe`; in the central column a larger subhalo, spanning a distance of about 1 Mpc, where only the central region is identified by `SubDLe`; and in the right column, a system of two close substructures that are combined into one subhalo in `SubDLe` catalog due to their short distance, compared with the density grid resolution.

resolution can indeed boost the percentage of matched subhalos, by separating close, but distinct substructures, in Sect. A.3.4.

A.3.2 *Stellar* `SubDLe`

Since `SubDLe` works inherently better in identifying highly contrasted clustered structures, I repeat the identification of substructures using stellar density maps instead of total density maps; I retrain `SubDLe` to do so and refer to this version as stellar `SubDLe`. Stellar substructures (i.e., galaxies) are indeed more clustered in comparison to DM and gas structures due to their formation through dissipative collapse of gas clouds. The training data set consisted of 700 grids of stellar mass densities extracted from cluster `CL-1`, while 300 grids are used for validation. Fig. A.8 shows the same comparison of Fig. A.5, this time with galaxies identified by `SubFind` (left panel) and by the stellar `SubDLe` (right panel) in the cluster `CL-2` at $z = 0$.
The discrepancy in substructure sizes is still visually noticeable, but less severe than in the former case, where all particles were considered. The matching between `SubFind` and stellar `SubDLe` substructures was performed as described in Section A.3.1 and the distribution of distances is shown in Fig. A.6, in orange, and compared to the distribution for the `SubDLe` run on all particles, in blue. The dashed lines define the softening lengths for

star (orange) and DM (blue) particles. As expected, the centering is slightly more accurate in terms of absolute distances, but overall the percentage of galaxies for which the distances between matched pairs is below the gravitational softening for star particles is still 91%, as for the run on all particles, with respect to the DM softening.

In total, the percentage of SubFind galaxies resolved with at least 100 star particles (corresponding to a total stellar mass of about 4×10^9 $M_\odot$), identified by the stellar SubDLe, is 93%, thus sensibly larger than in the previous case. Besides identifying a larger fraction of substructures, the stellar SubDLe performs better in reconstructing the stellar mass distribution in comparison to the reconstruction of the total mass distribution performed by the version based on all the particle species, as shown qualitatively in Fig. A.8.
In order to quantify this, I compute the ratio between the mass in star particles which both SubFind and SubDLe assign to a matched pair of galaxies, $M_{SF/SD}$, and the "true" mass. I assume the true mass to be the one given by SubFind catalogs (M_{SF}) for the purpose of this comparison. With the ratio $M_{SF/SD}/M_{SF}$ I quantify how complete the identification of a subhalo is in comparison with SubFind, in terms of its mass distribution, but not how pure it is. The mass of SubDLe's subhalo is generally smaller than the true mass, because, as already mentioned, SubDLe tends to identify only the cores of subhalos. Nevertheless, in cases similar to the one shown in the right column of Fig. A.7 which might still occasionally occur in the identification of stellar substructures, the mass of SubDLe "galaxy" is the sum of the two cores merged together. This is caused by the resolution of the grids and does not represent a conceptual flaw of this method. On the other hand, I am interested in quantifying the difference of subhalo masses caused by the intrinsic nature of SubDLe, which does not detect low-density regions in the d grids. For this reason, I remove the "over-merging" effect in the analysis of the reconstructed mass distribution, by considering only the mass of common particles in SubFind and SubDLe pairs ($M_{SF/SD}$), rather than the total mass of each (possibly composite) substructure found by SubDLe. In other words, at this stage of the analysis, I am not concerned with the "purity" of the mass distribution of SubDLe galaxies, but rather with its completeness.

The distribution of $M_{SF/SD}/M_{SF}$ is shown in Fig. A.9, for matched SubFind and SubDLe substructures in CL-2, when SubDLe is applied to all particles (in yellow) and to star particles only (in blue), in both cases considering matching SubFind substructures with at least 100 particles. The stellar SubDLe performs markedly better at identifying "complete" substructures, with a higher $M_{SF/SD}/M_{SF}$. The mean values for $M_{SF/SD}/M_{SF}$ are, indeed, 0.60 and 0.89, for the all-particle and stellar runs of SubDLe, respectively. Having established the better performance of the stellar SubDLe, I focus only on star particles in the remainder of this Section, though the general conclusions can be considered valid in both cases. I defer to future works for an equally detailed analysis of the all-particle version of SubDLe.

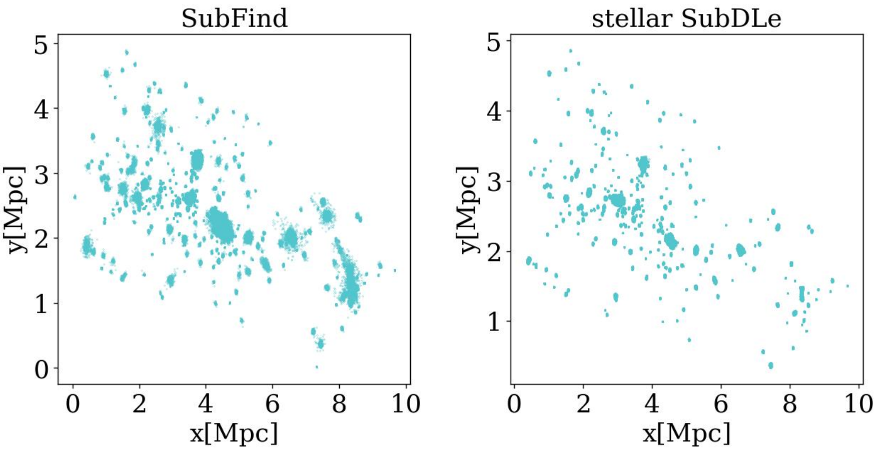


Figure A.8: Projected distribution of star particles belonging to different galaxies identified by `SubFind` (left panel) and by stellar `SubDLe` (right panel) in the cluster `CL-2` at $z = 0$.

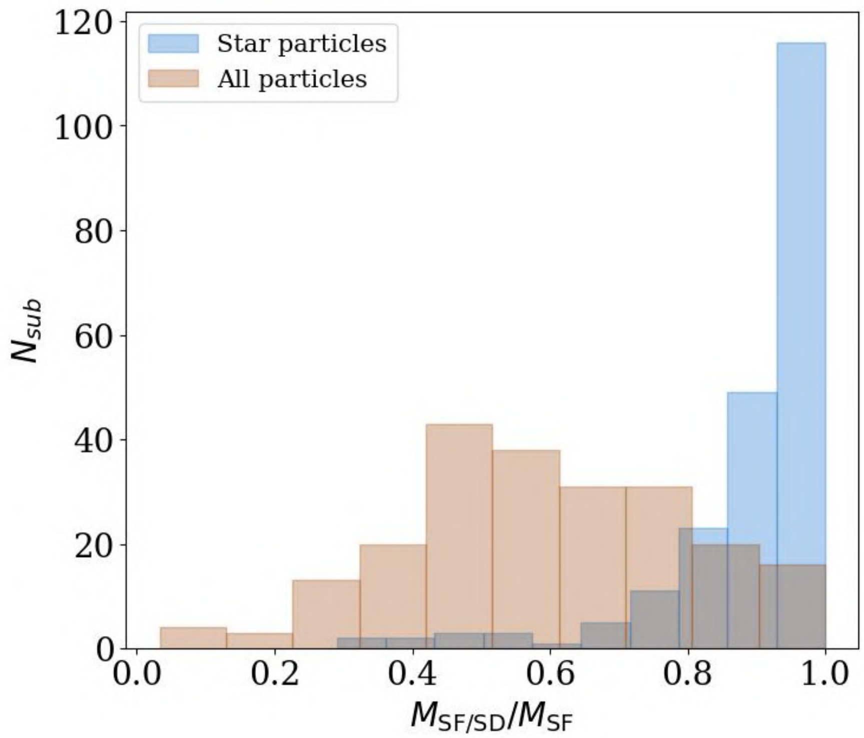


Figure A.9: Distribution of pairs of `SubFind` and `SubDLe` subhalos as a function of the ratio between the mass of common particles ($M_{SF/SD}$) and the mass given by `SubFind` (M_{SF}), as a tracer of the completeness of the mass reconstruction performed by `SubDLe`, for the run carried on the mass distribution of all particles (DM, gas, and stars), in orange, and for the run of stellar `SubDLe`, in blue, in the cluster `CL-2` at $z = 0$.

In order to further test the performance of the stellar SubDLe, I applied it to the identification of galaxies in 11 additional simulated galaxy clusters at $z = 0$ (identified as CL-[3-13] in Table A.1), spanning the virial mass range $10^{14} - 10^{15}$ $M_{\odot}$. Based on the previous analysis, SubDLe has most difficulty in recovering large (in comparison with the d grid size) substructures and in separating close (in comparison with the d grid resolution) ones. While the former does not compromise SubDLe's performance as a peak locator, the latter does. I expect galaxies in clusters to be closer on average than their counterparts in the field or in small groups, as per definition of cluster of galaxies. Thus the cluster environment that I probed in this work is the most challenging scenario for SubDLe.

The cluster analyzed so far, CL-2, is among the smallest in the sample. To analyze the most massive clusters, however, I had to address the significant memory usage of the d grids. For this reason, when necessary, I modify the sampling strategy of the cluster region, with overlapping d grids to avoid missing information at the borders, opting instead for a sort of sparse sampling. To this purpose, I build coarse d grids, each encompassing the whole cluster, and thus spanning regions of few megaparsecs per side (the exact size depending on the cluster), with a resolution of 640 kpc. I then define a threshold M_{th} for the mass assigned to each of those coarse pixels through the NGP scheme, below which the pixel is excluded from the analysis. In pixels assigned with a mass $M_{NGP} > M_{th}$, a 64^3 d sub-grid is built with 10 kpc resolution. In this analysis, the size of the clusters and the memory resources allowed to set $M_{th} = 0$, thus only empty regions are excluded. On the other hand, this simple scheme does not allow to analyze a given (higher-resolution) d grid of the cluster more than once, as in the previous case with overlapping d grids. If memory constraints require setting $M_{th} > 0$, the identification of substructures crossing the borders with a sub-grid excluded from the sampling will be affected, as only a fraction of the substructure will be processed by SubDLe. Memory usage is a limiting factor for SubDLe, but this is expected to be less problematic when running on multi-node high-performance computing systems with substantially larger accessible RAM.

Figure A.10 shows the results for all clusters analyzed at $z = 0$ (CL-[2-13]) in terms of $M_{SF/SD}/M_{SF}$; the distribution is consistent with that shown for the single cluster CL-2 in Fig. A.9, with the same mean value of 0.89 (black dashed line, the chosen y-scale here is logarithmic, due to the much higher number of galaxies when considering 12 clusters), significantly larger than the mean value for the all-particles SubDLe run (orange dashed line), 0.60. Table A.2 reports the percentages of galaxies identified by SubFind, which are detected by the stellar SubDLe (completeness) in the 11 clusters CL-[3-13], along with the percentage of the stellar SubDLe galaxies which have a SubFind counterparts (purity), and the wall-clock time requested by SubDLe runs. The latter includes the propagation of d grids through the network to recover the s grids, but does not take into account the time required for building the d grids, which might be already available during

Table A.2: Results of identification of galaxies in 11 simulated galaxy clusters analyzed with stellar SubDLe. The first column reports the cluster ID, the second and third ones show the completeness and purity of the stellar SubDLe catalogs, in comparison with SubFind, and the fourth one reports the wall clock time (in seconds) taken by SubDLe for the processing of the d grids.

ID	Completeness	Purity	SubDLe time [s]
CL-3	0.80	0.89	53
CL-4	0.78	0.86	21
CL-5	0.88	0.82	53
CL-6	0.84	0.88	41
CL-7	0.75	0.65	74
CL-8	0.81	0.86	105
CL-9	0.78	0.91	49
CL-10	0.84	0.92	146
CL-11	0.82	0.86	36
CL-12	0.83	0.90	152
CL-13	0.79	0.82	59

the evolution of a simulation (e.g., if gravity is solved through a Particle-Mesh approach) for stars, nor the final FoF run on pixels, which should be negligible anyway due to the relatively small number of active pixels in each s grid.

The average completeness of the stellar SubDLe, weighted by the number of resolved galaxies in each cluster (reported in the fourth column of Table A.1), is 0.82 in the set of 12 clusters at $z = 0$ (including CL-2), while the weighted average purity is 0.88 (including CL-2, in which SubDLe reported a purity of 0.82). The purity is sufficiently high, even higher than the completeness, meaning that the bulk of SubDLe galaxies have been identified by SubFind. SubFind does not necessarily represent the ideal identifier, though I assume its catalogs as ground truth in the framework of a "supervised" learning.

Nevertheless, I briefly investigate the nature of the "spurious" identification (in comparison with SubFind resolved substructures) of SubDLe galaxies. I relax the resolution limit for SubFind galaxies to a minimum of 20 star particles and repeat the stellar SubDLe and SubFind matching process, gaining an increase in weighted average purity from 0.88 to 0.96, confirming that almost all of the SubDLe galaxies correspond to gravitationally bound structures found by SubFind. At the same time, the completeness has decreased from 0.82 to 0.65, as expected since the stellar SubDLe might very well be able to identify substructures of very few stars where there is close to zero background, while this is more challenging in the denser central regions of clusters. I focus on how the galaxy position in the cluster affects the completeness in Sect. A.3.3.

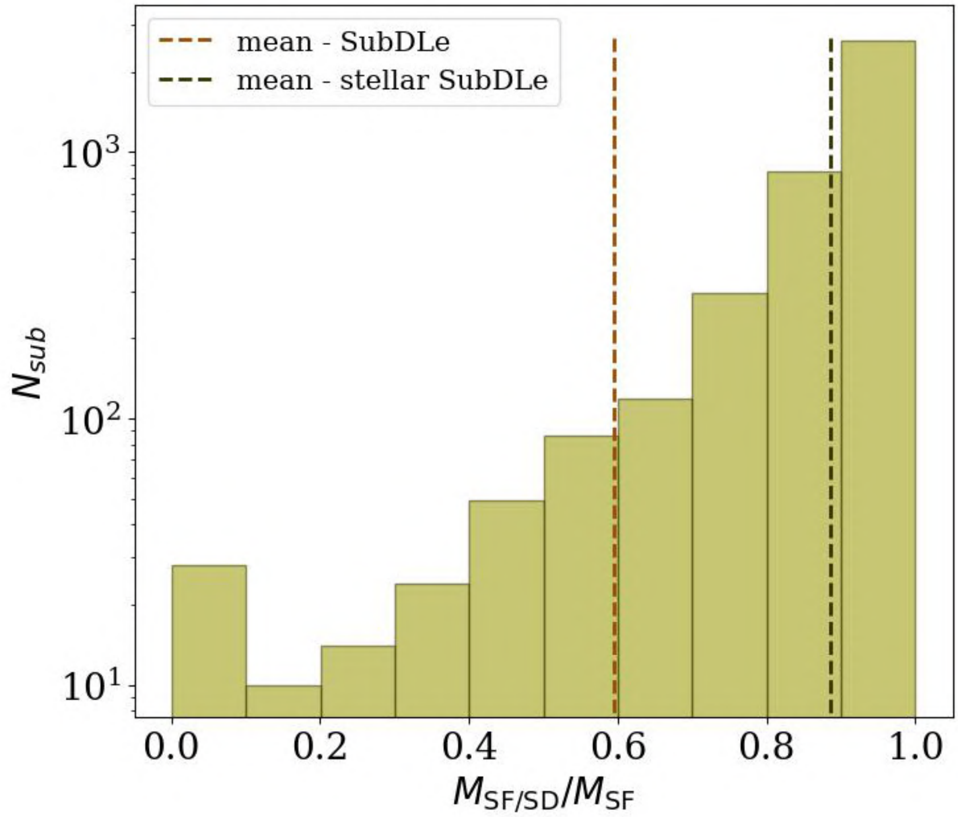


Figure A.10: Distribution of the ratio $M_{SF/SD}/M_{SF}$ in all galaxies identified by the stellar SubDLe and matched to a SubFind substructure for all the 12 clusters analyzed at $z = 0$. The mean value, across all 12 clusters, is traced by the black dashed line, while the orange dashed line represents the mean value for the all-particle run of SubDLe on cluster CL-2 (associated with the orange histogram shown in Fig. A.9).

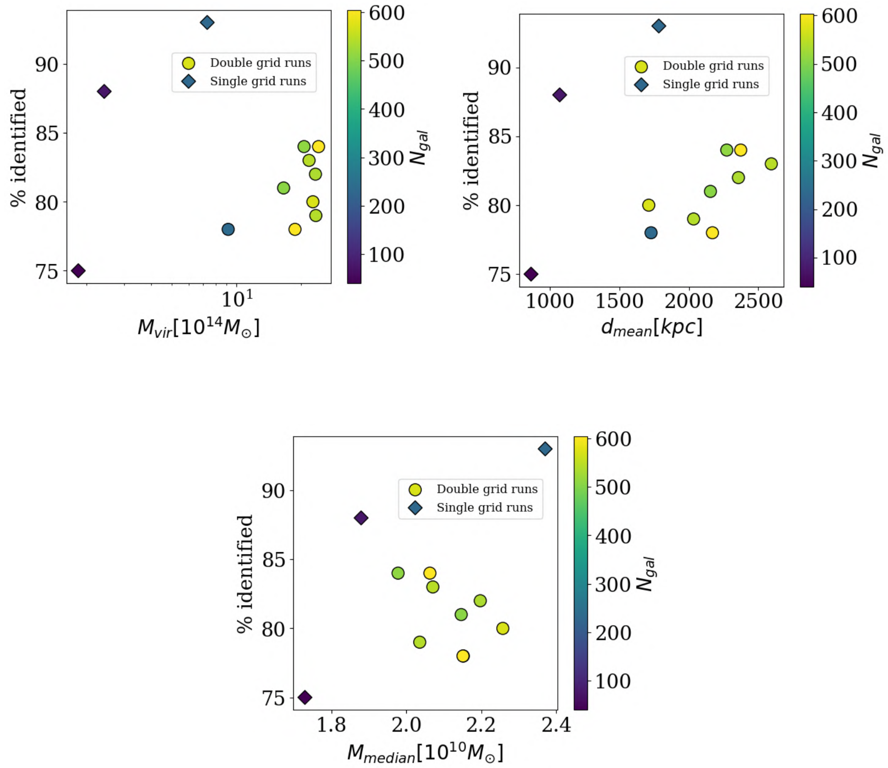


Figure A.11: Scaling of the percentage of resolved SubFind galaxies in the 12 clusters analyzed at $z = 0$ (CL-[2-13]), identified by the stellar SubDLe, with cluster properties: total virial mass (upper panel), mean inter-galaxy separation within one virial radius (middle panel), and median galaxy stellar mass (lower panel), color coded with the total number of SubFind galaxies (the ones deemed resolved enough, with at least 100 star particles). The circles represent clusters that required a treatment with sparsely distributed sub-grids (with $M_{th} = 0$) and the diamonds represent the three smaller clusters that did not require it.

The wall-clock time taken by SubDLe to perform galaxy identification in these clusters, shown in the fourth column of Table A.2, is quite interesting to assess whether SubDLe is suitable for frequent on-the-fly identifications of galaxies in cosmological simulations. With the above mentioned caveats about possible overheads, all the analyses required less than 3 minutes per cluster, which represents a remarkable improvement with respect to SubFind. For reference, running SubFind on one of such clusters required about half an hour on one node of the Marconi-100 machine at the CINECA Supercomputing Center[4].

A.3.3 *Testing performance against galaxy properties and environment*

To gain a clearer understanding of the performance of the stellar SubDLe, I investigate the origin of the cluster-to-cluster variations in the completeness

4 https://www.hpc.cineca.it/systems/hardware/marconi100/

of the SubDLe galaxy catalogs and explore possible correlations with cluster properties.
In this analysis, I include also cluster CL-2, which yielded the highest (93%) percentage of identified resolved galaxies. This higher value might be due to the fact that CL-2 has very similar properties to CL-1, which was used for the training of SubDLe, pointing to a slight over-fitting of the model. To investigate this, I look into possible correlations between the completeness of the stellar SubDLe catalogs and the properties of clusters in Fig. A.11. This shows the completeness of the identification in the 12 clusters at $z = 0$ as a function of the total virial mass (top-left panel), the mean inter-galaxy distance for SubFind galaxies within the virial radius (top-right panel) and the median galaxy stellar mass (bottom panel). The circles in the plot show clusters that required a sampling with sparsely distributed sub-grids (with $M_{th} = 0$, as described above) and the diamonds show the three smaller clusters that did not require such treatment. I note that clusters with a lower statistics of galaxies (the three diamonds, CL-[2,5,7], and the darker green circle, CL-4) appear to be more randomly scattered (thus disproving the possibility of over-fitting for less massive clusters), with an approximately increasing completeness with larger cluster mass, while the clusters with a larger number of galaxies follow a tighter locus with a possible dependence upon the total virial mass of the clusters, suggesting that the identification might be less complete in less massive clusters. The analyzed clusters, though, span a small range of virial masses (see Table A.1), so it is difficult to draw conclusions on the environment dependence of the model performance from this set of simulations.
There is an apparent correlation with the mean inter-galaxy distance, with two of the diamonds (CL-2 and CL-5) acting as outliers. Nevertheless, the statistical significance of these two clusters in defining a distinct locus of higher completeness for clusters identified with a different spatial sampling, is pretty low, in consideration also of the low completeness of CL-7 (0.75), which was analyzed in the same way. The general trend of increasing completeness with increasing inter-galaxy separation can be understood as the larger the separation, the lower the probability is for SubDLe to merge two substructures thus reducing the completeness.
On the other hand, the weak apparent anti-correlation between the completeness and the median galaxy stellar mass shown in the bottom panel for the bulk of clusters (here CL-4 is not visible because hidden behind one of the yellow circles) is less intuitive. This could be interpreted as an indirect effect of the correlation shown in the top-right panel, since a larger fraction of massive galaxies might be associated with a larger fraction of close galaxies on their way to merge, thus with a small separation. In general, the completeness of SubDLe identification does not seem to show a strong dependence upon global properties of the galaxy population in the analyzed simulated clusters.

One further step can be done to isolate the factors affecting the model performance, by studying the completeness as a function of SubFind stellar masses and distance from the center of the cluster (defined as the position of the most bound particle in the halo). This is shown in Fig. A.12; the top-left panel shows the percentage of SubFind galaxies identified by the stellar SubDLe as a function of SubFind stellar mass, for all the clusters analyzed at $z = 0$ (CL-[2-13]), in solid line, for the ones analyzed with a double grid (with $M_{th} = 0$, CL-3-4,6,8-13), in dashed line, and the ones analyzed with a single grid, in dotted line. As already discussed, the latter have lower statistical significance, with a total of 346 galaxies (see Table A.1). The completeness, for all clusters, increases with stellar mass, with an apparent difference below $\sim 6 \cdot 10^9$ $M_\odot$ for galaxies belonging to clusters processed with a single grid. Owing to the relatively poor statistics in this sample and the proximity to the resolution limit, I do not deem this difference significant. The increasing completeness with stellar mass is in line with the expectation that the most massive galaxies are better resolved against the background.

The top-right panel shows the completeness as a function of the cluster-centric radius, expressed in units of the virial radius, with the same distinction between different lines adopted in the top-left panel. As expected, I see an improvement in performance the farther the galaxies are from the over-crowded central regions of the clusters, up to $\sim 0.2 - 0.3$ R_{vir}, followed by a saturation at $\sim 80\%$ completeness in the outer regions. The dependence of completeness on the distance from the center is much stronger than what is observed with the stellar mass (see the different scales in the y-axis of the top-left and top-right panels), at least in the innermost regions, and is ubiquitous among galaxies processed in different ways. In any case, a larger sample of simulations uniformly sampling the halo mass range from groups to rich clusters would be required to draw a picture of how the external environment of galaxies affects the stellar SubDLe's performance.

As a further step, I try to disentangle the dependence of completeness on the stellar mass and on the distance from the center of the cluster, by separating galaxies in radial and stellar mass bins (bottom panels of Fig. A.12). The bottom-left panel shows the completeness of identification of galaxies in 5 radial bins, as a function of stellar mass. The bottom-right panel shows the completeness in 4 stellar mass bins, as a function of distance from the cluster center. Here, in order to provide a sufficient number of galaxies in each bin, I include all the clusters analyzed at $z = 0$. The separation in radial bins shows that the stellar-mass dependence is the strongest in the most central region (below 0.25 R_{vir}), where the overdense environment affects more significantly the identification of lower-mass galaxies. This is confirmed by the bottom-right panel, which shows an approximate hierarchy in the curves describing the completeness as a function of cluster-centric radius for different mass bins. In the central region of the clusters (below ~ 0.25 R_{vir}), the completeness increases with the distance from the cluster center and the curves representing less massive galaxies are lower (besides stochastic fluctuations due to small-number statistics). In conclusion, the main limitation of the stellar SubDLe is in the identification of small galaxies within the crowded core regions of galaxy clusters.

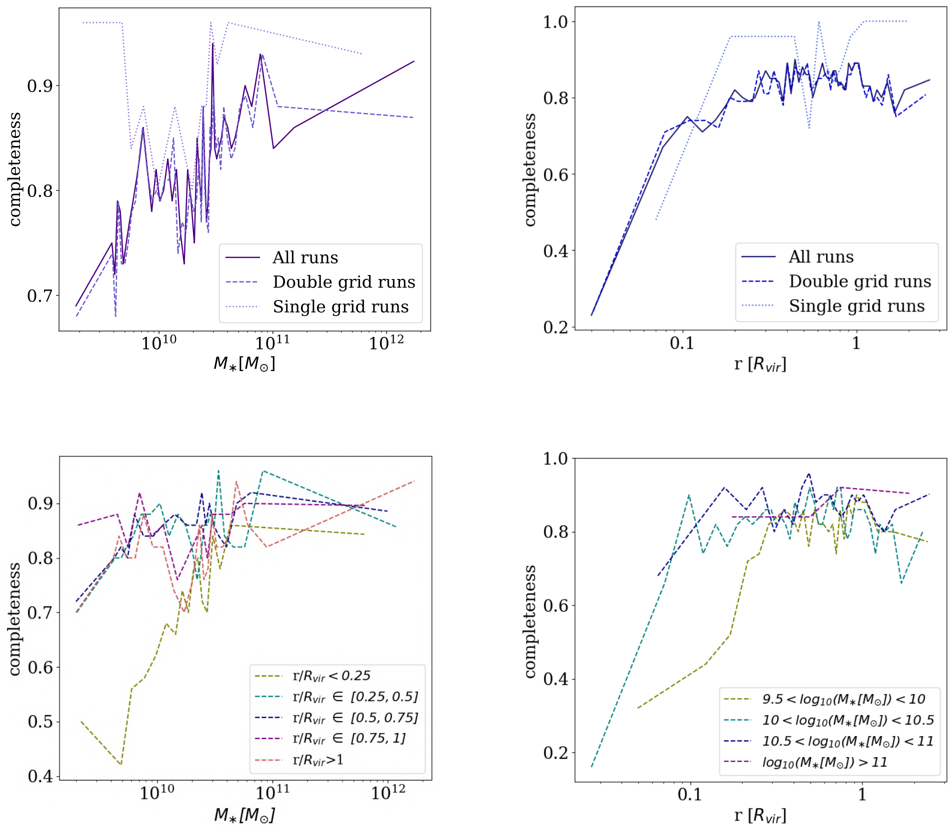


Figure A.12: Dependence of the stellar SubDLe completeness (defined as the fraction of SubFind galaxies identified by the stellar SubDLe) on SubFind stellar masses (left panels) and on 3D distance from the cluster center (in units of virial radii, right panels), for all the clusters analyzed at $z = 0$ (see Table A.2). In the top panels I show, with solid lines, the results for all the analyzed clusters, while dashed and dotted lines represent clusters analyzed with a single grid and a double grid ($M_{th} = 0$), respectively. In the bottom-left panel different colors of the dashed curves correspond to different radial bins. In the bottom-right panel different colors correspond instead to different intervals of the stellar mass of the galaxies.

Finally, I compare the distribution of stellar masses produced by SubFind and by the stellar SubDLe in clusters at $z = 0$ in Fig. A.13. The figure shows the median counts among different clusters, with error bars corresponding to the 25th-75th percentile ranges, in blue for SubDLe and in red for SubFind. In the low-mass end where the resolution limit of the simulation dominates the shape of the distribution, the SubDLe stellar mass distribution is slightly lower due to missing identifications and SubDLe galaxies being smaller than SubFind counterparts. In the well resolved high-mass end, though, there is broad agreement in shape and normalization among the two distributions.

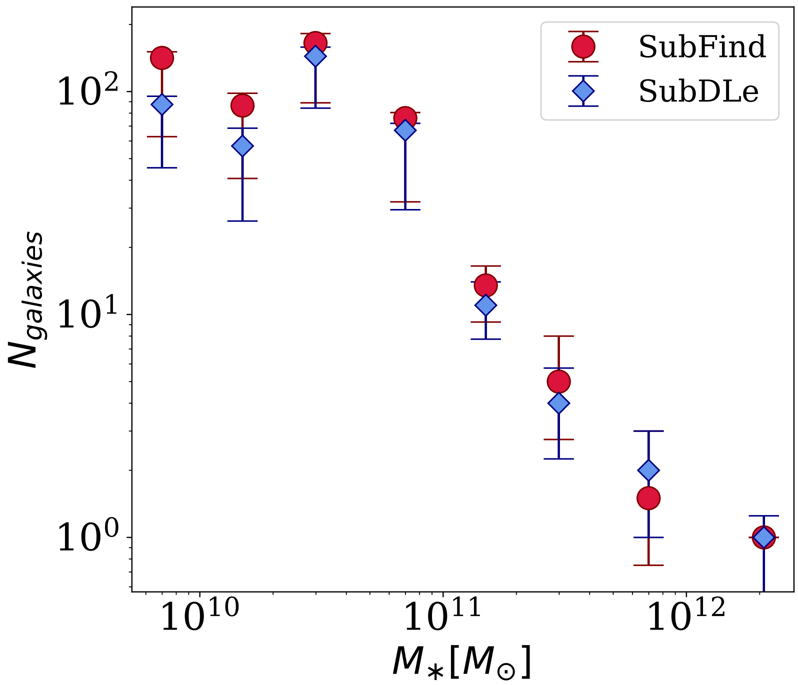


Figure A.13: Comparison of the distributions of stellar masses in the simulated galaxy clusters at $z = 0$, as predicted by the subhalo catalogs produced by `SubFind` (red dots) and by the stellar `SubDLe` (blue diamonds). The points show the median counts in each stellar mass bin among the 12 different clusters, while the error bars mark the 25th and 75th percentiles.

A.3.4 *Testing the robustness of the stellar* `SubDLe`

I now focus on testing the robustness of the stellar `SubDLe` on galaxy cluster `CL-13` at $z = 1$, using the model trained at $z = 0$. At $z = 1$ this cluster has a virial mass $M_{vir} \sim 7 \times 10^{14}\ M_{\odot}$. In principle more dynamically disturbed clusters are expected at $z = 1$, which could impact the performance of `SubDLe` in identifying galaxies. This complements the previous tests in assessing the capability of the model to extrapolate outside the training framework, which is always a primary concern in developing machine learning models. Here, with the uniform sampling of the cluster volume, I extract 1287 64^3 d grids with a resolution of 10 physical kpc. The percentage of resolved galaxies identified is 77%, consistent with the 79% percent of completeness obtained at $z = 0$ for the same cluster and with the average performance achieved by stellar `SubDLe` with a 10 kpc resolution at $z = 0$. This confirms the robustness of the method also when different evolutionary stages of galaxy clusters are considered.

As a last test, I assess the performance of the stellar `SubDLe` on higher resolution d grids. I perform this experiment on the cluster `CL-13` at $z = 1$. The stellar `SubDLe` is trained on galaxies of different sizes, so I expect it to be able to identify galaxies even when using a higher-resolution d grid, with respect to which their sizes will be larger. In general, I expect this method to be able to identify galaxies that were already identified at lower resolution. To increase resolution, I choose a grid size of 2 kpc and, due to memory constraints, perform a sparse sampling of the cluster volume as described before, still keeping $M_{\mathtt{thr}} = 0$. By doing this, I expect to: *(i)* increase the sensitivity of the network to small substructures; *(ii)* separate composite systems made up of close galaxies artificially merged by `SubDLe` due to the poor relative resolution of d grids in the denser regions. In general, I expect that the higher resolution d grids will increase the percentage of `SubFind` resolved galaxies found by the stellar `SubDLe`. On the other hand, by increasing the resolution, the average mass per grid pixel decreases, resulting in a more noisy discontinuous NGP density grid. This in turn might impact the model's performance, since CNNs are efficient on input data with strong spatial dependencies. Moreover, by reducing the pixel size, while keeping fixed the number of pixels in each grid, the whole d grid encompasses a smaller volume (128 kpc per side), which I expect to have a negative impact on the recovered spatial extent of the substructures.

The results of this test are shown in Fig.A.14. In the left panel I show the distribution of $M_{\mathtt{SF/SD}}/M_{\mathtt{SF}}$ for `SubFind` and stellar `SubDLe` matches, for the run at lower (10 kpc, orange histogram) and higher (2 kpc, blue histogram) resolution. The mean values vary from 0.9 at 10 kpc resolution (consistent with the mean value at $z = 0$), to 0.7 at 2 kpc resolution. As expected, the completeness of the galaxy mass distribution recovered by the 2 kpc resolution run of the stellar `SubDLe` is slightly lower. The most interesting aspect is though the increased percentage of identified peaks in the stellar mass density distribution when running the stellar `SubDLe` at higher resolution,

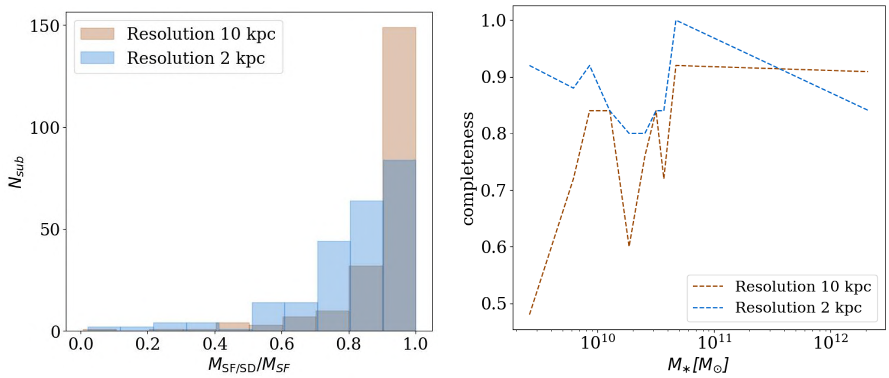


Figure A.14: Comparison of the performance of SubDLe on the same cluster, with different d grid resolutions. Left panel: Distributions of values of $M_{SF/SD}/M_{SF}$ for SubDLe galaxies matched to SubFind galaxies. Results are shown for the cluster CL-13 at $z = 1$, for SubDLe runs with the standard resolution of 10 kpc (in orange) and with the higher resolution of 2 kpc (in blue). Right panel: stellar SubDLe completeness as a function of stellar mass in cluster CL-13 at $z = 1$, for 2 kpc (in blue) and 10 kpc (in orange) resolution.

which proves that the model can reach peak performance by simply increasing the resolution. In fact, the percentage of SubFind resolved galaxies identified by the stellar SubDLe on cluster CL-13 at $z = 1$ increased from 77 to 87% when increasing the grid resolution. This is shown in the right panel where the completeness is plotted as a function of stellar mass (blue and orange lines for 2 and 10 kpc resolution, respectively). This figure highlights that the higher resolution adopted in the d grids boosts the performance of SubDLe mainly by recovering a larger fraction of low-mass galaxies, the ones that SubDLe struggles to disentangle from the background and from other close substructures in the central region of clusters, at lower resolution. This further demonstrates the effectiveness of SubDLe in detecting substructures within the simulated distribution of stars.

A.3.5 *Testing with a larger training set*

As already emphasized, I expect an improved performance from SubDLe once it is trained on a much larger data set on more performant computing facilities. In order to explicitly show that the model's performance is not saturated, I carry out a training on a data set which is twice as large as the original one, extracted from clusters CL-3 and CL-8, which is still feasible in the adopted computing framework. Then, I test the model on the remaining clusters at $z = 0$ and compare the results with the previously trained model, for those clusters that are not part of the training data set in neither models.

The results of this test are summarized in Figure A.15, where completeness and purity reached by the newly trained model are shown for each test cluster, against the corresponding quantities obtained from the version of

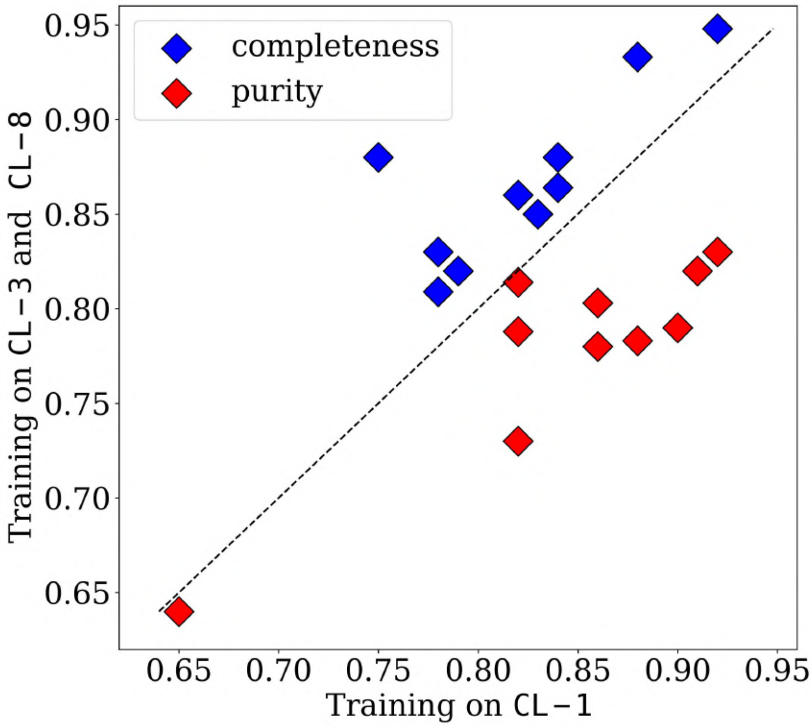


Figure A.15: Completeness (blue diamonds) and purity (red diamonds) of stellar `SubDLe` identifications, referred to `SubFind` catalogs, for galaxies in $z = 0$ clusters. The y-axis refers to metrics calculated on the catalog given by the model trained on two clusters (`CL-3` and `CL-8`), while the x-axis refers to results of the original stellar `SubDLe` trained on one cluster. The dashed line represents cases where the metrics are identical in the catalogs produced by the two models.

the model trained on a single cluster. I note that there is a systematic improvement in completeness (blue diamonds), with the overall completeness (weighted with the number of resolved galaxies in each cluster) increasing from 0.82 to 0.86. On the other hand, the purity is systematically lower, meaning that the new version of `SubDLe` identifies more galaxies which are either missed by `SubFind` or fall below the resolution limit. This improvement is quite promising in the perspective of a future use of a more extended training data set, spanning possibly different redshifts, implementations of the galaxy formation model and numerical resolutions, which I aim at carrying out in future developments of this work.

A.4 SUMMARY AND CONCLUSIONS

In this Appendix, I have presented the first application of deep learning techniques to perform identification of substructures in cosmological hydrodynamical simulations, with the goal of providing a much faster alternative to procedural algorithms typically used for this purpose (see Section A.1 for a brief review of some popular subhalo identifiers). Substructure identification represents a fundamental step to extract useful information from the raw output of cosmological hydrodynamical simulations and is typically performed by algorithms that, with different approaches, search for groups of gravitationally bound particles at the peaks of the density field sampled by the simulations, within a larger structure. This is computationally costly and results in long times-to-solution that are unfit for frequent identifications of galaxies during the evolution of simulations. The on-the-fly identification of substructures with a fine time cadence, on the other hand, could have interesting applications in the implementation of sub-resolution models of, for example, star formation and feedback, when they are parametrized on global properties of galaxies, such as their stellar masses.

Deep learning (and machine learning in general) consists of statistical methods to approximate solutions to many classes of problems and offers the advantage of typically short computational times. I used the architecture of an efficient fully convolutional network, `U-Net` (Ronneberger et al. 2015), developed for pixel-by-pixel identification of objects in 2D images, and extended it to the task of identifying 3D substructures in simulations.
In this work, I presented `SubDLe`, a substructure finder which incorporates this 3D generalization of `U-Net`. I analyzed the `DIANOGA` simulations (see Section A.2), a set of zoom-in cosmological hydrodynamical re-simulations (Tormen 1997) of galaxy clusters. As the ground truth for training `SubDLe`, I used the subhalo catalogs extracted from these simulations by the subhalo finder `SubFind` (Springel et al. 2001b; Dolag et al. 2009). I applied `SubDLe` to 3D Nearest Grid Point mass density grids extracted from the set of simulations, splitting the whole cluster regions in 640 kpc per side sub-grids with 10 kpc resolution (d grids). Twin grids (s grids) reporting a label "1" at grid meshes containing a majority of substructure-bound particles (and "0" elsewhere) were also produced for the training, based on `SubFind`'s catalogs.

I trained `SubDLe` on a NVidia Tesla P100 GPU on the Kaggle platform on a large number of representative d and s grids. The test on a different cluster, after matching `SubDLe` and `SubFind` subhalos, resulted in a completeness of the output subhalo catalogs of 0.80 with respect to `SubFind`. I then focused on the identification of galaxies, defined as concentration of star particles. As such, galaxies provide much more contrasted substructures against the background field, thus facilitating their identification performed by `SubDLe`. I refer to this newly trained version as "stellar `SubDLe`," to distinguish it from the previous version based on the density traced by all particles. The same test cluster analyzed with the standard `SubDLe` yielded a completeness of 0.93 with stellar `SubDLe`. Furthermore, averaging over a sample of 12 simulated clusters at $z = 0$, it reached 0.82 in completeness and an even higher purity of 0.88 (most of the spurious identifications actually correspond to under-resolved `SubFind` galaxies which were excluded from the input catalog). The average `SubDLe` time for processing one cluster is 72 seconds, much faster than the typical execution time of `SubFind`, thus matching the requirement for short computational times. In the estimate of the execution time I did not consider the time necessary to build mass density grids and the FoF run on grid meshes (see Section A.2). I expect the FoF runs to add a negligible overhead, while mass grids are typically already built during a simulation, for example, for gravity calculations.

I tested the robustness of stellar `SubDLe` by repeating the identification on one of the clusters analyzed at $z = 0$, at a different evolutionary stage, at $z = 1$, obtaining a completeness comparable to the result at $z = 0$. Furthermore, I tested the effect of increasing the resolution of the d grids by reducing the grid pixel size from 10 kpc to 2 kpc. This led to an increase of the percentage of `SubFind` galaxies that are identified by this method from 77 to 87%. This is due to the reduction of close substructures that were artificially merged together while adopting 10 kpc resolution grids. On the other hand, the d grids are 5 times smaller, which slightly impact the size of `SubDLe` galaxies in comparison with the lower resolution runs.

I found that the stellar `SubDLe` performs better than the all-particles version, not only in overall completeness, but also in reproducing the total mass of substructures. The most challenging aspect for the stellar `SubDLe` is the identification of galaxies in the innermost part of the clusters. I showed that the fraction of `SubFind` galaxies identified by `SubDLe` is decreasing toward the center of each cluster, within about 0.25 virial radii, especially for lower-mass galaxies (below few times 10^{10} $M_\odot$). The smaller galaxies are indeed easier to miss in the high-density cores of clusters, due to smaller separations and higher level of background density. As a consequence, I expect the stellar `SubDLe` to perform even better in different environments (groups and average fields).

In future developments of this study, I plan to perform a full assessment of how the stellar `SubDLe` performs in different environments, with different resolution and even different sub-resolution physics, which might affect the distribution of star particles. As far as the improvement of the algorithm is concerned, I aim to further refine it by taking advantage of high-performance computing nodes, where multiple GPUs can access a larger shared memory. This would largely relax the constraints I had to respect in this exploratory study, in terms of size and resolution of the grids that sample the density field. This will allow tests on how increasing the physical extent of d grids, while keeping the resolution fixed, can influence the mass distribution of the largest substructures which are currently cropped by the stellar `SubDLe`. I will be able to train on much larger data sets, which is expected to improve the performance of the model, as shown in the tests presented in Sect. A.3.5. I also plan on testing different mass assignment schemes, in order to enhance the contrast between substructures and the background field in the d grids, which could improve the completeness of the mass distribution reconstructed by `SubDLe` for each galaxy. As a further step, given the demonstrated improvement of completeness in the identifications with a higher resolution of the d grids, a possible development of this work would be a multi-resolution adaptive implementation of `SubDLe`, to zoom on higher-density regions in the simulation box and increase the resolution of the d grids, while keeping the average regions at a reasonable resolution.

This would require to access adaptively refined meshes to sample the mass density field in the simulation, which in many cases are already built by the gravity solvers of the simulation code. Another possible refinement might be the use of multiple channels in the input grids, including information on particle kinematics to improve the separation of close substructures.

In conclusion, I have developed SubDLe, a fast algorithm based on deep learning, capable of identifying a large fraction of galaxies in simulations of galaxy clusters, while its performance is expected to be even better within less dense environments. The final step, besides its possible improvements mentioned above, will be to integrate it into a cosmological code and test it for highly frequent, on-the-fly galaxy identifications.

B

BASICS OF COSMOLOGY

This Appendix summarizes some basics concepts of cosmology, starting with the description of the homogeneous expanding Universe (Sect. B.1) and then introducing cosmic perturbations and their statistical properties (Sect. B.2)

B.1 HOMOGENEOUS EXPANDING UNIVERSE

Modern cosmology is based on the evidence of the expansion of the Universe, expressed through Hubble-Lemaître law, and the *cosmological principle*.
The former is expressed through the observed relation between the receding velocity of galaxies v and their distance d, originally inferred through the period-luminosity relation of Cepheids in the original diagram by Hubble (1929) and from type Ia Supernovae more recently (see Fig. B.1):

$$v \approx H_0 d, \tag{35}$$

where H_0 is Hubble constant.
The cosmological principle states that there are no privileged observers in the Universe, which, combined with the observed isotropy of the Cosmic Microwave Background (CMB) radiation, implies the homogeneity of the Universe (on large scales)[1].
As a consequence, there is a class of *preferred* reference frames in cosmology, those for which the measured temperature map of the CMB is isotropic. In such frames, with the further request that the Universe is locally flat, the metric can be written in the form of *Robertson-Walker (RW) metric*:

$$ds^2 = -c^2 dt^2 + \frac{\mathcal{R}^2(t)}{\mathcal{R}_0}^2 \left[\frac{dr^2}{1 - \frac{k}{\mathcal{R}_0^2} r^2} + r^2 d\Omega^2 \right], \tag{36}$$

where $\mathcal{R}(t)$ is the *dimensional scale factor* at epoch t, whose evolution depends on the content of the Universe through Einstein's field equations, $\mathcal{R}_0 \doteq \mathcal{R}(t_0)$, where t_0 is the present epoch, t and r are temporal and radial coordinates in the chosen frame and $d\Omega^2$ is the element of solid angle. The RW metric is usually expressed with the dimensionless scale factor:

$$a(t) \doteq \frac{\mathcal{R}(t)}{\mathcal{R}_0}.$$

k is the curvature constant, dimensionless, that can be re-scaled to assume the values $\{-1, 0, +1\}$, corresponding to an open, flat or closed geometry of

1 Provided that the metric has an analytic solution.

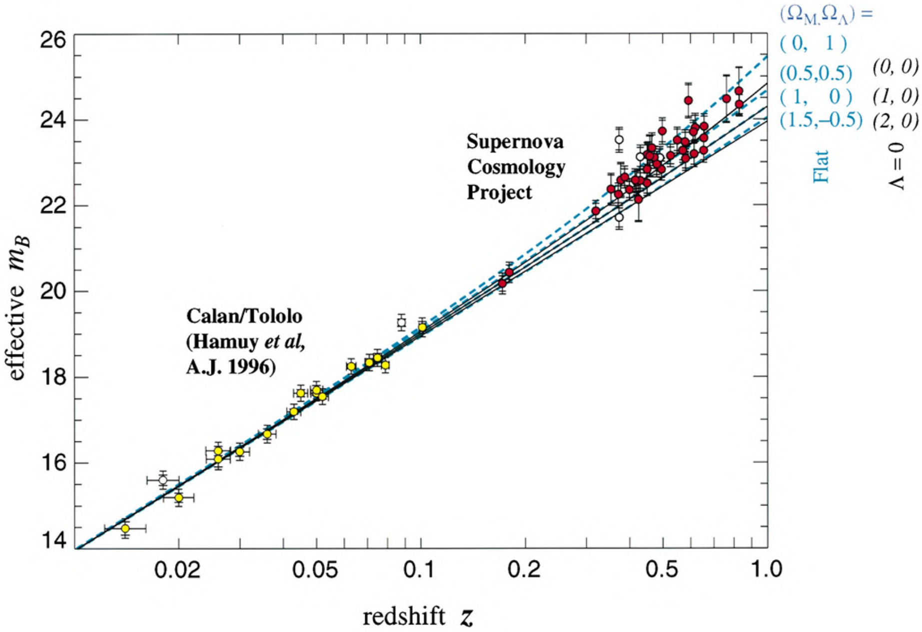


Figure B.1: Hubble diagram, obtained with 42 high-redshift type Ia Supernovae from the Supernova Cosmology Project and 18 low-redshift type Ia Supernovae from the Calán/Tololo Supernova Survey. Dashed and dotted lines correspond to different cosmological models, with the corresponding parameters reported at the top-right corner. Reproduced from Perlmutter et al. 1999.

the Universe.
A useful re-formulation of Eq. 36 can be obtained by introducing the distance along the line-of-sight (i.e. the radial direction from the origin of the reference frame) L_{los} and the distance projected on the sky L_{sky}, such that:

$$ds^2 = -c^2 dt^2 + a^2(t)\left[dL_{los}^2 + L_{sky}^2 d\Omega^2\right]. \tag{37}$$

Both L_{sky} and L_{los} are *comoving* distances, which means that they are defined along a path at the present epoch (i.e., for $a = 1$).

B.1.1 *Friedmann equations*

As already mentioned, the evolution of a as a function of the Universe content is given by Einstein's equations. In a homogeneous isotropic Universe, assuming it to be filled with an ideal fluid with matter density ρ and pressure p, they yield Friedmann equations:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3\frac{p}{c^2}) + \frac{\Lambda c^2}{3}, \tag{38}$$

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho - \frac{kc^2}{\mathcal{R}_0^2 a^2} + \frac{\Lambda c^2}{3}, \tag{39}$$

where G is the gravitational universal constant and Λ is the so-called *cosmological constant*, first introduced by Einstein to achieve a static solution for the evolution of the Universe. The third Friedmann equation is derived from the conservation of energy and momentum:

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\frac{p}{c^2} + \rho\right). \tag{40}$$

For a flat universe without a cosmological constant ($k, \Lambda = 0$), $\rho = \frac{3H(t)^2}{8\pi G}$; this allows for the introduction of the *critical density*

$$\rho_c(t) \doteq \frac{3H(t)^2}{8\pi G} \tag{41}$$

and the *density parameter*

$$\Omega(t) = \frac{\rho(t)}{\rho_c(t)}. \tag{42}$$

From Friedmann second equation (39), it follows that for a closed Universe $\Omega(t) > 1$, i.e. $\rho > \rho_c$, and for an open Universe $\Omega(t) < 1$, i.e. $\rho < \rho_c$. Hence, the critical density separates closed and open Universes.
Introducing the Hubble parameter:

$$H(t) \doteq \frac{\dot{a(t)}}{a}, \tag{43}$$

Friedmann second equation can be re-written in the more common form:

$$\frac{H(t)^2}{H_0^2} = \Omega_0\frac{\rho}{\rho_0} + (1 - \Omega_0)a^{-2}. \tag{44}$$

H_0, the already mentioned Hubble constant, is Hubble parameter evaluated at the present epoch and $\Omega_0 = \Omega(t_0)$.
Specifying the different components of the Universe, i.e. matter (with density parameter Ω_m), radiation (Ω_r) and Λ (Ω_Λ), Eq. 44 can be expressed as:

$$E^2(t) \doteq \frac{H(t)^2}{H_0^2} = \Omega_m(t_0)a^{-3} + \Omega_r(t_0)a^{-4} + \Omega_\Lambda + \Omega_k(t_0)a^{-2}, \tag{45}$$

where $\Omega_k = 1 - \Omega_m - \Omega_r - \Omega_\Lambda$ is a curvature term, equal to 0 if the Universe is flat.
The standard cosmological model, supported by multiple observations, is the ΛCDM (Λ Cold Dark Matter) model, which describes the Universe as flat and dominated by a *Dark Energy* (with a density parameter $\Omega_\Lambda \sim 0.68$) and a dynamically cold non-baryonic matter (Cold Dark Matter, $\Omega_{DM} \sim 0.27$).

B.2 STATISTICAL PROPERTIES OF THE PERTURBATIONS

In this Section, I introduce perturbations in the density field and describe their statistical.

The perturbation field (or density fluctuation) at the comoving position $\vec{x}$ and cosmic time t as:

$$\delta(\vec{x}, t) = \frac{\rho(\vec{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \tag{46}$$

where $\rho(\vec{x}, t)$ is the density field at $\vec{x}$ and at the epoch t and $\bar{\rho}(t)$ is its average over all possible positions.

Let us consider a partition of the "volume" of the Universe in cells centered at positions $\vec{x}_i$, for $i \in [1, n]$. A cosmological model predicts a multivariate probability function $\mathcal{P}$ such that $\mathcal{P}(\delta_1, \delta_2, ..., \delta_n))d\delta_1 d\delta_2 ... d\delta_n$ is the probability that the perturbation field δ has a value in the infinitesimal interval $[\delta_i, \delta_i + d\delta_i]$ at position $\vec{x}_i$, for $i \in [1, n]$.

In cosmology, the *ergodic hypothesis* is usually considered to be true; as a consequence, it is assumed that averages over the physical space are equivalent to ensemble averages.

Let us introduce some useful moments of the distribution function:

- first-order moment[2]:

  $$< \delta >= 0,$$

  because of the very definition of δ.

- Second-order moments:

  – the variance

  $$\sigma^2 \doteq < \delta^2(\vec{x}) >,$$

  independent of the position as a consequence of the ergodic hypothesis.

  – The two-point correlation function

  $$\xi(x) =< \delta(\vec{x}_1)\delta(\vec{x}_2) >,$$

  where $x = |\vec{x}_1 - \vec{x}_2|$. ξ is independent of the direction of the separation vector $\vec{x}_1 - \vec{x}_2$ as a consequence of the cosmological principle.

If the distribution function $\hat{\mathcal{P}}$ of the perturbation field is Gaussian, then first and second order moments are enough to define it completely.

I now consider the Fourier components of $\delta(\vec{x})$, δ_k, and its distribution function $\hat{\mathcal{P}}$. It can be proved that the second moments of $\mathcal{P}$ and $\hat{\mathcal{P}}$ are connected by a Fourier transform. Indeed, once defined the spectral density (also known as power spectrum) of the perturbation field as:

$$P(k) = V_u < |\delta_k|^2 >,$$

2 Here I use the notation $< \cdot >$ to refer to an ensemble average.

where $k = |\vec{k}|$ and V_u is the volume of the Universe, it can be easily proved that:

$$P(k) = \int_{V_u} \xi(x) e^{-i\vec{k}\cdot\vec{x}} d^3\vec{x}. \tag{47}$$

By anti-transforming Eq. 47, $\xi(x)$ can be written as:

$$\xi(x) = \frac{1}{2\pi^2} \int_0^\infty P(k) k^2 \frac{\sin(kx)}{kx} dk \tag{48}$$

In the case of a Gaussian random field, $\mathcal{P}$ is a multivariate Gaussian function:

$$\mathcal{P}(\delta_1, \delta_2, ..., \delta_n) = \frac{1}{[(2\pi)^n \det\mathcal{M}]^{1/2}} e^{-\frac{1}{2}\sum_{i,j}\delta_i \mathcal{M}_{ij}^{-1}\delta_j}, \tag{49}$$

where $\mathcal{M}$ is the covariance matrix, such that $\mathcal{M}_{ij} =< \delta_i \delta_j >$.
Separating $\delta_{\vec{k}}$ in its module $|\delta_{\vec{k}}|$ and phase $\varphi_{\vec{k}}$, their joint distribution function can be written as the product of a Rayleigh distribution function for $|\delta_{\vec{k}}|$ and a uniform distribution for $\varphi_{\vec{k}}$:

$$\hat{\mathcal{P}}(|\delta_{\vec{k}}|, \varphi_{\vec{k}}) d|\delta_{\vec{k}}| d\varphi_{\vec{k}} = \frac{d\varphi_{\vec{k}}}{2\pi} e^{-\frac{|\delta_{\vec{k}}|^2}{2V_u^{-1}P(k)}} \frac{|\delta_{\vec{k}}|}{V_u^{-1}P(k)} d|\delta_{\vec{k}}|. \tag{50}$$

This is exploited to generate initial conditions for cosmological simulations, once $P(k)$ is defined.

B.2.1 *Power spectrum of the perturbations*

The most simple model for the initial spectral density $P_i(k)$ is a power law:

$$P_i(k) \propto k^{n_s}, \tag{51}$$

where n_s is referred to as the primordial spectral index. Assuming that the initial perturbations were generated by *inflation*, a period of accelerated expansion of the Universe, $P_i(k)$ is expected to have the form of Eq. 51, with $n_s \sim 1$. The shape of $P(k)$ at the current epoch is determined by the causal processes which have acted on perturbations from the initial epoch to the present epoch. On the other hand, the normalization of $P(k)$ is defined observationally, since there is not a fundamental theory for inflation yet. Every practical measure is filtered on a given spatial scale. The perturbation field filtered on the scale R is then defined as:

$$\delta(\vec{x}; R) = \int \delta(\vec{x'}) W(\vec{x} + \vec{x'}, R) d^3\vec{x'}, \tag{52}$$

where W is a window function defined by the measurement process. Analogously, from Eq. 48, the filtered variance can be written as:

$$\sigma^2(R) = \frac{1}{2\pi^2} \int_0^\infty P(k) k^2 \tilde{W}^2(kR) dk, \tag{53}$$

where $\tilde{W}$ is the Fourier transform of W. Traditionally, the normalization of $P(k)$ has been expressed through the value $\sigma_8 = \sigma(8\ h^{-1}\mathrm{Mpc})$, which is inferred from the large-scale distribution of galaxies.

C

LINEAR EVOLUTION OF COSMIC PERTURBATIONS

Cosmological structures, such as galaxy clusters, form from the evolution of small density perturbations through gravitational instability in the highly homogeneous early Universe. In this Appendix, I introduce the theory of cosmic perturbations, applying the Newtonian approximation of small perturbations.
In the following, I denote with $\vec{x}$ the comoving coordinates and with $\vec{r}$ to proper coordinates, which are connected by the dimensionless scale factor a:

$$\vec{r} = a\vec{x}.$$

The evolution of perturbations in the linear regime can be described through Newtonian perturbation theory under the following hypotheses:

- the spatial scales of structures are much smaller than the horizon, so that causality can be considered instantaneous;
- $|\delta(\vec{x}, t)| << 1$;
- the matter content of perturbations can be treated as a non-relativistic fluid.

The fluid approximation requires the further assumption that the mean free path of particles is much smaller than the spatial scales of interest, which applies to both baryonic gas and collisionless Dark Matter (DM; for the latter, this is valid if the local velocity dispersion is sufficiently small to neglect diffusion on the scales of interest).

C.1 NEWTONIAN THEORY OF SMALL PERTURBATIONS

I consider a non-relativistic fluid with a proper velocity field $\vec{u}$ and a density field ρ, that generates a gravitational potential ϕ.
The evolution of the fluid, with pressure P, is described by the following set of equations:

- The continuity equation:

$$\frac{D\rho}{Dt} + \rho\nabla_{\vec{r}} \cdot \vec{u} = 0, \tag{54}$$

where $\frac{D}{Dt}$ is the Lagrangian derivative, defined as:

$$\frac{D}{Dt} = \frac{\partial}{\partial t}|_{\vec{r}} + \vec{u} \cdot \nabla_{\vec{r}}. \tag{55}$$

- Euler's equations:

$$\frac{D\vec{u}}{Dt} = -\frac{\nabla_{\vec{r}} P}{\rho} - \nabla_{\vec{r}} \phi. \tag{56}$$

- Poisson's equation:

$$\nabla^2_{\vec{r}} \phi = 4\pi G \rho. \tag{57}$$

This set of equations is closed by the equation of state that specifies the pressure P.

In terms of comoving coordinates $\vec{x}$, peculiar velocity[1] $\vec{v}$ ($\vec{v} = a\dot{\vec{x}}$), density contrast and peculiar potential Φ ($\Phi = \phi + \frac{1}{2} a\ddot{a}x^2$), and considering that $\bar{\rho} \propto a^{-3}$ for a non-relativistic fluid, the above equations become:

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot [(1+\delta)\vec{v}] = 0, \tag{58}$$

$$\frac{\partial \vec{v}}{\partial t} + \frac{\dot{a}}{a}\vec{v} + \frac{1}{a}(\vec{v} \cdot \nabla)\vec{v} = -\frac{\nabla \Phi}{a} - \frac{\nabla P}{a\bar{\rho}(1+\delta)}, \tag{59}$$

$$\nabla^2 \Phi = 4\pi G \bar{\rho} a^2 \delta, \tag{60}$$

here $\nabla = \nabla_{\vec{x}}$ and $\frac{\partial}{\partial t} = \frac{\partial}{\partial t}|_{\vec{x}}$.

This description can be extended to the case in which the Universe contains, beside non-relativistic matter, a smooth relativistic background and a vacuum component; in this case, only Poisson's equation should be modified to include contribution by radiation and vacuum:

$$\nabla^2_{\vec{r}} \phi = 4\pi G (\rho_m + \tilde{\rho_r} + \tilde{\rho_\Lambda}), \tag{61}$$

where $\tilde{\rho} = \rho + \frac{3P}{c^2}$ is the effective gravitational mass density. Equation 60 is still valid and the source term δ is still the perturbation in the non-relativistic component, whereas any other uniform component only contribute to the definition of the expanding background through Friedmann's equations.

When P depends only on density, the set of equations in closed, though generally P may depend on another quantity and can be expressed as a function of density and specific entropy S. Thus, another equation is required to describe the fluid, which comes from the definition of specific entropy S:

$$T\frac{dS}{dt} = \frac{\mathcal{H} - \mathcal{C}}{\rho}, \tag{62}$$

1 proper and peculiar velocities are related through: $\vec{u} = \dot{a}(t)\vec{x} + \vec{v}$

where $\mathcal{H}$ and $\mathcal{C}$ are heating and cooling rates per unit volume, respectively, and T is the temperature of the fluid. For non-relativistic monoatomic ideal gas Eq. 59 becomes:

$$\frac{\partial\vec{v}}{\partial t}+\frac{\dot{a}}{a}\vec{v}+\frac{1}{a}(\vec{v}\cdot\nabla)\vec{v}=-\frac{\nabla\Phi}{a}-\frac{c_s^2}{a}\frac{\nabla\delta}{1+\delta}-\frac{2\bar{T}}{3a}\nabla S, \quad (63)$$

where $c_s=\sqrt{\frac{\partial P}{\partial\rho}|_S}$ is the adiabatic sound speed and $\bar{T}$ is the temperature of the background. From Eq. 63 it can be seen that both density and entropy perturbations can be sources of density fluctuations; accordingly, there are two modes of initial perturbations, isentropic perturbations, for which the initial perturbations are only in ρ ($\delta_i \neq 0, \delta S_i = 0$) and isocurvature perturbations ($\delta_i = 0, \delta S_i \neq 0$), which can be both present in the initial conditions.

In the linear regime ($\delta << 1$ and $|\vec{v}| << c$) Eqs. 58 and 63 become:

$$\frac{\partial\delta}{\partial t}+\frac{1}{a}\nabla\cdot\vec{v}=0, \quad (64)$$

$$\frac{\partial\vec{v}}{\partial t}+\frac{\dot{a}}{a}\vec{v}=-\frac{\nabla\Phi}{a}-\frac{c_s^2}{a}\nabla\delta-\frac{2\bar{T}}{3a}\nabla S, \quad (65)$$

where $\bar{T}$ is the background temperature. From Eq. 65, it can be derived that there is no source for vorticity in the linear regime and $\nabla\times\vec{v}\propto a^{-1}$, which means that swirling peculiar motions die down as the Universe expands. Differentiating Eq. 64 and using Eqs. 60 and 65, I can derive a second order differential equation for the density contrast:

$$\frac{\partial^2\delta}{\partial t^2}+2\frac{\dot{a}}{a}\frac{\partial\delta}{\partial t}=4\pi G\bar{\rho}\delta+\frac{c_s^2}{a^2}\nabla^2\delta+\frac{2}{3}\frac{\bar{T}}{a^2}\nabla^2 S \quad (66)$$

The Hubble drag on the left-hand side of the equation creates a friction-like term which decelerates the otherwise exponential growth of perturbations that can arise from Jeans instability driven by gravity. The source terms in the right-hand size include the gravitational term depending on the amplitude of the perturbation and two pressure-like terms that depend on the spatial non-uniformity of the density contrast and of the specific entropy.

I now introduce the Fourier transform of the density contrast, $\delta_{\vec{k}}(t)$, such that:

$$\delta(\vec{x},t)=\sum_{\vec{k}}\delta_{\vec{k}}(t)e^{i\vec{k}\cdot\vec{x}}; \quad \delta_{\vec{k}}(t)=\frac{1}{V_u}\int\delta(\vec{x},t)e^{-i\vec{k}\cdot\vec{x}}d^3\vec{x}, \quad (67)$$

where V_u is a volume within which δ is assumed to be periodic.
In Fourier space Eq. 66 becomes:

$$\frac{d^2\delta_{\vec{k}}}{dt^2}+2\frac{\dot{a}}{a}\frac{d\delta_{\vec{k}}}{dt}=\left[4\pi G\bar{\rho}-\frac{k^2c_s^2}{a^2}\right]\delta_{\vec{k}}-\frac{2}{3}\frac{\bar{T}}{a^2}k^2S_{\vec{k}}, \quad (68)$$

and Eq. 60 becomes:

$$-k^2\Phi_{\vec{k}} = 4\pi G\bar{\rho}a^2\delta_{\vec{k}}. \tag{69}$$

Since $\nabla \times \vec{v} \approx 0$, the velocity field can be written as the gradient of a scalar field: $\vec{v} = \nabla\mathcal{V}$ or, equivalently, $\vec{v}_{\vec{k}} = i\vec{k}\mathcal{V}_{\vec{k}}$. Fourier-transforming Eq. 64, another relation in Fourier space can be derived:

$$\vec{v}_{\vec{k}} = \frac{ia\vec{k}}{k^2}\frac{d\delta_{\vec{k}}}{dt}. \tag{70}$$

For isentropic perturbations (i.e., with null initial entropy perturbation) with adiabatic evolution, and neglecting the Hubble drag term, Eq. 68 turns to:

$$\frac{d^2\delta_{\vec{k}}}{dt^2} = -\omega^2\delta_{\vec{k}}, \tag{71}$$

with $\omega^2 = \frac{k^2c_s^2}{a^2} - 4\pi G\bar{\rho}$.
When $\omega^2 > 0$, i.e. when $k > k_J = \frac{\sqrt{4\pi G\bar{\rho}}a}{c_s}$, the solution is a sound wave propagating at speed c_s, whereas for $k < k_J$, $\omega^2 < 0$ and the solution is a stationary wave whose amplitude either increases or decreases exponentially with time. The spatial scale corresponding to k_J is the Jeans length, $\lambda_J = c_s\sqrt{\frac{\pi}{G\bar{\rho}}}$.
Considering the Hubble drag, for $k > k_J$ ($\lambda < \lambda_J$) the acoustic waves are damped, while for $k < k_J$ the exponential growth of the growing mode of the solution is slowed down; in conclusion, only perturbations with $k < k_J$ can grow.

C.2 COLLISIONLESS GAS

In case of collisionless particles (such as DM particles), the fluid approximation is not valid, thus to resolve the evolution of perturbations one has to resort to Lagrangian mechanics. The Lagrangian of a particle of mass m in a gravitational potential ϕ:

$$\mathcal{L} = \frac{1}{2}mu^2 - m\phi, \tag{72}$$

which can be transformed, through a canonical transformation, into:

$$\mathcal{L} = \frac{1}{2}ma^2\dot{x}^2 - m\Phi. \tag{73}$$

For a collisionless gas, the phase-space density f obeys Boltzmann collisionless equation, which can be rewritten as Vlasov's equation:

$$\frac{\partial f(\vec{x},\vec{p},t)}{\partial t} + \frac{\vec{p}}{ma^2(t)} \cdot \nabla f(\vec{x},\vec{p},t) - m\nabla\phi(\vec{x},t) \cdot \frac{\partial f(\vec{x},\vec{p},t)}{\partial \vec{p}} = 0. \tag{74}$$

$\vec{p} = \frac{\partial \mathcal{L}}{\partial \dot{\vec{x}}} = ma\vec{v}$ is the canonical momentum conjugate to the comoving coordinates.
Observable quantities are defined through the coarse-grained $\vec{p}$-moments; given a quantity Q, its (first) $\vec{p}$-moment is:

$$< Q > (\vec{x}) = \frac{1}{n} \int Q f d^3\vec{p}, \quad n = \int f d^3\vec{p} = \bar{\rho} a^3 [1 + \delta(\vec{x})]/m. \tag{75}$$

Here, I assume that the density of the Universe is dominated by the collisionless particles under consideration.
Using the previous definitions and Eq. 74, one can obtain the collisionless equivalent of the continuity equation and Euler's equations in an expanding Universe:

$$\frac{\partial \delta}{\partial t} + \frac{1}{a} \sum_{j=1}^{3} \frac{\partial}{\partial x_j} [(1+\delta) < v_j >] = 0, \tag{76}$$

$$\frac{\partial < v_i >}{\partial t} + \frac{\dot{a}}{a} < v_i > + \frac{1}{a} \sum_j < v_j > \frac{\partial < v_i >}{\partial x_j} =$$
$$= -\frac{1}{a} \frac{\partial \Phi}{\partial x_i} - \frac{1}{a(1+\delta)} \sum_{j=1}^{3} \frac{\partial}{\partial x_j} [\sigma_{ij}^2 (1+\delta)], \tag{77}$$

where

$$\sigma_{ij}^2 = < v_i v_j > - < v_i >< v_j > \tag{78}$$

describes chaotic motions, that play the role of an anisotropic pressure in Eq. 77. The equations for first velocity moments involve the second moments and, in general, equations for the i-th moments involve the $(i+1)$-th, thus the dynamics can be fully resolved only if some assumptions are made to truncate the hierarchy of Euler-like equations.
Adopting a perturbative approach, one can solve Vlasov's equation splitting the phase-space density in an unperturbed term (f_0), which depends only on $|\vec{p}|$ in a homogeneous isotropic background, and a perturbation (f_1):

$$f = f_0 + f_1. \tag{79}$$

The equation for f_1, to first order in the perturbed quantities, is:

$$\frac{\partial f_1}{\partial t} + \frac{1}{ma^2} \vec{p} \cdot \nabla f_1 - m \nabla \Phi \cdot \frac{\partial f_0}{\partial \vec{p}} = 0. \tag{80}$$

Fourier transforming it and defining the variable ξ such that $\frac{dt}{d\xi} = a^2$, the solution for the Fourier transform of f_1, given the initial condition $f_{\vec{k}}(\vec{p}, \xi_i)$, is:

$$f_{\vec{k}}(\vec{p}, \xi) = f_{\vec{k}}(\vec{p}, \xi_i) e^{-\frac{i\vec{k}\cdot\vec{p}}{m}(\xi - \xi_i)} +$$
$$mi\vec{k} \cdot \left(\frac{\partial f_0}{\partial \vec{p}} \right) \int_{\xi_i}^{\xi} d\xi' a^2(\xi') \Phi_{\vec{k}}(\xi') e^{-\frac{i\vec{k}\cdot\vec{p}}{m}(\xi - \xi')}, \tag{81}$$

which can be solved iteratively. The first term of Eq. 81 is a kinematic term which describes the propagation of the initial condition, whereas the second one describes the dynamics through Φ.

C.3 FREE-STREAMING DAMPING

If $\vec{k} \cdot \vec{p}/m >> \xi^{-1}$ ($a/k << vt$), the integrand of the dynamical term in Eq. 81 varies so rapidly that the integral is negligible, so perturbations cannot grow with time and, since the density contrast is proportional to $\int f_1 d^3\vec{p}$, the rapid variation of the kinematic term in Eq. 81 implies that perturbations not only are prevented from growing, but they are damped as well. This is known as free-streaming damping and is due to the particles' random motions; the free-streaming length below which this effect is important is the comoving distance traveled by the particle till the cosmic time t:

$$\lambda_{fs} = \int_0^t \frac{v(t')}{a(t')} dt'. \tag{82}$$

Assumig that a particle becomes non-relativistic prior to the equivalence (t_{eq}), it can be shown that, at t_{eq}:

$$\lambda_{fs} = \frac{2ct_{nr}}{a(t_{nr})}[1 + \ln(a(t_{eq})/a(t_{nr}))]. \tag{83}$$

From this, one can obtain the free-streaming mass for light neutrinos of mass m_ν:

$$M_{fs} = 1.3 \times 10^{15} \left(\frac{m_\nu}{30\text{eV}}\right)^{-2} M_\odot. \tag{84}$$

Thus, in a Universe dominated by light neutrinos, the first objects to form should be superclusters.

D

NUMERICAL METHODS FOR COSMOLOGICAL SIMULATIONS

To describe the formation of cosmic structures in detail, particularly in the highly nonlinear regime where perturbations on different scales are coupled, numerical simulations are essential. In this Appendix, I give a brief overview of the main numerical techniques used to solve gravity (sect. D.1), followed by a description of hydrodynamic methods (sect. D.2), with a focus on Smoothed Particle Hydrodynamics (SPH).
For this Appendix, I refer to Appendix C of Mo et al. 2012, Dolag et al. (2008), Rosswog (2009), and Springel (2005).

D.1 N-BODY CODES

N-body techniques represent the most advanced tools available for solving gravitational dynamics in a cosmological context, where the linear approximation described in Appendix C breaks down, and the gravitational instability of DM perturbations in the strongly nonlinear regime must be treated numerically. In simulations, the DM density field is sampled either by particles or on a grid. We can therefore consider a system of N collisionless particles, defined by their positions $\vec{r}_i$ and velocities $\vec{u}_i$ ($i \in [1, N]$) at a given time, and write down the equations of motion for each of them. The equations of motion for the i-th particle are:

$$\frac{d\vec{r}_i}{dt} = \vec{u}_i \tag{85}$$

and

$$\frac{d\vec{u}_i}{dt} = -\vec{\nabla}_{\vec{r}}\phi_i. \tag{86}$$

In comoving coordinates and for peculiar velocity and potential, these become:

$$\frac{d\vec{x}_i}{dt} = \frac{\vec{v}_i}{a}, \quad \frac{d\vec{v}_i}{dt} + H(t)\vec{v}_i = -\frac{1}{a}\nabla\Phi_i, \tag{87}$$

where $H(t)$ is the Hubble parameter at cosmic time t.
The gravitational force acting on each particle depends on all the other particles' positions, so there are 6N coupled first-order ordinary differential equations that can be numerically integrated in finite time steps t_n. In order to solve these numerically, different time integrators are available. In the leap-frog time integrator, positions and (specific) forces ($\vec{F}_i$) are evolved in time by half time steps. For example, in the Drift-Kick-Drift version of the

method, the first step is a "drift", an update of the position at half time step, followed by a "kick", an update of velocity at full time step, and then by another drift:

$$\vec{r}_i(t_{n+1/2}) = \vec{r}_i(t_n) + \frac{\Delta t_n}{2}\,\vec{u}_i(t_n), \tag{88}$$

$$\vec{u}_i(t_{n+1}) = \vec{u}_i(t_n) + \vec{F}_i(t_{n+1/2})\,\Delta t_n, \tag{89}$$

$$\vec{r}_i(t_{n+1}) = \vec{r}_i(t_{n+1/2}) + \frac{\Delta t_{n+1}}{2}\,\vec{u}_i(t_{n+1}). \tag{90}$$

where $t_{n+1} = t_n + \Delta t_n$, $t_{n\pm1/2} = t_n \pm \frac{\Delta t_n}{2}$.

The crucial point in solving the evolution of a large number of particles is to find an efficient way of calculating the force (or the gradient of gravitational potential) on each particle. The different implementations of an N-body scheme thus correspond to different ways of computing forces (or potentials) and are briefly discussed in the following Sections.

D.1.1 *Direct Integration Codes*

The most simple and accurate way to describe gravitational interaction between mass points m_j is to calculate explicitly Newton's force:

$$\vec{F_i} = -G \sum_{j\neq i} \frac{m_j(\vec{r_i} - \vec{r_j})}{|\vec{r_i} - \vec{r_j}|^3}, \tag{91}$$

where $\vec{F_i}$ is the force per unit mass acting on particle i at $\vec{r_i}$ due to interaction with all the other particles, over which the summation is performed, and G is the gravitational constant.
At every time step, forces acting on each particle are recalculated and particles moved according to the adopted time integrator. During intervals between time steps, particles follow a uniform linear motion, thus to reduce errors in computed trajectories, time intervals need to be shorter where particles undergo stronger accelerations, which typically happens when close encounters occur.
The accuracy in controlling the time steps is fundamental in this scheme to reduce integration errors.

Using Eq. 91 to express the force between two extended pseudo-particles[1] provides a better approximation the farther apart the particles are. At small distances, the behavior of particles is not realistic, since the gravitational force between point-like particles described by Eq. 91 diverges as they approach, while this is not the case for extended mass elements. This problem, called "collisionality", causes unrealistic behavior in close encounters, with particles moving on hyperbolic orbits.

---

1 Here, I refer to *pseudo-particles* as the mass elements used to sample the DM density field in N-body simulations, emphasizing that these are not physical DM particles.

To avoid this problem, in direct codes, a "softening parameter"[2] ϵ is introduced in the denominator of Eq. 91. A possible choice, called the "Plummer" softening, is to add the square of ϵ to the squared module of inter-particle separation in Eq. 91, so that the implementation of the specific force experienced by particle i becomes:

$$\vec{F_i} = -G \sum_{j \neq i} \frac{m_j (\vec{r_i} - \vec{r_j})}{(|\vec{r_i} - \vec{r_j}|^2 + \epsilon^2)^{3/2}} \tag{92}$$

In this approximation, forces are non-Newtonian on scales smaller than few ϵ ($\sim 2.7 \times \epsilon$ for the specific case of Eq. 92). As a consequence, the trajectory of a single particle is not physically meaningful. Nevertheless, pseudo-particles tracing the DM density field are not real particles, as mentioned before, thus only their collective and statistical properties should be considered.
This method is considered the most accurate in solving an N-body problem and is applied when high precision is needed. On the other hand, it has the disadvantage that computational cost is proportional to N^2, where N is the total number of particles, since the computation of forces is a pairwise operation. This results in a prohibitive number of operations for large systems, making it not feasible for solving problems that cover a wide dynamical range, such as cosmological simulations.

D.1.2 *Particle-Mesh scheme*

In Particle-Mesh (PM) methods, gravitational potentials and forces are evaluated on a mesh and then interpolated at particle positions. At first, a scheme to assign density to mesh points is chosen, then the density is Fourier-transformed and potential at mesh points is computed from the density field by solving Poisson's equation in Fourier space. The next steps are anti-transforming the potentials, calculating forces at mesh points through a finite-difference approximation of derivatives, and then interpolating the forces at particle positions.
For explanatory purposes, consider a cubic domain of side length L, discretized into $N = M^3$ cubic cells (or meshes) of side $\delta = L/M$. The domain is assumed to be periodic, implying that any physical field — such as forces or potentials — evaluated at mesh points is periodic as well. Consequently, these fields admit a discrete representation in Fourier space. Every grid point is identified by a triplet of integers and its coordinates can be expressed as:

$$\vec{x}_{\vec{q}} = \delta \vec{q},$$

where $\vec{q} = (q_1, q_2, q_3)$ and $q_i \in [1, M]$, for $i = 1, 2, 3$. This indexing scheme defines the regular lattice on which the discrete fields are computed.
As previously discussed, the first step in solving an N-body problem with

2 ϵ is usually chosen to be 1/20-1/50 times the mean inter-particle separation in the simulation.

a PM method is defining an assignment scheme to map the mass density field on the grid. For a generic assignment scheme, if the mass of a particle is distributed to the closest n^3 meshes, then the contribution of the i-th particle to the mass density assigned to the grid point $\vec{x}_{\vec{q}}$ is described by a weighting function $W(\vec{x}_{\vec{q}} - \vec{x}_i)$, defined as:

$$W(\vec{x}_{\vec{q}} - \vec{x}_i) = \int \hat{W}\left(\frac{\vec{x} - \vec{x}_{\vec{q}}}{h}\right) S(\vec{x} - \vec{x}_i, h) d^3x, \tag{93}$$

where $h = n\delta$, $S(\vec{x}, h)$ is a shape function that defines the assignment method, and:

$$\hat{W}(\vec{x}) = \begin{cases} 1 & |\vec{x}| < 0.5 \\ 0 & otherwise. \end{cases} \tag{94}$$

At this point, all the necessary ingredients have been introduced to express the mass density assigned to the mesh point $\vec{x}_{\vec{q}}$, under a given assignment scheme (once S, and therefore W, are fixed):

$$\rho(\vec{q}) = \frac{1}{h^3} \sum_i m_i W(\vec{x}_i - \vec{x}_{\vec{q}}). \tag{95}$$

The weighting function, also referred to as the kernel function, is normalized to unity:

$$\int d^3\vec{x}\, W = 1. \tag{96}$$

In the following I list some examples of mass assignment schemes[3].

- Nearest Grid Point (NGP): the mass of a particle is assigned to the nearest mesh point, resulting in discontinuous forces. This corresponds, in the previously introduced notation, to $n = 1$ and $S(\vec{x}, h) = \delta^{(3)}(\vec{x})$, where $\delta^{(3)}$ here denotes the 3-dimensional Dirac δ-function.
- Cloud In Cell (CIC): the mass of a particle is distributed to the closest 8 mesh points, weighted by the intersection of the volume of the a cube centered on the particle position with side δ and the volume of the mesh. In this case, the forces are continuous but their first derivatives are not. Here $n = 2$ and $S(\vec{x}, h) = \hat{W}(\vec{x}/h)$.
- Triangular Shaped Cloud (TSC): $n = 3$ and $S(\vec{x}, h) = 1 - \frac{|\vec{x}|}{h}$, it is associated with a piece-wise quadratic interpolation function and provides both a continuous density field and a continuous first derivative of density.

Once the assignment scheme is chosen, the density field can be evaluated at mesh points and its Fourier components $\hat{\rho}(\vec{k})$ can be computed with a Fast Fourier Transform (FFT) technique.

3 I use the expressions "mass assignment" and "mass density assignment" interchangeably because all grid cells have the same volume δ^3, so mass and mass density assigned to a mesh point only differ by a constant factor

The gravitational potential of a given mass distribution can be written as a convolution of the density field and an appropriate Green's function, which depends on the boundary conditions:

$$\phi(\vec{x}) = \int d^3x' g(\vec{x} - \vec{x'})\rho(\vec{x'}). \tag{97}$$

Then, the Fourier transform of the potential can be expressed as:

$$\phi_{\vec{k}} = g_{\vec{k}} \cdot \rho_{\vec{k}}, \tag{98}$$

where $g_{\vec{k}}$ is the Fourier transform of Green's function, which needs to be computed only once, at the beginning of the simulation, from Poisson's equation in Fourier space:

$$-k^2\phi_{\vec{k}} = 4\pi G\rho_{\vec{k}} \tag{99}$$

where $k = |\vec{k}|$. This gives[4] $g_{\vec{k}} = -\frac{4\pi G}{k^2}$, from which $\phi_{\vec{k}}$ is obtained through Eq. 98 and then, by anti-transforming it, the potential in real space can be recovered at mesh points.

The next step is the calculation of forces at mesh points by computing the approximate gradient of the potential via finite-difference. To update positions and velocity of particles, forces must be evaluated at particles' positions $\vec{x}_i$ via interpolation:

$$\vec{F}(\vec{x_i}) = \sum_{\vec{q}} W(\vec{x}_i - \vec{x}_{\vec{q}})\vec{F}(\vec{x}_{\vec{q}}). \tag{100}$$

Here $W(\vec{x})$ should be the same kernel used for mass assignment, in order to guarantee symmetry in forces between pairs of particles, which in turn guarantees the conservation of momentum.

In this scheme, the force resolution is limited by the size of the grid, so a strong constraint is given by memory rather than computation time, which is the major limitation of this method. On the other hand, the computation time is dominated by the time required to perform a FFT, which gives the advantageous scaling of $N\log(N)$,where N here is the number of meshes.

D.1.3 *Tree codes*

In a tree code, the force acting on a particle is calculated from truncated multipole expansions[5] of the gravitational potential generated by other particles, which are hierarchically grouped according to their distance from the

---

4 This is referred to the Laplace operator in $\mathbb{R}^3$, the discretized spatial domain and the mass assignment scheme modify $g_{\vec{k}}$.

5 The potential of an arbitrary distribution of matter can be written as an expansion in multipoles as:

$$\phi(r, \theta, \varphi) = \sum_{l,m} \frac{Y_l^m(\theta, \varphi)}{r^{l+1}} Q_{lm}, \tag{101}$$

where Y_l^m are the spherical harmonics and Q_{lm} are the multipole moments of the system. See Binney & Tremaine (1987) for details.

particle under consideration.
In this scheme, one can either fix the order of the multipole expansion, in which case more distant particles can be grouped together and their contribution to gravity can be accounted for by considering a single multipole force, or one can fix the force accuracy instead. In the latter case, more distant groups require fewer terms in the expansion to achieve the same accuracy as nearby particles.
The central point of this method is the hierarchical grouping of particles, which can be done through a recursive space partition. in the version presented by Barnes & Hut (1986), this is done by building an oct-tree, in which the matter distribution is encompassed by a cubical root node, which is repeatedly partitioned in eight "children" cubic nodes with half-side length until each node either contains one particle or is parent to other nodes. To obtain the total force acting on the considered particle, one has to *walk* down the tree, summing contributions to the force by the different nodes as long as a fixed criterion - the "opening criterion", which defines the accuracy of the force computation - is met. At each node, starting from the root, an evaluation is made to check if the multipole expansion of that node guarantees enough precision in the force evaluation. If this is not the case, then the node is "opened", which means that one has to walk down the tree to its children nodes and check if they meet the accuracy criterion. In the standard formulation, given an accuracy parameter ϑ, a node of size l is opened if $r < l/\vartheta$, where r is the distance from the center of mass of the cell to the particle for which the force is computed. This means that the node is opened if it is seen by the particle under an angle larger than ϑ (typically $\vartheta \sim 0.5\,\mathrm{rad}$).
A simplified 2D example is shown in fig. D.1: a 2D box is hierarchically divided in squares with half the side of their parent until each of them only contains one particle.
In a tree code, the force evaluation requires a number of operations approximately proportional to $N\log(N)$, which is the exact scaling for a homogeneous distribution of particles, whereas in case of strong inhomogeneities the scaling is less efficient. The force error can be conveniently controlled through the accuracy criterion, for example reducing ϑ in the standard formulation, which means walking lower levels in the tree, opening nodes which are farther from the particle for which the force is computed. A drawback is that, at least in the standard formulation, the force between two particles is not symmetric, which means that momentum is not conserved; nevertheless, there are extensions of this scheme with improved conservation properties.

D.1.4 *Hybrid schemes*

Hybrid methods can be constructed by combining the ones previously described, like the TreePM (see, for example, Xu 1995), which merges a Tree code and a Particle-Mesh, or a P^3M (Particle-Particle-Particle-Mesh) code, which is a synthesis of a direct and a PM code.

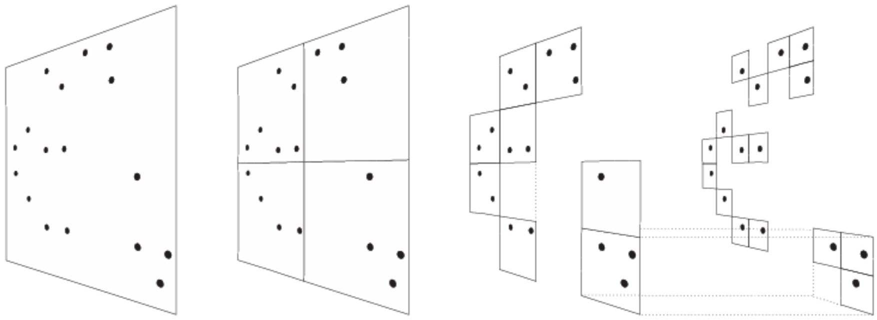


Figure D.1: A schematic illustration of the implementation of an oct-tree in two dimensions; the root node (top left) with side l is divided in sub-squares of sides $l/2$ (top right), then the same is done with each of the *children* nodes and with their children as well, until each square of the final partitioned domain contains only one particle (bottom right). Reproduced from Springel et al. (2001a).

In TreePM codes, the Fourier transform of the gravitational potential is split in a short-range and a long-range term based on a spatial scale r_s:

$$\phi_{\vec{k}} = \phi_{\vec{k}}^{short} + \phi_{\vec{k}}^{long}, \tag{102}$$

where

$$\phi_{\vec{k}}^{long} = \phi_{\vec{k}} e^{-k^2 r_s^2}. \tag{103}$$

If r_s is small compared to the size of the box, the solution for the short-range term in real space is obtained from Poisson's equation:

$$\phi^{short}(\vec{x}) = -G \sum_i \frac{m_i}{r_i} \mathrm{erfc}\left(\frac{r_i}{2r_s}\right), \tag{104}$$

where $r_i = |\vec{x}_i - \vec{x}|$ and $\mathrm{erfc}(x)$ is the complementary error function. The corresponding force term is evaluated in real space with a tree code, whereas the long-range term is calculated with a PM scheme. This hybrid method has the advantages of a tree code with an improved accuracy in the long-range force, which is no more an approximation as in a pure tree code. Furthermore, for r_s slightly larger than the mesh spacing, it avoids the anisotropies typical of PM codes on small scales.

In a P^3M code, long-range forces are calculated with a PM algorithm, while forces between particles closer than two times the mesh spacing are computed directly. To prevent high time consumption, since in the case of highly clustered particles the computation time is basically that of a direct code, an adaptive mesh can be applied. In this case, wherever there are many close particles in a parent grid, a finer grid is used, which defines an AP^3M (Adaptive P^3M) scheme.

D.1.5 *time step criterion*

The variation of the position vector $\vec{r}$ in a time step $\Delta t = t_{n+1} - t_n$ can be expanded as:

$$\Delta\vec{r} = \vec{u}(t_n)\Delta t + \frac{1}{2}\vec{F}(t_{n+1/2})\Delta t^2 + \cdots. \tag{105}$$

In order to have an accurate integration, the second term on the right-hand side of Eq. 105 must be sufficiently smaller than the first, which calls for a time step criterion of the form:

$$\Delta t = \alpha_{tol}\frac{\sigma}{|\vec{F}|}, \tag{106}$$

where σ is the typical velocity dispersion at the considered position and α_{tol} is a dimensionless tolerance parameter (with a typical value of ~ 0.1).

D.1.6 *Initial conditions*

To define the initial conditions for cosmological simulations, one needs to set up a uniform background for the unperturbed Universe and add initial perturbations with statistical properties depending on the cosmological parameters; the density contrast is typically assumed[6] to be a realization of a Gaussian random field with a *power spectrum* $P(k)$, which is the variance of the density fluctuation field in Fourier space, and completely describes the statistical properties of the Gaussian field.

As a consequence, the first step to generate initial conditions is to generate a set of complex numbers whose phases are uniformly distributed and whose amplitudes have a Rayleigh distribution with variance $P(k)$ (see, for example, Bardeen et al. 1986). Randomly drawing two real numbers r_1 and r_2 in the interval $]0,1]$, one gets the realization of the density contrast in Fourier space for a given $\vec{k}$:

$$\delta_{\vec{k}} = \sqrt{-2P(k)\ln r_1}\, e^{i2\pi r_2}. \tag{107}$$

The potential on a regular grid $\vec{q}$ in real space can be written as a Fourier transform:

$$\Phi(\vec{q}) = \sum_{\vec{k}} \frac{\hat{\delta}_{\vec{k}}}{k^2} e^{i\vec{k}\cdot\vec{q}}, \tag{108}$$

A uniform background represented through a regular grid of particles has the disadvantage of altering the statistical properties of perturbations on scales approaching the grid spacing and introducing strong anisotropies on all scales. As suggested by White (1996), one could use a uniform glass-like isotropic configuration, which can be created by generating a random distribution of particles in the simulation box and letting them evolve with the desired N-body integrator with an artificial repulsive gravity, resulting in a uniform background with no recognizable anisotropy on scales larger

6 This *assumption* is theoretically expected and confirmed with good precision by observations of the Cosmic Microwave Background.

than few inter-particle separations, from which small-scale structures do not grow until the expansion factor reaches large values.
Starting from the unperturbed Universe, positions and velocities of particles sampling a linearly evolved density field can be generated through Zel'dovich approximation (see Zel'Dovich 1970), which gives the following velocity and displacement:

$$\vec{u} = -\dot{D(t)}\vec{\nabla}\Phi(\vec{q}); \quad \vec{x}(\vec{q}, t) = \vec{q} - D(t)\Phi(\vec{q}), \tag{109}$$

where $D(t) = \delta_+(t)$, the linear growth rate, is the amplitude of the growing mode of the perturbations.

D.2 NUMERICAL HYDRODYNAMICS

The evolution of the collisional baryonic component of the Universe, in the gravitational field created by DM in nonlinear regime, is described through numerical hydrodynamic schemes. These consist of two main classes of methods, Lagrangian and Eulerian schemes. In Lagrangian methods, the evolution of fluid elements is traced, whereas the Eulerian approach follows the evolution of fields (e.g., density, velocity of the fluid etc.) at fixed positions.

The evolution of a collisional fluid is described by the set of the following six equations, which are expressed in a Lagrangian formulation, since the description of numerical hydrodynamics in this section will focus on Smoothed Particle Hydrodynamics, a widely used Lagrangian method:

- the continuity equation:

  $$\frac{d\rho}{dt} = -\rho\nabla \cdot \vec{u}, \tag{110}$$

  where ρ and $\vec{u}$ are density and velocity of the fluid, respectively.

- Euler's equations (for a non-viscous fluid):

  $$\frac{d\vec{u}}{dt} = -\frac{\nabla P}{\rho} + \vec{f}, \tag{111}$$

  where P is pressure and $\vec{f}$ is a possible external force per unit mass (e.g., gravitational or magnetic force);

- the first law of thermodynamics (adiabatic case):

  $$\frac{d\mathcal{E}}{dt} = -\frac{P}{\rho}\nabla \cdot \vec{u}, \tag{112}$$

  where $\mathcal{E}$ is the internal energy per unit mass;

- the equation of state, which for an ideal gas, takes the form

  $$P = (\gamma - 1)\rho\mathcal{E}, \tag{113}$$

  where γ is the polytropic index, equal to 5/3 for an adiabatic monoatomic gas.

D.2.1 *Smoothed Particle Hydrodynamics*

Smoothed Particle Hydrodynamics (SPH) is one of the most popular Lagrangian approaches to solve hydrodynamics numerically. The fluid is sampled with particles and hydrodynamic quantities, like density, are carried by each of them and smoothed over their neighbors. Here I describe the basic form of SPH equations, the *standard* SPH, and then give some hints of how to derive a more modern version with more "natural" conservation properties; to this purpose I will follow mainly the review by Rosswog (2009).

The standard SPH results from a direct discretization of hydrodynamics equations, with an ad-hoc symmetrization to guarantee momentum conservation.

In SPH, the already mentioned "particles" are not to be intended as physical constituent particles of the gas, but merely points at which thermodynamic quantities are interpolated, which are advected with the local velocity of the fluid.

The core part of SPH is the kernel approximation, which gives an approximate representation of a function $f(\vec{r})$ *smoothed* on a scale h, as:

$$\tilde{f_h}(\vec{r}) = \int f(\vec{r'})W(\vec{r}-\vec{r'},h)d^3r' , \tag{114}$$

where $W(\vec{r})$ is the smoothing kernel, that, in order to recover the exact function in the limit of a vanishing smoothing length h, has to satisfy:

$$\lim_{h\to 0} \tilde{f_h}(\vec{r}) = f(\vec{r}), \tag{115}$$

$$\int W(\vec{r}-\vec{r'},h)d^3r' = 1. \tag{116}$$

Equation 114 can be re-written as:

$$\tilde{f_h}(\vec{r}) = \int \frac{f(\vec{r'})}{\rho(\vec{r'})} W(\vec{r}-\vec{r'},h)\rho(\vec{r'})d^3r' , \tag{117}$$

that can be discretized as:

$$f(\vec{r}) = \sum_b \frac{m_b}{\rho_b} f_b W(\vec{r}-\vec{r_b},h), \tag{118}$$

where $f_b \coloneqq f(\vec{r_b})$. The kernel needs to have a compact support, to avoid inefficient N^2- method, that would be of no use in realistic hydrodynamics problems with large number of particles. From Eq. 118, the central equation of SPH formulation can be derived:

$$\rho(\vec{r}) = \sum_b m_b W(\vec{r}-\vec{r_b},h), \tag{119}$$

which expresses the density evaluated at position $\vec{r}$ as a weighted sum of masses of particles in the neighborhood of $\vec{r}$. In order to ensure the conservation of angular momentum, radial kernels are usually employed, with a popular choice being a cubic spline kernel (Monaghan 1992):

$$W(r,h) = \frac{1}{\pi h^3} \begin{cases} 1 - 3R^2/2 + 3R^3/4 & 0 \leqslant R \leqslant 1 \\ (2-R)^3/4 & 1 \leqslant R \leqslant 2 \\ 0 & \text{otherwise}, \end{cases} \tag{120}$$

where $R = r/h$. Other than its compact support and radial symmetry, the kernel defined in Eq. 120 has the advantage of having both first and second-order derivatives continuous.

By writing $f(\vec{r}')$ as its Taylor expansion around $\vec{r}$ and substituting Eq. 120 in Eq. 114, one can write explicitly the error made in this particular kernel approximation:

$$\tilde{f}_h(\vec{r}) = f(\vec{r}) + c_2 h^2 + O(h^4), \tag{121}$$

where the coefficient c_2 depends on the second-order derivatives of f, thus the smoothed representation is exact for constant or linear functions. Since, in addition to the interpolation, the equations of hydrodynamics have to be discretized, the final accuracy is also dependent on the distribution of particles.

If the particle masses are fixed, the mass conservation is perfect and the density can be computed from Eq. 119, without solving a discretized version of Eq. 110. The discretization of Eqs. 111-113 can be done without resorting to finite differences for the evaluation of derivatives thanks to the kernel approximation:

$$\nabla f(\vec{r}) = \sum_b \frac{m_b}{\rho_b} f_b \nabla W(\vec{r} - \vec{r_b}, h), \tag{122}$$

since the exact gradient of the kernel can be used.

Discretizing Eq. 111 with $\vec{f} = 0$ gives:

$$\frac{d\vec{u}_a}{dt} = -\frac{1}{\rho_a} \sum_b \frac{m_b}{\rho_b} P_b \nabla_a W_{ab}, \tag{123}$$

where $W_{ab} = W(\vec{r}_a - \vec{r}_b)$. We note that Eq. 123 breaks the conservation of momentum, since the forces exerted by particle a on particle b and vice versa read, respectively:

$$\vec{F}_{ab} = \frac{m_b m_a}{\rho_b \rho_a} P_a \nabla_a W_{ab}; \; \vec{F_{ba}} = -\frac{m_a m_b}{\rho_a \rho_b} P_b \nabla_a W_{ab}, \tag{124}$$

where the property $\nabla_a W_{ab} = -\nabla_b W_{ab}$ of radial kernels has been exploited; the presence of the pressure breaks the symmetry, since in general $P_a \neq P_b$, thus Newton's third law is not respected.

Nevertheless, we can take advantage of the freedom in the discretization to enforce the conservation of momentum, starting from:

$$\nabla\left(\frac{P}{\rho}\right) = \frac{\nabla P}{\rho} - P\frac{\nabla \rho}{\rho^2}, \tag{125}$$

isolating the term $\frac{\nabla P}{\rho}$ on the left-hand side and using the right-hand side for the discretization of Eq. 111. In the end, one obtains the manifestly symmetric equation:

$$\frac{d\vec{u}_a}{dt} = -\sum_b m_b \left(\frac{P_a}{\rho_a^2} + \frac{P_b}{\rho_b^2}\right) \nabla_a W_{ab}. \tag{126}$$

The implicit assumptions made till now are that h is constant and that $\nabla_a W_{ab}$ is symmetric in the smoothing lengths (i.e. $\nabla_a W(r_{ab}, h_a) = \nabla_a W(r_{ab}, h_b)$); while the latter can be imposed by enforcing the symmetry, the former is a simplification and h is usually different from particle to particle, which is the origin of the "grad-h-terms" that appear in the generalized equations of SPH, that are reported in the following. A variable h is used to exploit the Lagrangian nature of SPH, by increasing the (Eulerian) resolution where density is higher, i.e. where more particles are present.
The discretization of Eq. 112 gives:

$$\frac{d\mathcal{E}_a}{dt} = \frac{P_a}{\rho_a^2} \sum_b m_b \vec{u}_{ab} \cdot \nabla_a W_{ab}, \tag{127}$$

where $\vec{u}_{ab} = \vec{u}_a - \vec{u}_b$.
Together with the (straightforward) discretization of the equation of state, Eqs. 119, 126 and 127 form the complete set of equations that are solved in the standard SPH formulation.
For the choice of h, one would like to decrease it where density is higher, which can be done in several ways, like keeping the mass or the number of neighbors of each particle constant within a radius h or by solving:

$$\frac{dh_a}{dt} = \frac{1}{3} h_a (\nabla \cdot \vec{v})_a,$$

which comes from the ansatz $h(t) \propto \rho(t)^{-1/3}$ and the continuity equation.
The evolution of gas can easily develop in discontinuous solutions (on the scale of the simulation), known as "shocks", as a consequence of *physical* viscosity; this problem can be approached numerically by using the analytical solution to Riemann problem or by adding an *artificial* viscosity to broaden the discontinuity over numerically resolvable lengths, where the gradients of thermodynamic quantities can be calculated. Most astrophysical SPH implementations apply the second strategy, by adding an artificial contribution to the pressure term in Eq. 126:

$$\frac{d\vec{v}_a}{dt} = -\sum_b m_b \left(\frac{P_a}{\rho_a^2} + \frac{P_b}{\rho_b^2} + \Pi_{ab}\right) \nabla_a W_{ab}. \tag{128}$$

Equation 127 should be modified consistently:

$$\frac{d\mathcal{E}_a}{dt} = \frac{P_a}{\rho_a^2} \sum_b m_b \vec{v}_{ab} \cdot \nabla_a W_{ab} + \frac{1}{2} \sum_b m_b \Pi_{ab} \vec{v}_{ab} \cdot \nabla_a W_{ab}. \tag{129}$$

I refer to Monaghan & Gingold (1983) for a more detailed description of widely-used versions of artificial viscosity.

As already mentioned, symmetries need to be enforced *ad hoc* in the standard SPH to ensure conservation of energy, momentum and angular momentum, whereas a "clean" approach based on the variational principle can be adopted to derive SPH, based only on a Lagrangian with appropriate symmetries, the first law of thermodynamics and a prescription for density (e.g., Gingold & Monaghan 1982). The discretized Lagrangian of an ideal fluid reads:

$$\mathcal{L}_{SPH} = \sum_i m_i \left(\frac{u_i^2}{2} - \mathcal{E}(\rho_i, s_i), \right) \tag{130}$$

where s represents the entropy per unit mass. To obtain a prescription to compute density as a function of particle positions, Eq. 119 can be used, allowing for different smoothing lengths for different particles:

$$\rho(\vec{r_a}) = \sum_b m_b W(\vec{r_a} - \vec{r_b}, h_a), \tag{131}$$

where the adaptivity of h can be accounted for with $h_a = \eta(m_a/\rho_a)^{1/3}$ (with $\eta \in [1.2, 1.5]$, see Gingold & Monaghan 1982). In this way, ρ depends on h, which in turns depends on ρ, so iterations are needed to reach convergence.
From Eq. 131 and using the properties of radial kernels:

$$\frac{d\rho_a}{dt} = \frac{1}{\Omega_a} \sum_b m_b \vec{u}_{ab} \cdot \nabla_a W_{ab}(h_a), \tag{132}$$

where

$$\Omega_a := \left(1 - \frac{\partial h_a}{\partial \rho_a} \sum_b m_b \frac{\partial W_{ab}(h_a)}{\partial h_a} \right). \tag{133}$$

By using $\mathcal{L}_{SPH}$ in Euler-Lagrange equations, the first principle of thermodynamics and the new prescription for density assignment while still keeping the particle masses constant in time, the generalized SPH equations can be derived:

$$\frac{d\mathcal{E}_a}{dt} = \frac{P_a}{\Omega_a \rho_a^2} \sum_b m_b \vec{u}_{ab} \cdot \nabla_a W_{ab}(h_a), \tag{134}$$

$$\frac{d\vec{u}_a}{dt} = - \sum_b m_b \left(\frac{P_a}{\Omega_a \rho_a^2} \nabla_a W_{ab}(h_a) + \frac{P_b}{\Omega_b \rho_b^2} \nabla_a W_{ab}(h_b) \right). \tag{135}$$

By allowing a varying h, the terms on the right-hand side of the SPH equations have gained $1/\Omega$ factors, which are known as "grad-h" terms and ensure better accuracy in the solutions of SPH equations. Fixing the masses in time, the continuity equation does not need to be solved, whereas the others must be integrated with a time integrator. Because of the presence of spatial derivatives in the equations, the time step needs to fulfill the Courant condition $|v|\Delta t/\Delta r < 1$, which gives, in SPH:

$$\Delta t_i = \frac{\alpha_c h_i}{h_i|(\vec{\nabla} \cdot \vec{u})_i| + max(c_{s,i}, |\vec{u}_i|)}, \tag{136}$$

where $c_{s,i}$ is the sound speed at the position of the i-th particle and α_c is a tolerance parameter.
SPH performs poorly in low-density regions, being a Lagrangian, particle-based method, whereas it can achieve high resolution in dense regions; the conservation properties are its major strength and, in its "generalized" formulation, it is not affected by any arbitrariness.

ACT/DES/HSC Collaboration (July 2025). "The Atacama Cosmology Telescope: DR6 Sunyaev-Zel'dovich Selected Galaxy Clusters Catalog." In: *arXiv e-prints*, arXiv:2507.21459, arXiv:2507.21459. DOI: 10.48550/arXiv.2507.21459. arXiv: 2507.21459 [astro-ph.CO].

Abdul Karim, M. et al. (Oct. 2025). "DESI DR2 results. I. Baryon acoustic oscillations from the Lyman alpha forest." In: *Phys. Rev. D* 112.8, 083514, p. 083514. DOI: 10.1103/2wwn-xjm5. arXiv: 2503.14739 [astro-ph.CO].

Adame, A. G. et al. (Feb. 2025). "DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations." In: *JCAP* 2025.2, 021, p. 021. DOI: 10.1088/1475-7516/2025/02/021. arXiv: 2404.03002 [astro-ph.CO].

Aggarwal, C.C. (2018). *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing. ISBN: 9783319944647. URL: https://books.google.it/books?id=AsTswQEACAAJ.

Akins, Hollis B. et al. (Apr. 2022). "Quenching and the UVJ Diagram in the SIMBA Cosmological Simulation." In: *ApJ* 929.1, 94, p. 94. DOI: 10.3847/1538-4357/ac5d3a. arXiv: 2105.12748 [astro-ph.GA].

Alberts, Stacey & Allison Noble (Oct. 2022). "From Clusters to Proto-Clusters: The Infrared Perspective on Environmental Galaxy Evolution." In: *Universe* 8.11, 554, p. 554. DOI: 10.3390/universe8110554. arXiv: 2209.02781 [astro-ph.GA].

Allen, C. W. (1976). *Astrophysical Quantities*.

Allen, S. W. et al. (Apr. 2001). "ASCA and ROSAT observations of nearby cluster cooling flows." In: *MNRAS* 322.3, pp. 589–613. DOI: 10.1046/j.1365-8711.2001.04135.x. arXiv: astro-ph/9910188 [astro-ph].

Altamura, Edoardo et al. (Apr. 2023). "EAGLE-like simulation models do not solve the entropy core problem in groups and clusters of galaxies." In: *MNRAS* 520.2, pp. 3164–3186. DOI: 10.1093/mnras/stad342. arXiv: 2210.09978 [astro-ph.CO].

Ameglio, S. et al. (Nov. 2007). "Joint deprojection of Sunyaev-Zeldovich and X-ray images of galaxy clusters." In: *MNRAS* 382.1, pp. 397–411. DOI: 10.1111/j.1365-2966.2007.12384.x. arXiv: 0708.3737 [astro-ph].

Anbajagane, D. et al. (Sept. 2025). "The Dark Energy Camera All Data Everywhere cosmic shear project V: Constraints on cosmology and astrophysics from 270 million galaxies across 13,000 deg^2 of the sky." In: *arXiv e-prints*, arXiv:2509.03582, arXiv:2509.03582. DOI: 10.48550/arXiv.2509.03582. arXiv: 2509.03582 [astro-ph.CO].

Anderson, Lauren et al. (June 2014). "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples." In: *MN-*

RAS 441.1, pp. 24–62. DOI: 10.1093/mnras/stu523. arXiv: 1312.4877 [astro-ph.CO].

Andreon, S. et al. (Nov. 2009). "JKCS 041: a colour-detected galaxy cluster at z_{phot} ~1.9 with deep potential well as confirmed by X-ray data." In: *A&A* 507.1, pp. 147–157. DOI: 10.1051/0004-6361/200912299. arXiv: 0812.1699 [astro-ph].

Andreon, S. et al. (May 2014). "JKCS 041: a Coma cluster progenitor at z = 1.803." In: *A&A* 565, A120, A120. DOI: 10.1051/0004-6361/201323077. arXiv: 1311.4361 [astro-ph.CO].

Andrews, Moira et al. (Oct. 2024). "Galaxy populations in protoclusters at cosmic noon." In: *arXiv e-prints*, arXiv:2410.08412, arXiv:2410.08412. DOI: 10.48550/arXiv.2410.08412. arXiv: 2410.08412 [astro-ph.GA].

Angulo, R. E. et al. (Oct. 2012). "The journey of QSO haloes from $z \sim 6$ to the present." In: *MNRAS* 425.4, pp. 2722–2730. DOI: 10.1111/j.1365-2966.2012.21783.x. arXiv: 1203.5339 [astro-ph.CO].

Annunziatella, M. et al. (Nov. 2014). "CLASH-VLT: The stellar mass function and stellar mass density profile of the z = 0.44 cluster of galaxies MACS J1206.2-0847." In: *A&A* 571, A80, A80. DOI: 10.1051/0004-6361/201424102. arXiv: 1408.6356 [astro-ph.GA].

Annunziatella, M. et al. (Jan. 2016). "CLASH-VLT: Environment-driven evolution of galaxies in the z = 0.209 cluster Abell 209." In: *A&A* 585, A160, A160. DOI: 10.1051/0004-6361/201527399. arXiv: 1510.05659 [astro-ph.GA].

Arnaud, M. et al. (July 2010). "The universal galaxy cluster pressure profile from a representative sample of nearby systems (REXCESS) and the Y_{SZ} - M_{500} relation." In: *A&A* 517, A92, A92. DOI: 10.1051/0004-6361/200913416. arXiv: 0910.1234 [astro-ph.CO].

Baes, Maarten et al. (Mar. 2024). "The TNG50-SKIRT Atlas: Post-processing methodology and first data release." In: *A&A* 683, A181, A181. DOI: 10.1051/0004-6361/202348418. arXiv: 2401.04224 [astro-ph.GA].

Bahé, Yannick M. et al. (Oct. 2017). "The Hydrangea simulations: galaxy formation in and around massive clusters." In: *MNRAS* 470.4, pp. 4186–4208. DOI: 10.1093/mnras/stx1403. arXiv: 1703.10610 [astro-ph.GA].

Bahé, Yannick M. et al. (Oct. 2022). "The importance of black hole repositioning for galaxy formation simulations." In: *MNRAS* 516.1, pp. 167–184. DOI: 10.1093/mnras/stac1339. arXiv: 2109.01489 [astro-ph.GA].

Balogh, Michael L. et al. (Mar. 2016). "Evidence for a change in the dominant satellite galaxy quenching mechanism at z = 1." In: *MNRAS* 456.4, pp. 4364–4376. DOI: 10.1093/mnras/stv2949. arXiv: 1511.07344 [astro-ph.GA].

Balsara, D. S. (Jan. 1995). "von Neumann stability analysis of smooth particle hydrodynamics–suggestions for optimal algorithms." In: *Journal of Computational Physics* 121.2, pp. 357–372. DOI: 10.1016/S0021-9991(95)90221-X.

Bardeen, J. M. et al. (May 1986). "The Statistics of Peaks of Gaussian Random Fields." In: *Astrophysical Journal* 304, p. 15. DOI: 10.1086/164143.

Barnes, David J. et al. (Feb. 2017). "The redshift evolution of massive galaxy clusters in the MACSIS simulations." In: *MNRAS* 465.1, pp. 213–233. DOI: 10.1093/mnras/stw2722. arXiv: 1607.04569 [astro-ph.CO].

Barnes, David J. et al. (Dec. 2018). "A census of cool-core galaxy clusters in IllustrisTNG." In: *MNRAS* 481.2, pp. 1809–1831. DOI: 10.1093/mnras/sty2078. arXiv: 1710.08420 [astro-ph.CO].

Barnes, J. & P. Hut (Dec. 1986). "A hierarchical O(N log N) force-calculation algorithm." In: *Nature* 324.6096, pp. 446–449. DOI: 10.1038/324446a0.

Baron, Dalya (2019). *Machine Learning in Astronomy: a practical overview*. arXiv: 1904.07248 [astro-ph.IM].

Barret, Didier et al. (Apr. 2023). "The Athena X-ray Integral Field Unit: a consolidated design for the system requirement review of the preliminary definition phase." In: *Experimental Astronomy* 55.2, pp. 373–426. DOI: 10.1007/s10686-022-09880-7. arXiv: 2208.14562 [astro-ph.IM].

Bassini, L. et al. (Oct. 2020). "The DIANOGA simulations of galaxy clusters: characterising star formation in protoclusters." In: *A&A* 642, A37, A37. DOI: 10.1051/0004-6361/202038396. arXiv: 2006.13951 [astro-ph.GA].

Bate, Matthew R., Ian A. Bonnell & Nigel M. Price (Nov. 1995). "Modelling accretion in protobinary systems." In: *MNRAS* 277.2, pp. 362–376. DOI: 10.1093/mnras/277.2.362. arXiv: astro-ph/9510149 [astro-ph].

Beck, A. M. et al. (Jan. 2016). "An improved SPH scheme for cosmological simulations." In: *Monthly Notice of the RAS* 455.2, pp. 2110–2130. DOI: 10.1093/mnras/stv2443. arXiv: 1502.07358 [astro-ph.CO].

Behroozi, Peter S., Risa H. Wechsler & Hao-Yi Wu (Jan. 2013). "The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores." In: *The Astrophysical Journal* 762.2, 109, p. 109. DOI: 10.1088/0004-637X/762/2/109. arXiv: 1110.4372 [astro-ph.CO].

Bell, Eric F. et al. (June 2004). "Nearly 5000 Distant Early-Type Galaxies in COMBO-17: A Red Sequence and Its Evolution since z~1." In: *ApJ* 608.2, pp. 752–767. DOI: 10.1086/420778. arXiv: astro-ph/0303394 [astro-ph].

Best, P. N., M. S. Longair & H. J. A. Roettgering (Apr. 1998). "HST, radio and infrared observations of 28 3CR radio galaxies at redshift z~1 - II. Old stellar populations in central cluster galaxies." In: *MNRAS* 295.3, pp. 549–567. DOI: 10.1046/j.1365-8711.1998.01245.x. arXiv: astro-ph/9709195 [astro-ph].

Binney, J. & S. Tremaine (1987). *Galaxy Dynamics*. Princeton University Press.

Birkinshaw, M., S. F. Gull & H. Hardebeck (May 1984). "The Sunyaev-Zeldovich effect towards three clusters of galaxies." In: *Nature* 309, pp. 34–35. DOI: 10.1038/309034a0.

Bleem, L. E. et al. (Feb. 2015). "Galaxy Clusters Discovered via the Sunyaev-Zel'dovich Effect in the 2500-Square-Degree SPT-SZ Survey." In: *ApJ, Supplement* 216.2, 27, p. 27. DOI: 10.1088/0067-0049/216/2/27. arXiv: 1409.0850 [astro-ph.CO].

Bleem, L. E. et al. (Mar. 2020). "The SPTpol Extended Cluster Survey." In: *ApJ, Supplement* 247.1, 25, p. 25. DOI: 10.3847/1538-4365/ab6993. arXiv: 1910.04121 [astro-ph.CO].

Bleem, L. E. et al. (Feb. 2024). "Galaxy Clusters Discovered via the Thermal Sunyaev-Zel'dovich Effect in the 500-square-degree SPTpol Survey." In: *The Open Journal of Astrophysics* 7, 13, p. 13. DOI: 10.21105/astro.2311.07512. arXiv: 2311.07512 [astro-ph.CO].

Bluck, Asa F. L. et al. (Nov. 2020). "How do central and satellite galaxies quench? - Insights from spatially resolved spectroscopy in the MaNGA survey." In: *MNRAS* 499.1, pp. 230–268. DOI: 10.1093/mnras/staa2806. arXiv: 2009.05341 [astro-ph.GA].

Böhringer, H. et al. (Aug. 2000). "The Northern ROSAT All-Sky (NORAS) Galaxy Cluster Survey. I. X-Ray Properties of Clusters Detected as Extended X-Ray Sources." In: *ApJ, Supplement* 129.2, pp. 435–474. DOI: 10.1086/313427. arXiv: astro-ph/0003219 [astro-ph].

Böhringer, H. et al. (Oct. 2004). "The ROSAT-ESO Flux Limited X-ray (REFLEX) Galaxy cluster survey. V. The cluster catalogue." In: *A&A* 425, pp. 367–383. DOI: 10.1051/0004-6361:20034484. arXiv: astro-ph/0405546 [astro-ph].

Bonafede, A. et al. (Dec. 2011). "A non-ideal magnetohydrodynamic GADGET: simulating massive galaxy clusters." In: *Monthly Notice of the RAS* 418.4, pp. 2234–2250. DOI: 10.1111/j.1365-2966.2011.19523.x. arXiv: 1107.0968 [astro-ph.CO].

Bondi, H. (Jan. 1952). "On spherically symmetrical accretion." In: *MNRAS* 112, p. 195. DOI: 10.1093/mnras/112.2.195.

Bondi, H. & F. Hoyle (Jan. 1944). "On the mechanism of accretion by stars." In: *MNRAS* 104, p. 273. DOI: 10.1093/mnras/104.5.273.

Borgani, S. & A. Kravtsov (Feb. 2011). "Cosmological Simulations of Galaxy Clusters." In: *Advanced Science Letters* 4.2, pp. 204–227. DOI: 10.1166/asl.2011.1209. arXiv: 0906.4370 [astro-ph.CO].

Borgani, S. et al. (July 2005). "Entropy amplification from energy feedback in simulated galaxy groups and clusters." In: *MNRAS* 361.1, pp. 233–243. DOI: 10.1111/j.1365-2966.2005.09158.x. arXiv: astro-ph/0504265 [astro-ph].

Borgani, Stefano et al. (Nov. 2001). "Measuring Ω_m with the ROSAT Deep Cluster Survey." In: *ApJ* 561.1, pp. 13–21. DOI: 10.1086/323214. arXiv: astro-ph/0106428 [astro-ph].

Boselli, Alessandro & Giuseppe Gavazzi (Apr. 2006). "Environmental Effects on Late-Type Galaxies in Nearby Clusters." In: *Publications of the ASP* 118.842, pp. 517–559. DOI: 10.1086/500691. arXiv: astro-ph/0601108 [astro-ph].

Bouchet, P. et al. (Aug. 1985). "The visible and infrared extinction law and the gas-to-dust ratio in the Small Magellanic Cloud." In: *A&A* 149, pp. 330–336.

Boylan-Kolchin, Michael et al. (Sept. 2009). "Resolving cosmic structure formation with the Millennium-II Simulation." In: *MNRAS* 398.3, pp. 1150–1164. DOI: 10.1111/j.1365-2966.2009.15191.x. arXiv: 0903.3041 [astro-ph.CO].

Bruzual, G. & S. Charlot (Oct. 2003). "Stellar population synthesis at the resolution of 2003." In: *MNRAS* 344.4, pp. 1000–1028. DOI: 10.1046/j.1365-8711.2003.06897.x. arXiv: astro-ph/0309134 [astro-ph].

Bulbul, E. et al. (May 2024). "The SRG/eROSITA All-Sky Survey. The first catalog of galaxy clusters and groups in the Western Galactic Hemisphere." In: *A&A* 685, A106, A106. DOI: 10.1051/0004-6361/202348264. arXiv: 2402.08452 [astro-ph.CO].

Burenin, R. A. et al. (Oct. 2007). "The 400 Square Degree ROSAT PSPC Galaxy Cluster Survey: Catalog and Statistical Calibration." In: *ApJ, Supplement* 172.2, pp. 561–582. DOI: 10.1086/519457. arXiv: astro-ph/0610739 [astro-ph].

Burles, Scott, Kenneth M. Nollett & Michael S. Turner (May 2001). "Big Bang Nucleosynthesis Predictions for Precision Cosmology." In: *ApJ Letter* 552.1, pp. L1–L5. DOI: 10.1086/320251. arXiv: astro-ph/0010171 [astro-ph].

Burns, Jack O. et al. (Mar. 2008). "Why Do Only Some Galaxy Clusters Have Cool Cores?" In: *ApJ* 675.2, pp. 1125–1140. DOI: 10.1086/526514. arXiv: 0708.1954 [astro-ph].

Cai, Zheng et al. (Dec. 2016). "Mapping the Most Massive Overdensity Through Hydrogen (MAMMOTH) I: Methodology." In: *ApJ* 833.2, 135, p. 135. DOI: 10.3847/1538-4357/833/2/135. arXiv: 1512.06859 [astro-ph.GA].

Cai, Zheng et al. (Apr. 2017). "Mapping the Most Massive Overdensities through Hydrogen (MAMMOTH). II. Discovery of the Extremely Massive Overdensity BOSS1441 at z = 2.32." In: *ApJ* 839.2, 131, p. 131. DOI: 10.3847/1538-4357/aa6a1a. arXiv: 1609.02913 [astro-ph.GA].

Calzetti, Daniela et al. (Apr. 2000). "The Dust Content and Opacity of Actively Star-forming Galaxies." In: *ApJ* 533.2, pp. 682–695. DOI: 10.1086/308692. arXiv: astro-ph/9911459 [astro-ph].

Cameron, Ewan (June 2011). "On the Estimation of Confidence Intervals for Binomial Population Proportions in Astronomy: The Simplicity and Superiority of the Bayesian Approach." In: *Publ. Astron. Soc. Aust.* 28.2, pp. 128–139. DOI: 10.1071/AS10046. arXiv: 1012.0566 [astro-ph.IM].

Camps, P. & M. Baes (Apr. 2020). "SKIRT 9: Redesigning an advanced dust radiative transfer code to allow kinematics, line transfer and polarization by aligned dust grains." In: *Astronomy and Computing* 31, 100381, p. 100381. DOI: 10.1016/j.ascom.2020.100381. arXiv: 2003.00721 [astro-ph.GA].

Carlstrom, John E., Gilbert P. Holder & Erik D. Reese (Jan. 2002). "Cosmology with the Sunyaev-Zel'dovich Effect." In: *ARA&A* 40, pp. 643–680. DOI: 10.1146/annurev.astro.40.060401.093803. arXiv: astro-ph/0208192 [astro-ph].

Cattaneo, Andrea et al. (May 2007). "Accretion, feedback and galaxy bimodality: a comparison of the GalICS semi-analytic model and cosmological SPH simulations." In: *MNRAS* 377.1, pp. 63–76. DOI: 10.1111/j.1365-2966.2007.11597.x. arXiv: astro-ph/0605750 [astro-ph].

Cavaliere, A. & R. Fusco-Femiano (May 1976). "X-rays from hot plasma in clusters of galaxies." In: *A&A* 49, pp. 137–144.

Chabrier, Gilles (July 2003). "Galactic Stellar and Substellar Initial Mass Function." In: *Publications of the ASP* 115.809, pp. 763–795. DOI: 10.1086/376392. arXiv: astro-ph/0304382 [astro-ph].

Chadayammuri, Urmila et al. (June 2021). "Fountains and storms: the effects of AGN feedback and mergers on the evolution of the intracluster medium in the ROMULUSC simulation." In: *MNRAS* 504.3, pp. 3922–3937. DOI: 10.1093/mnras/stab1010. arXiv: 2001.06532 [astro-ph.GA].

Chen, Z. et al. (Jan. 2024). "COALAS II. Extended molecular gas reservoirs are common in a distant, forming galaxy cluster." In: *MNRAS* 527.3, pp. 8950–8972. DOI: 10.1093/mnras/stad3128. arXiv: 2310.05206 [astro-ph.GA].

Chiaberge, Marco et al. (Feb. 2010). "Three Candidate Clusters of Galaxies at Redshift ~1.8: The "Missing Link" Between Protoclusters and Local Clusters?" In: *ApJ Letter* 710.2, pp. L107–L110. DOI: 10.1088/2041-8205/710/2/L107. arXiv: 1001.2232 [astro-ph.CO].

Chiang, Yi-Kuan, Roderik Overzier & Karl Gebhardt (Dec. 2013). "Ancient Light from Young Cosmic Cities: Physical and Observational Signatures of Galaxy Proto-clusters." In: *ApJ* 779.2, 127, p. 127. DOI: 10.1088/0004-637X/779/2/127. arXiv: 1310.2938 [astro-ph.CO].

Choi, Jun-Hwan & Kentaro Nagamine (Feb. 2011). "Multicomponent and variable velocity galactic outflow in cosmological hydrodynamic simulations." In: *MNRAS* 410.4, pp. 2579–2592. DOI: 10.1111/j.1365-2966.2010.17632.x. arXiv: 1001.3525 [astro-ph.CO].

Chu, A., F. Durret & I. Márquez (May 2021). "Physical properties of brightest cluster galaxies up to redshift 1.80 based on HST data." In: *A&A* 649, A42, A42. DOI: 10.1051/0004-6361/202040245. arXiv: 2102.01557 [astro-ph.GA].

Chu, A. et al. (Oct. 2022). "Physical properties of more than one thousand brightest cluster galaxies detected in the Canada-France-Hawaii Telescope Legacy Survey." In: *A&A* 666, A54, A54. DOI: 10.1051/0004-6361/202243504. arXiv: 2206.14209 [astro-ph.GA].

Clements, D. L. et al. (Apr. 2014). "Herschel Multitiered Extragalactic Survey: clusters of dusty galaxies uncovered by Herschel and Planck." In: *MNRAS* 439.2, pp. 1193–1211. DOI: 10.1093/mnras/stt2253. arXiv: 1311.5758 [astro-ph.CO].

Collet, C. et al. (July 2015). "Defying jet-gas alignment in two radio galaxies at z ~2 with extended light profiles: Similarities to brightest cluster galaxies." In: *A&A* 579, A89, A89. DOI: 10.1051/0004-6361/201424544. arXiv: 1506.03725 [astro-ph.GA].

Collins, Ch. et al. (June 2003). "The Evolution of Brightest Cluster Galaxies." In: *Astrophysics and Space Science* 285.1, pp. 51–60. DOI: 10.1023/A:1024613910300.

Conroy, Charlie (Aug. 2013). "Modeling the Panchromatic Spectral Energy Distributions of Galaxies." In: *ARA&A* 51.1, pp. 393–455. DOI: 10.1146/annurev-astro-082812-141017. arXiv: 1301.7095 [astro-ph.CO].

Contini, Emanuele (Aug. 2021). "On the Origin and Evolution of the Intra-Cluster Light: A Brief Review of the Most Recent Developments." In: *Galaxies* 9.3, 60, p. 60. DOI: 10.3390/galaxies9030060. arXiv: 2107.04180 [astro-ph.GA].

Cooke, E. A. et al. (June 2014). "A z = 2.5 protocluster associated with the radio galaxy MRC 2104-242: star formation and differing mass functions in dense environments." In: *MNRAS* 440.4, pp. 3262–3274. DOI: 10.1093/mnras/stu522. arXiv: 1403.4259 [astro-ph.GA].

Cooke, E. A. et al. (Jan. 2016). "A Mature Galaxy Cluster at z=1.58 around the Radio Galaxy 7C1753+6311." In: *ApJ* 816.2, 83, p. 83. DOI: 10.3847/0004-637X/816/2/83. arXiv: 1511.05150 [astro-ph.GA].

Cowie, L. L. & A. Songaila (Apr. 1977). "Thermal evaporation of gas within galaxies by a hot intergalactic medium." In: *Nature* 266, pp. 501–503. DOI: 10.1038/266501a0.

Cucciati, O. et al. (Oct. 2014). "Discovery of a rich proto-cluster at z = 2.9 and associated diffuse cold gas in the VIMOS Ultra-Deep Survey (VUDS)." In: *A&A* 570, A16, A16. DOI: 10.1051/0004-6361/201423811. arXiv: 1403.3691 [astro-ph.CO].

Cucciati, O. et al. (Nov. 2018). "The progeny of a cosmic titan: a massive multi-component proto-supercluster in formation at z = 2.45 in VUDS." In: *A&A* 619, A49, A49. DOI: 10.1051/0004-6361/201833655. arXiv: 1806.06073 [astro-ph.CO].

Cui, Weiguang et al. (Nov. 2018). "The Three Hundred project: a large catalogue of theoretically modelled galaxy clusters for cosmological and astrophysical applications." In: *MNRAS* 480.3, pp. 2898–2915. DOI: 10.1093/mnras/sty2111. arXiv: 1809.04622 [astro-ph.GA].

Cullen, Lee & Walter Dehnen (Oct. 2010). "Inviscid smoothed particle hydrodynamics." In: *MNRAS* 408.2, pp. 669–683. DOI: 10.1111/j.1365-2966.2010.17158.x. arXiv: 1006.1524 [astro-ph.IM].

DES Collaboration (Mar. 2025). "Dark Energy Survey Year 3 Results: Cosmological Constraints from Cluster Abundances, Weak Lensing, and Galaxy Clustering." In: *arXiv e-prints*, arXiv:2503.13632, arXiv:2503.13632. DOI: 10.48550/arXiv.2503.13632. arXiv: 2503.13632 [astro-ph.CO].

Daddi, E. et al. (Sept. 2017). "Radio Selection of the Most Distant Galaxy Clusters." In: *ApJ Letter* 846.2, L31, p. L31. DOI: 10.3847/2041-8213/aa8808. arXiv: 1708.06910 [astro-ph.CO].

Daikuhara, Kazuki et al. (June 2024). "Star-formation activity of low-mass galaxies at the peak epoch of galaxy formation probed by deep narrow-band imaging." In: *MNRAS* 531.2, pp. 2335–2355. DOI: 10.1093/mnras/stae1243. arXiv: 2405.20663 [astro-ph.GA].

Damiano, Alice et al. (Dec. 2024). "Dynamical friction and the evolution of black holes in cosmological simulations: A new implementation in OpenGadget3." In: *A&A* 692, A81, A81. DOI: 10.1051/0004-6361/202450021. arXiv: 2403.12600 [astro-ph.CO].

Dannerbauer, H. et al. (Oct. 2014). "An excess of dusty starbursts related to the Spiderweb galaxy." In: *A&A* 570, A55, A55. DOI: 10.1051/0004-6361/201423771. arXiv: 1410.3730 [astro-ph.GA].

Darvish, Behnam et al. (July 2016). "The Effects of the Local Environment and Stellar Mass on Galaxy Quenching to z ~ 3." In: *ApJ* 825.2, 113, p. 113. DOI: 10.3847/0004-637X/825/2/113. arXiv: 1605.03182 [astro-ph.GA].

Davies, L. J. M. et al. (Mar. 2019). "Galaxy and Mass Assembly (GAMA): environmental quenching of centrals and satellites in groups." In: *MNRAS* 483.4, pp. 5444–5458. DOI: 10.1093/mnras/sty3393. arXiv: 1901.01640 [astro-ph.GA].

Davis, M. et al. (May 1985). "The evolution of large-scale structure in a universe dominated by cold dark matter." In: *ApJ* 292, pp. 371–394. DOI: 10.1086/163168.

Debuhr, Jackson, Eliot Quataert & Chung-Pei Ma (Apr. 2011). "The growth of massive black holes in galaxy merger simulations with feedback by radiation pressure." In: *MNRAS* 412.2, pp. 1341–1360. DOI: 10.1111/j.1365-2966.2010.17992.x. arXiv: 1006.3312 [astro-ph.CO].

Di Mascolo, Luca et al. (Mar. 2023). "Forming intracluster gas in a galaxy protocluster at a redshift of 2.16." In: *Nature* 615.7954, pp. 809–812. DOI: 10.1038/s41586-023-05761-x. arXiv: 2303.16226 [astro-ph.CO].

Di Matteo, Tiziana et al. (Mar. 2008). "Direct Cosmological Simulations of the Growth of Black Holes and Galaxies." In: *ApJ* 676.1, pp. 33–53. DOI: 10.1086/524921. arXiv: 0705.2269 [astro-ph].

Di Valentino, Eleonora et al. (Sept. 2021a). "Cosmology Intertwined II: The hubble constant tension." In: *Astroparticle Physics* 131, 102605, p. 102605. DOI: 10.1016/j.astropartphys.2021.102605. arXiv: 2008.11284 [astro-ph.CO].

Di Valentino, Eleonora et al. (Sept. 2021b). "Cosmology Intertwined III: $f\sigma_8$ and S_8." In: *Astroparticle Physics* 131, 102604, p. 102604. DOI: 10.1016/j.astropartphys.2021.102604. arXiv: 2008.11285 [astro-ph.CO].

Dolag, K. et al. (Oct. 2005). "The imprints of local superclusters on the Sunyaev-Zel'dovich signals and their detectability with Planck." In: *MNRAS* 363.1, pp. 29–39. DOI: 10.1111/j.1365-2966.2005.09452.x. arXiv: astro-ph/0505258 [astro-ph].

Dolag, K. et al. (Feb. 2008). "Simulation Techniques for Cosmological Simulations." In: *Space Science Reviews* 134.1-4, pp. 229–268. DOI: 10.1007/s11214-008-9316-5. arXiv: 0801.1023 [astro-ph].

Dolag, K. et al. (Oct. 2009). "Substructures in hydrodynamical cluster simulations." In: *MNRAS* 399.2, pp. 497–514. DOI: 10.1111/j.1365-2966.2009.15034.x. arXiv: 0808.3401 [astro-ph].

Donahue, Megan (1993). "Evolution of Clusters and the Intracluster Medium." In: *The Environment and Evolution of Galaxies: Proceedings of the Third Tetons Summer School Held in Grand Teton National Park, Wyoming, U.S.A., July 1992*. Ed. by J. Michael Shull & Harley A. Thronson. Dordrecht: Springer Netherlands, pp. 409–432. ISBN: 978-94-011-1882-8. DOI: 10.1007/978-94-011-1882-8_23. URL: https://doi.org/10.1007/978-94-011-1882-8_23.

Donnari, Martina et al. (June 2019). "The star formation activity of IllustrisTNG galaxies: main sequence, UVJ diagram, quenched fractions, and systematics." In: *MNRAS* 485.4, pp. 4817–4840. DOI: 10.1093/mnras/stz712. arXiv: 1812.07584 [astro-ph.GA].

Donnari, Martina et al. (Oct. 2021). "Quenched fractions in the IllustrisTNG simulations: comparison with observations and other theoretical models." In: *MNRAS* 506.4, pp. 4760–4780. DOI: 10.1093/mnras/stab1950. arXiv: 2008.00004 [astro-ph.GA].

Dressler, A. (Jan. 1984). "The Evolution of Galaxies in Clusters." In: *ARA&A* 22, pp. 185–222. DOI: 10.1146/annurev.astro.22.1.185.

Drouart, G. et al. (June 2014). "Rapidly growing black holes and host galaxies in the distant Universe from the Herschel Radio Galaxy Evolution Project." In: *A&A* 566, A53, A53. DOI: 10.1051/0004-6361/201323310. arXiv: 1404.1080 [astro-ph.GA].

Ebeling, H., A. C. Edge & J. P. Henry (June 2001). "MACS: A Quest for the Most Massive Galaxy Clusters in the Universe." In: *ApJ* 553.2, pp. 668–676. DOI: 10.1086/320958. arXiv: astro-ph/0009101 [astro-ph].

Eckert, D. et al. (Mar. 2017). "The XMM cluster outskirts project (X- COP)." In: *Astronomische Nachrichten* 338.293, pp. 293–298. DOI: 10.1002/asna.201713345. arXiv: 1611.05051 [astro-ph.CO].

Edward, Adit H. et al. (Jan. 2024). "The stellar mass function of quiescent galaxies in $2 < z < 2.5$ protoclusters." In: *MNRAS* 527.3, pp. 8598–8617. DOI: 10.1093/mnras/stad3751. arXiv: 2312.12380 [astro-ph.GA].

Eisenhardt, Peter R. M. et al. (Sept. 2008). "Clusters of Galaxies in the First Half of the Universe from the IRAC Shallow Survey." In: *ApJ* 684.2, pp. 905–932. DOI: 10.1086/590105. arXiv: 0804.4798 [astro-ph].

Eisenstein, Daniel J. et al. (Nov. 2005). "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies." In: *ApJ* 633.2, pp. 560–574. DOI: 10.1086/466512. arXiv: astro-ph/0501171 [astro-ph].

Elahi, Pascal J., Robert J. Thacker & Lawrence M. Widrow (Nov. 2011). "Peaks above the Maxwellian Sea: a new approach to finding substructures in N-body haloes." In: *Monthly Notices of the Royal Astronomical Society* 418.1, pp. 320–335. DOI: 10.1111/j.1365-2966.2011.19485.x. arXiv: 1107.4289 [astro-ph.CO].

Elbaz, D. et al. (June 2007). "The reversal of the star formation-density relation in the distant universe." In: *A&A* 468.1, pp. 33–48. DOI: 10.1051/0004-6361:20077525. arXiv: astro-ph/0703653 [astro-ph].

Esposito, Michela, Stefano Borgani & Giuseppe Murante (Sept. 2024). "SubDLe: Identification of substructures in cosmological simulations with deep learning. An image segmentation approach to substructure finding." In: *A&A* 689, A33, A33. DOI: 10.1051/0004-6361/202449961. arXiv: 2405.18257 [astro-ph.CO].

Esposito, Michela et al. (May 2025). "Galaxy populations of protoclusters in cosmological hydrodynamical simulations." In: *A&A* 697, A142, A142. DOI: 10.1051/0004-6361/202453507. arXiv: 2503.01674 [astro-ph.GA].

Fabian, A. C. (Jan. 1994). "Cooling Flows in Clusters of Galaxies." In: *ARA&A* 32, pp. 277–318. DOI: 10.1146/annurev.aa.32.090194.001425.

Fabjan, D. et al. (Jan. 2010). "Simulating the effect of active galactic nuclei feedback on the metal enrichment of galaxy clusters." In: *MNRAS* 401.3, pp. 1670–1690. DOI: 10.1111/j.1365-2966.2009.15794.x. arXiv: 0909.0664 [astro-ph.CO].

Fasano, G. et al. (Jan. 2006). "WINGS: a WIde-field Nearby Galaxy-cluster Survey. I. Optical imaging." In: *A&A* 445.3, pp. 805–817. DOI: 10.1051/0004-6361:20053816. arXiv: astro-ph/0507247 [astro-ph].

Ferland, G. J. et al. (July 1998). "CLOUDY 90: Numerical Simulation of Plasmas and Their Spectra." In: *Publications of the ASP* 110.749, pp. 761–778. DOI: 10.1086/316190.

Fontanot, Fabio et al. (Sept. 2024). "Galaxy clustering in a new implementation of the GAEA semi-analytical galaxy formation model." In: *arXiv e-prints*, arXiv:2409.02194, arXiv:2409.02194, A&A submitted. DOI: 10.48550/arXiv.2409.02194. arXiv: 2409.02194 [astro-ph.GA].

Forrest, Ben et al. (Aug. 2024). "Environmental Effects on the Stellar Mass Function in a $z \sim 3.3$ Overdensity of Galaxies in the COSMOS Field." In: *ApJ* 971.2, 169, p. 169. DOI: 10.3847/1538-4357/ad5e78. arXiv: 2405.18491 [astro-ph.GA].

Fossati, M. et al. (June 2018). "A Virgo Environmental Survey Tracing Ionised Gas Emission (VESTIGE). II. Constraining the quenching time in the stripped galaxy NGC 4330." In: *A&A* 614, A57, A57. DOI: 10.1051/0004-6361/201732373. arXiv: 1801.09685 [astro-ph.GA].

Franx, Marijn et al. (Dec. 2008). "Structure and Star Formation in Galaxies out to $z = 3$: Evidence for Surface Density Dependent Evolution and Upsizing." In: *ApJ* 688.2, pp. 770–788. DOI: 10.1086/592431. arXiv: 0808.2642 [astro-ph].

Fraser-McKelvie, Amelia et al. (Feb. 2018). "Multiple mechanisms quench passive spiral galaxies." In: *MNRAS* 474.2, pp. 1909–1921. DOI: 10.1093/mnras/stx2823. arXiv: 1710.10843 [astro-ph.GA].

Fujita, Yutaka, Tomoaki Matsumoto & Keiichi Wada (Sept. 2004). "Strong Turbulence in the Cool Cores of Galaxy Clusters: Can Tsunamis Solve the Cooling Flow Problem?" In: *ApJ Letter* 612.1, pp. L9–L12. DOI: 10.1086/424483. arXiv: astro-ph/0407368 [astro-ph].

Galbiati, M. et al. (Apr. 2025). "Connecting the growth of galaxies to the large-scale environment in a massive node of the Cosmic Web at $z \sim 3$." In: *A&A* 696, A95, A95. DOI: 10.1051/0004-6361/202452493. arXiv: 2410.03822 [astro-ph.GA].

Gaspari, M. (July 2015). "Shaping the X-ray spectrum of galaxy clusters with AGN feedback and turbulence." In: *MNRAS* 451, pp. L60–L64. DOI: 10.1093/mnrasl/slv067. arXiv: 1410.7769 [astro-ph.HE].

Gaspari, M. et al. (Oct. 2019). "The X-Ray Halo Scaling Relations of Supermassive Black Holes." In: *ApJ* 884.2, 169, p. 169. DOI: 10.3847/1538-4357/ab3c5d. arXiv: 1904.10972 [astro-ph.GA].

Gavazzi, G. et al. (July 2010). "A snapshot on galaxy evolution occurring in the Great Wall: the role of Nurture at $z = 0$." In: *A&A* 517, A73, A73. DOI: 10.1051/0004-6361/201014153. arXiv: 1003.3795 [astro-ph.CO].

Geha, M. et al. (Sept. 2012). "A Stellar Mass Threshold for Quenching of Field Galaxies." In: *ApJ* 757.1, 85, p. 85. DOI: 10.1088/0004-637X/757/1/85. arXiv: 1206.3573 [astro-ph.CO].

Ghirardini, Vittorio et al. (Jan. 2019). "Universal thermodynamic properties of the intracluster medium over two decades in radius in the X-COP sample." In: *A&A* 621, A41, A41. DOI: 10.1051/0004-6361/201833325. arXiv: 1805.00042 [astro-ph.CO].

Ghirardini, Vittorio et al. (Mar. 2021). "Evolution of the Thermodynamic Properties of Clusters of Galaxies out to Redshift of 1.8." In: *ApJ* 910.1, 14, p. 14. DOI: 10.3847/1538-4357/abc68d. arXiv: 2004.04747 [astro-ph.CO].

Gilbank, David G. et al. (Mar. 2011). "The Red-sequence Cluster Survey-2 (RCS-2): Survey Details and Photometric Catalog Construction." In: *The Astron. J.* 141.3, 94, p. 94. DOI: 10.1088/0004-6256/141/3/94. arXiv: 1012.3470 [astro-ph.CO].

Gingold, R. A. & J. J. Monaghan (June 1982). "Kernel estimates as a basis for general particle methods in hydrodynamics." In: *Journal of Computational Physics* 46 (3), pp. 429–453.

Gjergo, Eda et al. (Sept. 2018). "Dust evolution in galaxy cluster simulations." In: *MNRAS* 479.2, pp. 2588–2606. DOI: 10.1093/mnras/sty1564. arXiv: 1804.06855 [astro-ph.GA].

Gladders, Michael D. & H. K. C. Yee (Mar. 2005). "The Red-Sequence Cluster Survey. I. The Survey and Cluster Catalogs for Patches RCS 0926+37 and RCS 1327+29." In: *ApJ, Supplement* 157.1, pp. 1–29. DOI: 10.1086/427327. arXiv: astro-ph/0411075 [astro-ph].

Gobat, R. et al. (Feb. 2011). "A mature cluster with X-ray emission at z = 2.07." In: *A&A* 526, A133, A133. DOI: 10.1051/0004-6361/201016084. arXiv: 1011.1837 [astro-ph.CO].

Gobat, R. et al. (Oct. 2013). "WFC3 GRISM Confirmation of the Distant Cluster Cl J1449+0856 at langzrang = 2.00: Quiescent and Star-forming Galaxy Populations." In: *ApJ* 776.1, 9, p. 9. DOI: 10.1088/0004-637X/776/1/9. arXiv: 1305.3576 [astro-ph.CO].

González Villalba, Justo Antonio, Klaus Dolag & Veronica Biffi (Feb. 2025). "How the cool-core population transitions from galaxy groups to massive clusters: A comparison of the largest Magneticum simulation with eROSITA, XMM-Newton, Chandra and LOFAR observations." In: *A&A* 694, A232, A232. DOI: 10.1051/0004-6361/202452203. arXiv: 2412.13182 [astro-ph.CO].

Granato, G. L. et al. (Jan. 2001). "Modelling the Extinction Properties of Galaxies." In: *Astrophysics and Space Science Supplement* 277, pp. 79–82. DOI: 10.1023/A:1012783728339.

Granato, Gian Luigi et al. (June 2015). "The early phases of galaxy clusters formation in IR: coupling hydrodynamical simulations with GRASIL-3D." In: *MNRAS* 450.2, pp. 1320–1332. DOI: 10.1093/mnras/stv676. arXiv: 1412.6105 [astro-ph.GA].

Granato, Gian Luigi et al. (May 2021). "Dust evolution in zoom-in cosmological simulations of galaxy formation." In: *MNRAS* 503.1, pp. 511–532. DOI: 10.1093/mnras/stab362. arXiv: 2010.05919 [astro-ph.GA].

Greenslade, J. et al. (May 2018). "Candidate high-z protoclusters among the Planck compact sources, as revealed by Herschel-SPIRE." In: *MNRAS* 476.3, pp. 3336–3359. DOI: 10.1093/mnras/sty023. arXiv: 1712.07141 [astro-ph.GA].

Greggio, L. & A. Renzini (Feb. 1983). "The binary model for type I supernovae - Theoretical rates." In: *A&A* 118.2, pp. 217–222.

Groth, Frederick et al. (Nov. 2023). "The cosmological simulation code OPENGADGET3 - implementation of meshless finite mass." In: *MNRAS* 526.1, pp. 616–644. DOI: 10.1093/mnras/stad2717. arXiv: 2301.03612 [astro-ph.IM].

Grützbauch, Ruth et al. (Apr. 2011). "Galaxy properties in different environments up to z∼ 3 in the GOODS NICMOS Survey." In: *MNRAS* 412.4, pp. 2361–2375. DOI: 10.1111/j.1365-2966.2010.18060.x. arXiv: 1011.4846 [astro-ph.CO].

Guaita, L. et al. (Aug. 2020). "The VANDELS survey: Discovery of massive overdensities of galaxies at z > 2. Location of Lyα-emitting galaxies with respect to environment." In: *A&A* 640, A107, A107. DOI: 10.1051/0004-6361/201935855. arXiv: 2007.12314 [astro-ph.GA].

Gullieuszik, M. et al. (Sept. 2015). "OmegaWINGS: OmegaCAM-VST observations of WINGS galaxy clusters." In: *A&A* 581, A41, A41. DOI: 10.1051/0004-6361/201526061. arXiv: 1503.02628 [astro-ph.GA].

Gunn, James E. & J. Richard Gott III (Aug. 1972). "On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution." In: *ApJ* 176, p. 1. DOI: 10.1086/151605.

Guo, Fulai & S. Peng Oh (Dec. 2009). "Could AGN outbursts transform cool core clusters?" In: *MNRAS* 400.4, pp. 1992–1999. DOI: 10.1111/j.1365-2966.2009.15592.x. arXiv: 0907.3141 [astro-ph.CO].

Gursky, H. et al. (Aug. 1971). "A Strong X-Ray Source in the Coma Cluster Observed by UHURU." In: *ApJ Letter* 167, p. L81. DOI: 10.1086/180765.

Hahn, Oliver et al. (Sept. 2017). "rhapsody-g simulations - I. The cool cores, hot gas and stellar content of massive galaxy clusters." In: *MNRAS* 470.1, pp. 166–186. DOI: 10.1093/mnras/stx001. arXiv: 1509.04289 [astro-ph.CO].

Haines, C. P. et al. (June 2015). "LoCuSS: The Slow Quenching of Star Formation in Cluster Galaxies and the Need for Pre-processing." In: *ApJ* 806.1, 101, p. 101. DOI: 10.1088/0004-637X/806/1/101. arXiv: 1504.05604 [astro-ph.GA].

Hatch, N. A. et al. (Jan. 2011a). "Galaxy protocluster candidates around z∼ 2.4 radio galaxies." In: *MNRAS* 410.3, pp. 1537–1549. DOI: 10.1111/j.1365-2966.2010.17538.x. arXiv: 1008.4588 [astro-ph.CO].

Hatch, N. A. et al. (Aug. 2011b). "Hα emitters in z∼ 2 protoclusters: evidence for faster evolution in dense environments." In: *MNRAS* 415.4, pp. 2993–3005. DOI: 10.1111/j.1365-2966.2011.18735.x. arXiv: 1103.4364 [astro-ph.CO].

Hayashi, Masao et al. (Sept. 2012). "A Starbursting Proto-cluster in Making Associated with a Radio Galaxy at z = 2.53 Discovered by Hα Imaging." In: *ApJ* 757.1, 15, p. 15. DOI: 10.1088/0004-637X/757/1/15. arXiv: 1207.2614 [astro-ph.CO].

Hilton, M. et al. (Mar. 2021). "The Atacama Cosmology Telescope: A Catalog of >4000 Sunyaev-Zel'dovich Galaxy Clusters." In: *ApJ, Supplement* 253.1, 3, p. 3. DOI: 10.3847/1538-4365/abd023. arXiv: 2009.11043 [astro-ph.CO].

Hirschmann, Michaela, Gabriella De Lucia & Fabio Fontanot (Sept. 2016). "Galaxy assembly, stellar feedback and metal enrichment: the view from the GAEA model." In: *MNRAS* 461.2, pp. 1760–1785. DOI: 10.1093/mnras/stw1318. arXiv: 1512.04531 [astro-ph.GA].

Hirschmann, Michaela et al. (Aug. 2014). "Cosmological simulations of black hole growth: AGN luminosities and downsizing." In: *MNRAS* 442.3, pp. 2304–2324. DOI: 10.1093/mnras/stu1023. arXiv: 1308.0333 [astro-ph.CO].

Hoyle, F. & R. A. Lyttleton (Jan. 1939). "The effect of interstellar matter on climatic variation." In: *Proceedings of the Cambridge Philosophical Society* 35.3, p. 405. DOI: 10.1017/S0305004100021150.

Hubble, E. P. (Dec. 1926). "Extragalactic nebulae." In: *ApJ* 64, pp. 321–369. DOI: 10.1086/143018.

Hubble, Edwin (Mar. 1929). "A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae." In: *Proceedings of the National Academy of Science* 15.3, pp. 168–173. DOI: 10.1073/pnas.15.3.168.

Ito, Kei et al. (Mar. 2023). "COSMOS2020: Discovery of a Protocluster of Massive Quiescent Galaxies at z = 2.77." In: *ApJ Letter* 945.1, L9, p. L9. DOI: 10.3847/2041-8213/acb49b. arXiv: 2301.08845 [astro-ph.GA].

Jones, A. P. et al. (June 2017). "The global dust modelling framework THEMIS." In: *A&A* 602, A46, A46. DOI: 10.1051/0004-6361/201630225. arXiv: 1703.00775 [astro-ph.GA].

Kaiser, N. (Sept. 1986). "Evolution and clustering of rich clusters." In: *MNRAS* 222, pp. 323–345. DOI: 10.1093/mnras/222.2.323.

Karakas, A. I. (Apr. 2010). "Updated stellar yields from asymptotic giant branch models." In: *MNRAS* 403.3, pp. 1413–1425. DOI: 10.1111/j.1365-2966.2009.16198.x. arXiv: 0912.2142 [astro-ph.SR].

Kauffmann, Guinevere et al. (Sept. 2004). "The environmental dependence of the relations between stellar mass, structure, star formation and nuclear activity in galaxies." In: *MNRAS* 353.3, pp. 713–731. DOI: 10.1111/j.1365-2966.2004.08117.x. arXiv: astro-ph/0402030 [astro-ph].

Kawinwanichakij, Lalitwadee et al. (Oct. 2017). "Effect of Local Environment and Stellar Mass on Galaxy Quenching and Morphology at 0.5 < z < 2.0." In: *ApJ* 847.2, 134, p. 134. DOI: 10.3847/1538-4357/aa8b75. arXiv: 1706.03780 [astro-ph.GA].

Kennicutt Jr., Robert C. (Sept. 1989). "The Star Formation Law in Galactic Disks." In: *ApJ* 344, p. 685. DOI: 10.1086/167834.

Klein, M., J. J. Mohr & C. T. Davies (Oct. 2024a). "The ACT-DR5 MCMF galaxy cluster catalog." In: *A&A* 690, A322, A322. DOI: 10.1051/0004-6361/202451203. arXiv: 2406.14754 [astro-ph.CO].

Klein, M. et al. (July 2024b). "SPT-SZ MCMF: an extension of the SPT-SZ catalogue over the DES region." In: *MNRAS* 531.4, pp. 3973–3990. DOI: 10.1093/mnras/stae1359. arXiv: 2309.09908 [astro-ph.CO].

Kluge, M. et al. (Aug. 2024). "The SRG/eROSITA All-Sky Survey. Optical identification and properties of galaxy clusters and groups in the western galactic hemisphere." In: *A&A* 688, A210, A210. DOI: 10.1051/0004-6361/202349031. arXiv: 2402.08453 [astro-ph.CO].

Knebe, Alexander et al. (Aug. 2011). "Haloes gone MAD: The Halo-Finder Comparison Project." In: *MNRAS* 415.3, pp. 2293–2318. DOI: 10.1111/j.1365-2966.2011.18858.x. arXiv: 1104.0949 [astro-ph.CO].

Knebe, Alexander et al. (Jan. 2013a). "Galaxies going MAD: the Galaxy-Finder Comparison Project." In: *MNRAS* 428.3, pp. 2039–2052. DOI: 10.1093/mnras/sts173. arXiv: 1210.2578 [astro-ph.CO].

Knebe, Alexander et al. (Oct. 2013b). "Structure finding in cosmological simulations: the state of affairs." In: *MNRAS* 435.2, pp. 1618–1658. DOI: 10.1093/mnras/stt1403. arXiv: 1304.0585 [astro-ph.CO].

Koyama, Yusei et al. (Jan. 2013). "Massive starburst galaxies in a z = 2.16 proto-cluster unveiled by panoramic Hα mapping." In: *MNRAS* 428.2, pp. 1551–1564. DOI: 10.1093/mnras/sts133. arXiv: 1210.0972 [astro-ph.CO].

Koyama, Yusei et al. (May 2021). "A Planck-selected dusty proto-cluster at z = 2.16 associated with a strong overdensity of massive Hα-emitting galaxies." In: *MNRAS* 503.1, pp. L1–L5. DOI: 10.1093/mnrasl/slab013. arXiv: 2008.13614 [astro-ph.GA].

Kravtsov, Andrey V. & Stefano Borgani (Sept. 2012). "Formation of Galaxy Clusters." In: *ARA&A* 50, pp. 353–409. DOI: 10.1146/annurev-astro-081811-125502. arXiv: 1205.5556 [astro-ph.CO].

Kroupa, Pavel (Apr. 2001). "On the variation of the initial mass function." In: *MNRAS* 322.2, pp. 231–246. DOI: 10.1046/j.1365-8711.2001.04022.x. arXiv: astro-ph/0009005 [astro-ph].

Kugel, Roi et al. (Dec. 2023). "FLAMINGO: calibrating large cosmological hydrodynamical simulations with machine learning." In: *MNRAS* 526.4, pp. 6103–6127. DOI: 10.1093/mnras/stad2540. arXiv: 2306.05492 [astro-ph.CO].

Kuiper, E. et al. (Oct. 2011). "Discovery of a high-z protocluster with tunable filters: the case of 6C0140+326 at z= 4.4." In: *MNRAS* 417.2, pp. 1088–1097. DOI: 10.1111/j.1365-2966.2011.19324.x. arXiv: 1106.5495 [astro-ph.CO].

Kurk, J. D. et al. (June 2000). "A Search for clusters at high redshift. I. Candidate Lyalpha emitters near 1138-262 at z=2.2." In: *A&A* 358, pp. L1–L4. DOI: 10.48550/arXiv.astro-ph/0005058. arXiv: astro-ph/0005058 [astro-ph].

Kurk, J. D. et al. (Dec. 2004). "A search for clusters at high redshift. III. Candidate Hα emitters and EROs in the PKS 1138-262 proto-cluster at z

= 2.16.” In: *A&A* 428, pp. 793–815. DOI: 10.1051/0004-6361:20040075. arXiv: astro-ph/0410202 [astro-ph].

Labbé, Ivo et al. (May 2005). “IRAC Mid-Infrared Imaging of the Hubble Deep Field-South: Star Formation Histories and Stellar Masses of Red Galaxies at z>2.” In: *ApJ Letter* 624.2, pp. L81–L84. DOI: 10.1086/430700. arXiv: astro-ph/0504219 [astro-ph].

Lacey, C. & S. Cole (Dec. 1994). “Merger Rates in Hierarchical Models of Galaxy Formation - Part Two - Comparison with N-Body Simulations.” In: *MNRAS* 271, p. 676. DOI: 10.1093/mnras/271.3.676. arXiv: astro-ph/9402069 [astro-ph].

Larson, R. B., B. M. Tinsley & C. N. Caldwell (May 1980). “The evolution of disk galaxies and the origin of S0 galaxies.” In: *ApJ* 237, pp. 692–707. DOI: 10.1086/157917.

Le Fevre, O. et al. (Nov. 1996). “Clustering around the Radio Galaxy MRC 0316-257 at Z = 3.14.” In: *ApJ Letter* 471, p. L11. DOI: 10.1086/310319. arXiv: astro-ph/9609020 [astro-ph].

Leccardi, A. & S. Molendi (Aug. 2008). “Radial temperature profiles for a large sample of galaxy clusters observed with XMM-Newton.” In: *A&A* 486.2, pp. 359–373. DOI: 10.1051/0004-6361:200809538. arXiv: 0804.1909 [astro-ph].

Lee-Brown, Donald B. et al. (July 2017). “The Ages of Passive Galaxies in a z = 1.62 Protocluster.” In: *ApJ* 844.1, 43, p. 43. DOI: 10.3847/1538-4357/aa7948. arXiv: 1706.05017 [astro-ph.GA].

Lee, Khee-Gan et al. (Nov. 2014). “Lyα Forest Tomography from Background Galaxies: The First Megaparsec-resolution Large-scale Structure Map at z > 2.” In: *ApJ Letter* 795.1, L12, p. L12. DOI: 10.1088/2041-8205/795/1/L12. arXiv: 1409.5632 [astro-ph.CO].

Lehle, Katrin, Dylan Nelson & Annalisa Pillepich (Oct. 2025). “What drives cluster cool-core transformations? A population-level analysis using TNG-Cluster.” In: *A&A* 702, A25, A25. DOI: 10.1051/0004-6361/202554374.

Lemaux, B. C. et al. (July 2018). “The VIMOS Ultra-Deep Survey: Emerging from the dark, a massive proto-cluster at z 4.57.” In: *A&A* 615, A77, A77. DOI: 10.1051/0004-6361/201730870. arXiv: 1703.10170 [astro-ph.GA].

Lepore, M. et al. (Feb. 2024). “Feeding and feedback processes in the Spiderweb proto-intracluster medium.” In: *A&A* 682, A186, A186. DOI: 10.1051/0004-6361/202347538. arXiv: 2312.06392 [astro-ph.GA].

Leroy, L. et al. (Nov. 2024). “The role of stellar mass in the cosmic history of star formation as seen by Herschel and ALMA.” In: *A&A* 691, A248, A248. DOI: 10.1051/0004-6361/202348341.

Lewis, Ian et al. (Aug. 2002). “The 2dF Galaxy Redshift Survey: the environmental dependence of galaxy star formation rates near clusters.” In: *MNRAS* 334.3, pp. 673–683. DOI: 10.1046/j.1365-8711.2002.05558.x. arXiv: astro-ph/0203336 [astro-ph].

Li, Qingyang et al. (July 2020). “The Three Hundred project: the stellar and gas profiles.” In: *MNRAS* 495.3, pp. 2930–2948. DOI: 10.1093/mnras/staa1385. arXiv: 2005.06135 [astro-ph.GA].

Li, Qingyang et al. (July 2023). "THE THREE HUNDRED Project: the evolution of physical baryon profiles." In: *MNRAS* 523.1, pp. 1228–1246. DOI: 10.1093/mnras/stad1521. arXiv: 2305.09629 [astro-ph.GA].

Li, Yuan & Greg L. Bryan (Mar. 2012). "Simulating the Cooling Flow of Cool-core Clusters." In: *ApJ* 747.1, 26, p. 26. DOI: 10.1088/0004-637X/747/1/26. arXiv: 1112.2701 [astro-ph.CO].

Li, Yuan et al. (Oct. 2015). "Cooling, AGN Feedback, and Star Formation in Simulated Cool-core Galaxy Clusters." In: *ApJ* 811.2, 73, p. 73. DOI: 10.1088/0004-637X/811/2/73. arXiv: 1503.02660 [astro-ph.GA].

Lim, Seunghwan et al. (Feb. 2021). "Is there enough star formation in simulated protoclusters?" In: *MNRAS* 501.2, pp. 1803–1822. DOI: 10.1093/mnras/staa3693. arXiv: 2010.02259 [astro-ph.GA].

Lim, Seunghwan et al. (Aug. 2024). "The FLAMINGO simulation view of cluster progenitors observed in the epoch of reionization with JWST." In: *MNRAS* 532.4, pp. 4551–4569. DOI: 10.1093/mnras/stae1790. arXiv: 2402.17819 [astro-ph.GA].

Lin, Yen-Ting, Joseph J. Mohr & S. Adam Stanford (Aug. 2004). "K-Band Properties of Galaxy Clusters and Groups: Luminosity Function, Radial Distribution, and Halo Occupation Number." In: *ApJ* 610.2, pp. 745–761. DOI: 10.1086/421714. arXiv: astro-ph/0402308 [astro-ph].

Liu, Daizhong et al. (Dec. 2019). "Automated Mining of the ALMA Archive in the COSMOS Field (A^3COSMOS). II. Cold Molecular Gas Evolution out to Redshift 6." In: *ApJ* 887.2, 235, p. 235. DOI: 10.3847/1538-4357/ab578d. arXiv: 1910.12883 [astro-ph.GA].

Lower, Sidney et al. (Nov. 2020). "How Well Can We Measure the Stellar Mass of a Galaxy: The Impact of the Assumed Star Formation History Model in SED Fitting." In: *ApJ* 904.1, 33, p. 33. DOI: 10.3847/1538-4357/abbfa7. arXiv: 2006.03599 [astro-ph.GA].

Magdis, Georgios E. et al. (Nov. 2012). "The Evolving Interstellar Medium of Star-forming Galaxies since z = 2 as Probed by Their Infrared Spectral Energy Distributions." In: *ApJ* 760.1, 6, p. 6. DOI: 10.1088/0004-637X/760/1/6. arXiv: 1210.1035 [astro-ph.CO].

Magorrian, John et al. (June 1998). "The Demography of Massive Dark Objects in Galaxy Centers." In: *The Astron. J.* 115.6, pp. 2285–2305. DOI: 10.1086/300353. arXiv: astro-ph/9708072 [astro-ph].

Mantz, A. B. et al. (Nov. 2018). "The XXL Survey. XVII. X-ray and Sunyaev-Zel'dovich properties of the redshift 2.0 galaxy cluster XLSSC 122." In: *A&A* 620, A2. DOI: 10.1051/0004-6361/201630096. arXiv: 1703.08221 [astro-ph.CO].

Mantz, Adam B. et al. (Aug. 2020). "Deep XMM-Newton observations of the most distant SPT-SZ galaxy cluster." In: *MNRAS* 496.2, pp. 1554–1564. DOI: 10.1093/mnras/staa1581. arXiv: 2006.02009 [astro-ph.CO].

Marini, I. et al. (Nov. 2021). "Velocity dispersion of brightest cluster galaxies in cosmological simulations." In: *MNRAS* 507.4, pp. 5780–5795. DOI: 10.1093/mnras/stab2518. arXiv: 2109.00223 [astro-ph.GA].

Marini, I. et al. (Aug. 2022). “Machine learning to identify ICL and BCG in simulated galaxy clusters.” In: *MNRAS* 514.2, pp. 3082–3096. DOI: 10.1093/mnras/stac1558. arXiv: 2203.03360 [astro-ph.GA].
Matsuda, Y. et al. (Sept. 2012). “Diffuse Lyα haloes around Lyα emitters at z=3: do dark matter distributions determine the Lyα spatial extents?” In: *MNRAS* 425.2, pp. 878–883. DOI: 10.1111/j.1365-2966.2012.21143.x. arXiv: 1204.4934 [astro-ph.CO].
Matteucci, F. (2003). *The Chemical Evolution of the Galaxy*. Vol. 253.
Matteucci, F. & L. Greggio (Jan. 1986). “Relative roles of type I and II supernovae in the chemical enrichment of the interstellar gas.” In: *A&A* 154.1-2, pp. 279–287.
Matteucci, Francesca & Simone Recchi (Sept. 2001). “On the Typical Timescale for the Chemical Enrichment from Type Ia Supernovae in Galaxies.” In: *ApJ* 558.1, pp. 351–358. DOI: 10.1086/322472. arXiv: astro-ph/0105074 [astro-ph].
Mawatari, K. et al. (Nov. 2012). “Characterization of the Distribution of the Lyα Emitters in the 53w002 Field at z = 2.4.” In: *ApJ* 759.2, 133, p. 133. DOI: 10.1088/0004-637X/759/2/133. arXiv: 1209.1253 [astro-ph.CO].
Mazzotta, P. et al. (Oct. 2004). “Comparing the temperatures of galaxy clusters from hydrodynamical N-body simulations to Chandra and XMM-Newton observations.” In: *MNRAS* 354.1, pp. 10–24. DOI: 10.1111/j.1365-2966.2004.08167.x. arXiv: astro-ph/0404425 [astro-ph].
McCarthy, Ian G. et al. (Oct. 2004). “Models of the Intracluster Medium with Heating and Cooling: Explaining the Global and Structural X-Ray Properties of Clusters.” In: *ApJ* 613.2, pp. 811–830. DOI: 10.1086/423267. arXiv: astro-ph/0406329 [astro-ph].
McCarthy, Ian G. et al. (Mar. 2017). “The BAHAMAS project: calibrated hydrodynamical simulations for large-scale structure cosmology.” In: *MNRAS* 465.3, pp. 2936–2965. DOI: 10.1093/mnras/stw2792. arXiv: 1603.02702 [astro-ph.CO].
McConachie, Ian et al. (Feb. 2022). “Spectroscopic Confirmation of a Protocluster at z = 3.37 with a High Fraction of Quiescent Galaxies.” In: *ApJ* 926.1, 37, p. 37. DOI: 10.3847/1538-4357/ac2b9f. arXiv: 2109.07696 [astro-ph.GA].
McConnell, Nicholas J. & Chung-Pei Ma (Feb. 2013). “Revisiting the Scaling Relations of Black Hole Masses and Host Galaxy Properties.” In: *ApJ* 764.2, 184, p. 184. DOI: 10.1088/0004-637X/764/2/184. arXiv: 1211.2816 [astro-ph.CO].
McDonald, M. et al. (July 2017). “The Remarkable Similarity of Massive Galaxy Clusters from z ∼ 0 to z ∼ 1.9.” In: *ApJ* 843.1, 28, p. 28. DOI: 10.3847/1538-4357/aa7740. arXiv: 1702.05094 [astro-ph.CO].
McDonald, Michael et al. (Oct. 2010). “On the Origin of the Extended Hα Filaments in Cooling Flow Clusters.” In: *ApJ* 721.2, pp. 1262–1283. DOI: 10.1088/0004-637X/721/2/1262. arXiv: 1008.0392 [astro-ph.HE].
McDonald, Michael et al. (June 2011). “Star Formation Efficiency in the Cool Cores of Galaxy Clusters.” In: *ApJ* 734.2, 95, p. 95. DOI: 10.1088/0004-637X/734/2/95. arXiv: 1104.0665 [astro-ph.CO].

McGee, S. L., R. G. Bower & M. L. Balogh (July 2014). "Overconsumption, outflows and the quenching of satellite galaxies." In: *MNRAS* 442, pp. L105–L109. DOI: 10.1093/mnrasl/slu066. arXiv: 1404.6251 [astro-ph.GA].

McKee, C. F. & J. P. Ostriker (Nov. 1977). "A theory of the interstellar medium: three components regulated by supernova explosions in an inhomogeneous substrate." In: *ApJ* 218, pp. 148–169. DOI: 10.1086/155667.

McNamara, B. R. & P. E. J. Nulsen (Sept. 2007). "Heating Hot Atmospheres with Active Galactic Nuclei." In: *ARA&A* 45.1, pp. 117–175. DOI: 10.1146/annurev.astro.45.051806.110625. arXiv: 0709.2152 [astro-ph].

Meekins, John F. et al. (May 1971). "Physical Sciences: X-rays from the Coma Cluster of Galaxies." In: *Nature* 231.5298, pp. 107–108. DOI: 10.1038/231107a0.

Mehrtens, Nicola et al. (June 2012). "The XMM Cluster Survey: optical analysis methodology and the first data release." In: *MNRAS* 423.2, pp. 1024–1052. DOI: 10.1111/j.1365-2966.2012.20931.x. arXiv: 1106.3056 [astro-ph.CO].

Mérida, Rosa M. et al. (June 2024). "Measuring the gas reservoirs in $10^8 < M_\star < 10^{11}$ $M_\odot$ galaxies at $1 \leqslant z \leqslant 3$." In: *A&A* 686, A64, A64. DOI: 10.1051/0004-6361/202348661. arXiv: 2311.16279 [astro-ph.GA].

Miley, George K. et al. (Oct. 2006). "The Spiderweb Galaxy: A Forming Massive Cluster Galaxy at z ∼2." In: *ApJ Letter* 650.1, pp. L29–L32. DOI: 10.1086/508534. arXiv: astro-ph/0610909 [astro-ph].

Miley, George & Carlos De Breuck (Feb. 2008). "Distant radio galaxies and their environments." In: *A&ARv* 15.2, pp. 67–144. DOI: 10.1007/s00159-007-0008-z. arXiv: 0802.2770 [astro-ph].

Miller, T. B. et al. (Apr. 2018). "A massive core for a cluster of galaxies at a redshift of 4.3." In: *Nature* 556.7702, pp. 469–472. DOI: 10.1038/s41586-018-0025-2. arXiv: 1804.09231 [astro-ph.GA].

Mo, H., F. van den Bosch & S. White (2012). *Galaxy Formation and Evolution*. Cambridge University Press.

Mobasher, Bahram et al. (July 2015). "A Critical Assessment of Stellar Mass Measurement Methods." In: *ApJ* 808.1, 101, p. 101. DOI: 10.1088/0004-637X/808/1/101. arXiv: 1505.01501 [astro-ph.GA].

Molendi, Silvano & Fabio Pizzolato (Oct. 2001). "Is the Gas in Cooling Flows Multiphase?" In: *ApJ* 560.1, pp. 194–200. DOI: 10.1086/322387. arXiv: astro-ph/0106552 [astro-ph].

Monaghan, J. J. (Jan. 1992). "Smoothed particle hydrodynamics." In: *ARA&A* 30, pp. 543–574. DOI: 10.1146/annurev.aa.30.090192.002551.

— (Sept. 1997). "SPH and Riemann Solvers." In: *Journal of Computational Physics* 136.2, pp. 298–307. DOI: 10.1006/jcph.1997.5732.

Monaghan, J. J. & R. A. Gingold (Nov. 1983). "Shock simulation by the particle method SPH." In: *Journal of Computational Physics* 52 (2), pp. 374–389.

Monson, Erik B. et al. (Sept. 2021). "On the Nature of AGN and Star Formation Enhancement in the z = 3.1 SSA22 Protocluster: The HST WFC3 IR

View." In: *ApJ* 919.1, 51, p. 51. DOI: 10.3847/1538-4357/ac0f84. arXiv: 2106.13846 [astro-ph.GA].

Montes, Mireia (Mar. 2022). "The faint light in groups and clusters of galaxies." In: *Nature Astronomy* 6, pp. 308–316. DOI: 10.1038/s41550-022-01616-z. arXiv: 2203.06199 [astro-ph.GA].

Moore, Ben et al. (Feb. 1996). "Galaxy harassment and the evolution of clusters of galaxies." In: *Nature* 379.6566, pp. 613–616. DOI: 10.1038/379613a0. arXiv: astro-ph/9510034 [astro-ph].

Moretti, A. et al. (Apr. 2014). "WINGS Data Release: a database of galaxies in nearby clusters." In: *A&A* 564, A138, A138. DOI: 10.1051/0004-6361/201323098. arXiv: 1403.1408 [astro-ph.GA].

Mullis, C. R. et al. (Apr. 2005). "Discovery of an X-Ray-luminous Galaxy Cluster at z=1.4." In: *ApJ Letter* 623.2, pp. L85–L88. DOI: 10.1086/429801. arXiv: astro-ph/0503004 [astro-ph].

Murante, Giuseppe et al. (July 2010). "A subresolution multiphase interstellar medium model of star formation and supernova energy feedback." In: *MNRAS* 405.3, pp. 1491–1512. DOI: 10.1111/j.1365-2966.2010.16567.x. arXiv: 1002.4122 [astro-ph.CO].

Murante, Giuseppe et al. (Feb. 2015). "Simulating realistic disc galaxies with a novel sub-resolution ISM model." In: *MNRAS* 447.1, pp. 178–201. DOI: 10.1093/mnras/stu2400.

Muzzin, Adam et al. (June 2009). "Spectroscopic Confirmation of Two Massive Red-Sequence-Selected Galaxy Clusters at z ~1.2 in the SpARCS-North Cluster Survey." In: *ApJ* 698.2, pp. 1934–1942. DOI: 10.1088/0004-637X/698/2/1934. arXiv: 0810.0005 [astro-ph].

Muzzin, Adam et al. (Feb. 2012). "The Gemini Cluster Astrophysics Spectroscopic Survey (GCLASS): The Role of Environment and Self-regulation in Galaxy Evolution at z ~1." In: *ApJ* 746.2, 188, p. 188. DOI: 10.1088/0004-637X/746/2/188. arXiv: 1112.3655 [astro-ph.CO].

Muzzin, Adam et al. (Nov. 2013). "The Evolution of the Stellar Mass Functions of Star-forming and Quiescent Galaxies to z = 4 from the COSMOS/UltraVISTA Survey." In: *ApJ* 777.1, 18, p. 18. DOI: 10.1088/0004-637X/777/1/18. arXiv: 1303.4409 [astro-ph.CO].

Nantais, Julie B. et al. (Feb. 2017). "Evidence for strong evolution in galaxy environmental quenching efficiency between z = 1.6 and z = 0.9." In: *MNRAS* 465.1, pp. L104–L108. DOI: 10.1093/mnrasl/slw224. arXiv: 1610.08058 [astro-ph.GA].

Naufal, Abdurrahman et al. (Dec. 2024). "Revealing the Quiescent Galaxy Population in the Spiderweb Protocluster at z = 2.16 with Deep HST/WFC3 Slitless Spectroscopy." In: *ApJ* 977.1, 58, p. 58. DOI: 10.3847/1538-4357/ad8dcf. arXiv: 2410.16643 [astro-ph.GA].

Nelson, Dylan et al. (June 2024). "Introducing the TNG-Cluster simulation: Overview and the physical properties of the gaseous intracluster medium." In: *A&A* 686, A157, A157. DOI: 10.1051/0004-6361/202348608. arXiv: 2311.06338 [astro-ph.GA].

Newman, Andrew B. et al. (June 2014). "Spectroscopic Confirmation of the Rich z = 1.80 Galaxy Cluster JKCS 041 using the WFC3 Grism:

Environmental Trends in the Ages and Structure of Quiescent Galaxies." In: *ApJ* 788.1, 51, p. 51. DOI: 10.1088/0004-637X/788/1/51. arXiv: 1310.6754 [astro-ph.CO].

Noeske, K. G. et al. (May 2007). "Star Formation in AEGIS Field Galaxies since z=1.1: The Dominance of Gradually Declining Star Formation, and the Main Sequence of Star-forming Galaxies." In: *ApJ Letter* 660.1, pp. L43–L46. DOI: 10.1086/517926. arXiv: astro-ph/0701924 [astro-ph].

Nomoto, Ken'ichi, Chiaki Kobayashi & Nozomu Tominaga (Aug. 2013). "Nucleosynthesis in Stars and the Chemical Enrichment of Galaxies." In: *ARA&A* 51.1, pp. 457–509. DOI: 10.1146/annurev-astro-082812-140956.

Noordeh, E. et al. (Nov. 2021). "Quiescent galaxies in a virialized cluster at redshift 2: evidence for accelerated size growth." In: *MNRAS* 507.4, pp. 5272–5280. DOI: 10.1093/mnras/stab2459. arXiv: 2109.02200 [astro-ph.GA].

Nulsen, P. E. J. (Mar. 1982). "Transport processes and the stripping of cluster galaxies." In: *MNRAS* 198, pp. 1007–1016. DOI: 10.1093/mnras/198.4.1007.

Old, Lyndsay J. et al. (Apr. 2020). "The GOGREEN survey: the environmental dependence of the star-forming galaxy main sequence at $1.0 < z < 1.5$." In: *MNRAS* 493.4, pp. 5987–6000. DOI: 10.1093/mnras/staa579. arXiv: 2002.11735 [astro-ph.GA].

Onions, Julian et al. (June 2012). "Subhaloes going Notts: the subhalo-finder comparison project." In: *MNRAS* 423.2, pp. 1200–1214. DOI: 10.1111/j.1365-2966.2012.20947.x. arXiv: 1203.3695 [astro-ph.CO].

Onions, Julian et al. (Sept. 2025). "The life and times of dark matter haloes: what will I be when I grow up?" In: *MNRAS* 542.2, pp. 1477–1485. DOI: 10.1093/mnras/staf1293. arXiv: 2508.18778 [astro-ph.GA].

Oppenheimer, Benjamin D. & Romeel Davé (Dec. 2006). "Cosmological simulations of intergalactic medium enrichment from galactic outflows." In: *MNRAS* 373.4, pp. 1265–1292. DOI: 10.1111/j.1365-2966.2006.10989.x. arXiv: astro-ph/0605651 [astro-ph].

Oppenheimer, Benjamin D. et al. (Aug. 2010). "Feedback and recycled wind accretion: assembling the $z = 0$ galaxy mass function." In: *MNRAS* 406.4, pp. 2325–2338. DOI: 10.1111/j.1365-2966.2010.16872.x. arXiv: 0912.0519 [astro-ph.CO].

Oteo, I. et al. (Mar. 2018). "An Extreme Protocluster of Luminous Dusty Starbursts in the Early Universe." In: *ApJ* 856.1, 72, p. 72. DOI: 10.3847/1538-4357/aaa1f1. arXiv: 1709.02809 [astro-ph.GA].

Overzier, Roderik A. (Nov. 2016). "The realm of the galaxy protoclusters. A review." In: *A&ARv* 24.1, 14, p. 14. DOI: 10.1007/s00159-016-0100-3. arXiv: 1610.05201 [astro-ph.GA].

Overzier, Roderik A. et al. (Jan. 2008). "Lyman Break Galaxies, Lyα Emitters, and a Radio Galaxy in a Protocluster at $z = 4.1$." In: *ApJ* 673.1, pp. 143–162. DOI: 10.1086/524342. arXiv: astro-ph/0601223 [astro-ph].

Paccagnella, A. et al. (Jan. 2016). "Slow Quenching of Star Formation in OMEGAWINGS Clusters: Galaxies in Transition in the Local Universe." In: *ApJ Letter* 816.2, L25, p. L25. DOI: 10.3847/2041-8205/816/2/L25. arXiv: 1512.04549 [astro-ph.GA].

Padovani, Paolo & Francesca Matteucci (Oct. 1993). "Stellar Mass Loss in Elliptical Galaxies and the Fueling of Active Galactic Nuclei." In: *ApJ* 416, p. 26. DOI: 10.1086/173212.

Papovich, C. et al. (June 2010). "A Spitzer-selected Galaxy Cluster at z = 1.62." In: *ApJ* 716.2, pp. 1503–1513. DOI: 10.1088/0004-637X/716/2/1503. arXiv: 1002.3158 [astro-ph.CO].

Papovich, Casey et al. (Feb. 2018). "The Effects of Environment on the Evolution of the Galaxy Stellar Mass Function." In: *ApJ* 854.1, 30, p. 30. DOI: 10.3847/1538-4357/aaa766. arXiv: 1801.04934 [astro-ph.GA].

Park, Minjung et al. (Nov. 2024). "Widespread Rapid Quenching at Cosmic Noon Revealed by JWST Deep Spectroscopy." In: *ApJ* 976.1, 72, p. 72. DOI: 10.3847/1538-4357/ad7e15. arXiv: 2404.17945 [astro-ph.GA].

Pascarelle, S. M. et al. (Jan. 1996). "The Serendipitous Discovery of a Group or Cluster of Young Galaxies at z2.40 in Deep Hubble Space Telescope WFPC2 Images." In: *ApJ Letter* 456, p. L21. DOI: 10.1086/309852. arXiv: astro-ph/9512033 [astro-ph].

Pellissier, Alisson, Oliver Hahn & Chiara Ferrari (June 2023). "RHAPSODY-C simulations - anisotropic thermal conduction, black hole physics, and the robustness of massive galaxy cluster scaling relations." In: *MNRAS* 522.1, pp. 721–749. DOI: 10.1093/mnras/stad888. arXiv: 2301.02684 [astro-ph.CO].

Peng, Ying-jie et al. (Sept. 2010). "Mass and Environment as Drivers of Galaxy Evolution in SDSS and zCOSMOS and the Origin of the Schechter Function." In: *ApJ* 721.1, pp. 193–221. DOI: 10.1088/0004-637X/721/1/193. arXiv: 1003.4747 [astro-ph.CO].

Pensabene, A. et al. (Apr. 2024). "ALMA survey of a massive node of the Cosmic Web at z ~ 3. I. Discovery of a large overdensity of CO emitters." In: *A&A* 684, A119, A119. DOI: 10.1051/0004-6361/202348659. arXiv: 2401.04765 [astro-ph.GA].

Pentericci, L. et al. (Sept. 2000). "A search for clusters at high redshift. II. A proto cluster around a radio galaxy at z=2.16." In: *A&A* 361, pp. L25–L28. DOI: 10.48550/arXiv.astro-ph/0008143. arXiv: astro-ph/0008143 [astro-ph].

Pentericci, L. et al. (July 2001). "NICMOS Observations of High-Redshift Radio Galaxies: Witnessing the Formation of Bright Elliptical Galaxies?" In: *ApJ, Supplement* 135.1, pp. 63–85. DOI: 10.1086/321781. arXiv: astro-ph/0102323 [astro-ph].

Penzias, A. A. & R. W. Wilson (July 1965). "A Measurement of Excess Antenna Temperature at 4080 Mc/s." In: *ApJ* 142, pp. 419–421. DOI: 10.1086/148307.

Pérez-Martínez, J. M. et al. (Jan. 2023). "Signs of environmental effects on star-forming galaxies in the Spiderweb protocluster at z = 2.16." In:

MNRAS 518.2, pp. 1707–1734. DOI: 10.1093/mnras/stac2784. arXiv: 2209.13069 [astro-ph.GA].

Pérez-Martínez, J. M. et al. (Feb. 2024a). "Enhanced star formation and metallicity deficit in the USS 1558-003 forming protocluster at z = 2.53." In: *MNRAS* 527.4, pp. 10221–10238. DOI: 10.1093/mnras/stad3805. arXiv: 2312.03574 [astro-ph.GA].

Pérez-Martínez, J. M. et al. (Apr. 2025). "COALAS: III. The ATCA CO(1–0) look at the growth and death of Hα emitters in the Spiderweb protocluster at z = 2.16." In: *A&A* 696, A236, A236. DOI: 10.1051/0004-6361/202450785. arXiv: 2411.12138 [astro-ph.GA].

Pérez-Martínez, Jose Manuel et al. (Dec. 2024b). "JWST/NIRCam Paβ Narrowband Imaging Reveals Ordinary Dust Extinction for Hα Emitters within the Spiderweb Protocluster at z = 2.16." In: *ApJ* 977.1, 74, p. 74. DOI: 10.3847/1538-4357/ad8156. arXiv: 2410.03366 [astro-ph.GA].

Perlmutter, S. et al. (June 1999). "Measurements of Ω and Λ from 42 High-Redshift Supernovae." In: *The Astrophysical Journal* 517.2, pp. 565–586. DOI: 10.1086/307221. arXiv: astro-ph/9812133 [astro-ph].

Pierre, M. et al. (June 2016). "The XXL Survey. I. Scientific motivations - XMM-Newton observing plan - Follow-up observations and simulation programme." In: *A&A* 592, A1, A1. DOI: 10.1051/0004-6361/201526766. arXiv: 1512.04317 [astro-ph.CO].

Pierre, Marguerite et al. (Sept. 2004). "The XMM-LSS survey. Survey design and first results." In: *JCAP* 2004.9, 011, p. 011. DOI: 10.1088/1475-7516/2004/09/011. arXiv: astro-ph/0305191 [astro-ph].

Pintos-Castro, I. et al. (May 2019). "The Evolution of the Quenching of Star Formation in Cluster Galaxies since z ∼ 1." In: *ApJ* 876.1, 40, p. 40. DOI: 10.3847/1538-4357/ab14ee. arXiv: 1904.00023 [astro-ph.GA].

Planck Collaboration VI (Sept. 2020). "Planck 2018 results. VI. Cosmological parameters." In: *A&A* 641, A6, A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].

Planck Collaboration XX (Nov. 2014). "Planck 2013 results. XX. Cosmology from Sunyaev-Zeldovich cluster counts." In: *A&A* 571, A20, A20. DOI: 10.1051/0004-6361/201321521. arXiv: 1303.5080 [astro-ph.CO].

Planck Collaboration XXVII (Sept. 2016). "Planck 2015 results. XXVII. The second Planck catalogue of Sunyaev-Zeldovich sources." In: *A&A* 594, A27, A27. DOI: 10.1051/0004-6361/201525823. arXiv: 1502.01598 [astro-ph.CO].

Poggianti, Bianca M. et al. (Dec. 2019). "GASP XXIII: A Jellyfish Galaxy as an Astrophysical Laboratory of the Baryonic Cycle." In: *ApJ* 887.2, 155, p. 155. DOI: 10.3847/1538-4357/ab5224. arXiv: 1910.11622 [astro-ph.GA].

Polletta, M. et al. (Oct. 2021). "Spectroscopic observations of PHz G237.01+42.50: A galaxy protocluster at z = 2.16 in the Cosmos field." In: *A&A* 654, A121, A121. DOI: 10.1051/0004-6361/202140612. arXiv: 2109.04396 [astro-ph.GA].

Poole, Gregory B. et al. (Dec. 2008). "The impact of mergers on relaxed X-ray clusters - III. Effects on compact cool cores." In: *MNRAS* 391.3, pp. 1163–

1175. DOI: 10.1111/j.1365-2966.2008.14003.x. arXiv: 0804.1552 [astro-ph].

Pratt, G. W. et al. (Feb. 2010). "Gas entropy in a representative sample of nearby X-ray galaxy clusters (REXCESS): relationship to gas mass fraction." In: *A&A* 511, A85, A85. DOI: 10.1051/0004-6361/200913309. arXiv: 0909.3776 [astro-ph.CO].

Prescott, Moire K. M. et al. (May 2008). "The Overdense Environment of a Large Lyα Nebula at $z \approx 2.7$." In: *ApJ Letter* 678.2, p. L77. DOI: 10.1086/588606. arXiv: 0803.4230 [astro-ph].

Prevot, M. L. et al. (Mar. 1984). "The typical interstellar extinction in the Small Magellanic Cloud." In: *A&A* 132, pp. 389–392.

Puchwein, E., D. Sijacki & V. Springel (Nov. 2008). "Simulations of AGN Feedback in Galaxy Clusters and Groups: Impact on Gas Fractions and the L_X-T Scaling Relation." In: *ApJ Letter* 687.2, p. L53. DOI: 10.1086/593352. arXiv: 0808.0494 [astro-ph].

Quadri, Ryan F. et al. (Jan. 2012). "Tracing the Star-formation-Density Relation to $z \sim 2$." In: *ApJ* 744.2, 88, p. 88. DOI: 10.1088/0004-637X/744/2/88. arXiv: 1104.1426 [astro-ph.CO].

Ragagnin, Antonio et al. (Jan. 2021). "Cosmology dependence of halo masses and concentrations in hydrodynamic simulations." In: *MNRAS* 500.4, pp. 5056–5071. DOI: 10.1093/mnras/staa3523. arXiv: 2011.05345 [astro-ph.CO].

Ragone-Figueroa, C. et al. (Dec. 2013). "Brightest cluster galaxies in cosmological simulations: achievements and limitations of active galactic nuclei feedback models." In: *Monthly Notice of the RAS* 436.2, pp. 1750–1764. DOI: 10.1093/mnras/stt1693. arXiv: 1308.3246 [astro-ph.CO].

Ragone-Figueroa, C. et al. (Sept. 2018). "BCG mass evolution in cosmological hydro-simulations." In: *MNRAS* 479.1, pp. 1125–1136. DOI: 10.1093/mnras/sty1639. arXiv: 1803.08049 [astro-ph.GA].

Ragone-Figueroa, Cinthia et al. (Nov. 2024). "Intertwined formation of H_2, dust, and stars in cosmological simulations." In: *A&A* 691, A200, A200. DOI: 10.1051/0004-6361/202451344. arXiv: 2407.06269 [astro-ph.GA].

Rasia, E. et al. (July 2006). "Systematics in the X-ray cluster mass estimators." In: *MNRAS* 369.4, pp. 2013–2024. DOI: 10.1111/j.1365-2966.2006.10466.x. arXiv: astro-ph/0602434 [astro-ph].

Rasia, E. et al. (Nov. 2015). "Cool Core Clusters from Cosmological Simulations." In: *ApJ Letter* 813.1, L17, p. L17. DOI: 10.1088/2041-8205/813/1/L17. arXiv: 1509.04247 [astro-ph.CO].

Rasia, Elena et al. (Oct. 2025). "The Three Hundred Project: Modeling baryon and hot-gas fraction evolution in simulated clusters." In: *A&A* 702, A182, A182. DOI: 10.1051/0004-6361/202554283. arXiv: 2505.21624 [astro-ph.CO].

Remus, Rhea-Silvia, Klaus Dolag & Helmut Dannerbauer (June 2023). "The Young and the Wild: What Happens to Protoclusters Forming at Redshift $z \approx 4$?" In: *ApJ* 950.2, 191, p. 191. DOI: 10.3847/1538-4357/accb91. arXiv: 2208.01053 [astro-ph.CO].

Reynolds, Christopher S. et al. (Oct. 2023). "Overview of the advanced x-ray imaging satellite (AXIS)." In: *UV, X-Ray, and Gamma-Ray Space Instrumentation for Astronomy XXIII*. Ed. by Oswald H. Siegmund & Keri Hoadley. Vol. 12678. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 126781E, 126781E. DOI: 10.1117/12.2677468. arXiv: 2311.00780 [astro-ph.IM].

Riess, Adam G. et al. (Sept. 1998). "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant." In: *The Astron. J.* 116.3, pp. 1009–1038. DOI: 10.1086/300499. arXiv: astro-ph/9805201 [astro-ph].

Riva, G. et al. (Nov. 2024). "CHEX-MATE: The intracluster medium entropy distribution in the gravity-dominated regime." In: *A&A* 691, A340, A340. DOI: 10.1051/0004-6361/202451455. arXiv: 2410.11947 [astro-ph.CO].

Ronneberger, O., P.Fischer & T. Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation." In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Vol. 9351. LNCS. (available on arXiv:1505.04597 [cs.CV]). Springer, pp. 234–241. URL: http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a.

Rossetti, M. & S. Molendi (Feb. 2010). "Cool core remnants in galaxy clusters." In: *A&A* 510, A83, A83. DOI: 10.1051/0004-6361/200913156. arXiv: 0910.4900 [astro-ph.CO].

Rossetti, M. et al. (Aug. 2011). "Back and forth from cool core to non-cool core: clues from radio halos." In: *A&A* 532, A123, A123. DOI: 10.1051/0004-6361/201117306. arXiv: 1106.4563 [astro-ph.CO].

Rosswog, S. (Apr. 2009). "Astrophysical smooth particle hydrodynamics." In: *New Astronomy Review* 53.4-6, pp. 78–104. DOI: 10.1016/j.newar.2009.08.007. arXiv: 0903.5075 [astro-ph.IM].

Rykoff, E. S. et al. (Apr. 2014). "redMaPPer. I. Algorithm and SDSS DR8 Catalog." In: *ApJ* 785.2, 104, p. 104. DOI: 10.1088/0004-637X/785/2/104. arXiv: 1303.3562 [astro-ph.CO].

Saftly, W. et al. (June 2013). "Using hierarchical octrees in Monte Carlo radiative transfer simulations." In: *A&A* 554, A10, A10. DOI: 10.1051/0004-6361/201220854. arXiv: 1304.2896 [astro-ph.IM].

Salomé, P. et al. (Aug. 2006). "Cold molecular gas in the Perseus cluster core. Association with X-ray cavity, Hα filaments and cooling flow." In: *A&A* 454.2, pp. 437–445. DOI: 10.1051/0004-6361:20054745. arXiv: astro-ph/0603350 [astro-ph].

Sanders, J. S. et al. (Feb. 2018). "Hydrostatic Chandra X-ray analysis of SPT-selected galaxy clusters - I. Evolution of profiles and core properties." In: *MNRAS* 474.1, pp. 1065–1098. DOI: 10.1093/mnras/stx2796. arXiv: 1705.09299 [astro-ph.CO].

Sanders, Ryan L. et al. (Jan. 2023). "CO Emission, Molecular Gas, and Metallicity in Main-sequence Star-forming Galaxies at z ~ 2.3." In: *ApJ* 942.1, 24, p. 24. DOI: 10.3847/1538-4357/aca46f. arXiv: 2204.06937 [astro-ph.GA].

Sanderson, Alastair J. R., Trevor J. Ponman & Ewan O'Sullivan (Nov. 2006). "A statistically selected Chandra sample of 20 galaxy clusters - I. Temperature and cooling time profiles." In: *MNRAS* 372.4, pp. 1496–1508. DOI: 10.1111/j.1365-2966.2006.10956.x. arXiv: astro-ph/0608423 [astro-ph].

Santos, J. S. et al. (July 2011). "Discovery of a massive X-ray luminous galaxy cluster at z = 1.579." In: *A&A* 531, L15, p. L15. DOI: 10.1051/0004-6361/201117190. arXiv: 1105.5877 [astro-ph.CO].

Saro, A. et al. (Jan. 2009). "Simulating the formation of a protocluster at z ∼2." In: *MNRAS* 392.2, pp. 795–800. DOI: 10.1111/j.1365-2966.2008.14192.x. arXiv: 0810.2791 [astro-ph].

Saro, Alex et al. (July 2013). "Toward Unbiased Galaxy Cluster Masses from Line-of-sight Velocity Dispersions." In: *ApJ* 772.1, 47, p. 47. DOI: 10.1088/0004-637X/772/1/47. arXiv: 1203.5708 [astro-ph.CO].

Schaefer, A. L. et al. (Jan. 2017). "The SAMI Galaxy Survey: spatially resolving the environmental quenching of star formation in GAMA galaxies." In: *MNRAS* 464.1, pp. 121–142. DOI: 10.1093/mnras/stw2289. arXiv: 1609.02635 [astro-ph.GA].

Scharré, Lucie, Daniele Sorini & Romeel Davé (Oct. 2024). "The effects of stellar and AGN feedback on the cosmic star formation history in the SIMBA simulations." In: *MNRAS* 534.1, pp. 361–383. DOI: 10.1093/mnras/stae2098. arXiv: 2404.07252 [astro-ph.GA].

Schechter, P. (Jan. 1976). "An analytic expression for the luminosity function for galaxies." In: *Astrophysical Journal* 203, pp. 297–306. DOI: 10.1086/154079.

Schreiber, C. et al. (Mar. 2015). "The Herschel view of the dominant mode of galaxy growth from z = 4 to the present day." In: *A&A* 575, A74, A74. DOI: 10.1051/0004-6361/201425017. arXiv: 1409.5433 [astro-ph.GA].

Scoville, N. et al. (May 2013). "Evolution of Galaxies and Their Environments at z = 0.1-3 in COSMOS." In: *ApJ, Supplement* 206.1, 3, p. 3. DOI: 10.1088/0067-0049/206/1/3. arXiv: 1303.6689 [astro-ph.CO].

Seymour, N. et al. (Aug. 2012). "Rapid Coeval Black Hole and Host Galaxy Growth in MRC 1138-262: The Hungry Spider." In: *ApJ* 755.2, 146, p. 146. DOI: 10.1088/0004-637X/755/2/146. arXiv: 1206.5821 [astro-ph.CO].

Shakura, N. I. & R. A. Sunyaev (Jan. 1973). "Black holes in binary systems. Observational appearance." In: *A&A* 24, pp. 337–355.

Shelhamer, Evan, Jonathan Long & Trevor Darrell (2017). "Fully Convolutional Networks for Semantic Segmentation." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.4, pp. 640–651. DOI: 10.1109/TPAMI.2016.2572683.

Sherman, Sydney et al. (Dec. 2020). "Investigating the growing population of massive quiescent galaxies at cosmic noon." In: *MNRAS* 499.3, pp. 4239–4260. DOI: 10.1093/mnras/staa3167. arXiv: 2010.04741 [astro-ph.GA].

Shi, Ke et al. (Apr. 2021). "Accelerated Galaxy Growth and Environmental Quenching in a Protocluster at z = 3.24." In: *ApJ* 911.1, 46, p. 46. DOI: 10.3847/1538-4357/abe62e. arXiv: 2102.06499 [astro-ph.GA].

Shimakawa, R. et al. (June 2014). "Identification of the progenitors of rich clusters and member galaxies in rapid formation at z > 2." In: *MN-*

RAS 441, pp. L1–L5. DOI: 10.1093/mnrasl/slu029. arXiv: 1402.3568 [astro-ph.GA].

Shimakawa, Rhythm et al. (Jan. 2018a). "MAHALO Deep Cluster Survey I. Accelerated and enhanced galaxy formation in the densest regions of a protocluster at z = 2.5." In: *MNRAS* 473.2, pp. 1977–1999. DOI: 10.1093/mnras/stx2494. arXiv: 1708.06369 [astro-ph.GA].

Shimakawa, Rhythm et al. (Dec. 2018b). "MAHALO Deep Cluster Survey II. Characterizing massive forming galaxies in the Spiderweb protocluster at z = 2.2." In: *MNRAS* 481.4, pp. 5630–5650. DOI: 10.1093/mnras/sty2618. arXiv: 1809.08755 [astro-ph.GA].

Shimakawa, Rhythm et al. (Feb. 2024). "New insights into the role of AGNs in forming the cluster red sequence." In: *MNRAS* 528.2, pp. 3679–3695. DOI: 10.1093/mnras/stae118. arXiv: 2306.06392 [astro-ph.GA].

Sijacki, D. & V. Springel (Feb. 2006). "Hydrodynamical simulations of cluster formation with central AGN heating." In: *MNRAS* 366.2, pp. 397–416. DOI: 10.1111/j.1365-2966.2005.09860.x. arXiv: astro-ph/0509506 [astro-ph].

Sijacki, Debora et al. (Sept. 2007). "A unified model for AGN feedback in cosmological simulations of structure formation." In: *MNRAS* 380.3, pp. 877–900. DOI: 10.1111/j.1365-2966.2007.12153.x. arXiv: 0705.2238 [astro-ph].

Silva, Laura et al. (Dec. 1998). "Modeling the Effects of Dust on Galactic Spectral Energy Distributions from the Ultraviolet to the Millimeter Band." In: *ApJ* 509.1, pp. 103–117. DOI: 10.1086/306476.

Skelton, Rosalind E. et al. (Oct. 2014). "3D-HST WFC3-selected Photometric Catalogs in the Five CANDELS/3D-HST Fields: Photometry, Photometric Redshifts, and Stellar Masses." In: *ApJ, Supplement* 214.2, 24, p. 24. DOI: 10.1088/0067-0049/214/2/24. arXiv: 1403.3689 [astro-ph.GA].

Smail, Ian (Apr. 2024). "Obscured star formation in clusters at z = 1.6-2.0: massive galaxy formation and the reversal of the star formation-density relation." In: *MNRAS* 529.3, pp. 2290–2308. DOI: 10.1093/mnras/stae692. arXiv: 2401.08761 [astro-ph.GA].

Somerville, Rachel S. & Romeel Davé (Aug. 2015). "Physical Models of Galaxy Formation in a Cosmological Framework." In: *ARA&A* 53, pp. 51–113. DOI: 10.1146/annurev-astro-082812-140951. arXiv: 1412.2712 [astro-ph.GA].

Somerville, Rachel S. et al. (Dec. 2008). "A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei." In: *MNRAS* 391.2, pp. 481–506. DOI: 10.1111/j.1365-2966.2008.13805.x. arXiv: 0808.1227 [astro-ph].

Speagle, J. S. et al. (Oct. 2014). "A Highly Consistent Framework for the Evolution of the Star-Forming "Main Sequence" from z ~0-6." In: *ApJ, Supplement* 214.2, 15, p. 15. DOI: 10.1088/0067-0049/214/2/15. arXiv: 1405.2041 [astro-ph.GA].

Spergel, D. N. et al. (Sept. 2003). "First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters." In: *ApJ, Supplement* 148.1, pp. 175–194. DOI: 10.1086/377226. arXiv: astro-ph/0302209 [astro-ph].

Spitler, Lee R. et al. (Apr. 2012). "First Results from Z -FOURGE: Discovery of a Candidate Cluster at z = 2.2 in COSMOS." In: *ApJ Letter* 748.2, L21, p. L21. DOI: 10.1088/2041-8205/748/2/L21. arXiv: 1112.2691 [astro-ph.CO].

Springel, V. & L. Hernquist (Feb. 2003). "Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation." In: *Monthly Notice of the RAS* 339.2, pp. 289–311. DOI: 10.1046/j.1365-8711.2003.06206.x. arXiv: astro-ph/0206393 [astro-ph].

Springel, Volker (Dec. 2005). "The cosmological simulation code GADGET-2." In: *MNRAS* 364.4, pp. 1105–1134. DOI: 10.1111/j.1365-2966.2005.09655.x. arXiv: astro-ph/0505010 [astro-ph].

Springel, Volker, Tiziana Di Matteo & Lars Hernquist (Aug. 2005a). "Modelling feedback from stars and black holes in galaxy mergers." In: *MNRAS* 361.3, pp. 776–794. DOI: 10.1111/j.1365-2966.2005.09238.x. arXiv: astro-ph/0411108 [astro-ph].

Springel, Volker, Naoki Yoshida & Simon D. M. White (Apr. 2001a). "GADGET: a code for collisionless and gasdynamical cosmological simulations." In: *New Astronomy* 6.2, pp. 79–117. DOI: 10.1016/S1384-1076(01)00042-2. arXiv: astro-ph/0003162 [astro-ph].

Springel, Volker et al. (Dec. 2001b). "Populating a cluster of galaxies - I. Results at z=0." In: *MNRAS* 328.3, pp. 726–750. DOI: 10.1046/j.1365-8711.2001.04912.x. arXiv: astro-ph/0012055 [astro-ph].

Springel, Volker et al. (June 2005b). "Simulations of the formation, evolution and clustering of galaxies and quasars." In: *Nature* 435.7042, pp. 629–636. DOI: 10.1038/nature03597. arXiv: astro-ph/0504097 [astro-ph].

Staab, Priti et al. (Mar. 2024). "Protoclusters as drivers of stellar mass growth in the early Universe, a case study: Taralay - a massive protocluster at z 4.57." In: *MNRAS* 528.4, pp. 6934–6958. DOI: 10.1093/mnras/stae301. arXiv: 2312.11465 [astro-ph.GA].

Stadel, Joachim Gerhard (Jan. 2001). "Cosmological N-body simulations and their analysis." PhD thesis. University of Washington, Seattle.

Stanford, S. A. et al. (May 2001). "The Intracluster Medium in z>1 Galaxy Clusters." In: *ApJ* 552.2, pp. 504–507. DOI: 10.1086/320583. arXiv: astro-ph/0012250 [astro-ph].

Stanford, S. A. et al. (July 2012). "IDCS J1426.5+3508: Discovery of a Massive, Infrared-selected Galaxy Cluster at z = 1.75." In: *ApJ* 753.2, 164, p. 164. DOI: 10.1088/0004-637X/753/2/164. arXiv: 1205.3786 [astro-ph.CO].

Steidel, Charles C. et al. (Jan. 1998). "A Large Structure of Galaxies at Redshift Z approximately 3 and Its Cosmological Implications." In: *ApJ* 492.2, pp. 428–438. DOI: 10.1086/305073. arXiv: astro-ph/9708125 [astro-ph].

Steidel, Charles C. et al. (Mar. 2000). "Lyα Imaging of a Proto-Cluster Region at <z>=3.09." In: *ApJ* 532.1, pp. 170–182. DOI: 10.1086/308568. arXiv: astro-ph/9910144 [astro-ph].

Steidel, Charles C. et al. (June 2005). "Spectroscopic Identification of a Protocluster at z=2.300: Environmental Dependence of Galaxy Properties at High Redshift." In: *ApJ* 626.1, pp. 44–50. DOI: 10.1086/429989. arXiv: astro-ph/0502432 [astro-ph].

Steinborn, Lisa K. et al. (Apr. 2015). "A refined sub-grid model for black hole accretion and AGN feedback in large cosmological simulations." In: *MNRAS* 448.2, pp. 1504–1525. DOI: 10.1093/mnras/stv072. arXiv: 1409.3221 [astro-ph.GA].

Strazzullo, V. et al. (Aug. 2013). "Galaxy Evolution in Overdense Environments at High Redshift: Passive Early-type Galaxies in a Cluster at z ~2." In: *ApJ* 772.2, 118, p. 118. DOI: 10.1088/0004-637X/772/2/118. arXiv: 1305.3577 [astro-ph.CO].

Strazzullo, V. et al. (Apr. 2015). "Passive galaxies as tracers of cluster environments at z ~2." In: *A&A* 576, L6, p. L6. DOI: 10.1051/0004-6361/201425038. arXiv: 1501.05323 [astro-ph.CO].

Strazzullo, V. et al. (July 2018). "Deciphering the Activity and Quiescence of High-redshift Cluster Environments: ALMA Observations of Cl J1449+0856 at z = 2." In: *ApJ* 862.1, 64, p. 64. DOI: 10.3847/1538-4357/aacd10. arXiv: 1806.06856 [astro-ph.GA].

Strazzullo, V. et al. (Feb. 2019). "Galaxy populations in the most distant SPT-SZ clusters. I. Environmental quenching in massive clusters at $1.4 \lesssim z \lesssim 1.7$." In: *A&A* 622, A117, A117. DOI: 10.1051/0004-6361/201833944. arXiv: 1807.09768 [astro-ph.GA].

Sun, Hanwen et al. (June 2024). "JWST's First Glimpse of a z > 2 Forming Cluster Reveals a Top-heavy Stellar Mass Function." In: *ApJ Letter* 967.2, L34, p. L34. DOI: 10.3847/2041-8213/ad4986. arXiv: 2403.05248 [astro-ph.GA].

Sunyaev, R. A. & Ya. B. Zeldovich (Apr. 1970). "Small-Scale Fluctuations of Relic Radiation." In: *Astrophysics and Space Science* 7.1, pp. 3–19. DOI: 10.1007/BF00653471.

— (Nov. 1972). "The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies." In: *Comments on Astrophysics and Space Physics* 4, p. 173.

— (Jan. 1980). "Microwave background radiation as a probe of the contemporary structure and history of the universe." In: *Annual Review of Astronomy and Astrophysics* 18, pp. 537–560. DOI: 10.1146/annurev.aa.18.090180.002541.

Tanaka, Masayuki et al. (July 2024). "A Protocluster of Massive Quiescent Galaxies at z = 4." In: *ApJ* 970.1, 59, p. 59. DOI: 10.3847/1538-4357/ad5316. arXiv: 2311.11569 [astro-ph.GA].

Teodoro, Enrico M. Di, J. E. G. Peek & John F. Wu (Feb. 2023). "Identification of Galaxy Shreds in Large Photometric Catalogs Using Convolutional Neural Networks." In: *The Astronomical Journal* 165.3, p. 123. DOI: 10.3847/1538-3881/acb53a. URL: https://doi.org/10.3847%2F1538-3881%2Facb53a.

Thielemann, F. K. et al. (May 2003). "Nuclear cross sections, nuclear structure and stellar nucleosynthesis." In: *Nuclear Physics A* 718, pp. 139–146. DOI: 10.1016/S0375-9474(03)00704-8.

Thongkham, Khunanon et al. (June 2024). "The Massive and Distant Clusters of WISE Survey 2: Equatorial First Data Release." In: *ApJ* 967.2, 123, p. 123. DOI: 10.3847/1538-4357/ad3c44. arXiv: 2403.16976 [astro-ph.CO].

Tormen, Giuseppe (Sept. 1997). "The rise and fall of satellites in galaxy clusters." In: *MNRAS* 290.3, pp. 411–421. DOI: 10.1093/mnras/290.3.411. arXiv: astro-ph/9611078 [astro-ph].

Tornatore, L. et al. (Nov. 2007). "Chemical enrichment of galaxy clusters from hydrodynamical simulations." In: *Monthly Notices of the Royal Astronomical Society* 382.3, pp. 1050–1072. DOI: 10.1111/j.1365-2966.2007.12070.x. URL: https://doi.org/10.1111$\%$2Fj.1365-2966.2007.12070.x.

Toshikawa, Jun et al. (Aug. 2016). "A Systematic Survey of Protoclusters at z ~3-6 in the CFHTLS Deep Fields." In: *ApJ* 826.2, 114, p. 114. DOI: 10.3847/0004-637X/826/2/114. arXiv: 1605.01439 [astro-ph.GA].

Tozzi, P. et al. (June 2022a). "The 700 ks Chandra Spiderweb Field. I. Evidence for widespread nuclear activity in the protocluster." In: *A&A* 662, A54, A54. DOI: 10.1051/0004-6361/202142333. arXiv: 2203.02208 [astro-ph.GA].

Tozzi, P. et al. (Nov. 2022b). "The 700 ks Chandra Spiderweb Field. II. Evidence for inverse-Compton and thermal diffuse emission in the Spiderweb galaxy." In: *A&A* 667, A134, A134. DOI: 10.1051/0004-6361/202244337. arXiv: 2209.15467 [astro-ph.GA].

Tozzi, Paolo & Colin Norman (Jan. 2001). "The Evolution of X-Ray Clusters and the Entropy of the Intracluster Medium." In: *ApJ* 546.1, pp. 63–84. DOI: 10.1086/318237. arXiv: astro-ph/0003289 [astro-ph].

Travascio, A. et al. (Feb. 2025a). "X-ray view of a massive node of the Cosmic Web at $z \sim 3$: I. An exceptional overdensity of rapidly accreting supermassive black holes." In: *A&A* 694, A165, A165. DOI: 10.1051/0004-6361/202452179. arXiv: 2410.03933 [astro-ph.GA].

Travascio, Andrea et al. (Aug. 2025b). "X-ray view of a massive node of the Cosmic Web at z=3 II. Discovery of extended X-ray emission around a hyperluminous QSO." In: *arXiv e-prints*, arXiv:2508.20074, arXiv:2508.20074. DOI: 10.48550/arXiv.2508.20074. arXiv: 2508.20074 [astro-ph.GA].

Tremmel, M. et al. (Mar. 2019). "Introducing ROMULUSC: a cosmological simulation of a galaxy cluster with an unprecedented resolution." In: *MNRAS* 483.3, pp. 3336–3362. DOI: 10.1093/mnras/sty3336. arXiv: 1806.01282 [astro-ph.GA].

Troxel, M. A. et al. (Aug. 2018). "Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear." In: *Phys. Rev. D* 98.4, 043528, p. 043528. DOI: 10.1103/PhysRevD.98.043528. arXiv: 1708.01538 [astro-ph.CO].

Trčka, Ana et al. (Nov. 2022). "UV to submillimetre luminosity functions of TNG50 galaxies." In: *MNRAS* 516.3, pp. 3728–3749. DOI: 10.1093/mnras/stac2277. arXiv: 2208.06424 [astro-ph.GA].

Valentini, Milena & Klaus Dolag (Feb. 2025). "Hydrodynamic methods and sub-resolution models for cosmological simulations." In: *arXiv e-prints*, arXiv:2502.06954, arXiv:2502.06954. DOI: 10.48550/arXiv.2502.06954. arXiv: 2502.06954 [astro-ph.CO].

Valentini, Milena et al. (Jan. 2020). "Impact of AGN feedback on galaxies and their multiphase ISM across cosmic time." In: *MNRAS* 491.2, pp. 2779–2807. DOI: 10.1093/mnras/stz3131. arXiv: 1911.02572 [astro-ph.GA].

Valentini, Milena et al. (Jan. 2023). "Impact of H_2-driven star formation and stellar feedback from low-enrichment environments on the formation of spiral galaxies." In: *MNRAS* 518.1, pp. 1128–1147. DOI: 10.1093/mnras/stac2110. arXiv: 2207.13710 [astro-ph.GA].

Valentino, Francesco et al. (Sept. 2016). "A Giant Lyα Nebula in the Core of an X-Ray Cluster at Z = 1.99: Implications for Early Energy Injection." In: *ApJ* 829.1, 53, p. 53. DOI: 10.3847/0004-637X/829/1/53. arXiv: 1605.03194 [astro-ph.GA].

Venemans, B. P. et al. (Apr. 2002). "The Most Distant Structure of Galaxies Known: A Protocluster at z=4.1." In: *ApJ Letter* 569.1, pp. L11–L14. DOI: 10.1086/340563. arXiv: astro-ph/0203249 [astro-ph].

Venemans, B. P. et al. (Sept. 2004). "Discovery of six Lyα emitters near a radio galaxy at z ~ 5.2." In: *A&A* 424, pp. L17–L20. DOI: 10.1051/0004-6361:200400041. arXiv: astro-ph/0408520 [astro-ph].

Venemans, B. P. et al. (Mar. 2005). "Properties of Lyα emitters around the radio galaxy MRC 0316 257'." In: *A&A* 431.3, pp. 793–812. DOI: 10.1051/0004-6361:20042038. arXiv: astro-ph/0501259 [astro-ph].

Venemans, B. P. et al. (Jan. 2007). "Protoclusters associated with z > 2 radio galaxies . I. Characteristics of high redshift protoclusters." In: *A&A* 461.3, pp. 823–845. DOI: 10.1051/0004-6361:20053941. arXiv: astro-ph/0610567 [astro-ph].

Vikhlinin, A. et al. (Apr. 2006). "Chandra Sample of Nearby Relaxed Galaxy Clusters: Mass, Gas Fraction, and Mass-Temperature Relation." In: *ApJ* 640.2, pp. 691–709. DOI: 10.1086/500288. arXiv: astro-ph/0507092 [astro-ph].

Vikhlinin, A. et al. (Feb. 2009). "Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints." In: *ApJ* 692.2, pp. 1060–1074. DOI: 10.1088/0004-637X/692/2/1060. arXiv: 0812.2720 [astro-ph].

Vito, F. et al. (Sept. 2024). "Fast supermassive black hole growth in the SPT2349–56 protocluster at z = 4.3." In: *A&A* 689, A130, A130. DOI: 10.1051/0004-6361/202450225. arXiv: 2406.13005 [astro-ph.GA].

Vogelsberger, Mark et al. (Jan. 2020). "Cosmological simulations of galaxy formation." In: *Nature Reviews Physics* 2.1, pp. 42–66. DOI: 10.1038/s42254-019-0127-2. arXiv: 1909.07976 [astro-ph.GA].

Voit, G. Mark, Scott T. Kay & Greg L. Bryan (Dec. 2005). "The baseline intracluster entropy profile from gravitational structure formation." In:

MNRAS 364.3, pp. 909–916. DOI: 10.1111/j.1365-2966.2005.09621.x. arXiv: astro-ph/0511252 [astro-ph].

Vulcani, Benedetta et al. (Feb. 2010). "Comparing the Relation Between Star Formation and Galaxy Mass in Different Environments." In: *ApJ Letter* 710.1, pp. L1–L6. DOI: 10.1088/2041-8205/710/1/L1. arXiv: 0912.1180 [astro-ph.CO].

Vulcani, Benedetta et al. (Mar. 2011). "Galaxy stellar mass functions of different morphological types in clusters, and their evolution between z= 0.8 and 0." In: *MNRAS* 412.1, pp. 246–268. DOI: 10.1111/j.1365-2966.2010.17904.x. arXiv: 1010.4442 [astro-ph.CO].

Wagoner, Robert V., William A. Fowler & F. Hoyle (Apr. 1967). "On the Synthesis of Elements at Very High Temperatures." In: *ApJ* 148, p. 3. DOI: 10.1086/149126.

Wang, L. et al. (Oct. 2018a). "Galaxy and Mass Assembly (GAMA): The environmental dependence of the galaxy main sequence." In: *A&A* 618, A1, A1. DOI: 10.1051/0004-6361/201832697. arXiv: 1802.08456 [astro-ph.GA].

Wang, Tao et al. (Sept. 2016). "Discovery of a Galaxy Cluster with a Violently Starbursting Core at z = 2.506." In: *ApJ* 828.1, 56, p. 56. DOI: 10.3847/0004-637X/828/1/56. arXiv: 1604.07404 [astro-ph.GA].

Wang, Tao et al. (Nov. 2018b). "Revealing the Environmental Dependence of Molecular Gas Content in a Distant X-Ray Cluster at z = 2.51." In: *ApJ Letter* 867.2, L29, p. L29. DOI: 10.3847/2041-8213/aaeb2c. arXiv: 1810.10558 [astro-ph.GA].

Wang, Wenting & Simon D. M. White (Aug. 2012). "Satellite abundances around bright isolated galaxies." In: *MNRAS* 424.4, pp. 2574–2598. DOI: 10.1111/j.1365-2966.2012.21256.x. arXiv: 1203.0009 [astro-ph.CO].

Weaver, J. R. et al. (Sept. 2023). "COSMOS2020: The galaxy stellar mass function. The assembly and star formation cessation of galaxies at $0.2 < z \leqslant 7.5$." In: *A&A* 677, A184, A184. DOI: 10.1051/0004-6361/202245581. arXiv: 2212.02512 [astro-ph.GA].

Wetzel, Andrew R., Jeremy L. Tinker & Charlie Conroy (July 2012). "Galaxy evolution in groups and clusters: star formation rates, red sequence fractions and the persistent bimodality." In: *MNRAS* 424.1, pp. 232–243. DOI: 10.1111/j.1365-2966.2012.21188.x. arXiv: 1107.5311 [astro-ph.CO].

Whitaker, Katherine E. et al. (July 2011). "The NEWFIRM Medium-band Survey: Photometric Catalogs, Redshifts, and the Bimodal Color Distribution of Galaxies out to $z \sim 3$." In: *ApJ* 735.2, 86, p. 86. DOI: 10.1088/0004-637X/735/2/86. arXiv: 1105.4609 [astro-ph.CO].

Whitaker, Katherine E. et al. (Feb. 2012). "A Large Population of Massive Compact Post-starburst Galaxies at $z > 1$: Implications for the Size Evolution and Quenching Mechanism of Quiescent Galaxies." In: *ApJ* 745.2, 179, p. 179. DOI: 10.1088/0004-637X/745/2/179. arXiv: 1112.0313 [astro-ph.CO].

R., Schaeffer et al., eds. (Jan. 1996). *Cosmology and large scale structure*.

Wiersma, Robert P. C., Joop Schaye & Britton D. Smith (Feb. 2009). "The effect of photoionization on the cooling rates of enriched, astrophysical plasmas." In: *MNRAS* 393.1, pp. 99–107. DOI: 10.1111/j.1365-2966.2008.14191.x. arXiv: 0807.3748 [astro-ph].

Williams, Rik J. et al. (Feb. 2009). "Detection of Quiescent Galaxies in a Bicolor Sequence from Z = 0-2." In: *ApJ* 691.2, pp. 1879–1895. DOI: 10.1088/0004-637X/691/2/1879. arXiv: 0806.0625 [astro-ph].

Willis, J. P. et al. (Jan. 2020). "Spectroscopic confirmation of a mature galaxy cluster at a redshift of 2." In: *Nature* 577.7788, pp. 39–41. DOI: 10.1038/s41586-019-1829-4. arXiv: 2001.00549 [astro-ph.GA].

Wilson, Gillian et al. (June 2009). "Spectroscopic Confirmation of a Massive Red-Sequence-Selected Galaxy Cluster at z = 1.34 in the SpARCS-South Cluster Survey." In: *ApJ* 698.2, pp. 1943–1950. DOI: 10.1088/0004-637X/698/2/1943. arXiv: 0810.0036 [astro-ph].

Wisnioski, E. et al. (Dec. 2019). "The KMOS^{3D} Survey: Data Release and Final Survey Paper." In: *ApJ* 886.2, 124, p. 124. DOI: 10.3847/1538-4357/ab4db8. arXiv: 1909.11096 [astro-ph.GA].

Wylezalek, D. et al. (May 2013a). "Galaxy Clusters around Radio-loud Active Galactic Nuclei at $1.3 < z < 3.2$ as Seen by Spitzer." In: *ApJ* 769.1, 79, p. 79. DOI: 10.1088/0004-637X/769/1/79. arXiv: 1304.0770 [astro-ph.CO].

Wylezalek, D. et al. (Feb. 2013b). "The Herschel view of the environment of the radio galaxy 4C+41.17 at z = 3.8." In: *MNRAS* 428.4, pp. 3206–3219. DOI: 10.1093/mnras/sts264. arXiv: 1210.6361 [astro-ph.CO].

Wylezalek, Dominika & Nadia L. Zakamska (Oct. 2016). "Evidence of suppression of star formation by quasar-driven winds in gas-rich host galaxies at $z < 1$?" In: *MNRAS* 461.4, pp. 3724–3739. DOI: 10.1093/mnras/stw1557. arXiv: 1606.08442 [astro-ph.GA].

XRISM Collaboration (Feb. 2025). "The bulk motion of gas in the core of the Centaurus galaxy cluster." In: *Nature* 638.8050, pp. 365–369. DOI: 10.1038/s41586-024-08561-z. arXiv: 2502.08722 [astro-ph.HE].

Xu, G. (May 1995). "A New Parallel N-Body Gravity Solver: TPM." In: *Astrophysical Journal, Supplement* 98, p. 355. DOI: 10.1086/192166. arXiv: astro-ph/9409021 [astro-ph].

Xu, Ke et al. (Mar. 2025). "Ram-pressure stripping caught in action in a forming galaxy cluster 3 billion years after the Big Bang." In: *arXiv e-prints*, arXiv:2503.21724, arXiv:2503.21724. DOI: 10.48550/arXiv.2503.21724. arXiv: 2503.21724 [astro-ph.GA].

Zavala, J. A. et al. (Dec. 2019). "On the Gas Content, Star Formation Efficiency, and Environmental Quenching of Massive Galaxies in Protoclusters at $z \approx 2.0$-2.5." In: *ApJ* 887.2, 183, p. 183. DOI: 10.3847/1538-4357/ab5302. arXiv: 1910.13457 [astro-ph.GA].

Zel'Dovich, Y. B. (Mar. 1970). "Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations." In: *Astronomy and Astrophysics* 500, pp. 13–18.

Zhang, Y. H. et al. (Dec. 2024). "ASW^{2}DF: Census of the obscured star formation in a galaxy cluster in formation at z = 2.2." In: *A&A* 692, A22, A22. DOI: 10.1051/0004-6361/202451379. arXiv: 2410.10169 [astro-ph.GA].

Zhang, Y. H. et al. (Dec. 2025). "JWST reveals extended stellar disks for ALMA-bright dusty star-forming galaxies in the Spiderweb protocluster." In: *arXiv e-prints*, arXiv:2512.07050, arXiv:2512.07050. arXiv: 2512.07050 [astro-ph.GA].

Zheng, Xian Zhong et al. (Jan. 2021). "MAMMOTH: confirmation of two massive galaxy overdensities at z = 2.24 with Hα emitters." In: *MNRAS* 500.4, pp. 4354–4364. DOI: 10.1093/mnras/staa2882. arXiv: 2009.08068 [astro-ph.GA].

Zirm, Andrew W., Mark Dickinson & Arjun Dey (Mar. 2003). "Massive Elliptical Galaxies at High Redshift: NICMOS Imaging of z∼1 Radio Galaxies." In: *ApJ* 585.1, pp. 90–111. DOI: 10.1086/346021. arXiv: astro-ph/0211469 [astro-ph].

ZuHone, J. A., M. Markevitch & R. E. Johnson (July 2010). "Stirring Up the Pot: Can Cooling Flows in Galaxy Clusters be Quenched by Gas Sloshing?" In: *ApJ* 717.2, pp. 908–928. DOI: 10.1088/0004-637X/717/2/908. arXiv: 0912.0237 [astro-ph.CO].

Zwicky, F. (Jan. 1933). "Die Rotverschiebung von extragalaktischen Nebeln." In: *Helvetica Physica Acta* 6, pp. 110–127.

de Bernardis, P. et al. (Apr. 2000). "A flat Universe from high-resolution maps of the cosmic microwave background radiation." In: *Nature* 404.6781, pp. 955–959. DOI: 10.1038/35010035. arXiv: astro-ph/0004404 [astro-ph].

van Marrewijk, J. et al. (Sept. 2024). "XLSSC 122 caught in the act of growing up: Spatially resolved SZ observations of a z = 1.98 galaxy cluster." In: *A&A* 689, A41, A41. DOI: 10.1051/0004-6361/202348213. arXiv: 2310.06120 [astro-ph.CO].

van der Burg, Remco F. J. et al. (Sept. 2013). "The environmental dependence of the stellar mass function at z ∼1. Comparing cluster and field between the GCLASS and UltraVISTA surveys." In: *A&A* 557, A15, A15. DOI: 10.1051/0004-6361/201321237. arXiv: 1304.5525 [astro-ph.CO].

van der Burg, Remco F. J. et al. (Oct. 2018). "The stellar mass function of galaxies in Planck-selected clusters at 0.5 < z < 0.7: new constraints on the timescale and location of satellite quenching." In: *A&A* 618, A140, A140. DOI: 10.1051/0004-6361/201833572. arXiv: 1807.00820 [astro-ph.GA].

van der Burg, Remco F. J. et al. (June 2020). "The GOGREEN Survey: A deep stellar mass function of cluster galaxies at 1.0 < z < 1.4 and the complex nature of satellite quenching." In: *A&A* 638, A112, A112. DOI: 10.1051/0004-6361/202037754. arXiv: 2004.10757 [astro-ph.GA].

van der Wel, A. et al. (June 2014). "3D-HST+CANDELS: The Evolution of the Galaxy Size-Mass Distribution since z = 3." In: *ApJ* 788.1, 28, p. 28. DOI: 10.1088/0004-637X/788/1/28. arXiv: 1404.2844 [astro-ph.GA].