THE UNIVERSITY OF CHICAGO

MODELING GALAXIES IN COSMOLOGICAL SIMULATIONS: USING HALO CORES TO MODEL GALAXY DISTRIBUTIONS IN CLUSTERS, AND CONSTRUCTING THE SYNTHETIC GALAXY CATALOG COSMODC2

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
DANILA KORYTOV

CHICAGO, ILLINOIS
JUNE 2020

TABLE OF CONTENTS

LIST OF FIGURES v

LIST OF TABLES x

ACKNOWLEDGMENTS xi

ABSTRACT xii

1 INTRODUCTION 1
1.1 Large Scale Structure Probes 3
1.2 Galaxy Modeling from Simulations 3
1.3 Dissertation Outline 6

2 HALO CORE TRACKING 7
2.1 Introduction 7
2.2 Methodology and Modeling 9
2.2.1 N-body Simulation 10
2.2.2 Identifying and Tracking Individual Halo Cores 11
2.2.3 Tracking Halo Cores with Halo Evolution 12
2.2.4 Comparison of Cores to Subhalos 17
2.2.5 Modeling the Galaxy Population in Halos 19
2.3 Observations: SDSS Data Reduction 21
2.3.1 redMaPPer Clusters 22
2.3.2 SPIDERS Clusters 22
2.3.3 Obtaining SDSS Galaxies 23
2.3.4 Galaxy Radial Profile Stacks 25
2.3.5 Background Estimation and Subtraction 25
2.4 Model Fitting 26
2.4.1 Modeling Cores as Galaxies 27
2.4.2 Modeling Galaxy Profiles from Cores 31
2.4.3 Fitting Profiles to Observations 32
2.5 Results and Discussion 34
2.5.1 Results for the Different Model Types 35
2.5.2 Luminosity Dependence 35
2.5.3 Robustness and Convergence Tests 39
2.5.4 Predicted Disruption Rates 42
2.6 Conclusion 43

3 SYNTHETIC GALAXY CATALOG 44
3.1 Introduction 44
3.2 CosmoDC2 Production Overview 49
3.3 The Underlying Simulations 53

3.3.1 The Outer Rim Simulation . . . 53
3.3.2 The AlphaQ Simulation . . . 54
3.3.3 Halo Catalogs and Merger Trees . . . 55
3.3.4 Lightcone Generation . . . 55
3.3.5 Workflow . . . 59
3.4 Weak Lensing . . . 59
3.5 The Galaxy Catalog . . . 64
3.5.1 Empirical Galaxy Catalog Generation . . . 64
3.5.2 The Galacticus Library . . . 73
3.5.3 Galaxy Catalog and Galaxy Library Matching . . . 77
3.5.4 Additional Modeling . . . 84
3.5.5 Galaxy Catalog Content . . . 86
3.6 Selected Validation Results . . . 87
3.6.1 Cumulative Number Counts as a Function of Magnitude . . . 87
3.6.2 Redshift Distribution . . . 89
3.6.3 Color Distributions . . . 91
3.6.4 Two-Point Correlation Function . . . 93
3.6.5 Galaxy-galaxy Lensing . . . 94
3.7 Summary and Future Directions . . . 96

4 CONCLUSION . . . 100
4.1 Halo Core Tracking . . . 100
4.1.1 Primary Results . . . 100
4.1.2 Future Work . . . 101
4.2 Synthetic Galaxy Catalog . . . 102
4.2.1 Primary Results . . . 102
4.2.2 Future Work . . . 103

REFERENCES . . . 105

LIST OF FIGURES

1.1 The dark matter density from a slice of the Outer Rim simulation. 2

2.1 Cartoon of the method of assigning cores to halos using the merger trees. Circles are DM halos and the arrows represent halo descents across adjacent simulation snapshots. Colored dots represent halo cores, each color a separate core. Cores that are linked by a dashed line are represented by the same set of DM simulation particles. Each halo can host up to one central core and any number of satellite cores. All cores start out as isolated centrals in halos above the mass threshold (solid circles). Halos below the mass threshold (dashed circles) do not have their halo core tracked. Central cores are updated with a new set of particles from the halo center at each simulation snapshot. Central cores become satellites cores when they merge with a heavier halo. As a result every halo will have exactly the same number of cores as in it's merger tree branches above the mass threshold. 13

2.2 *Top panel:* Population density as function of core radii and core infall mass in the AlphaQ simulation. Several distinct features are visible. The bottom group with the smallest radii are central cores. The particles representing the core have been refreshed to be the spatially closest of the current host halo. The population above the central are satellite cores which had time for the core particles to spread out. The population to the left of the vertical ridge are cores which originate from halos below the 100 particle threshold but have a progenitor that was above the threshold and inherited the core. *Bottom pane:* Core radius as a function of core infall redshift for all cores in AlphaQ at $z = 0$. Satellite cores that have recently merged still have a small radius as the cores have not experienced strong tidal forces in the host halo and other harassment. 16

2.3 Core distribution and assembly for two representative halos at $z = 0$. The left panels show the distribution of cores and subhalos. Subhalos are resolved down to masses equal to $\sim 1.610 \times h^{-1}\mathrm{M}_{\odot}$, while cores only originate from halos which have FoF halo mass of at least $\sim 1.6{\times}\mathrm{h}^{-1}\mathrm{M}_{\odot}$. The right panels show the trajectories of cores throughout the assembly history of the halos. The black circle is the R_{200c} at $z = 0$. 18

2.4 Fraction of subhalos with a matched core as a function of subhalo mass in the AlphaQ simulation. Subhalos below 100 particles (dashed line) 19

2.5 Core infall mass against subhalo mass in the AlphaQ simulation at $z = 0$. The dashed black line is the one-to-one relationship line. 20

2.6 Halo mass as determined in the redMaPPer and SPIDERS catalog, converted to M_{200m} 23

2.7 *Top panels:* Stacked surface density profiles of L_* galaxies for the redMaPPer and SPIDERS cluster catalogs separated into halo mass bins. *Bottom panels:* Stacked profiles normalized to the redMaPPer cluster sample. 26

2.8 Galaxy surface density profiles of cores as a function the M_{infall} threshold value for halos $14.0 < log_{10}(M_{200m}) < 14.25$ in the Outer Rim simulation at $z = 0.24$. Varying M_{infall} significantly modifies the overall normalization of the profile but also creates a more peak profile shape for higher M_{infall}. 28
2.9 Galaxy surface density profiles of cores as a function the $R_{disrupt}$ threshold value for halos $14.0 < log_{10}(M_{200m}) < 14.25$. $R_{disrupt}$ modifies the shape of the profile. 30
2.10 Galaxy surface density profiles of cores as a function the R_{merge} for halos $14.0 < log_{10}(M_{200m}) < 14.25$. Because the percolating nature of the FoF merging, the cores have been truncated to have an abundance similar to L_* galaxies: M_{infall} threshold of 10^12 h^{-1} $M_\odot$and $R_{disrupt}$ threshold of 20 h^{-1} kpc. 31
2.11 Grid likelihood evaluation of the Rd L_* galaxy model in the Outer Rim simulation. The left triangle plot shows the negative log likelihood calculated on a wide parameter grid scan. The right triangle plot shows the likelihood calculated on a zoomed-in region identified in the left panel. 32
2.12 Core model parameter best fit and 1σ likelihood bounds for the Outer Rim halos for redMaPPer L_* galaxies. The $\tilde{\chi}^2$ for each model flavor is shown in the right most panel. 34
2.13 Radial L_* galaxy surface density profiles from observations and core galaxy models for several halo mass bins. The closed and open markers are SDSS galaxy surface density profiles centered on redMaPPer and SPIDERS clusters, respectively. The lines and shaded regions are the average and the error on the mean, respectively, of the profiles of modeled galaxies in the Outer Rim simulation using the best fit to the observed redMaPPer clusters. The markers for redMaPPer and SPIDERS have been offset for clarity. The SPIDERS cluster catalog does not have any clusters in the top panel mass range. 36
2.14 N_{gal} as a function of cluster mass for the redMaPPer and SPIDERS cluster catalogs and for the Rd galaxy model of cores. The cluster N_{gal} is calculated by integrating the background subtracted galaxy counts within R_{200m} of the cluster center. The Rd model uses the best fit model parameters to the redMaPPer cluster profiles. The markers position for redMaPPer and SPIDERS have been offset in the x-axis for clarity. 37
2.15 Best fit galaxy model parameters and $\tilde{\chi}^2$ for the Outer Rim simulation and redMaPPer cluster catalog as a function of galaxy luminosity. Each model flavor is a separate color. The the lines are best fit model parameters and the shaded regions are the 1σ statistical bounds. 38
2.16 Best fit model parameters and $X^2_{reduced}$ for the Rd L_* galaxy model as a function of projection radius. All cores within the projection radius are used to compute the projected modeled galaxy surface density. 40
2.17 Best fit parameters for the Rd L_* galaxy model as a function of the central cluster radius excised from the likelihood calculation. 41
2.18 Modeled galaxy survival rate for Rd model in the Outer Rim simulation. 42

3.1 Conceptual overview of the workflow to produce cosmoDC2. Data products are shown as rectangles in dark and light purple for data derived from the Outer Rim and the smaller auxiliary simulations, respectively. Pipelines are shown as ovals in light orange. Numbers in parentheses refer to the sections and figures in this work where a detailed descriptions are given. 50
3.2 Projected particle overdensity power spectrum (top) and residuals with respect to theory (bottom) for a shell of width of approximately 450Mpc at $z \approx 1.7$, after shot-noise corrections. The theory curve is obtained from the CosmicEmu power spectrum emulator, and has been corrected for expected levels of smoothing due to finite shell-width effects. The shaded region in the residual plot shows the approximate 1σ level of cosmic variance given the Polspice kernels. The up-turn in the residuals at small scales is due to the interpolation errors, these are sub-percent below $\ell \approx 3000$, rising to approximately 3% by $\ell = 6000$. 57
3.3 Schematic of the interpolation process which fills the cosmoDC2 halo lightcone. Each plane represents a projected simulation snapshot, and time increases vertically, with the observer located at $\mathbf{o}$. A merger tree branch including halo h is seen crossing the observer's lightcone between snapshots j and $j+1$ (the purple worldlines of each halo are unknown between the snapshots). Interpolation between halo h and its most massive progenitor $h_{\rm mmp}$ (orange dashed line) is used to solve for the temporal and spatial components of event $\mathbf{h}'$, where we place an object with properties (mass, etc.) identical to halo h. 60
3.4 Workflow to produce the Outer Rim and AlphaQ simulation data products used as inputs to the cosmoDC2 production pipeline. Data products are shown as rectangles in dark and light purple for data derived from the Outer Rim and AlphaQ simulations, respectively. Code modules are shown as ovals in dark orange. Numbers in parentheses refer to the sections in the paper where a detailed description of the workflow component is given. 61
3.5 Workflow for the shear pipeline. Data products derived from the Outer Rim simulation are shown as boxes in dark purple and code modules are shown as ovals in dark orange. 63
3.6 Left image: Cartesian projection of a patch of the convergence source plane, in observer coordinates, at $z \approx 1.0$. Right image: Zoom-in of a box within this patch, with the cosmic shear field overlaid. For visualization purposes the lengths of the shear vectors are truncated to a maximum value above $|\gamma| = 0.025$. 64
3.7 E-mode shear power spectrum (top) and residuals with respect to theory (bottom) for a selection of source planes. Theory curves (black dashed lines) are computed using the Born approximation on the CosmicEmu power spectrum emulator, extrapolated to high k values using Padé approximants, the residual points are slightly offset for visualization purposes. The gray shaded region in the residual panel indicates the advertised 4% accuracy of the power spectrum, transparency of the theory curves and residual points indicate that high-k extrapolation accounts for more than 10% of the total power in the theory curve. 65

3.8 Comparison of the $z = 0.15$ HOD in the UniverseMachine mock (dashed curves) with those of baseDC2 (solid curves). The HOD quantifies $P(N_{\rm gal}|M_{\rm halo})$, the probability that a halo of mass $M_{\rm halo}$ hosts $N_{\rm gal}$ galaxies with stellar mass $M_\star$ greater than the threshold indicated in the panel. The first moment of the HOD, $\langle N_{\rm gal}|M_{\rm halo}\rangle$, is shown on the vertical axis of each panel as a function of $M_{\rm halo}$. Solid curves show the corresponding HOD in the baseDC2 model that is the foundation of cosmoDC2. We tune our HOD to match UniverseMachine so that our model can inherit the observational realism of the galaxy-halo connection shown in [13]. 73
3.9 Workflow for the Galacticus match-up pipeline to produce the final output for cosmoDC2. Data products are shown as rectangles in dark and light purple for data derived from the Outer Rim and auxiliary simulations, respectively. Code modules are shown as ovals in dark orange. 78
3.10 Observed $g - r$ color distribution of cosmoDC2 galaxies with $M_{\rm r} < -19$ as a function of redshift, up to $z = 1$. The smooth distribution is obtained through the interpolation procedure described in Sec. 3.5.3. Filamentary structures visible in this figure arise from repeated sampling of library galaxies in sparsely populated color-magnitude space. 82
3.11 Distributions of rest-frame color (top panel) and color-magnitude (bottom panel) of $0.5 < z < 0.54$ galaxies in baseDC2, the Galacticus library and cosmoDC2. The top panel shows that the color distribution of baseDC2 is well recovered in cosmoDC2 by selecting specific Galacticus galaxies, with only a small discrepancy for red galaxies. The contours shown in the bottom panel encompass 75% and 99% of the galaxy populations. For bright galaxies, the luminosity adjustment prioritizes the match with baseDC2 colors while the luminosity rescaling extends the coverage of library galaxies (See Sec. 3.5.3). The brightest and reddest galaxies do not reproduce the baseDC2 colors exactly because there are few library galaxies with similar colors. The luminosity adjustment is applied for faint galaxies, so for faint red galaxies the match-up procedure compromises galaxy color in order to match luminosity better. 83
3.12 Top panel: Observed cumulative i-band number counts per square degree as a function of magnitude from cosmoDC2 (blue) and extrapolated from the HSC survey (black) (see text for more details). The grey shaded band shows a $\pm 40\%$ uncertainty around the HSC extrapolation. The vertical shaded region shows the magnitude range within which the two curves are compared. Bottom panel: Relative difference between the two curves in the top panel. 89
3.13 Redshift distribution of cosmoDC2 galaxies compared with fits to DEEP2 data for a redshift range of $0 < z < 1.5$ for three magnitude-limited samples. The selection cuts are LSST-$r < 22.0$, $r < 23.0$ and $r < 24.0$ as indicated in the legend. 92
3.14 SDSS color distributions of cosmoDC2 galaxies compared with those of SDSS data. The cosmoDC2 galaxies comprise a magnitude- and redshift-limited sample with SDSS r-band magnitude < 17.7 and a redshift range of $0.05 < z < 0.1$. The colors shown in the figure are SDSS $u - g$, $g - r$, $r - i$ and $i - z$. 93

3.15 Validation test for the two-point correlation function, computed on the 440 sq. deg. cosmoDC2 catalog for cuts in SDSS r-band magnitude as given in the legend. The points correspond to measurements on the cosmoDC2 catalog, with error bars obtained through jackknife resampling, and solid lines to SDSS measurements from Table 2 of [153]. 95
3.16 Validation test for galaxy-galaxy lensing, computed on the cosmoDC2 image area with cuts chosen to match those of the SDSS LOWZ sample of [133]. The test returns a total SDSS LOWZ galaxy number density of 58.8 per square degree for the cosmoDC2 image area, compared to the observed value of 57.0 per square degree from [119]. The solid blue line in the figure corresponds to the measurement performed on the cosmoDC2 catalog, and the black points to the measurement on the SDSS sample. At small scales the signal is smoothed due to lensing map resolution limits. 96

LIST OF TABLES

2.1 N-body simulation parameters 10
2.2 Number of clusters sized halos in the simulations and number of clusters in each cluster catalog. 11
2.3 Properties identified or computed for each core. 15
2.4 The best fit parameters for the Rd model of L_* galaxies using infall and peak mass definitions of cores. The error bounds only include statistical error. 42

3.1 Color and A_v cuts applied to the Galacticus library to remove unphysical galaxies 77

ACKNOWLEDGMENTS

I would like to thank my advisor Salman Habib for providing guidance in my transition into cosmology and for always willing to dive into the gritty details when I was stuck. I would also like to thank all past and present members of the Cosmological Physics and Advanced Computing group at Argonne National Laboratory, for without their work which I've built off, their insight and advice, much of the work this work would not have possible. I would also like to thank my thesis committee members: Juan Collar, Josh Frieman, Daniel Holz and Emil Martinec.

I would like to to thank the Department of Energy Advanced Scientific Computing Research (ASCR) and Argonne National Laboratory Department of High Energy Physics for support this work.

Lastly, I would like to thank Julia for sticking through all these years with my "Soon$^{\text{TM}}$".

ABSTRACT

Modern cosmology is entering a new era of precision thanks in part to the next generation of imaging and spectroscopic surveys. These surveys will provide an unprecedented volume of high quality observations. As a result, cosmological analyses will no longer be limited by statistical uncertainty, but instead be limited by the ability to characterize and control systematic errors. Cosmological simulations and the models derived from these simulations play an irreplaceable role in our ability to test, validate and control systematics in analyses. This thesis presents a novel method of modeling galaxies within halos, and the construction of a realistic synthetic galaxy catalog.

We first describe a novel method of tracking substructure within dark matter halos to be used in conjunction with galaxy models. Many galaxy models, including empirical and semi-analytical (SAM), rely on subhalos and subhalo merger trees for determining galaxy positions and evolutions within simulated dark matter halos. Subhalos and subhalo merger trees are computationally expensive and difficult to robustly define. Furthermore, the abundance of subhalos does not match the abundance of galaxies within halos. To fix this mismatch, some galaxy models introduce galaxies, called 'orphan' galaxies, that are not hosted in a subhalo. In our approach, instead of relying on subhalos and 'orphan' galaxies, we introduce a new method, called 'halo core tracking', to track substructure and its evolution within halos. To test core tracking utility and viability in galaxy modeling, we model the distribution of galaxies within galaxy clusters.

In the second part, we discuss the methods used to construct a highly realistic, and yet tunable, synthetic galaxy catalog, named cosmoDC2. Large galaxy catalogs generally use two broad types of models: empirical models and SAMs. CosmoDC2 is built by hybridizing an empirical model with a SAM, and so taking the advantages of each and minimizing the disadvantages. The backbone of cosmoDC2 is an empirical model that is able to precisely capture the distribution of key galaxy properties, such as luminosity and galaxy color. By

matching each 'backbone' galaxy to an individual galaxy in the SAM catalog, we are able to include detailed and consistent galaxy properties that the empirical model cannot provide.

CHAPTER 1

INTRODUCTION

The structure of the universe is well described by the Lambda Cold Dark Matter (ΛCDM) model. In the fiducial ΛCDM model, the energy density of today's universe can be broken down into three main components [68, 112]. The majority (∼69%) of energy density is in dark energy which acts as the energy density of space itself and can be expressed as a constant in Albert Einstein's field equation of gravity [46]. The second largest component (∼26%) is cold dark matter. Dark matter is effectively collisionless and only has significant interactions through gravity. The third largest component (∼5%) is baryonic matter from the standard model (in cosmology, electrons are included in the category of baryonic matter).

The ΛCDM model is supported by a myriad of observations: the structure of the cosmic microwave background [68, 92, 111, 112], the large scale distribution of matter as measured by baryonic acoustic oscillation [22, 47, 70, 93, 129, 130], galaxy distributions [45, 145, 144, 141, 152], gravitational weak lensing [73, 76, 97, 120], the expansion rate of the universe measured by Type Ia supernovae [25, 44, 121, 140] and by the abundances of primordial hydrogen, helium and lithium [38, 72, 139, 150].

Dark energy represents the energy density of vacuous space itself. This vacuum energy exerts a negative pressure, causing the acceleration of the expansion rate of the universe. As the universe expands, the energy densities of all other components decrease as the volume increases. However, according to general relativity, the dark energy density stays constant with time. One of the most important open questions in cosmology is about the nature of dark energy. Any deviation from the expectations of general relativity, such as dark energy evolving with time, would indicate exciting new physics.

Unlike the dark energy that seems to be uniform across the universe, matter is extremely inhomogeneously distributed on scales smaller than about a hundred megaparsecs (Mpc) [68, 112]. The initial matter distribution after the big bang was almost homogeneous–with

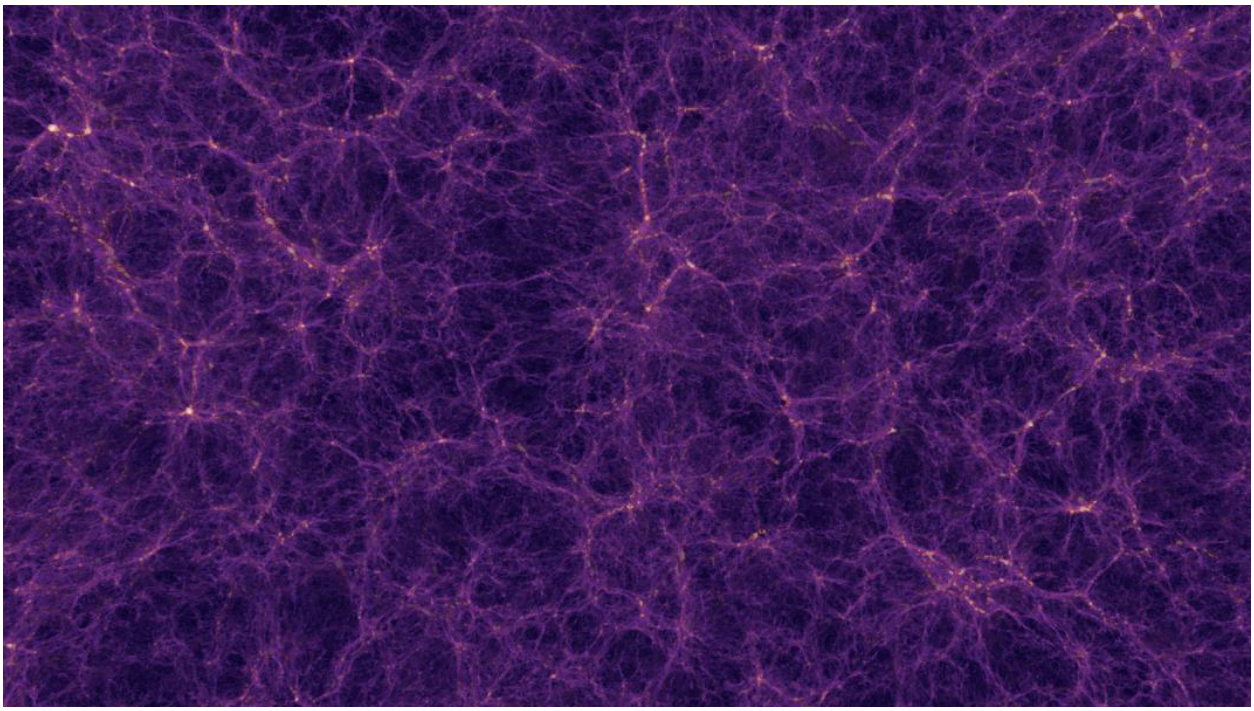

Figure 1.1: The dark matter density from a slice of the Outer Rim simulation.

only quantum fluctuations causing variations in density that were on the order of one part in a hundred thousand. These quantum fluctuations were blown up to cosmological scales during inflation. Due to the gravitational attraction, the regions that were initially more dense gathered more matter and grew in density. This, in turn, attracted even more matter in a run away process. Matter first collapsed into single dimensional sheets. Those sheets collapsed into two dimensional filaments, which in turn collapsed into three dimensional halos. The large scale structures of sheets, filaments and halos create the "cosmic web" shown in Fig. 1.1. Large halos are formed hierarchically from the merging of smaller halos and smooth accretion of matter. The smaller merged halos survive intact as subhalos or substructures within the larger hosting halo.

Unfortunately, $\sim 4/5$ of the matter, aptly called dark matter, has not been directly observed. As far as currently known, dark matter only interacts gravitationally: it produces neither photons nor other standard model particles. Fortunately, the remaining $\sim 1/5$ is

baryonic matter. It can be easily observed as optically bright galaxies, x-ray emitting gas, as well as other forms. By studying these observables, we can understand the distribution of dark matter.

1.1 Large Scale Structure Probes

By studying the large scale structure (LSS) in the universe and its evolution with time, we can gain insight and constrain the natures of dark energy and dark matter. The LSS can be determined by studying the galaxies and the gas that trace dark matter, as well as the light that is gravitationally deflected around structures. Some of the most important probes of LSS include the baryonic acoustic oscillations (BOA), weak gravitational lensing of galaxies, galaxy clusters and their cross correlation with the cosmic microwave background.

Upcoming next generation surveys, such as the Large Synoptic Survey Telescope (LSST) [95], Dark Energy Spectroscopic Instrument (DESI) [42], South Pole Telescope [29], Wide Field Infrared Survey Telescope [7], Euclid [89] and the Subaru Prime Focus Spectrograph (PFS) [142], will provide an unprecedented amount and quality of observational data. As a result, cosmological constraints from these observational surveys will be the smallest statistical errors ever reached. In order to constrain cosmology further, we must now understand and control the systemic errors to a higher level. Models starting from computational simulations instead of observations can provide a lever to understanding and characterizing these errors that can otherwise bleed into analyses and the final cosmological inferences.

1.2 Galaxy Modeling from Simulations

Galaxies serve as an irreplaceable corner stone in observations for cosmology. Modeling of galaxies in simulations plays a critical role in cosmological analysis. Modeling galaxies directly from first principles of physics is not feasible in cosmological simulations. The

physics of galaxy formations occur on the sub-parsec scales while cosmological simulations run on scales of gigaparsecs (Gpc). There is a spectrum of approaches to modeling galaxies on cosmological scales for gravity-only simulations. Broadly speaking, on one end of the spectrum are empirical models that focus on capturing distributions and correlations of observed galaxies without necessarily modeling the physics that drive those properties. On the other end of the spectrum are semi-analytical models (SAMs) that aim to model the physics of galaxy formation itself and attempt to recover observed properties.

There are many flavors of empirical galaxy models. A straight forward class of empirical galaxy models is the Halo-Occupation Distribution model (HOD) [18, 86, 157]. HOD models the number of a certain type of galaxies in a halo as purely a function of host halo mass, typically with about four parameters. HOD had a wide success in reproducing galaxy abundances and two-point clustering statistics. This approach does not take the individual halo assembly history into account nor does it describe the position of the galaxies within halos. Frequently galaxies are distributed either according to the Navarro-Frenk-White (NFW) profile [104] or according to the dark matter profile. Subhalo-abundance matching (SHAM) models also have a wide range of success in reproducing the two-point clustering statistics [146, 36, 131, 16, 103, 58]. Unlike the HOD, SHAMs assign galaxies to subhalos based on peak masses of these subhalos. Their masses are not constant, as the subhalos are stripped of mass due to tidal forces, dynamic friction and subhalo harassment [55, 148, 149]. As a consequence, the low mass of subhalos does not necessarily indicate that these subhalos originated from small merging halos. Furthermore, some subhalos can be completely stripped of any bound mass or be merged into the main body of the halo. SHAMs rely on robust subhalo identification and the construction of subhalo merger trees to identify the peak mass of each subhalo. The computational cost of robustly identifying subhalos and constructing the merger trees is high. SHAMs also suffer from the 'orphan' galaxy problem [103, 27]. Relying on present subhalos does not recreate correct statistics unless additional 'orphan'

galaxies, that are not hosted by any identified subhalos, are included.

This dissertation introduces a procedure to track substructures for galaxy modeling. The substructure tracking method, which we call 'halo core tracking', attempts to alleviate the need for robust subhalo identification and subhalo merger trees, as well as the need to model 'orphan' galaxies. This is accomplished by directly tracking simulation particles of the halo 'core', the central region of the halo, as it merges with larger halos. Halo core tracking can be used for both empirical models and SAMs. To validate the method, we apply a physically motivated empirical model to reconstruct the profiles of galaxy clusters.

SAMs are inherently more complex than empirical models because they attempt to model the highly non-linear physical process of galaxy formation. The benefit of SAMs is that they are able to predict galaxy properties in regimes where observations have not yet been performed. The drawbacks are that SAMs are computationally expensive to run and that they model physical processes that are not present in the simulations or cannot be resolved within the simulations. These processes require to be parameterized in terms of the available quantities in the simulations, leading to a large number of free parameters. SAMs may have up to 200 parameters [17]. As a result of the complexity of galaxy formation, the large number parameters in SAMs, and the computational expense, it becomes difficult to tune the model parameters to reproduce precisely observed statistics. We have developed a method of combining empirical models and SAMs to overcome the weaknesses but keep the strengths of each approach.

This thesis presents a synthetic galaxy catalog and the methods used in its construction. The catalog, named cosmoDC2, is the result of empirical modeling, resampling a galaxy model into a larger simulation and the combination of the empirical model and a SAM. The purpose of the catalog is to prepare the Dark Energy Science Collaboration (DESC) for the LSST data once it begins operation in 2022. The catalog will allow collaboration members to develop, test, run and characterize their analysis pipelines. Unlike real observations,

simulated observations have an accessible truth which is vital to understanding systematics. The synthetic catalog will serve the collaboration in two main avenues. First, it will be used in simulated telescope imaging of galaxies, called LSST Data Challenge 2. These simulations will test and prepare the data management pipeline of processing raw images and science analyses based on these images. Second, the catalog can be used directly without the additional complications of the image simulations to develop and test science analyses. The catalog contains about $\sim$2.6 billion galaxies and each galaxy has about 500 properties.

1.3 Dissertation Outline

We first address the task of modeling galaxies within halos simulation: Chapter 2 introduces a novel method of tracking substructure within gravity-only halos for galaxy modeling and other uses. In Chapter 3, we discuss the construction of the synthetic galaxy catalog for the Large Synoptic Survey Telescope-Dark Energy Science Collaboration (LSST-DESC). Finally in Chapter 4, we summarize the work preformed, the results and the conclusions. We also discuss future work that can stem from the work presented.

CHAPTER 2

HALO CORE TRACKING

The galaxy distribution in dark matter-dominated halos is expected to trace the details of the underlying dark matter substructure. In this paper we introduce halo 'core-tracking' as a way to efficiently follow the small-scale substructure in cosmological simulations and apply it towards modeling the galaxy distribution in clusters. The method relies on explicitly tracking the set of particles identified as belonging to a halo's central density core, once a halo has attained a certain threshold mass. The halo cores are then followed throughout the entire evolution of the simulation. The aim is to simplify analysis by avoiding the use of subhalos and, at the same time, to more easily account for the so-called "orphan" galaxies, which have lost substantial dark mass to tidal stripping. We show that simple models based on halo cores reproduce the number and spatial distribution of SDSS galaxies in redMaPPer and SPIDERS clusters.

2.1 Introduction

In the ΛCDM model , which well describes the structure of the universe , small initial density fluctuations in the matter distribution after inflation gravitationally collapse into localized clumps, called halos. Halos are dominated by dark matter (DM) and are the nurseries from which galaxies form . The formation of halos is a hierarchical process: larger halos are the result of mergers of smaller halos. The history of a halo, its "merger tree", also helps to determine its galaxy distribution.

As halos merge, the remnants of previously accumulated halos can survive as localized over densities called subhalos. Subhalos are subjected to tidal forces, dynamic friction and ram-pressure stripping, which strips them of bound mass and may lead to their eventual disruption. These effects are particularly pronounced inside galaxy cluster sized halos, the

most massive gravitationally bound objects in the universe.

In the case of an observed cluster at some redshift, many of the galaxies within the cluster halo did not form inside it, but were gathered through merger events with other halos. Individual galaxies form at the center of halos where the gas density would be the highest for star formation [155, 23]; the most massive galaxies are the result of multiple galaxies merging into one object.

The spherically-averaged DM halo density profile can be reasonably well-described by a broken power law with a dense center ($\rho \sim 1/r$), but with much less dense outskirts ($\rho \sim 1/r^3$) [105], the so-called NFW profile. The NFW profile is determined by two parameters, which can be taken to be the halo mass and the halo concentration, a density scale parameter. The observed distribution of galaxies tends to follow the NFW profile . In simulations, subhalos do not follow an NFW profile, as their population density, relative to DM, is suppressed at small radius [58, 54]. This breakdown of a simple galaxy-subhalo identification is the origin of "orphan" galaxies invoked in semi-analytic models of galaxy formation. As the above indicates, directly associating galaxies with subhalos can be problematic for several reasons. First, the current subhalo mass does not directly reflect the mass of the original halo when it was accreted by the main halo (the "infall" mass). Up to 90% of the infall mass can be lost from tidal stripping, dynamic friction, or other interactions. Second, finding subhalos and determining their properties robustly in simulations is not straightforward [107, 83]. Third, subhalos themselves can be lost from a low force and mass resolution in the simulation [149].

From the point of view of establishing a simple description of galaxy populations in halos, we introduce a model that, instead of relying on subhalos, follows a group of simulation particles that can be explicitly tracked and serve as a proxy for a galaxy position. Several assumptions underlie this procedure. As in subhalo abundance matching (SHAM), we first assume that all galaxies form as central galaxies in halos and become satellites only through

merging with other halos. Second, that once a galaxy is formed, the main force that dictates the bulk motion of the galaxy is gravity. (The stellar component is dense enough that hydrodynamic forces are negligible and the stars are sparse enough that they are effectively collisionless.) With these two assumptions, the trajectory of the central or 'core' set of simulation particles of a halo should have a very similar trajectory to that of a galaxy that originated at the center of the same halo. We call the central set of simulation particles the halo core. Following the halo core not only yields the galaxy position and velocity inside the halo, but also informs us if the galaxy experienced strong tidal forces that may have led to its disruption.

In this paper, we describe our method for identifying and following the halo cores using two cosmological simulations; a smaller one for subhalo comparisons, development, and testing, and a very large run that has sufficient volume for carrying out cluster analyses. Halo core modeling, combined with a SHAM approach to populating galaxies in halos, is used to reconstruct the radial profiles of galaxies in SDSS clusters, with very good results.

The paper is organized as follows. The simulations and the halo core tracking algorithm are discussed in Sec. 2.2.2 and in Sec. 2.2.3. Sec. 2.3 covers how observed data from SDSS is converted to galaxy cluster surface density profiles. We go over how we apply several simplistic galaxy models to the cores in Sec. 2.4. In Sec. 3.6, we discuss the fitting results and the various robustness tests performed. The concluding remarks and future outlook are given in Sec. 2.6.

2.2 Methodology and Modeling

In this section we describe the cosmological simulations used for the halo core analysis, and present the procedures employed for core identification and tracking, along with a description of measured core properties. A comparison of the core distribution against that of subhalos is investigated and discussed; in this context a short description of galaxy modeling with

Simulation	Box Length	Particles	Mass Resolution
AlphaQ	256 h^{-1} Mpc	1024^3	$\sim 1.6 \times 10^9$ h^{-1} $M_\odot$
Outer Rim	3000 h^{-1} Mpc	10240^3	$\sim 1.7 \times 10^9$ h^{-1} $M_\odot$

Table 2.1: N-body simulation parameters

cores as contrasted with HOD and SHAM techniques is also provided.

2.2.1 N-body Simulation

This work employs multiple N-body gravity-only simulations, generated using the Hardware/Hybrid Accelerated Cosmology Code (HACC) [60]. The properties (box sizes, number of particles) of the simulations are listed in Table. 2.1. The simulations share the same flat ΛCDM cosmology with parameters similar to WMAP7: $\Omega_{\rm tot} = 0.265$, $\Omega_b = 0.045$, $h = 0.71$, $\sigma_8 = 0.8$, and $n_s = 0.963$. All calculations in this paper are carried out with these cosmological parameters.

The small AlphaQ simulation was used for development of methods and codes, as well for the subhalo-core comparisons. The much larger Outer Rim simulation [63] provided the primary data source to perform fits to cluster profiles.

All the simulations were run and processed similarly: the simulations start at z_{in} ∼200 and evolve to $z = 0$. About 100 snapshots of the simulation are saved to disk from $z \sim 10$ to $z = 0$. Halos in each snapshot are identified using the Friends-of-Friends (FoF) halo finder algorithm [40] with a link length $b = 0.168$. The halo centers are defined as the most gravitationally bound FoF particle. Spherical over-density (SOD) halos [114] with density 200 times the critical density (M_{200c}) are identified using the FoF halo minimum potential as the SOD halo center. Cluster M_{200c} are converted to SOD masses with density 200 time the mean density (M_{200m}) assuming an NFW profile and a mass-concentration relationship specified in [31]. The number cluster sized halos in each simulation is given in Table 2.2.1.

Merger trees are constructed by comparing common particles between FoF halos in adja-

Mass	AlphaQ	Outer Rim	redMaPPer	SPIDERS
$14.00 < \log(M_{200c}) < 14.25$	92	175735	2746	2
$14.25 < \log(M_{200c}) < 14.50$	26	60719	888	39
$14.50 < \log(M_{200c}) < 14.75$	7	16203	230	91
$14.75 < \log(M_{200c}) < 15.00$	1	3018	49	22
$15.00 < \log(M_{200c}) < 15.25$	1	318	5	5

Table 2.2: Number of clusters sized halos in the simulations and number of clusters in each cluster catalog.

cent time steps to establish progenitor and descent relationships. In most cases, halos merge into a single halo but some individual halos split into multiple halos, which is a physical process in the simulation [148] . Any halo that splits into two or more descent halos at a later time step is separated into a corresponding number of "fragment" halos equal to the number of descent halos. The "fragmentation" of halos effectively undoes the halo merger event until the halos permanently merge. The resulting halo merger trees strictly have mergers and each tree is considered fully independent of any other tree. Further details on the employed merger tree technique can be found in [117].

2.2.2 Identifying and Tracking Individual Halo Cores

Halo core tracking is preformed over two stages. The first stage identifies particles that will represent cores of individual halos at each simulation snapshot and tracks those particles for the remainder of the simulation. A second stage uses the halo merger trees and the tracked particles to follow the evolution of cores as halos evolve and merge. The first stage is expensive to run as it requires the full particle data of the simulation. The second stage works with the truncated set of particles from the first stage and is relatively less expensive to run.

In the first stage, every FoF halo above a certain particle number or mass threshold has its halo core identified as the spatially closest 20 simulation particles to the halo center. For the threshold we use 100 particles corresponding to $M_{FoF} \sim 2 \times 10^{11} h^{-1} M_{\odot}$for AlphaQ and

OuterRim. This processes is repeated for each FoF halo at each simulation snapshot. Note that a passively evolving halo with no mergers will have a new set of core particles identified for every simulation snapshot.

Every particle that has been flagged as belong to a core at one point is added to a running list of particles, starting from early snapshots. Particles in this set can belong to multiple halo cores and so the membership to halo cores is not stored. At every snapshot, the position and velocity of these accumulated particles is updated from the simulation. In the second stage, we use this data set to track the evolution of cores. The accumulation of particles in this manner avoid the need to load the full simulation particle data in the second stage, which would be prohibitively expensive.

2.2.3 Tracking Halo Cores with Halo Evolution

The second stage links cores across snapshots to track the evolution of halos and substructure within halos. Cores are linked by the use of the halo merger trees. A cartoon of the process is shown in Fig. 2.1. We will first discuss the general case of construction and then discuss the complications related to fragment halos and other effects.

In the merger trees, we can define three basic types of halos: halos below the chosen mass threshold of 100 particles, isolated halos above the mass threshold, and composite halos above the mass threshold. Halos below the threshold do not have their core track until they gain enough mass to surpass the threshold hold. Once a halo does, we start tracking its core. At that stage it's an isolated halo as it only has one core at the center and no satellites. Isolated halo cores are assigned the set of particles that were most recently identified as being the closest to the halo center at that time, at usually the same time step. When two or more halos containing central cores merge into one halo at the next time step, the core belonging to the most massive merging halo is assigned as the central core for the merged halo at next time step. The other central cores then become satellites. Satellite cores are permanently

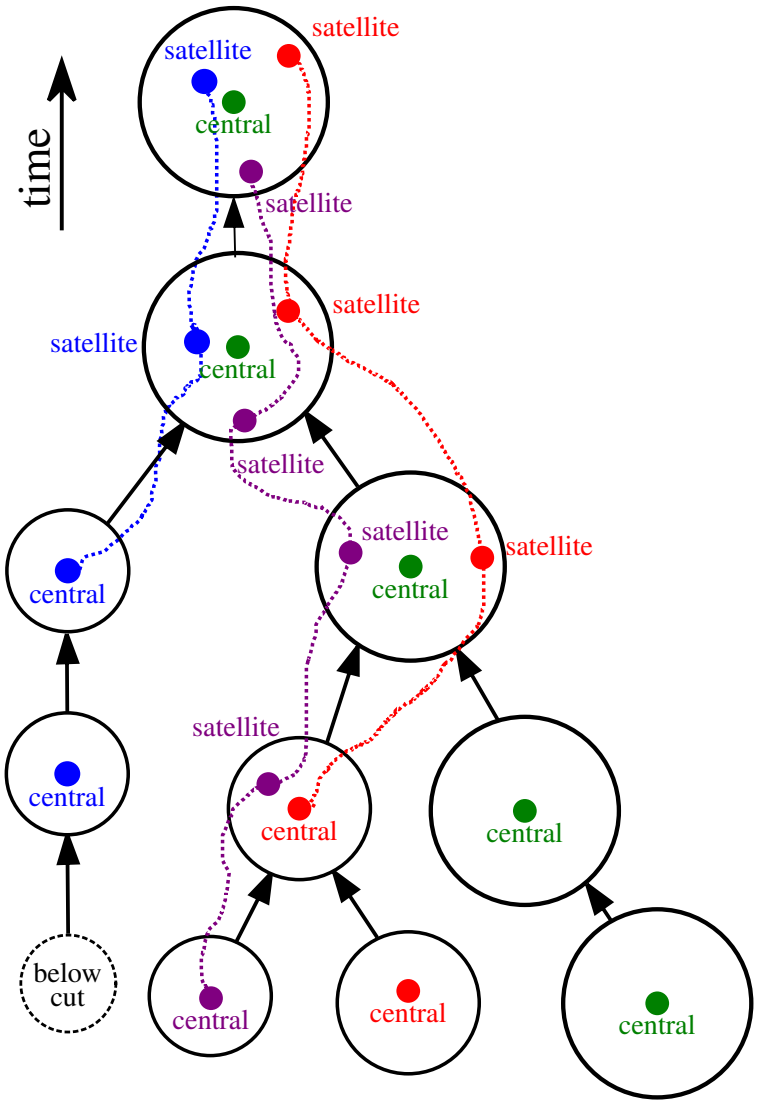

Figure 2.1: Cartoon of the method of assigning cores to halos using the merger trees. Circles are DM halos and the arrows represent halo descents across adjacent simulation snapshots. Colored dots represent halo cores, each color a separate core. Cores that are linked by a dashed line are represented by the same set of DM simulation particles. Each halo can host up to one central core and any number of satellite cores. All cores start out as isolated centrals in halos above the mass threshold (solid circles). Halos below the mass threshold (dashed circles) do not have their halo core tracked. Central cores are updated with a new set of particles from the halo center at each simulation snapshot. Central cores become satellites cores when they merge with a heavier halo. As a result every halo will have exactly the same number of cores as in it's merger tree branches above the mass threshold.

locked to the last set of particles assigned right before the merger. While the central core will be updated with new particles from the halo as if it was an isolated halo.

There are several edges cases that we handle. Halos may have two or more potential minimums that are comparability deep and the halo potential center may flip-flop between them across snapshots. Each potential should be tracked by a core. If we in this case naïvely update the central core with the particles at potential center, one potential will not be tracked and the other will be represented by two cores. To avoid this type of jump, we require the distance between the central core and the halo potential to below a threshold.

In the merger tree construction, halos that have any splits are separated into a number of fragment halos. Most fragment halos do not have well defined positions–with the exception of the largest fragment which is assumed share the same potential center as halo it was fragmented from. Fragment halos are not present in the FoF halo catalogs nor do they have well defined centers so no core particles are associated with them for that time step (unless it's the largest fragment halo). Instead they adopt the core particles from their main progenitor halo, which may be another fragment halo. In very rare cases, this adoption method does produce any particles as each progenitor is either a fragment or below the mass threshold.

Halos that have passed the mass threshold may fluctuate below the threshold at some future time step. For these halos, we follow a similar procedure of adopting previous core particles and include them in our analysis. As a result some cores have a recorded infall mass (described below) that is less the mass threshold for tracking cores.

Once all the halo cores have their particles assigned, we compute the properties of each core. The full list of properties are in Table 2.3. Satellite cores have additional information from when they were last a central galaxy in a host halo, right before their host halo infallen into a more massive host halo. The infall redshift, identify of the infall halo and mass of the infall halo are recorded. For central galaxies, the infall mass is simply the current host halo.

Property	Description
Core tag	unique identifying number
Host halo tag	identifies the current hosting halo
Core position	position of center most particle
Core velocity	median velocity of particle set
Core radius	compactness measure
Peak core radius	maximum radius obtained
Infall halo tag	identifies infall halo
Infall mass	mass of the infall halo
Infall time	time step of infall halo
Infall peak mass	the maximum mass infall halo obtained
Central flag	identifies it the core is currently a central

Table 2.3: Properties identified or computed for each core.

These infall masses are important in our modeling of galaxy luminosity using cores.

For all cores, the position of the core is defined to be the position of the core particle that has the minimum distance to all the other core particles. The core velocity is taken to be the median velocity of all core particles, calculated in each dimension separately. The effective radius, a measure of compactness, is the root mean squared of the standard deviation of particle positions in each dimension:

$$\mathrm{R_{eff}} = \frac{1}{3}\sqrt{\frac{1}{N}\sum_{i=0}^{N}(\bar{x}-x_i)^2+(\bar{y}-y_i)^2+(\bar{z}-z_i)^2}, \tag{2.1}$$

where N is the number of particles in the core, x_i is the x position of i^{th} particle, and $\bar{x}, \bar{y}, \bar{z}$ is the average x, y, z position of all the particles in the core, respectively. The distribution of core radii and infall masses for AlphaQ at z=0 is shown in Fig. 2.2.

While several definitions of effective radius were explored, they all preformed similarly and correlated well ($r > 0.9$) with cores sampled with 50 particles. The other definitions explored were the average particle distance from the central particle, 50th, 60th and 80th percentile of particle distances from the center. Applying Chauvent's criterion outlier rejection for the all methods did not improve the convergence properties of the radius definition.

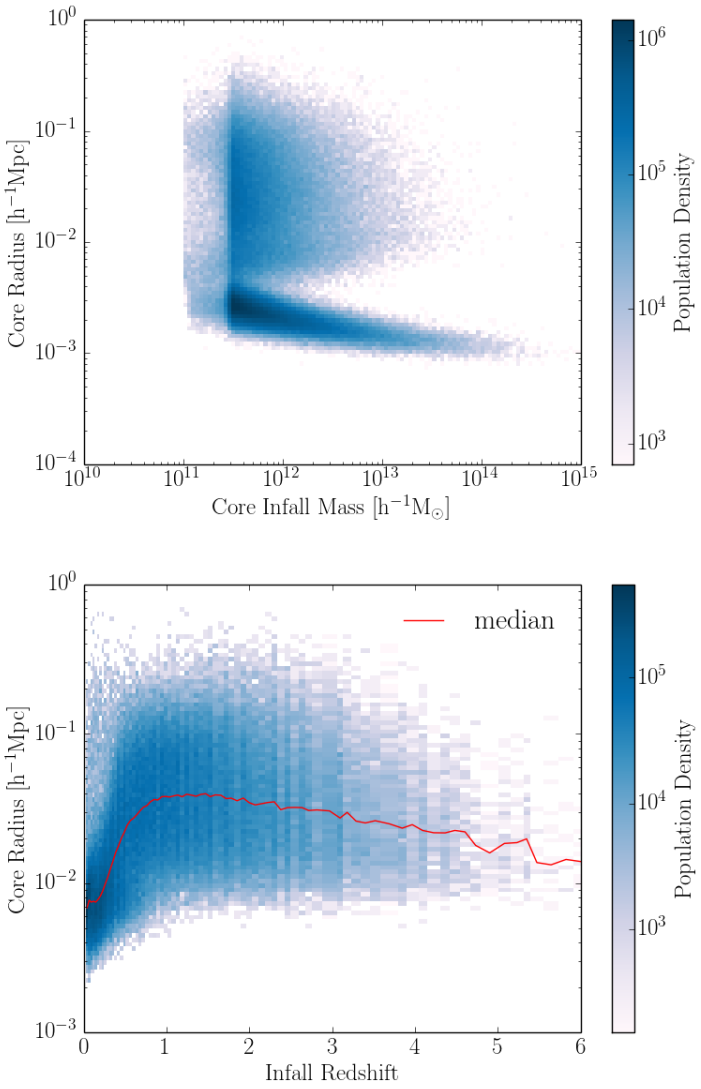

Figure 2.2: *Top panel:* Population density as function of core radii and core infall mass in the AlphaQ simulation. Several distinct features are visible. The bottom group with the smallest radii are central cores. The particles representing the core have been refreshed to be the spatially closest of the current host halo. The population above the central are satellite cores which had time for the core particles to spread out. The population to the left of the vertical ridge are cores which originate from halos below the 100 particle threshold but have a progenitor that was above the threshold and inherited the core. *Bottom pane:* Core radius as a function of core infall redshift for all cores in AlphaQ at $z = 0$. Satellite cores that have recently merged still have a small radius as the cores have not experienced strong tidal forces in the host halo and other harassment.

Once the core the properties are calculated, the core catalog is assembled. In the catalog, it is possible to determine which cores belong in which halo and to follow the assembly history of the cores of interest. The z=0 distribution of cores and the assembly core trajectories for two example halos are shown in Fig. 2.3. Halos have a tight cluster of cores that have fallen into the potential center of the halo. In observations, the centers of galaxy clusters are often populated by the brightest central galaxy (BCG) . These BCG are elliptical galaxies are often the result of many galaxies falling into the potential center and merging into one massive galaxy . Cores follow a similar pattern where many cores from mergers concentrate at the potential center. We will be modeling cores merging into a single galaxy proxy in Sec. 2.4.1.

2.2.4 Comparison of Cores to Subhalos

We compare cores to subhalos as found by a subhalo finder that combines spatial density and phase space information. The halo particle are organized in Barnes-Hut (BH) tree and an SPH kernel is used to compute local densities, similar to the approach in SUBFIND [138], followed by density sorting, phase space neighbor finding, and iterative gravitational unbinding.

The position and velocity of the subhalo are calculated as the mean of the position and velocities of the subhalo particles.

In order to compare the subhalo distribution to that of cores without resorting to subhalo merger trees, we match the two spatially. For each subhalo, we find the closest halo core within 0.1 h^{-1} Mpc. Extending the search radius can pair uncorrelated subhalos and cores. We find that $\sim$99% of the subhalos above 100 particles (1.6×10^{10} h^{-1} $M_\odot$) are matched to cores, see Fig. 2.4. The remaining small fraction of unmatched 100+ particle subhalos can be explained by the difference in the position definition: Unrelaxed subhalos can have a center of mass position that is significantly differs from their own potential minimum, which

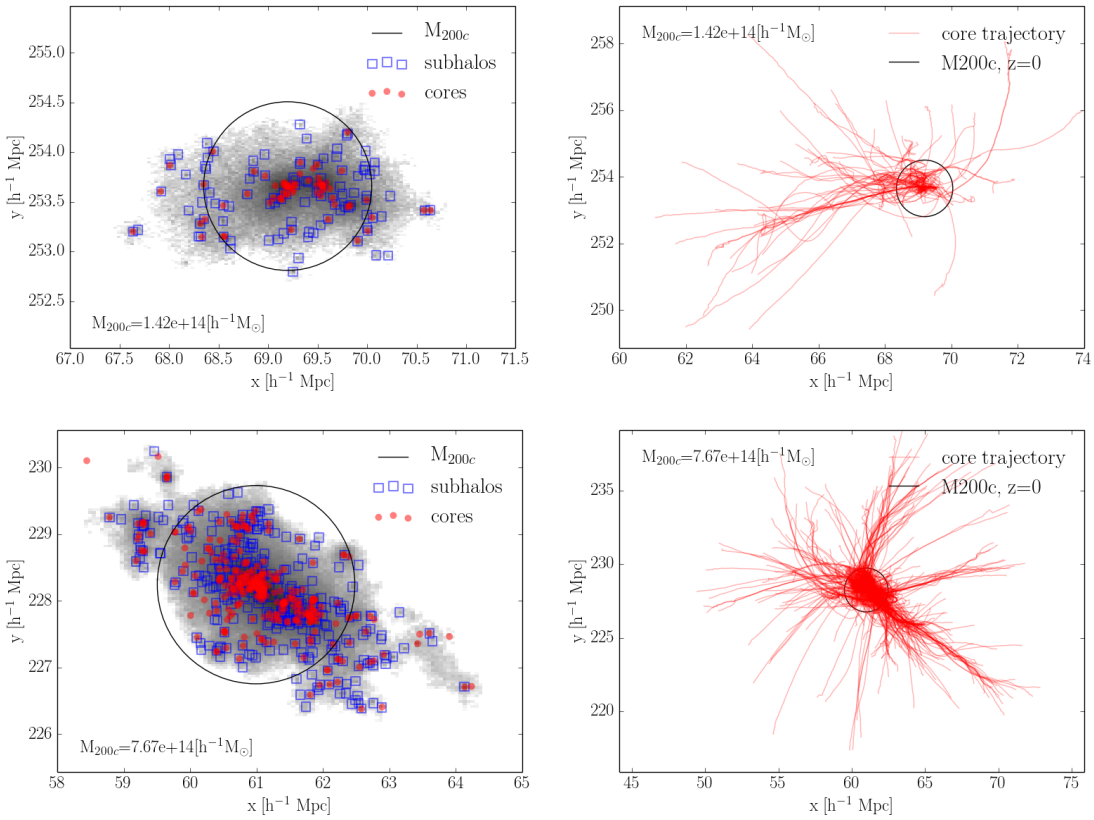

Figure 2.3: Core distribution and assembly for two representative halos at $z = 0$. The left panels show the distribution of cores and subhalos. Subhalos are resolved down to masses equal to $\sim 1.610 \times h^{-1}\mathrm{M}_{\odot}$, while cores only originate from halos which have FoF halo mass of at least $\sim 1.6{\times}\mathrm{h}^{-1}\mathrm{M}_{\odot}$. The right panels show the trajectories of cores throughout the assembly history of the halos. The black circle is the R_{200c} at $z = 0$.

is where cores typically reside inside subhalos. Subhalos below 100 particles can originate from halos that never had more than 100 particles. These low mass halo do not have their core tracked and so we would not necessarily expect each subhalo of these subhalo to have a core.

As a smaller halo merges with a larger halo, with Due to tidal stripping and other mechanisms [149] , the current subhalo mass should be lower than the peak mass or the infall mass of the (sub)halo.

The ratio of current subhalo mass to core infall mass is almost always less than one, as

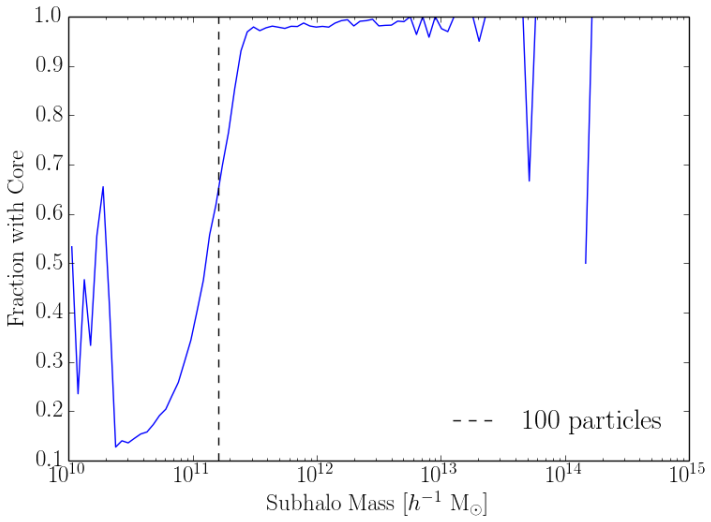

Figure 2.4: Fraction of subhalos with a matched core as a function of subhalo mass in the AlphaQ simulation. Subhalos below 100 particles (dashed line)

expected. The mass ratio does correlate with the infall time of the cores as well as with the effective radius. Cores that merged significantly earlier, or that have a larger radius, on average have relatively more mass stripping. Matching cores to subhalos is complicated by the fact that cores are able to group at density potentials peaks in halos where there maybe be only one subhalo, see Fig. 2.3. If we preform FoF merging of cores with a link length 0.1 h^{-1} $M_\odot$, we can match these FoF core groups to subhalos. We find that most groups $\sim 97\%$ were able to match to a subhalo.

2.2.5 Modeling the Galaxy Population in Halos

There are a number of empirical models for populating halos with galaxies, two popular classes are halo occupation distribution (HOD) approaches [110, 19, 158] and subhalo abundance matching (SHAM) [86, 36, 16, 103] methods. The simplest HOD model describes how many galaxies are expected in a halo as a function of halo mass alone. Usually the modeled galaxy population are galaxies above a fixed luminosity threshold. The functional forms in

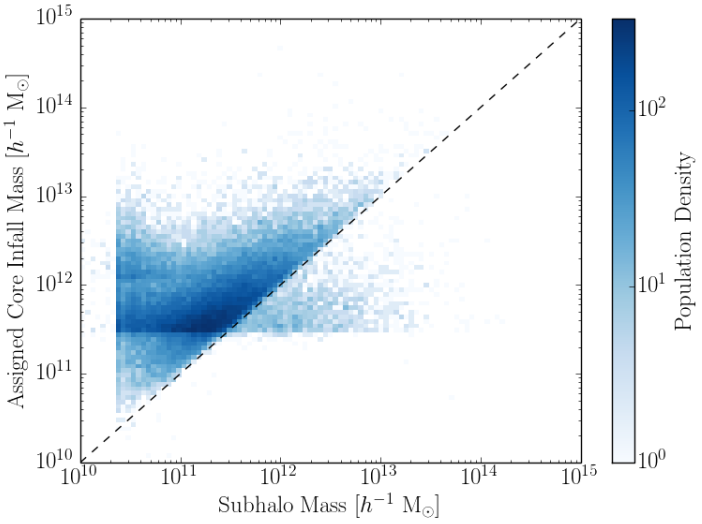

Figure 2.5: Core infall mass against subhalo mass in the AlphaQ simulation at $z = 0$. The dashed black line is the one-to-one relationship line.

HOD models have a number of free parameters and these are tuned to match the density and two point correlation function of the specific galaxy populations of interest. The galaxy population is separated into central and satellite galaxies: The central galaxy is located at the center of the halo and the satellites are distributed spatially following a prescription related to the halo's DM profile or another NFW profile.

SHAMs rely on subhalos to populate halos with satellite galaxies. In the simplest SHAM model, there is a monotonic one-to-one relationship between subhalo masses and the galaxy luminosity. Typically the subhalo peak mass (M_{peak}) or peak maximum circular velocity (V_{peak}) is used instead of the current mass and maximum circular velocity as (sub)halos may loose up to 90% of their initial bound mass after merging with a larger halo . Due to the subhalo matching, galaxies can be tracked across simulation snapshots. Compared to HODs, a SHAM model requires much higher resolution simulations as the subhalos need to sufficiently resolved and tracked through the simulation while the galaxy hosted by that subhalo is not as disturbed . Because galaxies are tied to persistent subhalos, galaxies can be traced across simulation snap shots. Subhalo methods have a number of short comings. As

mentioned earlier, subhalo identification and construction of subhalo merger have additional complexities that must be solved , require high simulation resolution , and are computationally expensive to run on large simulations. Additionally, subhalo based models do not recover the correct small scale clustering unless an 'orphan' galaxies are included into model . 'Orphan' galaxies are not hosted by a subhalo and their position and evolution must be modeled using an alternative method.

The approach taken in this paper is in the spirit of a SHAM model but relies on halo cores to model galaxies rather than subhalos. The simulations used here simultaneously possess sufficient mass resolution to model bright galaxies in clusters, as well as large enough volumes to provide excellent statistics; additionally the robustness of cores relative to subhalos and the increased ability to model galaxies within the halo's virial radius provide a set of natural parameters (e.g., describing core disruption and central/satellite core merging) that can be used to extend the single mass threshold parameter of the SHAM approach. The details underlying our core-based galaxy modeling methodology are presented in Section 2.4.

2.3 Observations: SDSS Data Reduction

Empirical methods, by construction, rely on observational data to determine modeling parameters. In this paper, we use the number of cluster galaxies above a luminosity threshold, as well as the spatial distribution of these galaxies, as a function of halo mass as the input information. Using this, we investigate how well the core-tracking approach works, which of its parameters are most effective, and how robustly they can be determined. The source of the optical galaxy information here is the SDSS, which restricts us to relatively low redshifts.

We use optically identified redMaPPer clusters [127] and x-ray identified SPIDERS clusters [32] as our observational cluster catalogs. For each cluster within $0.1 < z < 0.35$, we obtain all galaxies within R_{200} from the SDSS DR15 main galaxy sample to construct radial galaxy surface density profiles. We utilize both M_{200c} and M_{200m} mass definitions in our

fits. To convert from one mass definition to the other, we assume an NFW profile and a mass-concentration relationship as specified in [31].

2.3.1 redMaPPer Clusters

We use redMaPPer catalog v6.3 [1] provides optically selected galaxy clusters from SDSS photometric data. In brief, redMaPPer finds clusters of red galaxies and calculates a color and position weighted galaxy richness. The richness can be converted to give an estimate of the cluster mass. From the cluster richness and redshift, we calculate the cluster M_{200m} mass using the weak lensing calibration determined in [132]:

$$M_{200m} = 10^{14.344}(\frac{\lambda}{40})^{1.33}, \tag{2.2}$$

where λ is the cluster richness calculated by redMaPPer. The probability of each cluster galaxy member to be the central galaxy is weighed by the luminosity of the galaxy, color, photo-z and local environment. The number of clusters in our in redMaPPer sample is given in Table 2.2.1.

2.3.2 SPIDERS Clusters

The SPIDERS (SPectroscopic IDentification of eROSITA Sources) program within SDSS-IV eBOSS aims to provide a complete and homogeneous optical follow-up to X-ray clusters [32]. Before the availability of eROSITA data, the cluster targets are taken from the ROSAT all-sky survey (RASS) and from serendipitous XMM-Newton surveys. In particular, we use RASS CODEX sample [2]. The catalog provides an x-ray selected sample of galaxy clusters and estimates the M_{200c} cluster mass from the x-ray luminosity and temperature. We use

1. http://risa.stanford.edu/redmapper/

2. https://data.sdss.org/sas/dr14/eboss/spiders/analysis

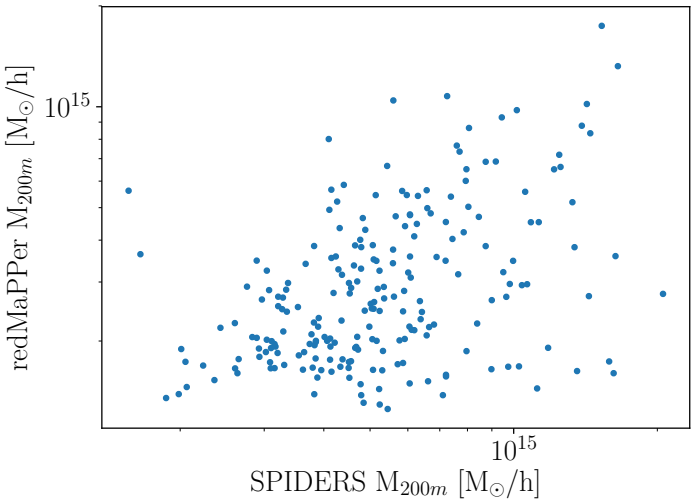

Figure 2.6: Halo mass as determined in the redMaPPer and SPIDERS catalog, converted to M_{200m}

their pre-computed value in our analysis. Instead of using the x-ray center, we use the brightest central galaxies (BCG) identified for each cluster [49] as the cluster center.

The SPIDERS clusters overlap with redMaPPer clusters is about 76% for clusters $z < 0.35$. The overlap is computed by spatially matching cluster centers in RA, Dec and redshift space. To be considered the same cluster, the angular separation must be less than 0.1° and redshift difference must less 10%. Although both catalogs share the same clusters, the mass of clusters in each catalog are estimated using different methods: richness for redMaPPer and x-ray luminosity and temperature for SPIDERS. The comparison of halo mass for the matched clusters is shown in Fig. 2.6.

2.3.3 Obtaining SDSS Galaxies

In order to construct individual galaxy surface density profiles, we first must determine the angular size of R_{200} on the sky for each cluster. By definition, the SOD halos directly link

the physical mass and physical radius by:

$$\mathrm{R}_{200m/c} = \sqrt[3]{\frac{3\mathrm{M}_{200m/c}}{4\pi 200\rho_{m/c}}}, \tag{2.3}$$

where $\rho_{m/c}$ is the mean/critical density at the cluster's redshift. Given the physical radius of the cluster, the angular radius on the sky (knowing the cluster's redshift) is calculated using the Astropy cosmology package [9].

For each cluster, all galaxies in the main SDSS DR15 main galaxy sample that are within the angular size of R_{200} from the cluster center are gathered. Galaxies with flags BRIGHT, SATURATED, SATUR-CENTER, NOPETRO, DEBLENDED-AS_MOVING set in the g, r or i bands are removed. Any cluster with a BRIGHT_STAR or BLEED mask that overlaps with the cluster's R_{200} is excluded from the analysis. A total of 18.5% of redMaPPer and 26.6% of SPIDERS clusters are removed due to an interloping mask. While there is correlation between cluster sky area and clusters being masked, there is no significant bias in cluster mass. These masked are often a result of bright stars in our own Milky Way galaxy that are of course uncorrelated with external galaxies.

We construct profiles for several rest frame SDSS r-band luminosity thresholds. The four galaxy luminosity thresholds used in this paper are: $2.5\mathrm{L}_\star$, $\mathrm{L}_\star$, $0.63\mathrm{L}_\star$ and $0.4\mathrm{L}_\star$ which correspond to r-band magnitudes $\mathrm{M}_\star$-1, $\mathrm{M}_\star$, $\mathrm{M}_\star$+0.5, $\mathrm{M}_\star$+1, respectively. $\mathrm{L}_\star/\mathrm{M}_\star$ corresponds to the bend in the galaxy luminosity/magnitude function. Galaxies brighter than the threshold are selected by a k-corrected r-band magnitude cut. We use the same polynomial approximation for $\mathrm{M}_\star(z)$ as in [127]:

$$M_\star(z) = 12.27 + 62.36z - 289.79z^2 + 729.69z^3 - 709.42z^4. \tag{2.4}$$

2.3.4 Galaxy Radial Profile Stacks

We first separate clusters into mass bins into seven mass bins from 14.0 to 15.75 $log_{10}(\mathrm{M}_{200m}/\mathrm{h}^{-1}\mathrm{M}_{\odot})$. The number of cluster halos in each mass bin for each sample is given in Table 2.2.1. For each cluster, we bin galaxy into 15 linearly spaced radial bins from $r = 0$ to $r = \mathrm{R}_{200m}$, calculate the surface density for each bin and subtract the expected background surface density (see following section). The surface density is calculated in units $(\mathrm{R}_{200m})^{-2}$. These processed radial profiles are stacked according to their mass bin. The estimated error for each stacked radial bin is a combination of Poisson counting error on the sum of galaxies in the stack and the error on the mean from different clusters.

2.3.5 Background Estimation and Subtraction

Not all galaxies projected within the galaxy cluster R_{200} belong to the galaxy cluster. Galaxies uncorrelated to the halo can appear either in front or behind the cluster at different redshifts. To remove the interloper galaxy contribution to the cluster galaxy profile, we assume that the uncorrelated galaxies have a constant and flat spatial distribution that can be estimated and subtracted out. Any excess galaxy counts above the background level should correspond to the contribution from the cluster. We calculate the expected background galaxy density per steradian as a function of the magnitude threshold and subtract it from individual cluster profiles, similar to process in [56, 61].

While the background galaxies have a constant observed magnitude distribution, the magnitude cut applied to clusters depends on both on the cluster redshift and rest-frame luminosity threshold we are using to construct the galaxy profiles. To this end, we require a background galaxy density estimate for each cluster to take into account of both the luminosity threshold and redshift. To calculate background galaxy surface density, we use redMaPPer's random catalog, which points to fairly sampled random locations in the SDSS footprint. For each luminosity threshold, we calculated the average counts of galaxies per

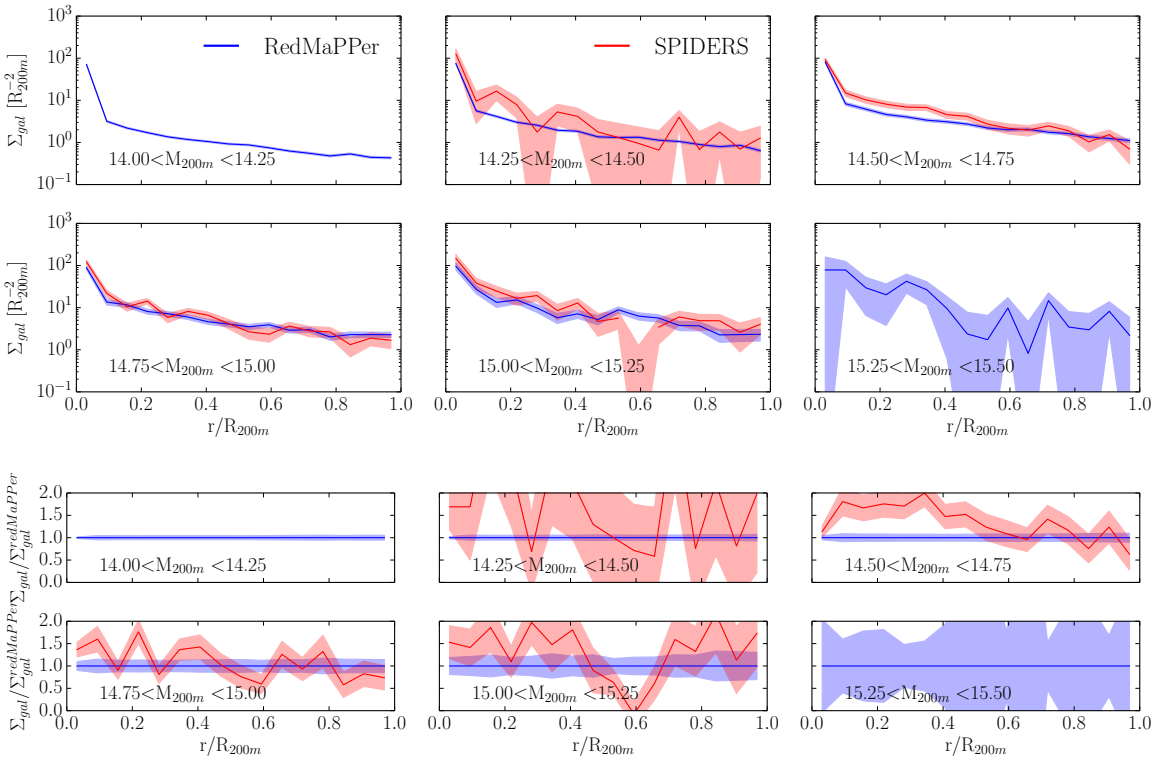

Figure 2.7: *Top panels:* Stacked surface density profiles of L_* galaxies for the redMaPPer and SPIDERS cluster catalogs separated into halo mass bins. *Bottom panels:* Stacked profiles normalized to the redMaPPer cluster sample.

steradian using Eq. 2.4 at 12 linearly spaced redshifts spanning from $z = 0$ to $z = 0.4$. The background density at a specific redshift is interpolated using a spline from these 12 redshifts.

Once the raw galaxy counts are obtained for each cluster in our sample, the count of galaxies in each radial bin is reduced by the product of background surface density at the cluster redshift and the sky area of the radial bin.

2.4 Model Fitting

In this section, we present a detailed description of our approach to modeling the galaxy distribution in halos using halo cores. We discuss several different options that are available

in such a modeling process and justify the choices made.

2.4.1 Modeling Cores as Galaxies

Once the simulation output has been reduced (Section 2.2) to the core catalog, we aim to match the halo cores to galaxies above a given luminosity threshold. In order to do this, we explored a few options as described further below. All of these share a common aspect, viz., the core needs to originate from an infall halo above an associated mass threshold to be considered as a candidate galaxy position. We also allowed two non-exclusive mechanisms: the disruption or loss of cores, and the merging of cores into a single core, and hence, galaxy. In sum, we have four models, each of which converts cores into galaxies, labeled as:

- **Mi**: Cores with infall mass greater than M_{infall} are labeled as galaxies.
- **Rd**: Cores with infall mass greater than M_{infall} and radius less than $R_{disrupt}$ are labeled as galaxies.
- **Rm**: Cores with infall mass greater than M_{infall} that can merge with other cores above M_{infall} if they are within a distance R_{merge} of each other. The unmerged and merged cores are labeled as galaxies.
- **RdRm** : Cores with infall mass greater than M_{infall} and radius less than $R_{disrupt}$are kept. The cores that pass both thresholds can merge into a single object if they are within R_{merge} of each other.

Core Infall Mass Threshold

Similar to the concept in SHAM, the infall mass of the core will serve as a proxy for the luminosity of the galaxy modeled. Since we are modeling profiles of galaxies above a luminosity threshold, cores need to be above a mass threshold, labeled as M_{infall}, to be considered

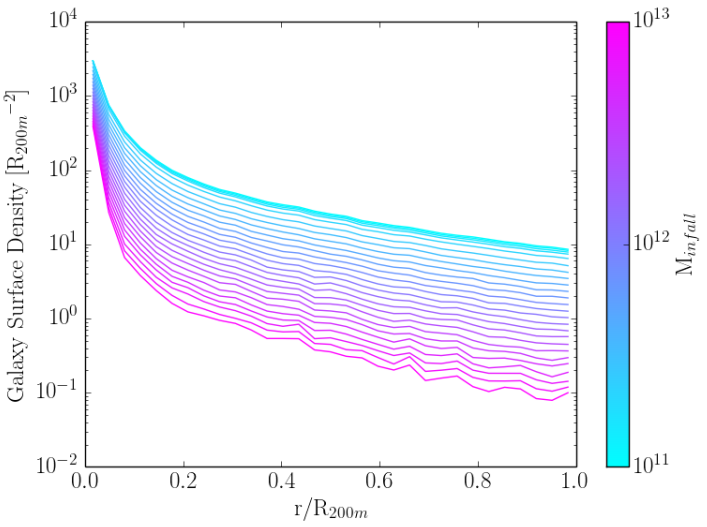

Figure 2.8: Galaxy surface density profiles of cores as a function the M_{infall} threshold value for halos $14.0 < log_{10}(M_{200m}) < 14.25$ in the Outer Rim simulation at $z = 0.24$. Varying M_{infall} significantly modifies the overall normalization of the profile but also creates a more peak profile shape for higher M_{infall}.

as a galaxy above the luminosity threshold. Depending on the model, a galaxy above the threshold can be lost either due to disruption or merging with another galaxy.

There is considerable freedom in choosing the particular quantity that represents the best infall mass proxy for the purpose of galaxy modeling. Aside from the direct halo mass at infall, one may use the maximum circular velocity, or the maximum mass attained by the infalling halo at any time in its past history (the “peak" mass). We use the FoF halo mass as defined in our merger trees for our infall mass. Using peak mass along the merger tree gave similar results as infall mass. We compare the two options is in Sec. 2.5.3.

The effects of varying the M_{infall} threshold on clusters profiles is shown in Fig. 2.8. The total number of modeled galaxies in clusters depends strongly on the threshold. The profile shape is affect as well. For higher M_{infall} thresholds, the profile is more peaked at the center. The cores from more massive halos, including the central core of host halo itself, tend to reside at the potential center of the halo and increasing the threshold does remove as many galaxies from the center as it does from the outskirts.

Core Disruption

Strong gravitational tidal fields within a cluster environment can cause a galaxy to be stripped of its stellar component (e.g. [10]). (Ram-pressure stripping of halo gas is a relevant physical effect but is missing in gravity-only simulations.) Because stars are subject primarily to gravitational forces, the simulation particles that represent the core experience similar gravitational forces that the stellar component of galaxies would. An important caveat is that galaxies can be more compact than the cores we use to represent them, so mapping between galaxy dynamics and core dynamics is expected to be somewhat approximate.

We apply a simple model of galaxy disruption using the cores as proxies. If the core particles remain tightly clustered, we assume the galaxy the core represents has not lost enough stellar mass to fall below the luminosity threshold or completely disrupted. If the particles are widely spread apart, the modeled galaxy is assumed to be disrupted or below the luminosity threshold. The spread of the particles is measured using the core effective radius as defined in Eq. 2.1. The threshold value distinguishing compact galaxies and disrupted galaxies is a modeling parameter we label as $\mathrm{R}_{disrupt}$. The parameter is in comoving distance and has no redshift or infall mass dependence.

Investigating the distribution of core radii at fixed redshift, there is no clear distinguishing feature in the satellite core distribution that would naturally set a value for $\mathrm{R}_{disrupt}$ (see Fig. 2.2, top panel) This is to be expected since the final set of cores has a complex history with individual core members having gone through a large number of interactions as they traversed their branches in the associated merger tree. Cores that have recently infallen into have smaller radius than those have spent more time spent as satellite in their host halo (see Fig. 2.2, bottom panel).

The effect of varying $\mathrm{R}_{disrupt}$ on galaxy profiles is shown in Fig. 2.9. The $\mathrm{R}_{disrupt}$ parameter affects the total number of galaxies, but it also affects the shape of the profile more strongly than varying M_{infall}. The change in the profile is the greatest near the halo

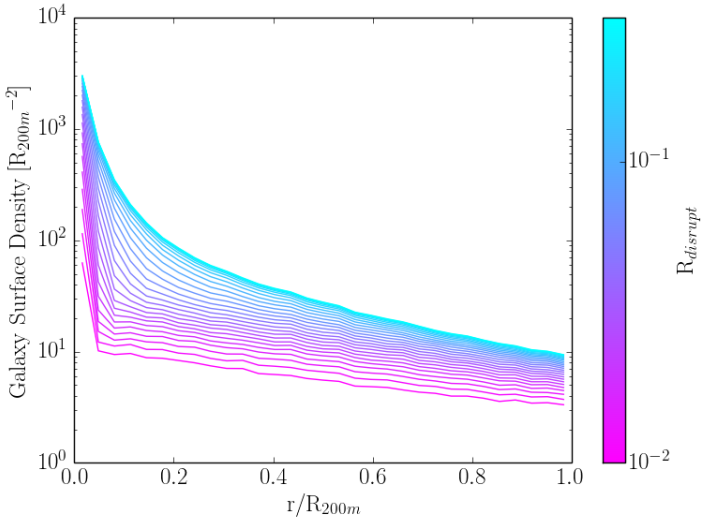

Figure 2.9: Galaxy surface density profiles of cores as a function the $R_{disrupt}$ threshold value for halos $14.0 < log_{10}(M_{200m}) < 14.25$. $R_{disrupt}$ modifies the shape of the profile.

center where the cores have spend more time interacting with the strong tidal field and other harassment. Core in the halo outskirt have are more likely to only have recently merged. Due to projection, the outskirts of the halos have more compact centrals from small interloping halos.

Core Mergers

We allow core mergers to stand in for galaxy mergers. If two compact candidate cores are within an R_{merge} comoving 3D distance, they are merged into one modeled galaxy. R_{merge} is a free parameter in the model. Multiple cores can merge to form one object: a FoF algorithm is used to determine the mergers. The position of the new object is taken as the mean position of all the FoF members. We do not allow cores with infall mass below M_{infall} to merge into more massive objects. Additionally, R_{merge} has no redshift or core mass dependence. We do include any temporal or velocity information for determining mergers.

The effect of varying the parameter on the galaxy profile in clusters is shown in Fig. 2.10. At very high R_{merge}, the FoF linking of cores can percolate to almost full core population

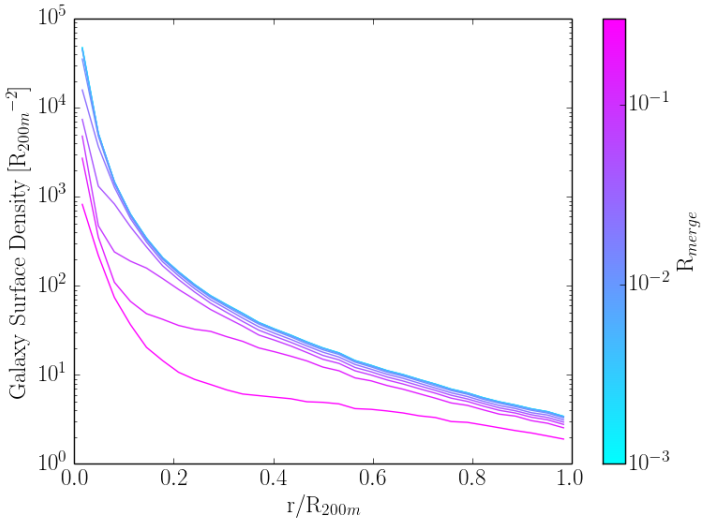

Figure 2.10: Galaxy surface density profiles of cores as a function the R_{merge} for halos $14.0 < log_{10}(M_{200m}) < 14.25$. Because the percolating nature of the FoF merging, the cores have been truncated to have an abundance similar to L_* galaxies: M_{infall} threshold of 10^12 h^{-1} $M_\odot$and $R_{disrupt}$ threshold of 20 h^{-1} kpc.

and remove a significant fraction of the cores. At smaller R_{merge}, most the merging occurs at the center of the halo where the density of cores is the highest.

2.4.2 Modeling Galaxy Profiles from Cores

Surface density radial profiles of cores are constructed for each simulation halo with mass above 10^{14} $M_\odot$ h^{-1}. To include projection effects, radial profiles are built using all cores within a 10 Mpc h^{-1} sphere around the halo cluster center. We have found that varying the inclusion radius does not change the best fit parameters when fitting to observational data, as long as the radius is chosen to be greater than 4 Mpc h^{-1}. Cores are modeled as galaxy proxies using the different model choices described above.

The radial surface density profile is calculated as follows. For each galaxy and radial bin, a weight proportional to the probability of the galaxy being projected into that radial bin is determined. The probability is proportional to the surface area of a sphere that would be

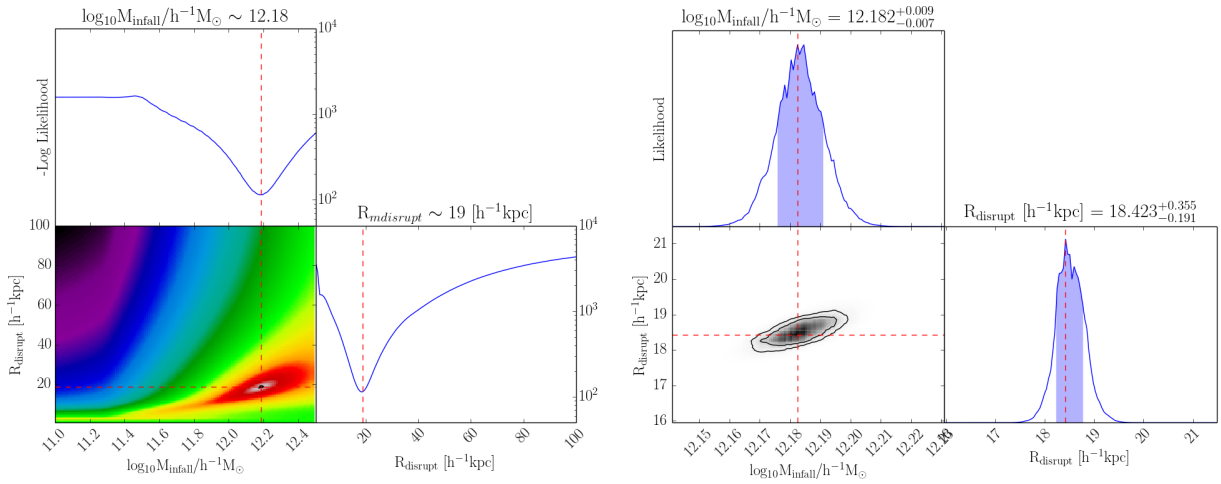

Figure 2.11: Grid likelihood evaluation of the Rd L_* galaxy model in the Outer Rim simulation. The left triangle plot shows the negative log likelihood calculated on a wide parameter grid scan. The right triangle plot shows the likelihood calculated on a zoomed-in region identified in the left panel.

projected onto the two dimensional radial bins:

$$Rad_i = \Sigma_j \begin{cases} \cos\theta_{ij} - \cos\theta_{(i+1)j} & r_i < R_j \\ \cos\theta_i & r_i \geq R_j \end{cases} \tag{2.5}$$

where Rad_i is the ith cluster radial density bin with a lower and upper bin edge of r_i and r_{i+1}, respectively, R_j is the real distance of the core from the cluster center and $\theta_{ij} = \arcsin r_i/R_j$.

2.4.3 Fitting Profiles to Observations

The observed clusters are binned into seven logarithmic mass bins from $\mathrm{M}_{200c/m} = 10^{14} h^{-1}\mathrm{M}_\odot$ to $\mathrm{M}_{200c/m} = 10^{15.75} h^{-1}\mathrm{M}_\odot$. The number of halos in each simulation and catalog are given in Table 2.2.1. The galaxy surface density is calculated in 15 projected linear radial bins from $r = 0$ to $r = \mathrm{R}_{200c/m}$. The radial galaxy density profile of the clusters are stacked within each mass bin. The error in each radial bin is taken as the Poisson fluctuation of galaxy counts in the bin.

The simulation profiles are constructed on the fixed redshift outputs of the simulation. The selected snapshot is close to the mean redshift of the observed galaxies, $z =\sim 0.24$. For each cluster above $M_{200c/m} = 10^{14}$ $M_\odot$ h^{-1}, all the cores within 10 Mpc h^{-1} are assigned to that cluster. As discussed previously, the extended volume is used to capture projection effects. The cores are then converted into model galaxies using the different models described in Section 2.4.

The individual radial profiles are stacked in mass bins. The error is calculated as the combination of Poisson error and error on the mean in each radial bin. The model likelihood takes into account the difference in radial profiles with the errors incorporated. For the SPIDERS clusters a few modifications were required. The lowest mass bin with two SPIDERS clusters were excluded from the likelihood evaluation. The central radial bin is also excluded. This is because the pointing accuracy of the X-ray data used in SPIDERS is large enough that it does not guarantee that a BCG-like galaxy will fall into the central radial bin. The method we use for modeling galaxies with cores has, by construction, a core at the halo center. To avoid skewing the fitting results, we exclude the central radial bin from the likelihood evaluation; the effect of excluding radial bins is further discussed in Section 2.5.3. Clusters found via redMaPPer are centered on a BCG-like galaxy by construction and do not require any special treatment.

The parameter likelihoods are calculated from a grid sampling of the model parameters. The models have a maximum of three parameters and a grid scan of parameter values is therefore sufficient. We have compared our grid scan results against MCMC runs and found that the posterior likelihoods agree very closely, but the computational time for MCMCs is significantly higher.

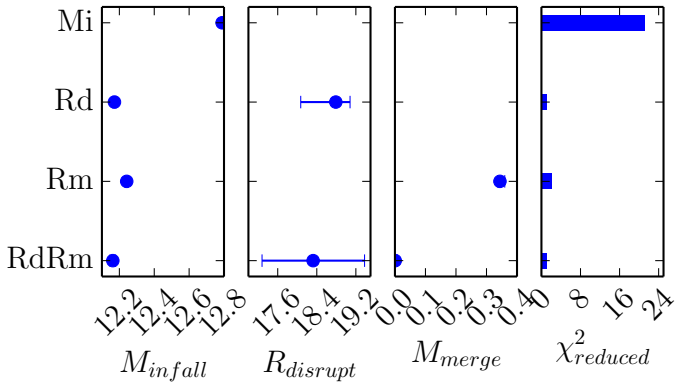

Figure 2.12: Core model parameter best fit and 1σ likelihood bounds for the Outer Rim halos for redMaPPer L_* galaxies. The $\tilde{\chi}^2$ for each model flavor is shown in the right most panel.

2.5 Results and Discussion

With the exception of the Mi model (single-parameter SHAM), all the other models were able to reconstruct the galaxy profiles of both redMaPPer and SPIDERS cluster samples. The $\tilde{\chi}^2$ for non-Mi models ranged from $\sim$0.5 to a few. For the Mi model, $\tilde{\chi}^2$ was as high as 16. The parameter likelihoods for modeling redMaPPer L_* galaxies in the Outer Rim simulation are given in Fig. 2.12, while the best fit parameters as a function galaxy luminosity model are given in Fig. 2.15. In all cases, the models that include the disruption mechanism (Rd, RdRm) preformed better than when it was excluded. There were no large qualitative differences in model performance as a function of galaxy luminosity.

Sec. 2.5.1 goes further into detail in comparing performance of model flavors. Luminosity dependence is covered in Sec. 2.5.2. Finally, we discuss various convergence and robustness checks in Sec. 2.5.3

2.5.1 Results for the Different Model Types

We fit four subtypes of the core-based galaxy model to the redMaPPer cluster data. For comparison, the redMaPPer and the four core model profiles for three cluster mass bins are shown in Fig 2.13. The simple SHAM-like model (Mi), fails to reproduce the galaxy surface density profile. As shown in Fig. 2.13 This clearly demonstrates that a pure abundance matching technique applied to the core distribution cannot reproduce cluster galaxy profiles.

The Rm model (core merging only) performs significantly better than the Mi model but is second worst, with $\tilde{\chi}^2$ ranging from 2 to 7.5. The Rm model does well in reproducing the outskirts of halos and the central bin density but fails to reproduce the profile near the center. At around 0.2 r/R_{200}, the Rm model profile dips down and there is such dip in the data. The dip is most visible in the top panel of Fig. 2.13. The Mi and Rm models preform similarly for $2.5L_*$ galaxy profiles, indicating that merging of $2.5L_*$ galaxies does not play a significant role in the profile shape.

The Rd and RdRm models perform similarly well – the Rd model is a subset of the RdRm model where the merging radius is reduce to zero. The change in $\tilde{\chi}^2$ in including the merging mechanism is small. This may point to that satellite core (or subhalo) mergers of infall mass we are investigating are relatively rare events. For the L_* and the $0.4L_*$ profiles, the best fit RdRm model has a small merging radius, not statically significant different from zero. A zero merging length would indicate no merging and the RdRm would be effectively the same as the Rd model.

2.5.2 Luminosity Dependence

We fit our models to several luminosity threshold, ranging from 0.4 L_* to 2.5 L_*. The model best fit parameters and $\tilde{\chi}^2$ as a function of luminosity are shown in Fig. 2.15. As expected, the M_{infall} parameter for all models increases with luminosity being modeled– brighter galaxies tend to be more rare and originate from more massive halos. Mi model

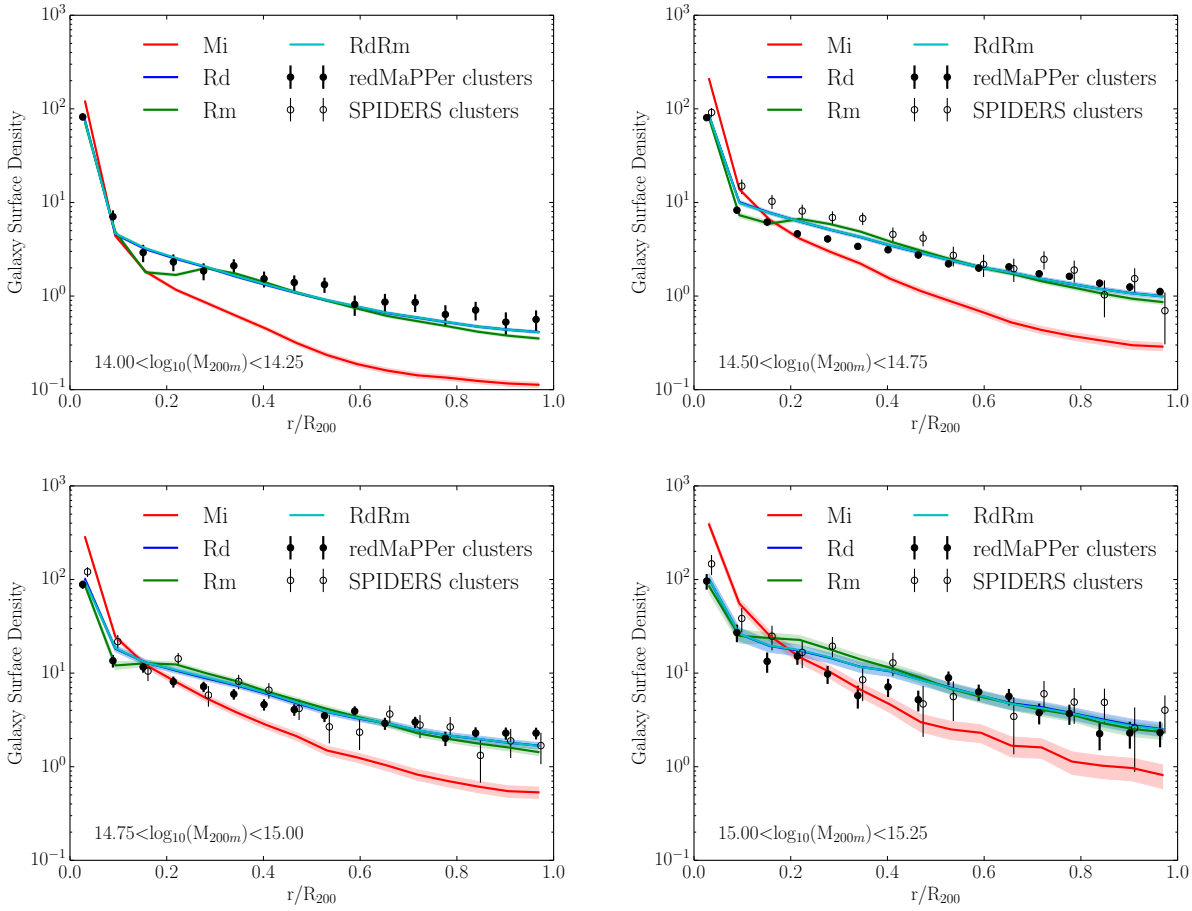

Figure 2.13: Radial L_* galaxy surface density profiles from observations and core galaxy models for several halo mass bins. The closed and open markers are SDSS galaxy surface density profiles centered on redMaPPer and SPIDERS clusters, respectively. The lines and shaded regions are the average and the error on the mean, respectively, of the profiles of modeled galaxies in the Outer Rim simulation using the best fit to the observed redMaPPer clusters. The markers for redMaPPer and SPIDERS have been offset for clarity. The SPIDERS cluster catalog does not have any clusters in the top panel mass range.

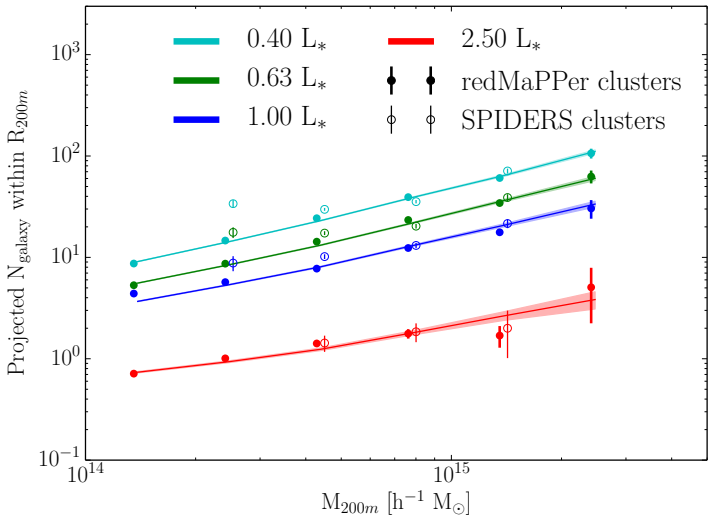

Figure 2.14: N_{gal} as a function of cluster mass for the redMaPPer and SPIDERS cluster catalogs and for the Rd galaxy model of cores. The cluster N_{gal} is calculated by integrating the background subtracted galaxy counts within R_{200m} of the cluster center. The Rd model uses the best fit model parameters to the redMaPPer cluster profiles. The markers position for redMaPPer and SPIDERS have been offset in the x-axis for clarity.

performs significantly worse at all luminosity ranges as compared to other models. The Rm model does not preform as well as the Rd and the RdRm models. As the luminosity threshold is increased, the Mi model performs better. At the highest threshold, the Mi and Rm models have the same M_{infall} and $\tilde{\chi}^2 \sim 4$. This may indicate that loss mechanisms for these bright galaxies do not play as an important role.

The $R_{disrupt}$ parameter in the Rd model also increases with galaxy luminosity. A higher $R_{disrupt}$ parameter can be interpreted that it is more difficult to disrupt these galaxies. The RdRm model does not have significant change in the $R_{disrupt}$ parameter across the full luminosity range and has a much wider uncertainty bounds as compared to the Rd model. Most the change in loss mechanisms in the RdRm model is restricted to merging. It seems that the merging and disruption mechanism have a degenerate effect on the galaxy profiles.

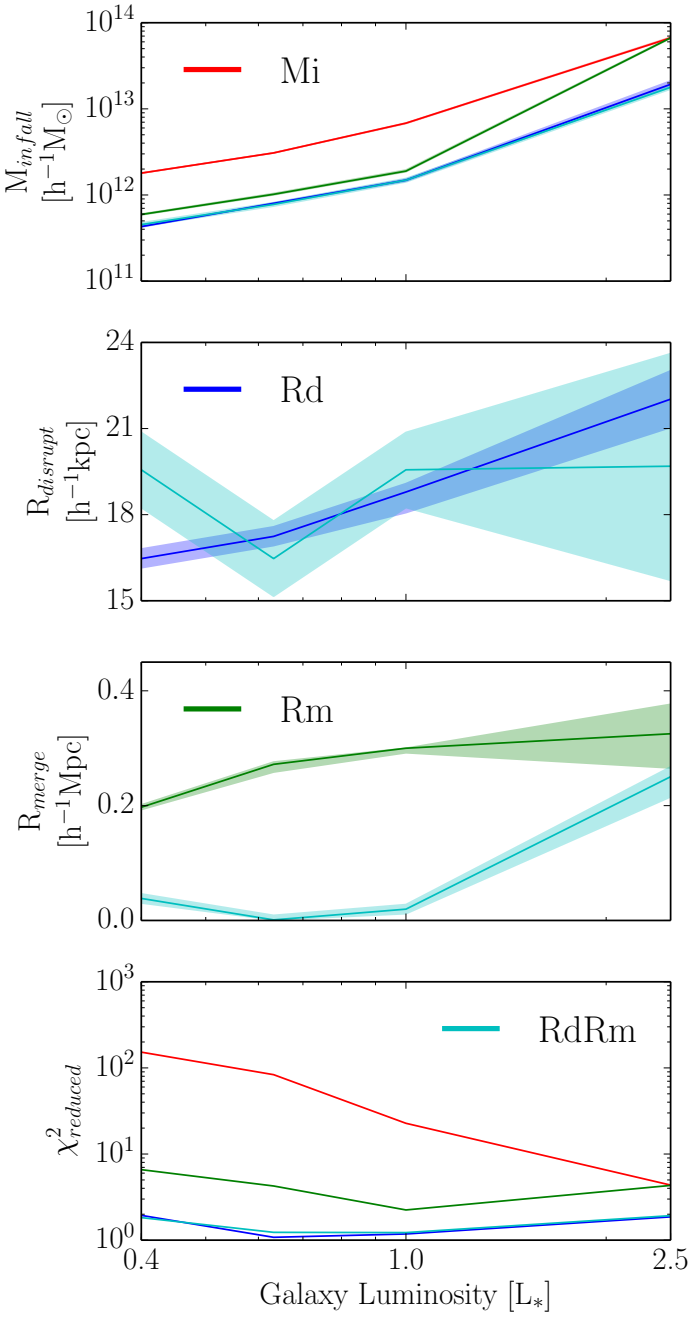

Figure 2.15: Best fit galaxy model parameters and $\tilde{\chi}^2$ for the Outer Rim simulation and redMaPPer cluster catalog as a function of galaxy luminosity. Each model flavor is a separate color. The the lines are best fit model parameters and the shaded regions are the 1σ statistical bounds.

2.5.3 Robustness and Convergence Tests

We performed several checks to investigate the robustness of the results. These include choices made in our analysis such as volume to consider or the definition of core mass.

-

Cluster Volume

The SDSS galaxy profiles from redMaPPer and SPIDERS cluster catalogs are calculated using projected galaxies. While we do a background subtraction (Sec. 2.3.5), the environment around clusters is more dense with both matter and galaxies . To include these project affects into core model as well, we include all cores within a fixed distance of the cluster center. By increasing the volume used for projection, a larger fraction of galaxies within the projected R_{200} will be interloper galaxies and not within the physical R_{200} of the galaxy cluster. If we increase the projected radius too large we would be including modeled galaxies that would have been subtracted out by the background subtraction in the simulation.

We measure the effect of varying the cluster projection volume on the best fit parameters for the Rd model for L_* galaxies. The best fit parameters as function of projection volume is shown in Fig. 2.16. Increasing the projection radius, will always increase the number of interloper galaxies, so the parameters will shift to decrease the fraction of cores that pass both the mass and radius cut. The best fit $R_{disrupt}$ model parameter does not change significantly past 4 h^{-1} Mpc. Most of the compensation for the increased volume occurs in the M_{infall} parameter. From 4 h^{-1} Mpc to 10 h^{-1} Mpc, M_{infall}increases by 0.06 dex, which is significant greater than the typical statistical uncertainty of $\sim$0.01 dex. Increasing M_{infall}, as shown in Sec. 2.4.1, strongly affects the normalization of the of the profile, but does not change the shape as strongly. The choice of cluster volume introduces a systematic uncertainty greater than the statistical uncertainty on the precise value M_{infall} in the core galaxy model.

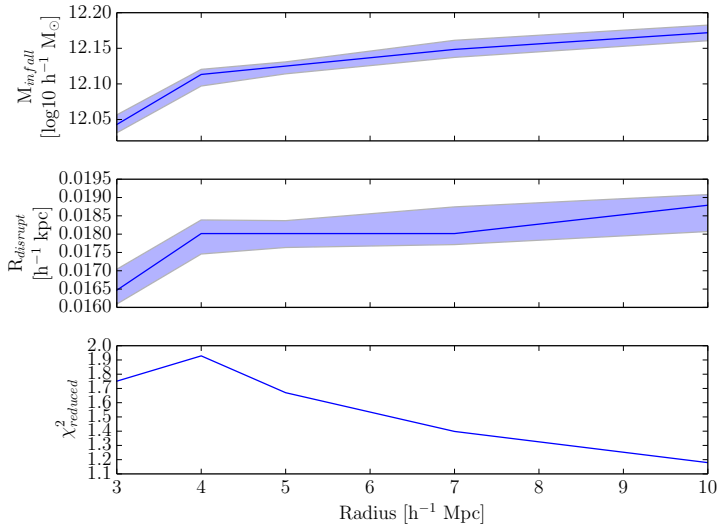

Figure 2.16: Best fit model parameters and $X^2_{reduced}$ for the Rd L_* galaxy model as a function of projection radius. All cores within the projection radius are used to compute the projected modeled galaxy surface density.

RedMapper vs SPIDERS

We test our model against an X-ray selected cluster sample to ensure that the model is able to generalize to a differently selected cluster catalogs. The difference in SDSS galaxy profiles is shown in Fig. 2.7. In particular redMaPPer will preferentially select clusters that align with line of sight, while X-ray selected clusters are less biased in this way.

Excluding the Central Cluster Region

To reduce the affect of mis-centering of observed cluster profiles, we exclude the central radial sections of the profiles. The value of the best fit parameters does not change until we exclude up to $0.3R_{200m}$. After which, the best fit values plateau off.

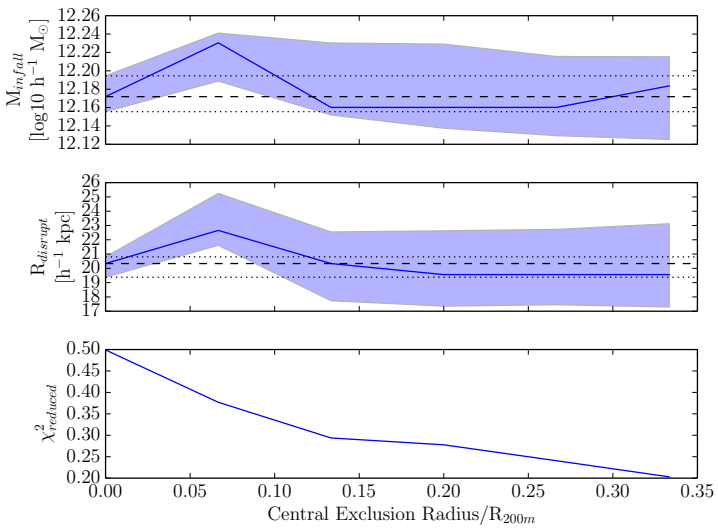

Figure 2.17: Best fit parameters for the Rd L_* galaxy model as a function of the central cluster radius excised from the likelihood calculation.

Comparison of Peak Mass and Infall Mass

We investigated using the peak mass of the halo along its past history just before it merges, instead of the halo mass at infall and found the difference to be minimal in the halos of interest. When we applied the peak mass instead of infall mass in the AlphaQ simulation, and found that the average mass associated with cores above an infall mass of $10^{12}h^{-1}$ $M_\odot$ increased by 1.5%, with a standard deviation of 5%. Such a small change is not unexpected. The difference between infall mass and peak mass become significant when a (sub)halo has multiple merger events. With each merger, the (sub)halo can be stripped a significant fraction of it mass while inside another halo. The final infall event may have a significantly lighter (sub)halo. Our merger trees by construction only have at most a single infall event for each main halo. For the peak and infall mass to differ, the main halo itself must lost mass by some mechanism such as glancing fly-bys.

Fitting the cores to L_* galaxies with the Rd model, we found no significant difference in the profile shapes. The best fit parameters were very similar, given in Table. 2.4. The best fit $R_{disrupt}$ did not have a significant change, the change was well within the statistical error

Mass Type	$\log10(M_{infall}/h^{-1}\ M_\odot)$	$R_{disrupt}$	$\tilde{\chi}^2$
Infall	$12.168^{+0.006}_{-0.013}$	$18.62^{+0.20}_{-0.57}\ h^{-1}$ kpc	a
Peak	$12.218^{+0.018}_{-0.011}$	$18.79^{+0.79}_{-0.38}\ h^{-1}$ kpc	b

Table 2.4: The best fit parameters for the Rd model of L_* galaxies using infall and peak mass definitions of cores. The error bounds only include statistical error.

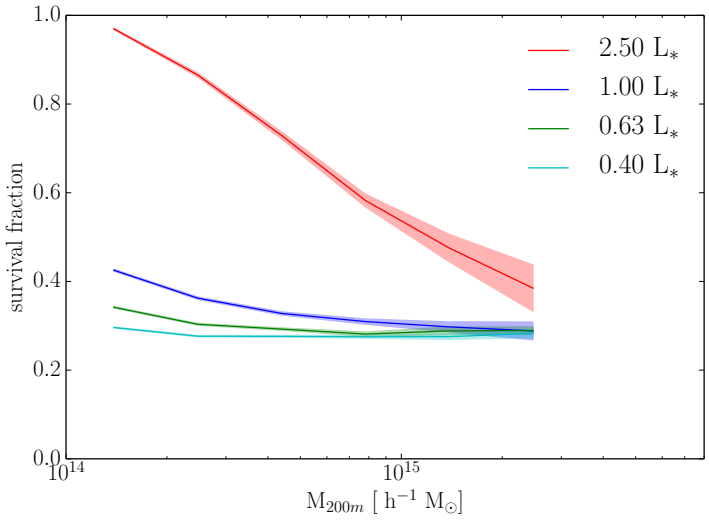

Figure 2.18: Modeled galaxy survival rate for Rd model in the Outer Rim simulation.

bounds. The best fit M_{infall} changed increased by about $\sim 12\%$ which is a statistical 4.0σ change. The systematic uncertainty is much greater for M_{infall} from the choice of project volume (Sec. 2.5.3).

2.5.4 Predicted Disruption Rates

[10] found that about 50% of galaxies survive within halos at z=0.21. In our model, we have much lower survival rates about 20-30% depending on the model flavor and galaxy luminosity being modeled. Modeled brighter galaxies have higher survival rates than fainter galaxies. The discrepancy can point to requiring a more complicated model if we want to be more consistent with the dynamics that other people are finding.

2.6 Conclusion

In this paper, we present a novel method to identify and track substructure within halos. The advantage of tracking halo cores is that it is computationally more efficient than identifying and tracking subhalos. Nor does it suffer from the complexity and challenges associated with robustly identifying, characterizing subhalos and constructing subhalo merger trees. We use halo core tracking to generate several models of galaxies above luminosity threshold within galaxy cluster sized halos. The models parameters were fit to match galaxy surface density profiles of clusters observed in SDSS and find excellent agreement for most model variants.

- As expected, the best fit minimum infall mass to reach the luminosity threshold of the galaxy scales with luminosity. The Mi model lacks any mechanism to remove galaxies, so the only method left is to decrease the initial population by increasing the mass threshold.

- The best preforming model flavors allowed the disruption of model galaxies, with $\tilde{\chi}^2 \leq\sim 1$ for the luminosity range modeled.

- The Mi model which does not allow the removal of galaxies either through preformed disruption or merging preformed the worst out of all the model flavors. While it strongly disagreed with data for fainter galaxies, for bright galaxies ($2.5L_*$), the model was able to match data which would indicate that disruption and merger events are more rare or less important for brighter galaxies.

- The best fit disruption criteria depends on the luminosity of modeled galaxy. This may point that a better model would include a varying disruption criteria with modeled luminosity: more massive and brighter galaxies should be difficult to disrupt.

- Allowing merging between galaxies with disruption turned on barely improved $\tilde{\chi}^2$.

CHAPTER 3

SYNTHETIC GALAXY CATALOG

3.1 Introduction

The next generation of large imaging and spectroscopic survey projects including LSST [95], the Wide Field Infrared Survey Telescope (WFIRST) [7], the Dark Energy Spectroscopic Instrument (DESI) [42], Euclid [89] and the Subaru Prime Focus Spectrograph (PFS) [142] will provide a wealth of data for modern cosmological analyses that seek to probe the nature of dark energy. The advent of these large data sets heralds a new era in cosmology that is characterized by small statistical uncertainties, improved control of systematic errors, and unique opportunities to combine multiple probes of dark energy, all leading to much improved constraints on this little understood component of the Universe. The LSST Dark Energy Science Collaboration (LSST DESC) [94] has convened to prepare for the large and complex data set that will arrive with the commencement of LSST operations in 2022. Since the data will become publicly available almost immediately, scientists must have robust, well-understood analysis pipelines in place. This goal would not be achievable without extensive simulation campaigns aimed at providing simulated skies in the form of synthetic galaxy catalogs.

These galaxy catalogs play a number of essential roles in cosmological surveys. They serve as testbeds to study questions of survey design, to enable studies of possible systematics (e.g., fiber collisions, telescope point-spread functions, shape measurement assumptions), to facilitate tests of data reduction pipelines and to provide reference data that can be used to validate analysis pipelines for cosmological inference. The demands of contemporary surveys for high-quality synthetic skies are only becoming more stringent as the field progresses further into an era in which cosmological inference is systematics-limited.

Given sufficient understanding of the underlying physics, such catalogs can be thought

of as the result of forward models that provide approximate theoretical predictions for the galaxies inhabiting a physical universe. As the models improve over time, the predictions become ever more faithful to the properties of the observed Universe. Consequently, synthetic catalogs, and summary statistics derived from them, enable cosmological inference and function as tests of robustness for these methods.

A wide variety of techniques is currently being used to construct large-volume synthetic galaxy catalogs for contemporary and near-future galaxy surveys [see 154, for an overview]. At the extreme end of computational expense, hydrodynamical simulations are the closest approximation in the field to an *ab initio* model of galaxy formation. These simulations track the evolution of the dark matter and the baryons under the influence of gravity and gas physics and include sub-grid models for a number of astrophysical processes such as gas cooling and heating, star formation, and supernova and AGN feedback. Contemporary approaches include simulated cosmological volumes of the scale of $\mathcal{O}(100)$ Mpc [79, 128, 137] to $\mathcal{O}(1)$ Gpc [43, 48] as well as zoom-in simulations at much higher resolution [2, 26, 80, 151, 69]. While hydrodynamical simulations are indispensable to the study of systematic effects associated with baryonic physics, these simulations are not typically used to produce large-scale galaxy catalogs for sky surveys due to their computational expense, which arises from the necessity of solving for gas dynamics at high spatial and mass resolution, and resolving small spatio-temporal scales when including sub-grid models. Additionally, parameterized sub-grid models are typically integrated within the main simulation and cannot be conveniently "bolted on" after the simulation is completed. This significantly increases the number of simulations needed to explore the associated parameter space and to calibrate the models against observations.

A less computationally expensive approach involves the use of semi-analytic models of galaxy formation (SAMs). In these models, the synthetic catalog is generated using gravity-only simulations coupled with additional modeling for the baryonic physics not contained in

the underlying simulation. This additional modeling is based on the mass-assembly history of each halo [77, 37, 17, 87]. Hence, all SAMs are predicated upon connecting the properties of individual galaxies with individual dark matter halos. The SAM approach is to parameterize baryon-specific processes as functions of the halos and their evolution; on a halo-by-halo basis SAMs seek to model directly how baryons would have evolved had they been included in the N-body simulation [see 135, for a recent review]. Using SAMs to generate synthetic catalogs for large surveys is still computationally demanding, particularly for imaging surveys in need of large quantities of synthetic data for faint galaxies. Furthermore, while considerable recent progress has been made in our ability to explore the parameter space of contemporary SAMs [65, 147, 124], calibrating a SAM to high accuracy is nonetheless computationally expensive [although see 108, 21, for a closely related effort to the present work].

When generating synthetic data for contemporary galaxy surveys, such difficulties are compounded by the evolving nature of the validation criteria used to evaluate the fitness of the catalog. As additional scientists join and become active in a collaboration, their expertise informs changes or refinements to the criteria; adjustments may also be made as surveys release new data. The development of new analysis techniques may also require synthetic data to have properties in addition to those originally planned. This evolving nature is fundamental to the operating mode of collaborations conducting large galaxy surveys. As a consequence, when generating simulated catalogs to complement a survey, the underlying model should be straightforward to calibrate, since in practice the parameters may need to be refit many times over the course of the survey.

To address the aforementioned challenges associated with using SAMs, many collaborations have instead used empirical models to generate their synthetic data. In this approach, one assumes the existence of simple scaling relations between dark matter halos and galaxies, and fits the parameters of these scaling relations to observational data. Empirical models are computationally more efficient than SAMs; therefore, the parameters of these models are less

expensive to fit. When generating synthetic catalogs with empirical methods the challenge lies in attaining sufficient realism to reflect the level of complexity needed to satisfy survey requirements. For example, in the MICE catalogs [53] created for the Dark Energy Survey (DES), the authors begin with a standard form of the Halo Occupation Distribution [HOD, 18, 86, 157] to inject synthetic galaxies into simulated host halos. After tuning their baseline HOD model, the properties of galaxies they augmented with additional attributes, such as broadband color, using data-driven methods. A similar methodology is used to generate the Euclid Flagship simulation galaxy catalog[1] based on a large gravity-only simulation described in [113]. Similarly, the Buzzard catalogs developed for DES [41] use a baseline abundance-matching approach [86, 36, 14, 102, 118] on a high-resolution simulation to generate a tuning catalog. The ADDGALS technique (R. Wechsler et al. 2019, in preparation) is used to populate lower-resolution simulations with galaxies based on this tuning catalog. The galaxies are then assigned template spectral energy distributions (SEDs) using empirical techniques.

This paper presents the cosmoDC2 synthetic galaxy catalog, an ambitious synthetic catalog with physically motivated, highly nonlinear correlations between a large number of galaxy properties spanning a multidimensional parameter space. Our approach combines empirical methods with semi-analytic modeling using the Galacticus code [17] to produce a catalog which is easily calibrated while maintaining a high degree of physical realism. The input to cosmoDC2 is a large-volume, N-body (gravity-only) cosmological simulation, the Outer Rim run, carried out using the Hybrid/Hardware Accelerated Cosmology Code (HACC) [59]. For the construction of the catalog, data products from the smaller AlphaQ simulation, and the UniverseMachine [13] modelling approach are used in the semi-analytic and empirical modeling components, respectively.

CosmoDC2 was designed for the second LSST DESC Data Challenge (DC2), an ambitious

1. http://sci.esa.int/euclid/59348-euclid-flagship-mock-galaxy-catalogue

effort to generate a data set that is very similar to a multi-year LSST data release (R. Armstrong et al in preparation). DC2 not only aims to improve analysis pipelines in advance of the arrival of LSST imaging, but also serves the important function of uniting hundreds of LSST DESC members with a specific and concrete data set to analyze with common software tools. The design of cosmoDC2 was driven by a number of science goals that are applicable to any synthetic catalog intended to provide simulated data for a wide and deep sky survey such as LSST. These goals are quantified by specific criteria applied to a large number of tests that have been developed by the LSST DESC analysis working groups and incorporated into DESCQA [98], the LSST DESC automated validation framework for synthetic sky catalogs. These tests compare the catalog contents with judiciously chosen observational data, enabling an assessment of the synthetic catalog's fidelity to the statistics of the real sky. The validated catalog can then be used to provide realistic inputs for scientific analyses as described above and to serve as the input for extragalactic component of DC2 image simulations .

Broadly speaking, the need to test analysis pipelines for probes involving static measurements such as weak lensing, large-scale structure, galaxy clusters, and photometric redshifts leads to a requirement that the full catalog cover an area of thousands of square degrees, which in turn demands that a very large input simulation underlies the catalog. However the area requirement for the image simulations is relatively modest in comparison. The catalog is, therefore, being released in 2 stages: the first stage provides enough area (440 $\deg^2$) to run the DC2 image simulations; the second stage, to be completed in the near future, will provide the larger area (5000 $\deg^2$) needed to enable analysis tests requiring high statistics. The complexity of the validation tests also leads to stringent requirements on the properties of the galaxy distribution and their connection to the underlying dark matter distribution. In particular, for weak lensing quantities to be correctly estimated at each galaxy position a full ray-tracing code based on the Outer Rim particle lightcones is needed. As described

below, all of these considerations informed the methodology used to produce the cosmoDC2 catalog.

The rest of the paper is organized as follows. In Sec. 3.2 we provide a general overview of the end-to-end workflow for generating the synthetic sky catalog. We briefly describe the gravity-only simulations and related data products underlying the catalog (halos, merger trees and lightcones) in Sec. 3.3. Sec. 3.4 describes the generation of the weak lensing catalog. Sec. 3.5 discusses the galaxy catalog, where we give a brief summary of the catalog generation and the catalog content. The release of this paper is coincident with the public release of the 440 sq. deg. cosmoDC2 catalog. Selected results from the catalog-validation tests are shown in Sec. 3.6. A concluding discussion and future outlook is provided in Sec. 3.7.

3.2 CosmoDC2 Production Overview

The key aspects of our production pipeline are shown in Figure 3.1, which provides a conceptual overview of how we use the simulation data products available from both the large Outer Rim simulation and the smaller UniverseMachine and AlphaQ simulations to produce the cosmoDC2 extragalactic catalog. As noted in the figure, the data products and pipelines are described in detail in subsequent sections of the paper. The first stage of the workflow shows how the data products from each simulation are used as inputs to separate pipelines to produce corresponding intermediate data products. Next, these intermediate data products are combined in a final pipeline to produce the cosmoDC2 extragalactic catalog.

The Lensing Pipeline (Sec. 3.4) inputs particle data from the Outer Rim simulation to a lightcone generator, projects these particles onto discrete redshift shells and computes the surface density using a density estimator, and traces photons through these shells using a ray-tracing code, to generate curved-sky lensing maps. These lensing maps include the deflections experienced by photons as well as the image distortions in the weak lensing regime.

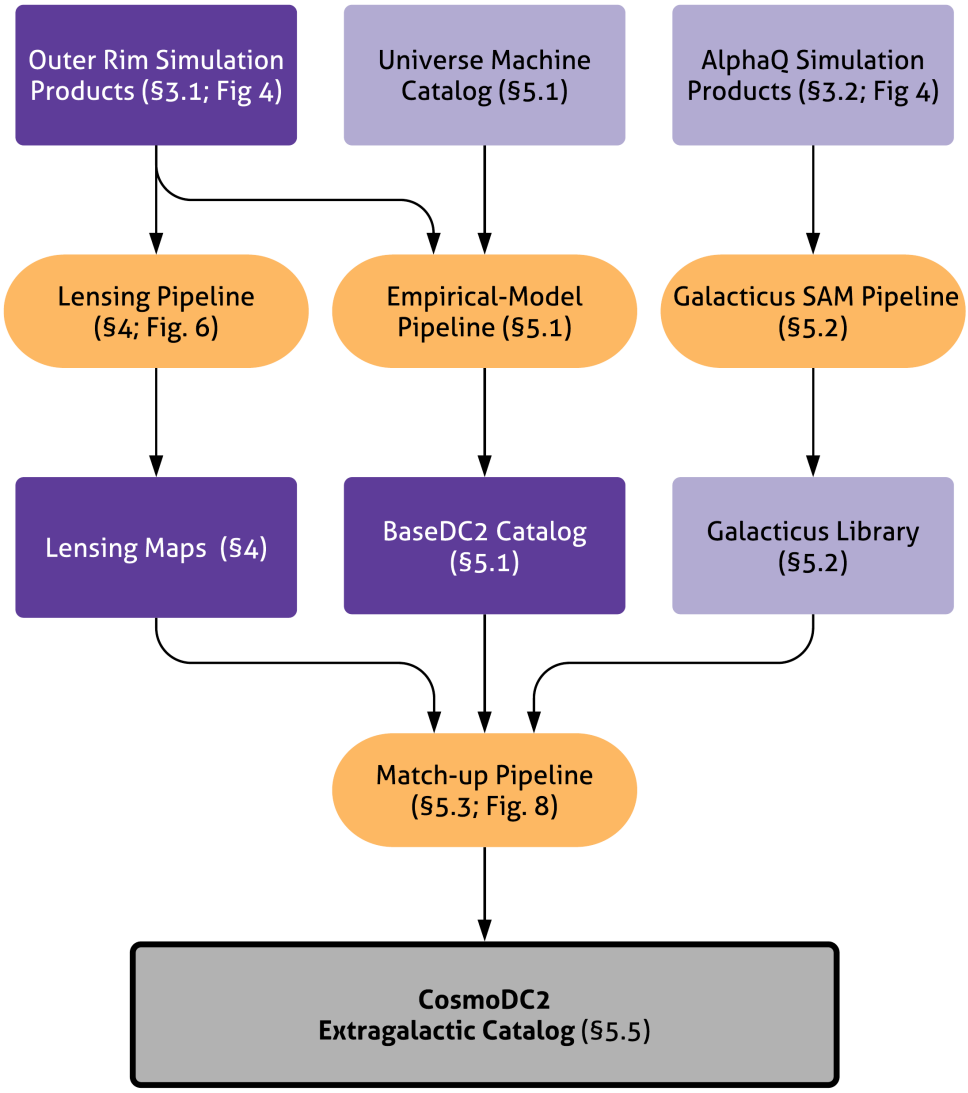

Figure 3.1: Conceptual overview of the workflow to produce cosmoDC2. Data products are shown as rectangles in dark and light purple for data derived from the Outer Rim and the smaller auxiliary simulations, respectively. Pipelines are shown as ovals in light orange. Numbers in parentheses refer to the sections and figures in this work where a detailed descriptions are given.

The Empirical Model Pipeline (Sec. 3.5.1) uses halo and merger-tree data from the Outer Rim simulation combined with Monte Carlo resampling of galaxies from the UniverseMachine simulation (version ERD) to produce the baseDC2 catalog. BaseDC2 contains all of the cosmoDC2 galaxies but with a limited set of properties. The UniverseMachine simulation consists of galaxies that have been assigned to halos in the MultiDark Planck 2 (MDPL2) N-body simulation [82] using the UniverseMachine prescription described in [13]. The numbers, the positions and properties (stellar mass, $M_\star$, and star-formation rate, SFR) for these galaxies are determined by tuning the UniverseMachine-model parameters to observational data. The tuning procedure takes into account errors in the observational data by marginalizing over nuisance parameters that represent the additional sources of systematic errors required to achieve global fits to multiple observational data sets. We use further empirical modeling to augment the properties of these galaxies to include a limited selection of LSST rest-frame colors and magnitudes. These properties are assigned probabilistically.

The SAM pipeline (Sec. 3.5.2) uses merger trees built from the small AlphaQ simulation as inputs to the Galacticus SAM [17] to produce the Galacticus library of galaxies. These galaxies have an extensive set of self-consistent properties that have been obtained by solving a set of ordinary differential equations (ODEs) following the mass assembly history of each galaxy's host halo. Galacticus nonetheless has simulation-dependent model parameters that must be tuned in order to produce results that match observational data. Since this tuning was beyond the scope of our work, the statistical ensemble of properties for the galaxies modeled by Galacticus does not fully meet our strict validation requirements. In our hybrid approach, however, we can use this set of galaxies as a library from which to select appropriate galaxies with a complex set of properties. The additional properties obtained from the Galacticus library are required by the validation criteria that have been supplied by the DESC science working groups.

The selection of library galaxies is performed in the Match-up Pipeline which uses the

three intermediate data products as inputs. The lensing maps are used to interpolate the values of the weak lensing quantities (shear, convergence and deflection angles) to the position of each baseDC2 galaxy. Then the pipeline finds a suitable matching galaxy from the Galacticus Library for each galaxy in the baseDC2 catalog and augments the properties of the baseDC2 galaxies with those of the library galaxy, thus adding realistic complexity to the galaxy model. Additional empirical modeling is done after the match-up.

The workflow shows how our hybrid method combining empirical modeling and SAMs leverages the data from both small and large simulations to produce a realistically complex synthetic catalog. It is important to note that in order to tune the properties of the galaxies to better match any validation data, it is only necessary to change the empirical model and rerun the Match-up Pipeline, both of which are relatively inexpensive. Since we are using the results from the Galacticus SAM as a library, that computationally expensive step in the workflow need only be carried out once. It is much less expensive to iterate over parameters in the empirical model than to tune the parameters for the SAM. Our workflow makes it feasible to iterate the galaxy properties multiple times to achieve good agreement with the validation data. Critically, the distributions of galaxy properties of the SAM galaxies *span* the range of values required by the mock to meet observational validation criteria. If this is not the case, the matching procedure will fail in some regions and some additional strategies will be required to generate a good match. (See Sec. 3.5.3.)

For the program outlined above to be successful, it is important that the empirical model has the flexibility to produce an accurate realization of the limited galaxy properties, and that these properties are sufficiently correlated with other galaxy properties of interest. In our case, the UniverseMachine model generates stellar mass ($M_\star$) and star-formation rate (SFR) distributions that match observational data. Additional empirical modeling generates selected colors and luminosities that are based on these $M_\star$ and SFR values and exploits the inherent correlations between stellar mass, star-formation rates and the brightness and

color of galaxies. Matching to a SAM-library galaxy to obtain a complex set of properties assumes that the stellar mass and selected brightness and colors are sufficiently correlated with the full set of galaxy properties to obtain realistic distributions of the latter. If all of these assumptions are fulfilled, then the combination of empirical modeling with SAM resampling should generate a catalog of galaxies with a complex set of properties whose summary statistics satisfy the validation criteria.

3.3 The Underlying Simulations

Two gravity-only simulations underlie the construction of the cosmoDC2 catalog. The Outer Rim simulation serves as the basis of the final synthetic galaxy catalog including weak lensing quantities. A smaller simulation, called AlphaQ, was used for generating the Galacticus-based galaxy library. We now provide a description of both simulations as well as the related data products, such as halo catalogs, merger trees, and lightcones.

3.3.1 The Outer Rim Simulation

The Outer Rim simulation (for details and results on basic statistics such as power spectra and mass functions at different redshifts, see [64]) is one of the largest simulations at its mass resolution available world-wide and was carried out with the Hardware/Hybrid Accelerated Cosmology Code (HACC) [59]. HACC has been developed to run on all currently available and planned supercomputer architectures, including many-core machines and GPU-accelerated systems. The Outer Rim run was carried out on 32 racks of the IBM BG/Q system Mira (two-thirds of the entire machine) at the Argonne Leadership Computing Facility (ALCF). On many-core systems, such as the BG/Q, HACC uses a high-order spectral particle mesh method, combined with a tree calculation for the short-range forces [59]. For the analysis of the simulation we employed Mira itself as well as Cooley, a powerful GPU-enhanced analysis cluster at the ALCF.

The Outer Rim simulation covers a volume of $(4.225\ \mathrm{Gpc})^3$ sampled with $10{,}240^3$ tracer particles, leading to a mass resolution of $m_p = 2.6 \cdot 10^9 \mathrm{M}_\odot$. We used a cosmological model close to the best-fit WMAP-7 parameter set [84] given by $\omega_{\mathrm{cdm}} = 0.1109$, $\omega_{\mathrm{b}} = 0.02258$, $n_s = 0.963$, $h = 0.71$, $\sigma_8 = 0.8$, and $w = -1.0$.

We saved 101 time snapshots from $z \sim 10$ down to $z = 0$, evenly spaced in $\log_{10}(a)$, where a is the scale factor. (Two snapshots were corrupted on disk before we were able to fully analyze them, leading to a final number of 99 snapshots[2]). The spacing between snapshots varies from ~ 300 Myr for snapshots with $z \sim 0$, ~ 200 Myr for snapshots with $z \sim 1$, and ~ 80 Myr for snapshots with $z \sim 3$. For each snapshot we saved the full particle output and 1% randomly downsampled particles. The full particle outputs are stored on tape, while the downsampled outputs are kept on disk for easy accessibility. An extensive analysis for each snapshot was carried out as detailed in [64]. The analysis steps relevant for the present work are summarized in Sec. 3.3.3.

3.3.2 The AlphaQ Simulation

The AlphaQ simulation covers a volume of $(360.56\ \mathrm{Mpc})^3$ and evolves 1024^3 particles, with the same cosmology as the Outer Rim simulation. This yields a mass resolution of $m_p = 1.6 \cdot 10^9 \mathrm{M}_\odot$, close to that of the Outer Rim simulation, but with a volume that is approximately 1600 times smaller. We carry out exactly the same analyses as for the Outer Rim run, and save data products at the same time steps. Initially, the AlphaQ simulation was used for prototyping the workflow to generate the synthetic galaxy catalog in order to avoid having to handle large amounts of data from the very start. This resulted in the 'protoDC2' catalog, which was used for DC2 pipeline development. The AlphaQ simulation is also the source of the Galacticus-generated galaxy library that is used to assign complex galaxy properties to the final cosmoDC2 catalog.

2. The corrupted snapshots correspond to redshifts of z=0.914 and z=0.434.

3.3.3 Halo Catalogs and Merger Trees

The halo catalogs are generated using a parallel Friends-of-Friends (FOF)-based halo finder with a linking length of $b = 0.168$ and a minimum requirement of 20 particles per halo. Halo merger trees (based on the FOF catalogs) are constructed using a newly developed particle-membership algorithm [116] aiming to address the inconsistencies that arise from temporary mergers. Working in reverse sequential order, pairs of temporally adjacent snapshots are processed to identify halo progenitor/descendant relationships, simultaneously replacing halos that have split, i.e., halos with multiple descendants, with their individual "fragment" components. In this way, the fragment halos propagate the splitting event through the analysis and ensure at most a single merging event for every halo.

3.3.4 Lightcone Generation

We build particle and halo lightcones by tiling the Outer Rim simulation box in space, and applying a parallel solver which interpolates objects between adjacent snapshot positions to find their location of null spacetime separation from an observer at the origin. The cosmoDC2 observed density field is created by building a particle lightcone from Outer Rim snapshots randomly downsampled to 1%, through which lensing observables are simulated (see Sec. 3.4). We also construct an accompanying halo lightcone built from the simulation's FOF merger tree, upon which galaxies are later placed (see Sec. 3.5.1). Each of these lightcones fills one octant (~5,000 deg^2) of the sky, and has a depth of $z = 3$. A high-level description of the particle and halo lightcone generation is given below in Sec. 3.3.4 and 3.3.4, respectively.

Particle Lightcone

A lightcone, as built from N-body simulations, can be thought of as a set of concentric shells centered on the observer, where the boundaries of those shells are determined by the discrete redshifts of each simulation output snapshot. In filling the lightcone, we require a general

prescription to solve for the contents of each shell. Numerous such methods are described in the literature; previous efforts have chosen to fill the lightcone volume with simulation objects which retain their snapshot positions and velocities [21, 81], while others implement linear and higher order schemes to interpolate particle positions between snapshots [50, 99, 134].

The optimal choice between these approaches depends on the characteristics of the underlying simulation; if the time resolution of the outputs is coarse, then interpolation may cause excessive smoothing of the density field on small scales. On the other hand, simply concatenating shells extracted from snapshots introduces discontinuities in the correlation function across redshift. Given Outer Rim's relatively high spatial and temporal resolution (57 snapshot outputs to z=3), we find that a linear interpolation method is sufficient.

The angular overdensity power spectrum of a section of the particle lightcone at $z \approx 1.7$ resulting from our procedure being applied to downsampled Outer Rim snapshots is shown in Figure 3.2. This is computed using the Polspice code [30], and corrected for shot noise. We compare this to a theory prediction based on the CosmicEmu power spectrum [90] corrected for finite shell-width effects [see, e.g., 143, for a description of these effects]. At large scales there is good agreement with the model to within the cosmic variance and expected levels of model inaccuracy. On small scales where we expect the particle interpolation to increase the power (see, e.g., [99]), we see biases of $\tilde{}1 - 3\%$.

The extremely large volume that will be probed by LSST poses a challenge when solving for an observer's past lightcone. In particular, the DC2 catalog effort needs to model galaxies out to redshift $z \lesssim 3$, where the comoving distance is in excess of 6 Gpc, while Outer Rim extends only to 4.225 Gpc in each dimension. Our strategy, then, is to build a lattice of replicated simulation volumes which is large enough to host the cosmoDC2 lightcone. To prevent duplicate structures from being projected atop of one another, we follow the approach of previous sky simulations [21, 81, 20], and choose to randomly rotate each box replication.

While this strategy does decorrelate particle pairs at each box edge, we only need to

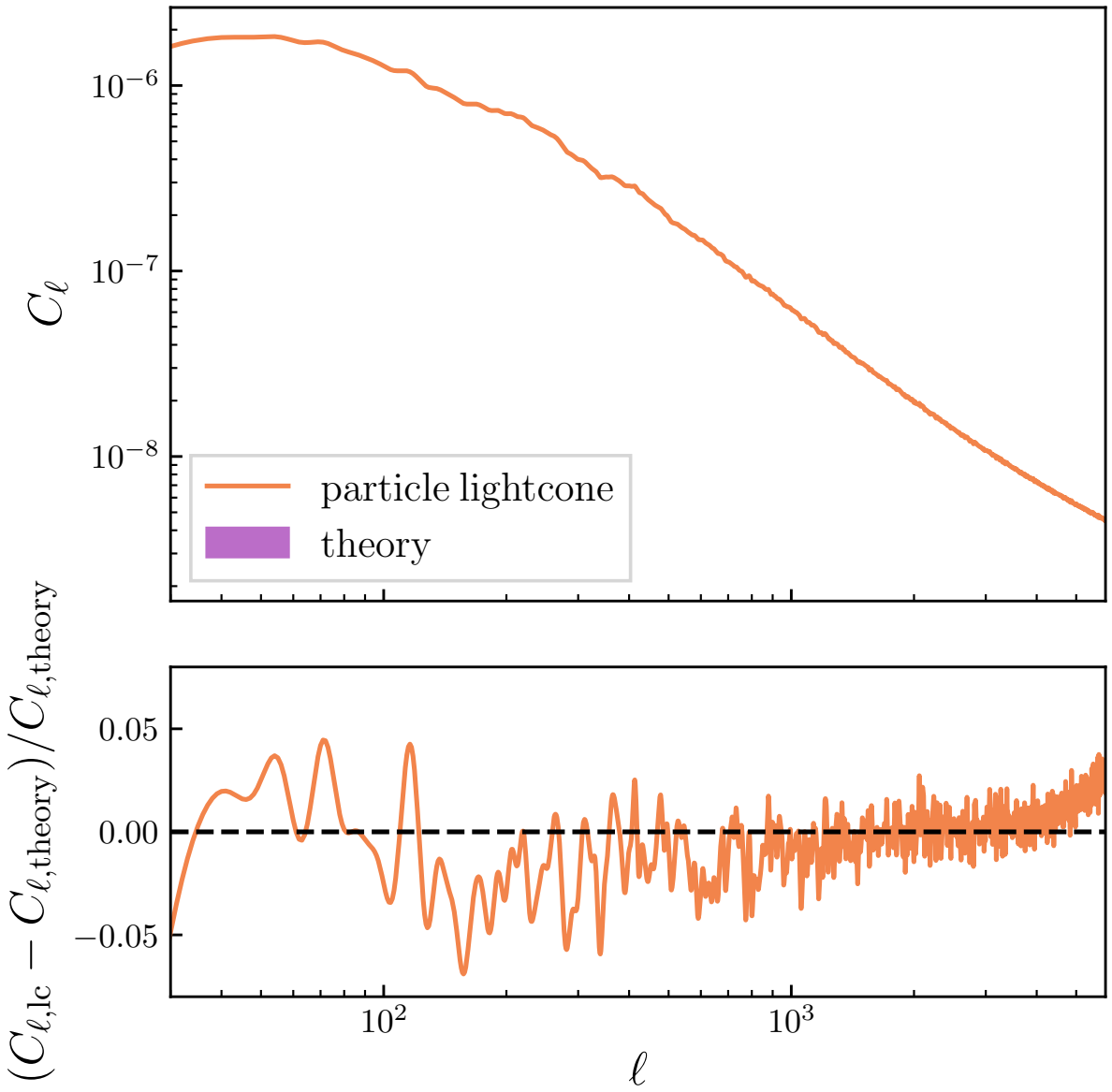

Figure 3.2: Projected particle overdensity power spectrum (top) and residuals with respect to theory (bottom) for a shell of width of approximately 450Mpc at $z \approx 1.7$, after shot-noise corrections. The theory curve is obtained from the CosmicEmu power spectrum emulator, and has been corrected for expected levels of smoothing due to finite shell-width effects. The shaded region in the residual plot shows the approximate 1σ level of cosmic variance given the Polspice kernels. The up-turn in the residuals at small scales is due to the interpolation errors, these are sub-percent below $\ell \approx 3000$, rising to approximately 3% by $\ell = 6000$.

replicate the Outer Rim volume once per axis at a depth of $z \approx 1.3$, so the impact of this decorrelation is relatively minimal. Though, in principle, this effect is present in the results shown in Figure 3.2, we find it to be negligible in practice. We refer the reader to [21] for an in-depth study of these and other considerations related to generating lightcones from cosmological simulations.

Halo Lightcone

To construct a catalog of halos on the observer's lightcone, an intuitive solution would be to re-run the halo-finding algorithm on the Outer Rim particle lightcone. Doing so, however, would be excessive in terms of computational cost, given that we already have a pre-constructed snapshot-based FOF halo catalog and associated merger tree (Sec. 3.3.3). Therefore, we choose to pass the contents of the halo merger tree through the lightcone solver as described in Sec. 3.3.4, resulting in a halo data set which is spatially commensurate with the particle lightcone.

In implementing the lightcone solver for the halo case, we use the same simulation volume replication and rotation strategy as previously described, and we also adopt the same linear approximation when interpolating spatial positions between snapshots. However, when generating halo lightcones, we conduct the interpolation proceeding backwards in time, interpolating halo positions in the direction of increasing redshift.

To understand why this time reversal is performed, it is helpful to imagine a branch of the halo merger tree spanning some time interval, which the lightcone surface "slices through" near the time of a halo merger event We see that the adjacent snapshots j and $j+1$ (grey planes), which bound the intersection of the lightcone and the merger tree branch, each host different objects – the extent of our knowledge is that a merger happened *somewhere* in the interval $t_{j-1} < t_{\mathrm{merge}} < t_j$.

Various prescriptions for assigning t_{merge}, and defining halo properties at that time,

have been described in the literature. For example, in building their Millennium-XXL-based lightcone, [134] choose the merger time randomly per halo progenitor, and interpolate masses between snapshots. For cosmoDC2, we take a simpler approach, and assume that the merger has *always* happened prior to it intersecting the lightcone surface (that is, for a merging merger tree branch that crosses the lightcone at t_e, we assert $t_{\rm merge} < t_e$ in all cases). We set the position of each halo within the lightcone by interpolating between the current halo position and that of its most massive progenitor, retaining all halo properties (mass, radius, etc.) as they appear in the later snapshot at t_{j+1}.

3.3.5 Workflow

Having described the simulations and the data products that are generated, we now provide a final summary by discussing the workflow for producing the inputs to the cosmoDC2 production pipeline. The workflow diagram is shown in Figure 3.4 and begins with the particle catalogs from the smaller AlphaQ simulation and the larger Outer Rim simulation. These are both processed by the halo finder to construct halo catalogs which are then input into the merger tree builder. In the case of the Outer Rim simulation, the merger trees are used to build halo lightcones (see Sec. 3.3.4) that serve as inputs for the Empirical-Model Pipeline and provide host halos for the galaxies in cosmoDC2. For the AlphaQ simulation the merger trees are used as inputs to the Galacticus SAM that is subsequently used to build the Galacticus Library. The particle snapshots from the Outer Rim simulation are also input into the particle-lightcone generator to produce the inputs required for the Lensing Pipeline.

3.4 Weak Lensing

Weak lensing distortions are key observables of the LSST survey, providing constraints on the growth of cosmic structure and therefore dark energy [e.g., 96]. These distortions, which take the form of an isotropic change in area (convergence) and an area-preserving change in

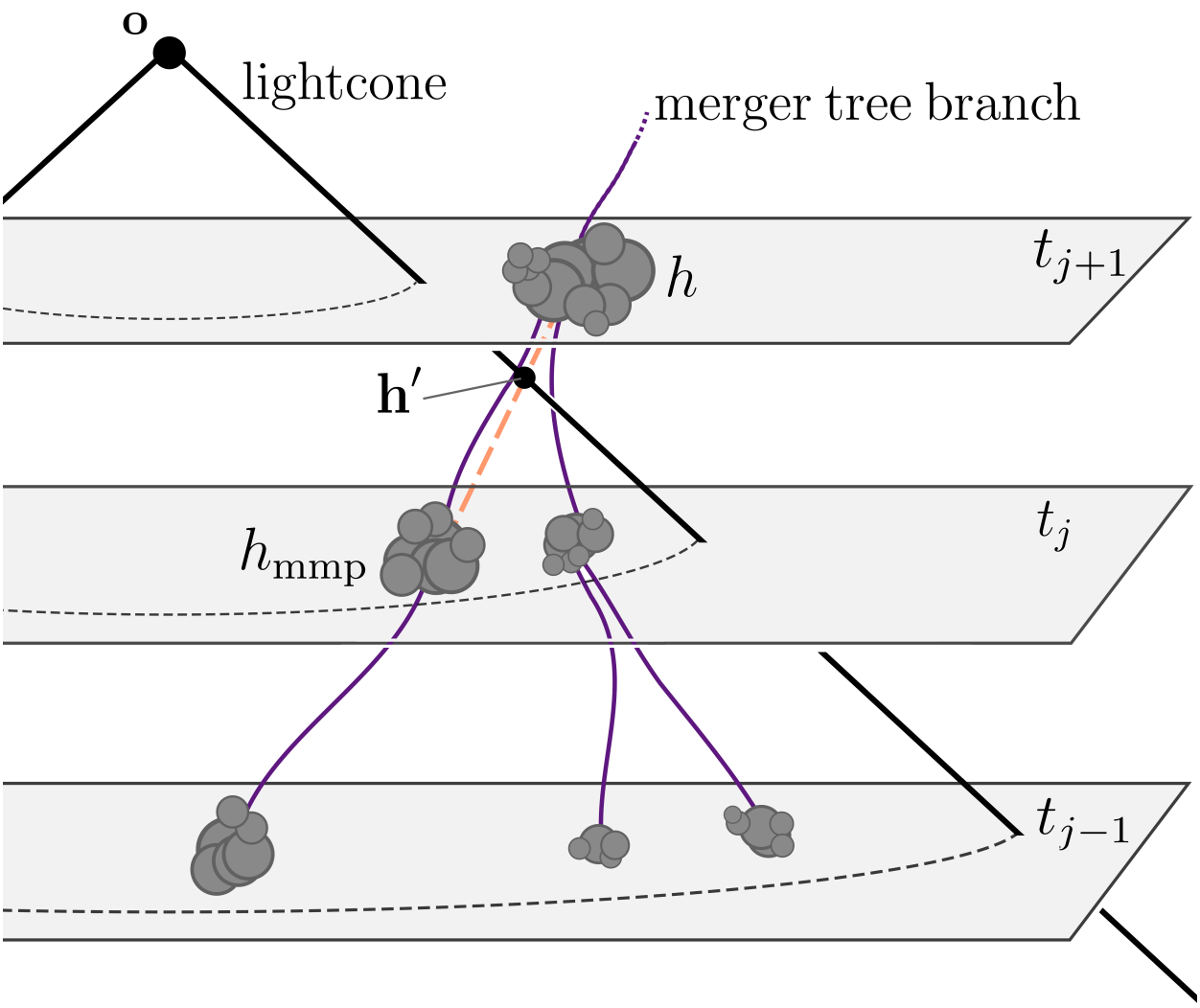

Figure 3.3: Schematic of the interpolation process which fills the cosmoDC2 halo lightcone. Each plane represents a projected simulation snapshot, and time increases vertically, with the observer located at **o**. A merger tree branch including halo h is seen crossing the observer's lightcone between snapshots j and $j+1$ (the purple worldlines of each halo are unknown between the snapshots). Interpolation between halo h and its most massive progenitor h_{mmp} (orange dashed line) is used to solve for the temporal and spatial components of event $\mathbf{h}'$, where we place an object with properties (mass, etc.) identical to halo h.

shape (shear), can be mimicked in simulations by following the paths of photon rays as they traverse the matter field. In practice, maps of the lensing quantities are obtained as follows: the particle lightcone is divided into discrete shells, then photon paths are traced backwards in time from an observer grid to a 'source' shell, with deflections applied corresponding to

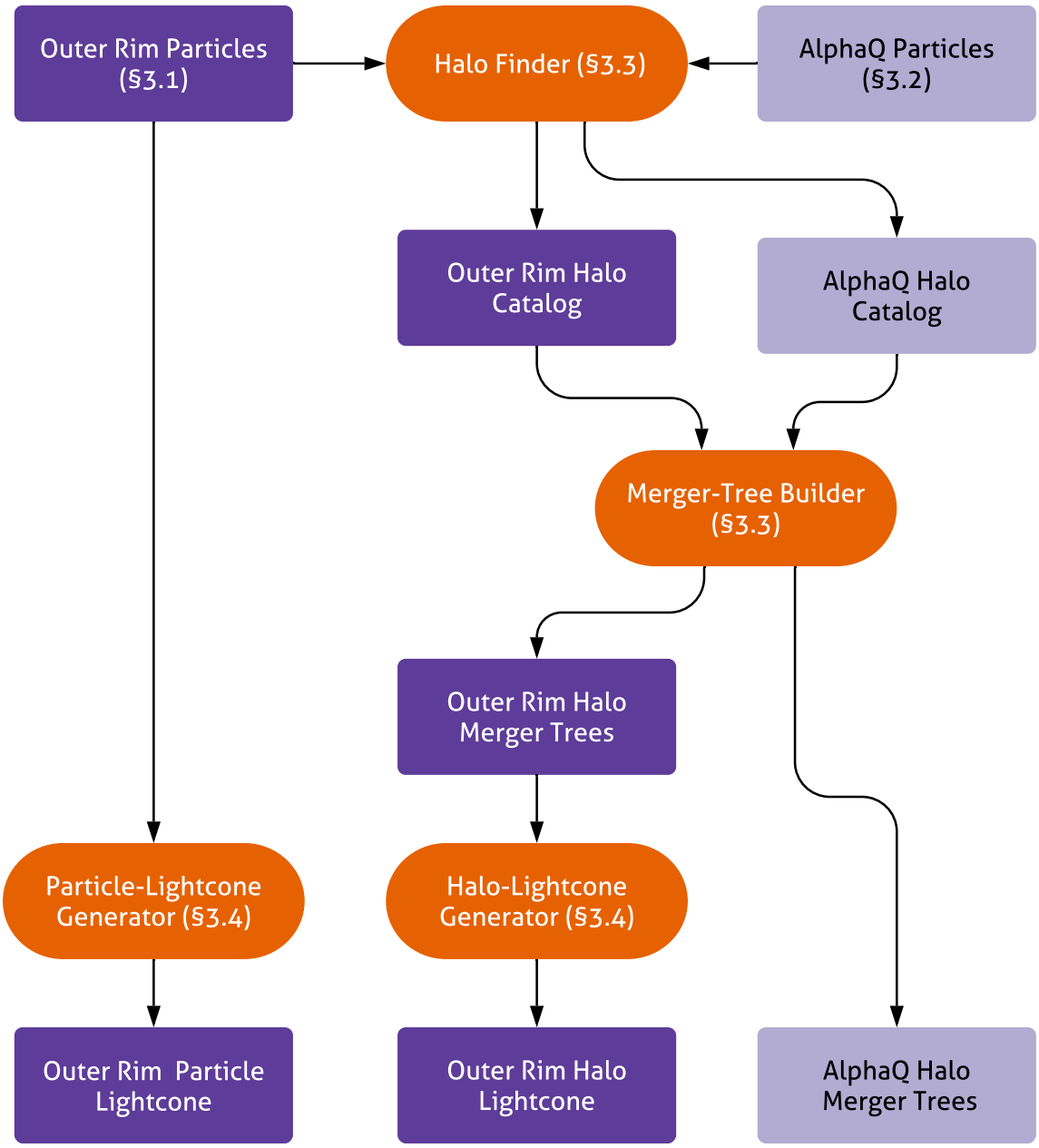

Figure 3.4: Workflow to produce the Outer Rim and AlphaQ simulation data products used as inputs to the cosmoDC2 production pipeline. Data products are shown as rectangles in dark and light purple for data derived from the Outer Rim and AlphaQ simulations, respectively. Code modules are shown as ovals in dark orange. Numbers in parentheses refer to the sections in the paper where a detailed description of the workflow component is given.

the surface density of particles at each 'lens' shell between the source and observer using a ray-tracing algorithm [e.g., 39, 67].

The baseDC2 lensing maps are built with the pipeline presented in P. Larsen et al. (2019, in preparation). The full workflow is illustrated in Figure 3.5. After we create a downsampled particle lightcone using the techniques described in Sec. 3.3.4 and divide it into discrete shells, we compute the surface densities on a HEALPix[3] [57] grid of Nside=4096 using a modified Delaunay Tesselation Field Estimator based on the code of [115]. We then conduct ray tracing using the standard iterative equations of [67], computing gradients and applying deflections on the full-sky maps with Lenspix routines [91].[4] The shells are chosen to cover the line-of-sight distance between adjacent simulation outputs, as described in Sec. 3.3.1, with a median width of approximately 114Mpc. Figure 3.6 displays a cartesian projection of a 100 deg^2 patch of a resulting convergence source map at $z \approx 1$. The inset shows a sub-region with an expanded scale and with the shear field overlaid to show the tangential shearing around massive structures. Figure 3.7 compares the cosmic shear E-mode power spectra in three source maps to theoretical expectations; power spectra in the source maps are computed using the Polspice code with error bars obtained from jackknife sampling; theory curves are derived from the CosmicEmu power spectrum emulator using the Born approximation. These appear to agree to within the 4% model errors anticipated from the power spectra on scales below $\ell \approx 2000$, and to within 10% below $\ell \approx 4000$. As described in Sec. 3.3.4, we note that cosmic shear on small scales is affected by interpolation of the downsampled particle lightcones, as well as limitations of the theoretical model and density estimation, and so the $\lesssim 10\%$ level of agreement shown in Figure 3.7 is expected.

Galaxies are assigned the lensing quantities of their source shell. We note that the

3. https://sourceforge.net/projects/healpix/

4. In particular we compute derivatives of spherical harmonics using the `HealpixAlm2GradientMap` routine, and deflect the mass shell to the observed plane using the `HealpixInterpLensedMap` routine; these functions use cubic interpolation after a high resolution equi-cylindrical grid repixelization, with a cutoff in multipole of $\ell_{\rm max} = 8000$.

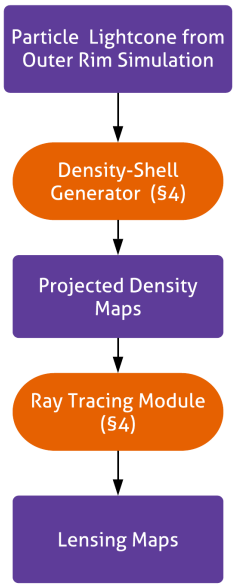

Figure 3.5: Workflow for the shear pipeline. Data products derived from the Outer Rim simulation are shown as boxes in dark purple and code modules are shown as ovals in dark orange.

maps resulting from the iterative ray-tracing equations of [67] are on an observer grid, so that galaxies must be shifted to their observed positions via the lensing deflections before computing the distortions. As the total deflection angles are small compared to the scale of pixelization, we use a first-order approximation to shift the positions from the source shell to the observed grid, given by $\hat{n}_{\rm obs} \approx \hat{n}_{\rm source} - \nabla\phi(\hat{n}_{\rm source})$. We then assign lensing quantities by bilinear interpolation to the source maps.

Several complications arise when making simulation-based lensing predictions with ray-tracing techniques. These include the effects of shot noise, noise associated with triangulation in the density estimation of the downsampled particle data, particle lightcone interpolation errors, ringing effects from the use of spherical harmonics, and artificial smoothing due to pixelization and finite shell-width effects. We give a detailed discussion of these and other issues in P. Larsen et al. (2019, in preparation).

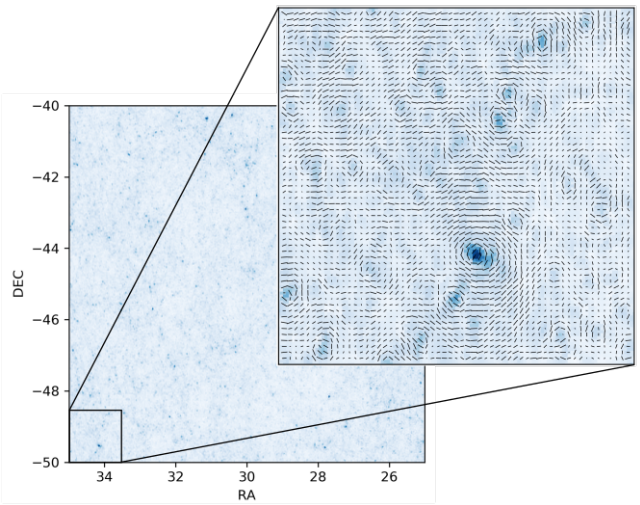

Figure 3.6: Left image: Cartesian projection of a patch of the convergence source plane, in observer coordinates, at $z \approx 1.0$. Right image: Zoom-in of a box within this patch, with the cosmic shear field overlaid. For visualization purposes the lengths of the shear vectors are truncated to a maximum value above $|\gamma| = 0.025$.

3.5 The Galaxy Catalog

In this section we describe the method by which we populate the halo lightcone of the Outer Rim simulation with galaxies. First we produce the baseDC2 catalog by resampling galaxies from the UniverseMachine catalog and applying additional empirical modeling as described in Sec. 3.5.1. Next we incorporate the weak lensing properties determined from the Outer Rim particle lightcone. Finally, we match the empirically-modeled galaxies to those in the Galacticus library to complete the assignment of the complex set of properties.

3.5.1 Empirical Galaxy Catalog Generation

Here we describe the ingredients of the empirical model that we use to construct the baseDC2 catalog. The model parameters are simulation dependent and do not have a simple connection to observables. They have been chosen by an iterative procedure to produce a galaxy catalog whose distributions of observable properties are sufficiently close to that of the observational data to pass the DC2 validation criteria. It is a non-trivial task to calibrate these

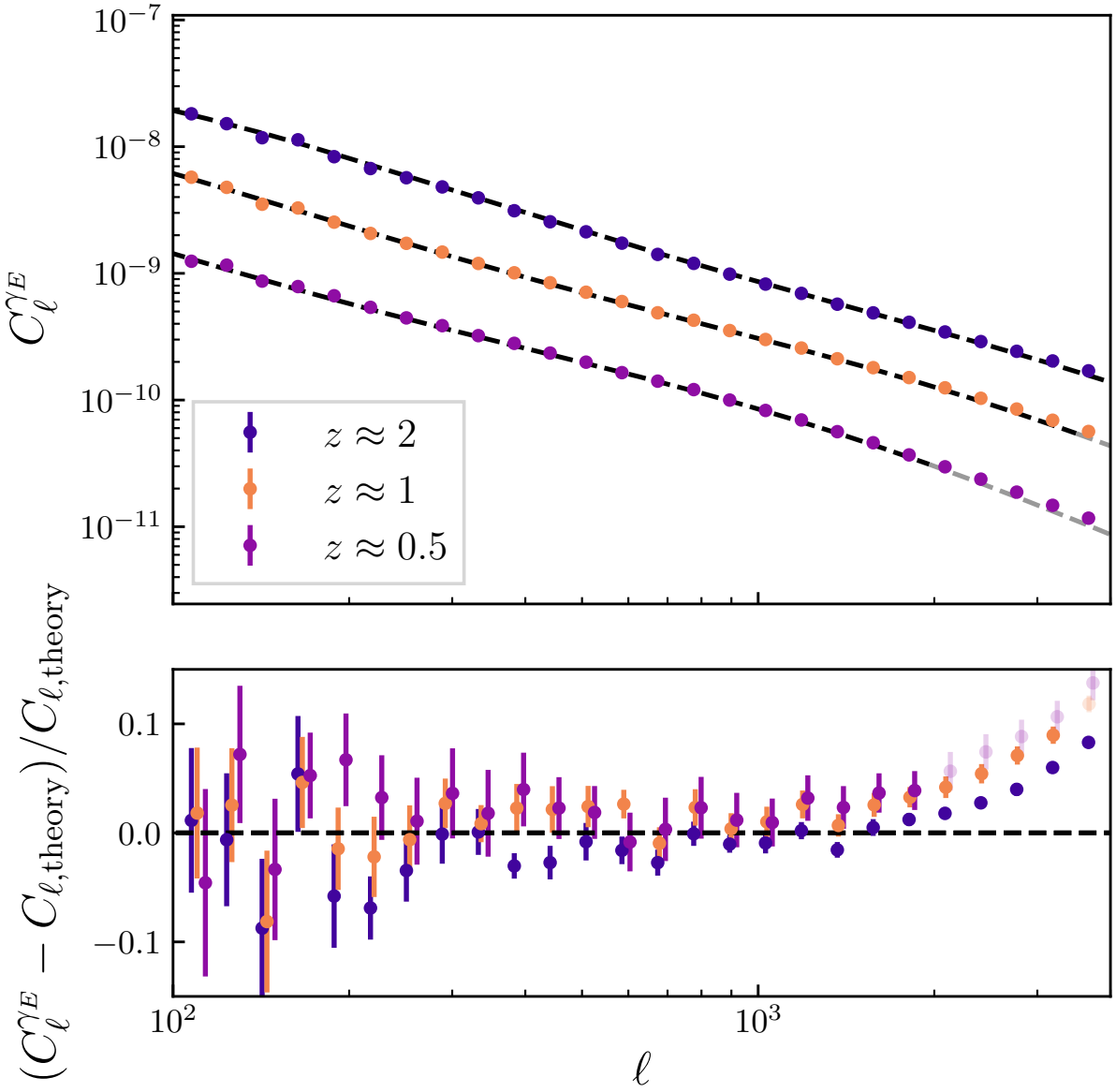

Figure 3.7: E-mode shear power spectrum (top) and residuals with respect to theory (bottom) for a selection of source planes. Theory curves (black dashed lines) are computed using the Born approximation on the CosmicEmu power spectrum emulator, extrapolated to high k values using Padé approximants, the residual points are slightly offset for visualization purposes. The gray shaded region in the residual panel indicates the advertised 4% accuracy of the power spectrum, transparency of the theory curves and residual points indicate that high-k extrapolation accounts for more than 10% of the total power in the theory curve.

model parameters because they describe the behavior of rest-frame quantities in the baseDC2 model, rather than the observer-frame quantities used in the validation tests. For example, the validation test for galaxy number density as a function of magnitude (see Sec. 3.6.1) impacts the parameters used for modeling ultra-faint galaxies to be discussed in Sec. 3.5.1. All of the model-parameter values can be obtained from the public code release.[5]

Restframe Colors

The starting point for baseDC2 is the publicly available UniverseMachine synthetic galaxy catalog [13]. The UniverseMachine is an empirical model for predicting the star-formation history of galaxies; the model is predicated upon the assumption that the mass assembly of a galaxy is correlated with the assembly of its underlying dark matter halo. While this is a longstanding assumption of the semi-analytic modeling approach to galaxy formation [e.g., 78, 136, 17, 135], recent theoretical developments have enabled a new generation of models to leverage this assumption in a way that significantly improves the complexity that can be captured with empirical techniques [11, 33, 101, 122]. The UniverseMachine model has been shown to faithfully capture a wide range of statistics summarizing the observed galaxy distribution across redshift, including stellar mass functions, quenched fractions, and the SFR-dependence of two-point clustering.

The UniverseMachine catalog we use contains synthetic galaxies populating snapshots in the MDPL2 simulation [82], such that every subhalo identified by Rockstar [15] in MDPL2 is populated with a synthetic galaxy. For the purposes of baseDC2, we restrict attention to just two attributes of these synthetic galaxies, stellar mass $M_\star$, and star-formation rate SFR. In particular, the value of $M_\star$ that we use is defined as the total surviving stellar mass belonging to the galaxy, excluding contributions from intra-cluster light. We use the GalSampler technique to transfer the UniverseMachine galaxy population in MDPL2 to the Outer Rim

5. https://github.com/LSSTDESC/cosmodc2

simulation (for technical details, see A. Hearin et al, 2019 in preparation). Briefly, for every host halo in the Outer Rim, we randomly select a host halo in MDPL2 of similar mass, and map the galaxy content of the selected MDPL2 halo into the Outer Rim halo, preserving the halo-centric positions and velocities of the galaxies. By construction, the GalSampler technique preserves the conditional distribution $P(\mathrm{SFR}, \mathrm{M}_\star|\mathrm{M}_{\rm halo})$, as well as the halo mass dependence of the UniverseMachine halo occupation statistics, $P(N_{\rm gal}|M_{\rm halo})$. For the most massive halos in the Outer Rim simulation, which have no counterparts in MDPL2, using the GalSampler random selection procedure ensures that we are not repeatedly resampling galaxies from the same MDPL2 halo. We take the larger halo mass of the Outer Rim halo into account by applying a redshift-dependent boost to the UniverseMachine value of $M_\star$.

At this stage, every halo in the Outer Rim lightcone has been populated with synthetic galaxies with $M_\star$ and SFR. We model restframe absolute magnitude $M_{\rm r}$ as a function of $x \equiv \log_{10} M_\star$ and redshift z, we map M_r onto every synthetic galaxy using the following model:

$$\langle M_{\rm r}|M_\star, z\rangle = \Big[M_r^0 - x\alpha(x)\Big] \times (1 + f(x, z))\,, \tag{3.1}$$

where $M_r^0 = -20.1$ is a constant, and where we model both $\alpha(x)$ and $f(x, z)$ using $S(x)$, a sigmoid function:

$$S(x) = y_{\rm min} + \frac{y_{\rm max} - y_{\rm min}}{1 + \exp(-k(x - x_0))}. \tag{3.2}$$

The function $\alpha(x)$ controls the $M_\star$-dependence of the power-law slope. For $\alpha(x)$, we use a low-mass slope of $y_{\rm min} = 1.75$ and a high-mass slope of $y_{\rm min} = 1.8$, with a transition speed of $k = 2.5$ and a pivot mass of $\log_{10} M_{\rm pivot}/\mathrm{M}_\odot \equiv \mathrm{x}_0 = 10$.

The function $f(x, z)$ controls the redshift evolution of the $\langle M_{\rm r}|M_\star, z\rangle$ relation. Because galaxies at higher redshift are generally composed of younger stellar populations, we expect that the median value $\langle M_{\rm r}|M_\star, z\rangle$ brightens with redshift, and that this brightening is

stronger for lower-mass galaxies. We capture this complexity by modeling

$$f(x,z) = \frac{\delta(x)}{1+\exp(-k(z-z_0))}, \tag{3.3}$$

with a transition speed $k = 10$, and a pivot redshift $z_0 = 0.7$, and a third-order polynomial for $\delta(x)$ defined to pass through the points $\{(6,-2),(8,-1.5),(10,-0.5),(12,0)\}$. The preceding pairs of numbers give values for $(\log(M_\star), \langle \Delta M_r \rangle)$, where $\langle \Delta M_r \rangle$ is the average brightening for galaxies for a given stellar mass.

Having mapped $M_\star$ and M_r onto every synthetic galaxy, we proceed to model restframe colors $g-r$ and $r-i$. For each distribution we use a double Gaussian, with statistically distinct star-forming and red sequence populations:

$$\begin{aligned} P(g-r|M_\mathrm{r},z) &= F_\mathrm{q}(M_\mathrm{r},z) \times \mathcal{N}^\mathrm{q}_\mathrm{g-r}(\mu;\sigma) \\ &+ \big(1-F_\mathrm{q}(M_\mathrm{r},z)\big) \times \mathcal{N}^\mathrm{sf}_\mathrm{g-r}(\mu;\sigma), \end{aligned} \tag{3.4}$$

and likewise for $P(r-i|M_\mathrm{r},z)$. We model the two-dimensional dependence of $F_\mathrm{q}(M_\mathrm{r},z)$ using a composition of sigmoid functions:

$$F_\mathrm{q}(M_\mathrm{r},z) = F_\mathrm{q}^{z_0}(M_\mathrm{r}) + \frac{F_\mathrm{q}^{z_1}(M_\mathrm{r}) - F_\mathrm{q}^{z_0}(M_\mathrm{r})}{1+\exp(-k_\mathrm{z}(z-z_\mathrm{c}))}, \tag{3.5}$$

where for both $g-r$ and $r-i$ colors we use $z_\mathrm{c} = 0.5$ and $k_\mathrm{z} = 12$. The $F_\mathrm{q}^{z_\mathrm{i}}$ functions characterize the M_r-dependence of the quenched fraction at the two asymptotic redshifts, z_0 and z_1. We model $F_\mathrm{q}^{z_\mathrm{i}}(M_\mathrm{r})$ as

$$F_\mathrm{q}^{z_\mathrm{i}}(M_\mathrm{r}) = f_\mathrm{q}^\mathrm{faint} + \frac{f_\mathrm{q}^\mathrm{bright} - f_\mathrm{q}^\mathrm{faint}}{1+\exp(-k_\mathrm{r}(M_\mathrm{r}-M_\mathrm{r}^\mathrm{c}))}. \tag{3.6}$$

In cosmoDC2, the widths of the quenched and star-forming sequences are constant, but the centroids depend on both mass and redshift, i.e., in Eq. 3.4, for each sequence and broadband color, $\mu = \mu(M_{\rm r}, z)$. We model this simultaneous dependence as a composition of sigmoid functions.

$$\mu(M_{\rm r}, z) = \mu_{z_0}(M_{\rm r}) + \frac{\mu_{z_1}(M_{\rm r}) - \mu_{z_0}(M_{\rm r})}{1 + \exp(-\alpha_{\rm z}(z - z_{\rm c}))}, \tag{3.7}$$

where

$$\mu_{z_i}(M_{\rm r}) = \mu_{\rm faint} + \frac{\mu_{\rm bright} - \mu_{\rm faint}}{1 + \exp(-\alpha_{\rm r}(M_{\rm r} - M_{\rm r}^{\rm c}))}. \tag{3.8}$$

We arrived at these functional forms and best-fit values after considerable experimentation and iteration with the DESCQA color validation tests to be presented in Sec. 3.6.3.

Cluster Environment

While the methods used in Sec. 3.5.1 produce model galaxies with realistic stellar mass and broadband flux, as well as reasonably accurate two-point clustering, two additional ingredients are needed in order to meet validation requirements in cluster environments. First, more stellar mass is required of central galaxies in very massive halos, $M_{\rm halo} \gtrsim 10^{14}{\rm M}_{\odot}$[71]. Second, the normalization of the mass-richness relation in cluster-mass halos appears to be $\sim 20\%$ low relative to expectations based on DES data. In principle, the UniverseMachine model could capture these effects if suitable observational data for galaxies in cluster environments were used to tune the UniverseMachine model parameters. In practice, the necessity of these modifications is not surprising because the observational constraints used to fit the UniverseMachine model are relatively insensitive to the behavior of the galaxy-halo connection in the statistically rare environment of very massive halos.

To address the boost to the stellar mass of cluster centrals, we remapped the $\langle M_\star | M_{\rm halo} \rangle$ relation for $M_{\rm halo} > M_{\rm halo}^c = 10^{13.5}{\rm M}_{\odot}$ according to $M_\star \propto \left(M_{\rm halo}/M_{\rm halo}^c\right)^\alpha$, using $\alpha = 0.65$,

and set the normalization according to the existing value for $\langle M_\star | M^c_{\rm halo} \rangle$. To boost the mass-richness relation, we generate an additional Monte Carlo realization of cluster satellites so that the total number of objects increases by 20% in all halos $M_{\rm halo} > M^c_{\rm halo}$, decreasing this boost factor linearly in $\log M_{\rm halo}$ so that the boost is zero for halos with $M_{\rm halo} \leq 10^{13} {\rm M}_\odot$.

Ultra-Faint Galaxies

A range of science goals related to weak lensing and deblending benefit from a synthetic catalog that is complete down to galaxy masses below the resolution limit of the simulation. As discussed in Sec. 3.6.1, the primary DESCQA validation requirement that drives the need for this additional modeling is quantified in the test for the cumulative galaxy number counts as a function of apparent magnitude. To meet this requirement, part of the cosmoDC2 model involves the inclusion of galaxies that are disconnected from resolved halos in the Outer Rim simulation. In this section and below, we will refer to this population as "ultra-faint galaxies", using this term to refer to the population of galaxies added to the mock to extend beyond the resolution limit of the simulation.[6]

Our modeling for the ultra-faint population begins by defining how many galaxies should be included in order to boost the abundance of faint galaxies to meet the DESCQA validation criteria. We make two physical assumptions to determine the abundance of ultra-faint galaxies at each redshift. The assumptions are that there is one-to-one correspondence between galaxies and (sub)halos, accounting for both distinct host halos and for subhalos within them, and that the (sub)halo mass function ${\rm d}n/{\rm d}M_{\rm peak}$ should exhibit power-law behavior at the low-mass end. These assumptions are used to extend the reach of the simulation in the following way. Let $M^{\rm lim}_{\rm peak}$ denote the resolution limit of the simulation, i.e., below

6. We note in particular that our usage of the term "ultra-faint galaxies" has no connection to low-mass galaxy populations with similar names in the literature such as "ultra-faint dwarfs" or "ultra-diffuse galaxies". Our usage of this terminology is simply shorthand for the population of synthetic galaxies in cosmoDC2 for which there is no host halo.

$M_{\rm peak}^{\rm lim}$ we begin to see departures from power-law behavior of $\mathrm{d}n/\mathrm{d}M_{\rm peak}$. Then for any particular value $M_{\rm peak}^{\rm ext} < M_{\rm peak}^{\rm lim}$, we can simply extrapolate the power-law approximation to $\mathrm{d}n/\mathrm{d}M_{\rm peak}$ to estimate how many (sub)halos are missing due to the finite mass resolution of the simulation. We elaborate upon this procedure below.

In each snapshot of the MDPL2 subhalo catalogs used in baseDC2, we identify a value of $M_{\rm peak}$ that is sufficiently larger than $M_{\rm peak}^{\rm lim}$, and fit a power law to $\mathrm{d}n/\mathrm{d}M_{\rm peak}$ in the neighborhood of this mass. Selecting $M_{\rm peak}^{\rm ext} = 10^{9.8}\mathrm{M}_{\odot}$, we then calculate $n_{\rm ultra-faints}(z|M_{\rm peak} > M_{\rm peak}^{\rm ext})$, the cumulative comoving number density of (sub)halos that would be present if MDPl2 were $M_{\rm peak}$-complete to this mass:

$$n_{\rm ultra-faints}(z|M_{\rm peak} > M_{\rm peak}^{\rm ext}) \equiv \int_{M_{\rm peak}^{\rm ext}}^{M_{\rm peak}^{\rm lim}} \mathrm{d}M_{\rm peak} \mathrm{d}n/\mathrm{d}M_{\rm peak}^{\rm fit}(z) \tag{3.9}$$

Equation 3.9 defines the volume number density of synthetic ultra-faint galaxies to which we will apply the GalSampler technique to add them to baseDC2. We also use $\mathrm{d}n/\mathrm{d}M_{\rm peak}^{\rm fit}$ to draw values of $M_{\rm peak}$ for each ultra-faint galaxy we add. Once we have a sample of synthetic ultra-faint galaxies that have values of $M_{\rm peak}$, we assign stellar masses according to a power-law fit to the faint end of the UniverseMachine relation $\langle M_{\star}|M_{\rm peak}, z\rangle$, and we assign uniform random values to $P(< \mathrm{SFR}|M_{\star})$, the CDF of the SFR conditional probability distribution. Based on the assignments of $M_{\star}$ and SFR, we assign colors as described in Sec. 3.5.1. We then assign spatially random locations to these galaxies, computing redshifts from the corresponding comoving distances. At this point, the synthetic ultra-faint galaxies have all the attributes needed to treat them as ordinary UniverseMachine galaxies in the baseDC2 pipeline.

As in the case of the rest-frame color model, the parameter values presented here are obtained by boosting the number of ultra-faint galaxies until there are sufficiently many to pass the cumulative number-density validation test to be presented in Sec. 3.6.1. In the

future, for use-cases requiring more realistic spatial distributions of the ultra-faint population, it would be necessary to incorporate expected correlations between the density field and the positions of very low-mass galaxies.

HOD Comparison with UniverseMachine

One of the simplest ways to quantify the relationship between galaxies and the cosmic density field is through the Halo Occupation Distribution (HOD), $P(N_{\rm gal}|M_{\rm halo})$, the probability that a halo of mass $M_{\rm halo}$ hosts $N_{\rm gal}$ galaxies that meet some selection criteria. In order to demonstrate that the $P(N_{\rm gal}|M_{\rm halo})$ in cosmoDC2 is reasonably realistic, we compare our HOD to that seen in mock catalogs made with the UniverseMachine. Two-point projected clustering in the UniverseMachine model has been shown to exhibit close agreement with SDSS [13], and so for purposes of ensuring reasonably accurate correlations between galaxies and the density field, we compare our redshift-zero HOD to that seen in UniverseMachine.

In Figure 3.8, the dashed curves shows the HOD of $z = 0.15$ UniverseMachine galaxies, with different stellar mass thresholds as indicated in each panel. To calculate the corresponding quantity in our model, we use the $z = 0.15$ snapshot of the Outer Rim halo catalog populated with baseDC2 galaxies. The good agreement between the dashed and solid curves in each panel of Figure 3.8 should be sufficient to ensure that cosmoDC2 has reasonably accurate relationships between stellar mass and halo mass. Since stellar mass and luminosity are tightly correlated in the cosmoDC2 model (see Sec. 3.5.1), then the HOD in cosmoDC2 will naturally inherit dependence upon broadband magnitude. As discussed in Sec. 3.6, our tuning of this technique was fairly coarse relative to the accuracy with which the HOD has been shown to recover the clustering of specific galaxy samples [e.g., 157]. To build the cosmoDC2 model, we instead prioritized modeling galaxies with HODs that exhibit the expected scaling with a wide variety of complex observational selection functions, and we have found the level of agreement shown in Figure 3.8 to be sufficient to serve a broad range of

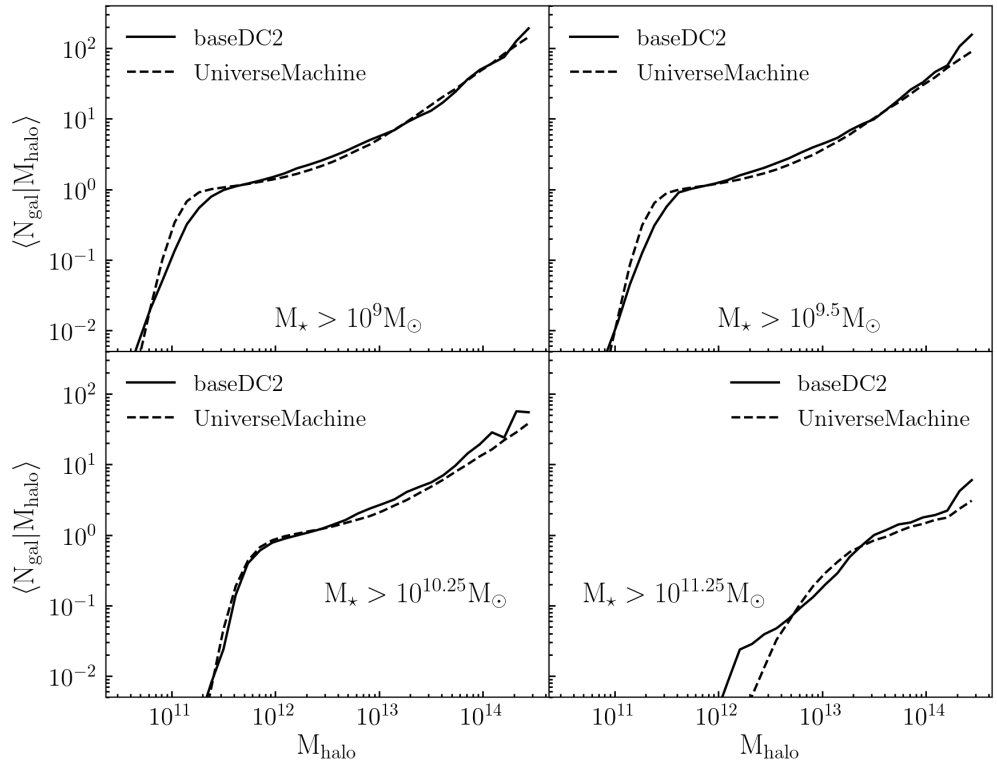

Figure 3.8: Comparison of the $z = 0.15$ HOD in the UniverseMachine mock (dashed curves) with those of baseDC2 (solid curves). The HOD quantifies $P(N_{\rm gal}|M_{\rm halo})$, the probability that a halo of mass $M_{\rm halo}$ hosts $N_{\rm gal}$ galaxies with stellar mass $M_\star$ greater than the threshold indicated in the panel. The first moment of the HOD, $\langle N_{\rm gal}|M_{\rm halo}\rangle$, is shown on the vertical axis of each panel as a function of $M_{\rm halo}$. Solid curves show the corresponding HOD in the baseDC2 model that is the foundation of cosmoDC2. We tune our HOD to match UniverseMachine so that our model can inherit the observational realism of the galaxy-halo connection shown in [13].

science applications of DC2.

3.5.2 The Galacticus Library

The Galacticus SAM solves, along each branch of an input merger tree, a set of ODEs which describe the evolution of coarse-grained properties of the galaxy that forms within that branch. This differential evolution is supplemented with impulsive events that describe galaxy mergers. In addition to merging, the physics modeled by Galacticus also includes the cooling of gas in the circum-galactic medium and its inflow into galaxies, star formation,

supermassive black hole growth, feedback processes powered by both supernovae and active galactic nuclei, and metal production. Galacticus follows the evolution of the disk and bulge components separately. The latter component forms as a result of galaxy mergers or via dynamical instabilities of the galactic disk. In either case, some fraction of the disk's mass is transferred to the bulge component.

Our Galacticus library was generated by running version 0.9.4 of the Galacticus code[7] on merger trees built from the AlphaQ simulation. Galacticus outputs were requested at each snapshot redshift of the input merger trees so that a complete history of each galaxy would be available in the Galacticus Library. We use the default 0.9.4 parameter file to specify the inputs to the Galacticus model. Since these input parameters are tuned to observational data using extended Press-Schechter trees rather than merger trees from N-body simulations, we do not expect our results to match these observational data perfectly. However, as mentioned in Sec.3.2, the collection of Galacticus galaxies is serving as a library from which to select galaxies with suitable properties, so that a complete match to the data is not required.

In addition to the basic properties such as stellar mass, star-formation rate, gas and stellar metallicity, the user can supply filters for which Galacticus then calculates luminosities. In principle, these luminosities can be determined by modeling the spectral-energy distribution (SED) of a galaxy with a stellar-population synthesis (SPS) model [35], and integrating the resulting SED over the band-pass for the selected filter. The SPS model provides a library of single-stellar-populations (SSP) SEDs that depend on metallicity and time. The SED of a galaxy at a given redshift is obtained by convolving over time and metallicity, the SFR as calculated by Galacticus, with these SSP SEDs. In practice, it is too time consuming to calculate these integrals for each SED on the fly, so Galacticus pre-computes a table of integrals of the SSP SEDs over the desired filter band-passes and uses these as coefficients in the convolution integral with the SFR. Luminosities for the desired band passes are available

7. https://bitbucket.org/galacticusdev/galacticus/wiki

in either rest or observer frame and are computed separately for the disk and bulge component of each galaxy. Observer-frame luminosities are calculated by appropriately blue-shifting the filter transmission function depending on the cosmological redshift of the galaxy. (Recall that output redshifts are specified so that these required blue-shifts are known in advance.) Note however, that the blue-shifts that account for the line-of-sight peculiar velocities are not included. Although the direct computation of observer-frame magnitudes obviates the need to perform any k-corrections on the rest-frame magnitudes, we still need to interpolate both sets of magnitudes to the galaxy lightcone redshifts. This will be discussed in Sec. 3.5.3.

Galacticus luminosities are computed in the AB-magnitude system. We convert the rest-frame luminosities L to magnitudes using $M = -2.5\log_{10}(L)$. Observer-frame luminosities are converted to apparent magnitudes by including additional factors of $-2.5\log_{10}(1+z)$ and $\mu(z)$, where z is the redshift and μ is the distance modulus of the galaxy. These factors account for the compression of photon frequencies in the observer frame and the luminosity distance, respectively.

The luminosities provided by these user-selected filters are critical for providing the galaxy properties required by the validation tests. For example, many of these tests make cuts on observer-frame LSST magnitudes, which are not obtained from empirical modeling. In addition to the LSST *ugrizy* filters, we included SDSS *ugriz* filters to facilitate the validation of cosmoDC2 against SDSS data and the Johnson B and V filters to provide inputs for the image simulations. We also define a set of 30 top-hat filters spanning the range from 100 nm to 2 μm. The width of these filters grow with wavelength but are designed to provide roughly constant resolutions $\lambda/\Delta\lambda$ that vary from ~ 4 to ~ 7. These top-hat filters provide a coarsely binned estimate of the galaxy's SED that is based on the star-formation history of each galaxy. They are required inputs for the image simulations and are critical for evaluating the accuracy of photometric-redshift determinations. Finally, to estimate emission-line luminosities (described below), we add three continuum filters that compute

the ionizing luminosity for H I, He II, and O II, respectively.

Dust corrections and emission lines are added in post-processing. The dust model is from [52] who used ray-tracing simulations to calculate dust-attenuation curves as a function of inclination, dust distribution and other dust properties. In post processing, a random inclination is generated for each galaxy and the attenuation for each filter is computed by interpolating (or extrapolating) the attenuation at the effective wavelength of the filter from the tabulated values. The effective wavelength is defined as $\int \lambda T(\lambda) d\lambda / \int T(\lambda) d\lambda$, the wavelength averaged over the filter transmission, $T(\lambda)$. No corrections are included for attentuation by the IGM or by Milky Way dust. (The latter are included in the image simulations.)

The emission-line model is based on [109], and its specific implementation within Galacticus is described in detail by [100]. Using this model, emission line luminosities are computed using the photo-ionization code Cloudy [51] as a function of a grid of values for H II region densities, metallicities and the ionizing luminosities for H I, He II, and O II. In post-processing, for each galaxy, the emission-line luminosities are interpolated from this pre-computed grid using the measured values of the ionizing luminosities that are supplied by Galacticus.

As a final step, we remove galaxies with extreme values in the dust-corrected magnitudes, colors and total extinction from the Galacticus library. Any galaxies with any totally extincted dust-corrected magnitude (due to extrapolation failures in the dust model), values of the rest-frame $\mathrm{B}-\mathrm{V}$ excess color or the ratio of total to selective extinction, R_V, close to zero, or values of dust-corrected rest- or observer-frame colors or values of the V-band total extinction, A_V, falling outside of the limits shown in Table 3.5.2 are cut from the library.

Table 3.1: Color and A_v cuts applied to the Galacticus library to remove unphysical galaxies

Quantity	Lower Limit	Upper Limit
Rest-frame $g-r$	-0.5	1.5
Rest-frame $r-i$	-0.5	1.5
Rest-frame $i-z$	-0.5	2.0
Observer-frame $g-r$	-0.5	2.5
Observer-frame $r-i$	-0.5	2.0
Observer-frame $i-z$	-1.0	2.5
A_V	-0.1	3.1

3.5.3 Galaxy Catalog and Galaxy Library Matching

The goal of the Match-Up Pipeline, shown in Figure 3.9, is to find a Galacticus library galaxy for each baseDC2 galaxy that best preserves the careful tuning of existing properties while incorporating additional information provided by the library. The key galaxy properties to match between baseDC2 and the Galacticus library are rest frame r-band magnitude, and rest frame $g-r$ and $r-i$ color. All other baseDC2 properties, including stellar mass and SFR, are copied directly into cosmoDC2. Weak lensing properties (see Sec. 3.4) are incorporated into the baseDC2 catalog before the matching.

Quality of Match Metric

The quality of the match between a library galaxy and a baseDC2 galaxy is measured by the Euclidean distance between their 3-dimensional property vectors $P^{\rm lib}$ and $P^{\rm base}$, respectively, where the components of P are M_r, rest frame $g-r$ and $r-i$ color:

$$D^2 = \sum_j (P_j^{\rm lib} - P_j^{\rm base})^2. \tag{3.10}$$

Cluster red sequence galaxies have a tight relationship between color and redshift [24, 123]. Optical cluster finders such as redMaPPer [125] rely on this tight relationship to help isolate cluster members, and so in order to ensure that cosmoDC2 possesses a tight

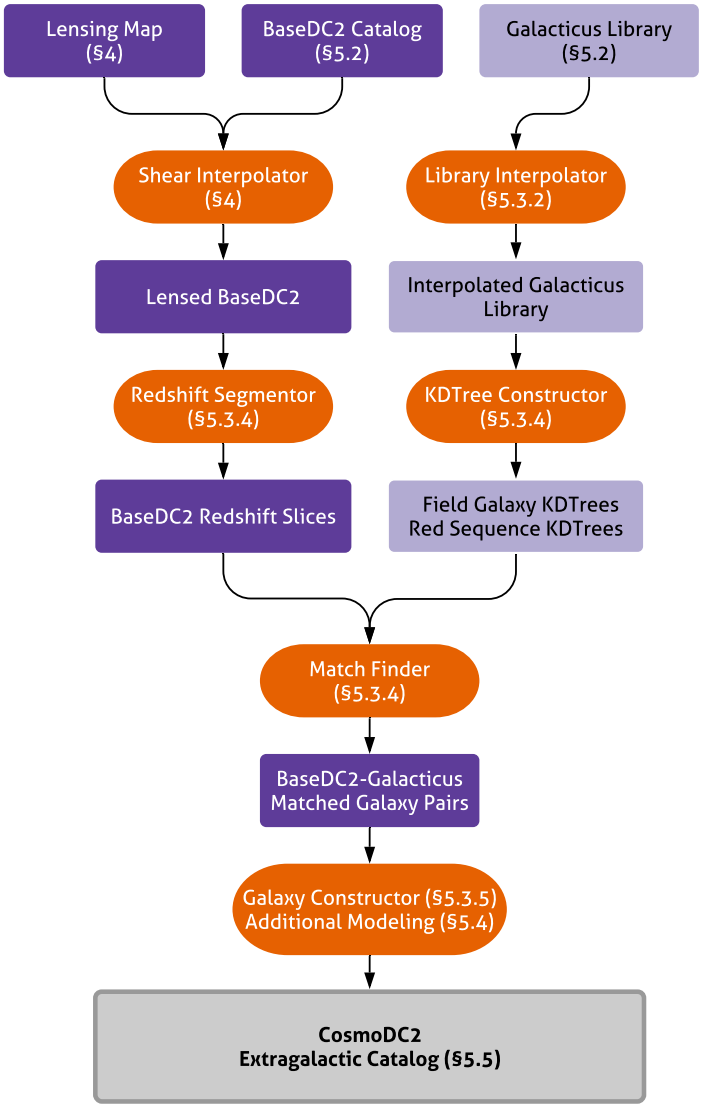

Figure 3.9: Workflow for the Galacticus match-up pipeline to produce the final output for cosmoDC2. Data products are shown as rectangles in dark and light purple for data derived from the Outer Rim and auxiliary simulations, respectively. Code modules are shown as ovals in dark orange.

relation between redshift and observed galaxy color we include a second term in the distance calculation for cluster red sequence galaxies:

$$D^2 = \sum_j \left(P_j^{\rm lib} - P_j^{\rm base}\right)^2 + \sum_j \left(Q_j^{\rm lib} - Q_j^{\rm DES}(z)\right)^2, \tag{3.11}$$

where the color vector Q has components of $g-r,\ r-i$, and $i-z$ in the observer frame, and $Q^{\rm DES}(z)$ is the expected mean red sequence color as a function of redshift observed in DES [126]. Cluster red sequence galaxies are defined as all red sequence galaxies with $M_{\rm halo} > 10^{13}h^{-1}{\rm M}_\odot$. To ensure a smooth transition between the use of the two metrics, we also apply Eq. 3.11 to a fraction of the galaxies for which $10^{12}h^{-1}{\rm M}_\odot < {\rm M}_{\rm halo} < 10^{13}{\rm h}^{-1}{\rm M}_\odot$. This fraction increases log-linearly with halo mass from zero to one. We discuss this modeling choice further in Sec. 3.7.

Redshift Interpolation of Library Galaxies

Whereas baseDC2 is constructed on a lightcone, the Galacticus library has been constructed at discrete redshift snapshots. If galaxies in the baseDC2 lightcone are naively matched to galaxies in the Galacticus library, then discrete bands in observed-color and redshift space will be clearly visible. These bands appear because the color distribution of galaxies changes noticeably between the redshifts of two adjacent snapshots. This is due both to redshifting of the galaxy SED and, to a lesser extent, evolution of the galaxy population. We use interpolation to compute the properties of galaxies lying between the snapshot redshifts. Our procedure substantially reduces the discreteness effects that would otherwise be present.

The Galacticus library contains the full history of each galaxy at every snapshot redshift. To shift a Galacticus galaxy to an intermediate redshift z, we linearly interpolate the

properties of the galaxy between adjacent snapshots, expressed as:

$$g(z) = f(z_t) + (z - z_t)\,\frac{f(z_{t+1}) - f(z_t)}{z_{t+1} - z_t}, \tag{3.12}$$

where z_t and z_{t+1} are the discrete snapshot redshifts of the library that bracket z, $g(z)$ is the interpolated function of a galaxy property and $f(z_i)$ is the value of a galaxy library property at redshift z_i.

Not all galaxies in the library are suited for interpolation between snapshots. For example, some galaxies in Galacticus either formed recently or merged with another galaxy and are therefore missing from a snapshot; galaxies may have evolved in to or out of regions of color-magnitude space that fail to pass quality cuts applied to the library. In such cases we exclude the candidate galaxy from the library, so that only suitable galaxies are selected.

Luminosity Adjustment

A complication in the Match-Up Pipeline is that the baseDC2 galaxies are more luminous than the ones present in the Galacticus library. To remedy the luminosity mismatch at the bright end, the magnitudes for both sets of galaxies are compressed at the brighter end into a smaller and overlapping range:

$$M_r'(M_r) = \begin{cases} \alpha \tanh\left((M_r - M_{r0})/\alpha\right) + M_{r0}, & \text{if } M_r \leq M_{r0} \\ Mr, & \text{Otherwise} \end{cases} \tag{3.13}$$

where M_r and M_r' are, respectively, the original and compressed r-band magnitudes, M_{r0} is the threshold where the compression starts, and α is the range of the compression. Luminosities brighter than M_{r0} are compressed into a range that is strictly less bright than $M_{r0} - \alpha$. The compression effectively downweights the luminosity matching requirements at the brighter end while keeping the color matching requirement fixed. Once a matching

library galaxy is found, the luminosity in each bandpass is rescaled by the same constant factor that forces the r-band luminosity for the library and baseDC2 galaxy to agree by construction.

Assigning a Match

The matching between baseDC2 and the library is done by constructing k-d trees for the library and querying these trees for the nearest neighbors to each baseDC2 galaxy. Since the library galaxy properties change with redshift, the k-d trees have to be reconstructed each time the library is interpolated to a new redshift. It would be computationally impractical to construct the k-d trees for each individual redshift value in the baseDC2 lightcone catalog. Instead, we use sets of k-d trees at narrow redshift slices of the lightcone as described below.

We divide the redshift range between adjacent snapshots into five slices. For each of these slices, the galaxy library is interpolated to its median redshift and two k-d trees are constructed from the library. The first tree, which is used for most galaxies, only uses rest frame r-band, $g-r$ and $r-i$ color and uses the distance metric of Eq. 3.10. The second tree, which is used for red sequence cluster members, additionally uses observer frame $g-r$, $r-i$ and $i-z$ colors and uses the distance metric of Eq. 3.11. For each baseDC2 galaxy in the slice, we use the k-d tree to find the ten closest neighbors in the galaxy library and assign a match randomly from those ten, which smooths out discreteness effects otherwise caused by repeated selection of the same library galaxy in a sparsely populated area of color-magnitude space. Figure 3.10 shows the color-redshift evolution of galaxies resulting from our match-up procedure. The color distribution before and after the match-up procedure is shown in Figure 3.11.

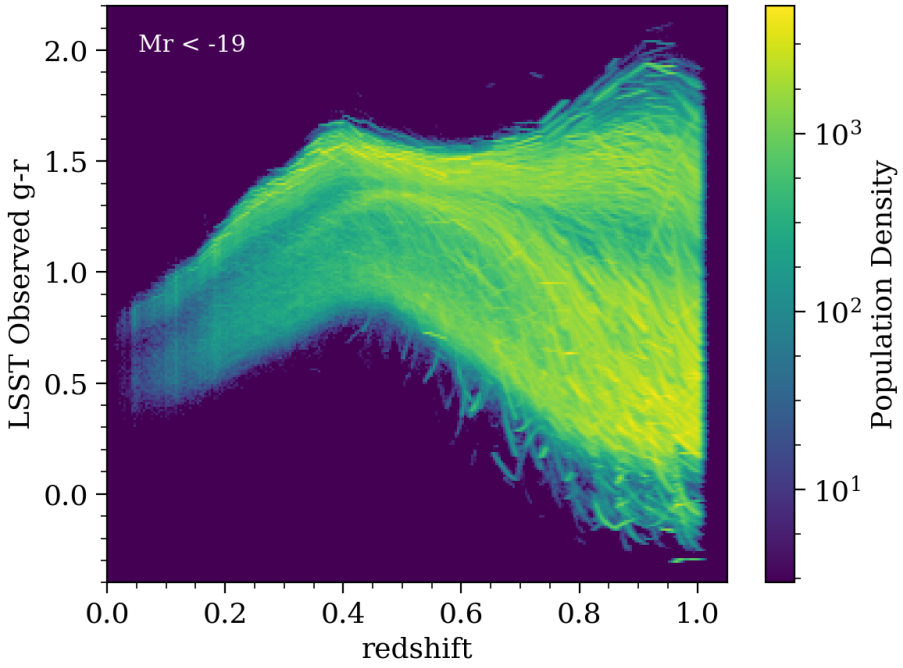

Figure 3.10: Observed $g - r$ color distribution of cosmoDC2 galaxies with $M_{\rm r} < -19$ as a function of redshift, up to $z = 1$. The smooth distribution is obtained through the interpolation procedure described in Sec. 3.5.3. Filamentary structures visible in this figure arise from repeated sampling of library galaxies in sparsely populated color-magnitude space.

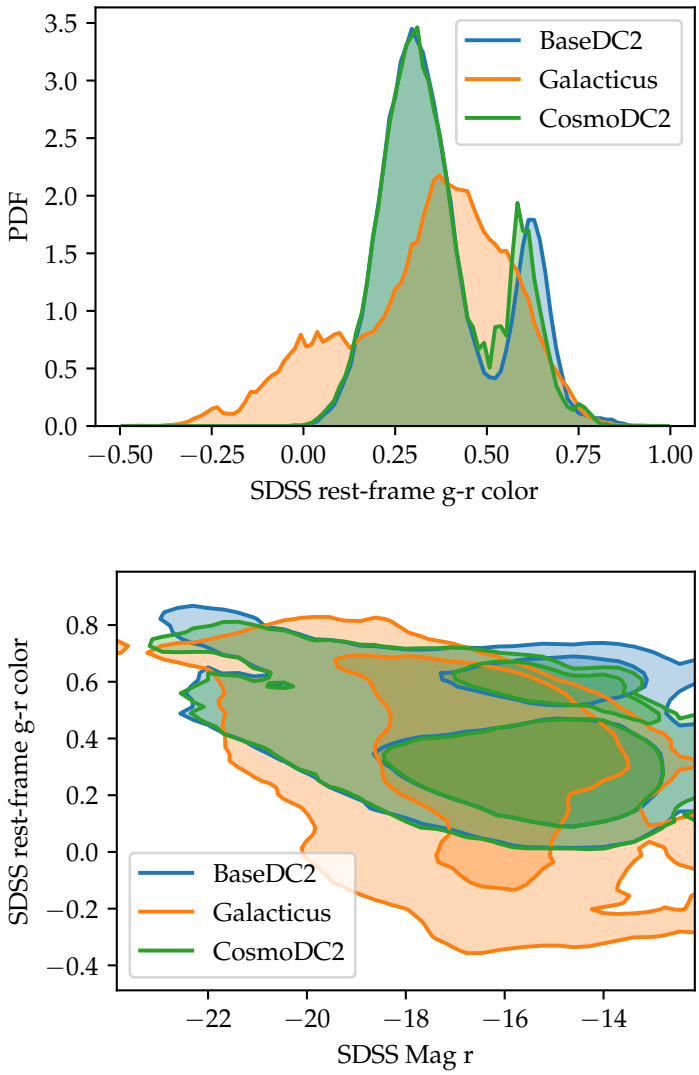

Figure 3.11: Distributions of rest-frame color (top panel) and color-magnitude (bottom panel) of $0.5 < z < 0.54$ galaxies in baseDC2, the Galacticus library and cosmoDC2. The top panel shows that the color distribution of baseDC2 is well recovered in cosmoDC2 by selecting specific Galacticus galaxies, with only a small discrepancy for red galaxies. The contours shown in the bottom panel encompass 75% and 99% of the galaxy populations. For bright galaxies, the luminosity adjustment prioritizes the match with baseDC2 colors while the luminosity rescaling extends the coverage of library galaxies (See Sec. 3.5.3). The brightest and reddest galaxies do not reproduce the baseDC2 colors exactly because there are few library galaxies with similar colors. The luminosity adjustment is applied for faint galaxies, so for faint red galaxies the match-up procedure compromises galaxy color in order to match luminosity better.

3.5.4 Additional Modeling

The final code module of Figure 3.9 shows the additional empirical modeling performed after the Galacticus library match-up. This modeling, which is required to meet validation criteria, relies on information obtained from the match-up. Galaxy profiles are assumed to be given by an $n = 1$ Sersic (exponential) profile for the disk component of each galaxy and an $n = 4$ Sersic profile for the bulge component.

Disk and Bulge Size

The stellar mass of each galaxy in the Galacticus library is divided into disk and bulge components, from which we determine B/T, the ratio of the stellar mass in the bulge to the total stellar mass. We use this quantity to model galaxy sizes in order to meet DESCQA validation criteria for the size-luminosity relation.

We separately model the size-luminosity relation for disk and bulge components, in both cases using the functional form introduced in [156] to obtain R_{50}, the Petrosian half-light radius:

$$R_{50} = \gamma L^{\alpha}(1 + L)^{\beta-\alpha}. \tag{3.14}$$

Here $L = 10^{-0.4(M_r - M_r^0)}$, and γ, α, β, and M_r^0 are fitting parameters that depend on the classification scheme used to determine the subtype of the fitted galaxy sample. [156] characterize the size-luminosity relation for several classification schemes including a morphological classification into either elliptical or spiral types and a classification that uses the B/T value. We choose the latter scheme to obtain values of the fit parameters in Eq. 3.14. Specifically, for the disk and bulge components of cosmoDC2 galaxies at $z = 0$, we use the parameters in Table 1 of [156] pertaining to SDSS galaxies with B/T < 0.5 and B/T > 0.5, respectively.

We incorporate redshift dependence in the galaxy-size distribution by setting $\gamma = \gamma(z)$. We choose the functional form of $\gamma(z)$ to be a sigmoid function, as defined by Eq. 3.2,

which regulates the redshift evolution such that size decreases with increasing redshift. The parameters in Eq. 3.2 are chosen such that we recover the [156] parameters at $z = 0$, the sizes of both disks and bulges are reduced by a factor of two at $z = 1$ and k is set to 4.

Black Hole Mass and Accretion Rate

A black hole resides at the center of every cosmoDC2 galaxy. The properties of these black holes can be used to model time-varying active galactic nuclei. For the mass of the black hole, $M_\bullet$, we adopt the power-law scaling relation reported in [85]:

$$M_\bullet = 0.0049 M_{\rm bulge} \left(M_{\rm bulge}/M_0\right)^\alpha , \tag{3.15}$$

where $\alpha = 0.15$ and $M_0 = 10^{11}\mathrm{M}_\odot$.

For the mass accretion rate of the black hole, $\mathrm{d}M_\bullet/\mathrm{dt}$, we define $\mathrm{dlog}M_\bullet/\mathrm{dt} \equiv \lambda_{\rm edd}\dot{M}_{\rm edd}$, where $\lambda_{\rm edd}$ is the Eddington ratio and we assume an Eddington rate of $\dot{M}_{\rm edd} = 0.022\mathrm{M}_\odot/\mathrm{Myr}/M_\bullet$. We model $\lambda_{\rm edd}$ according to the redshift-dependent probability distribution reported in [6]:

$$P(\lambda_{\rm edd}|z) = A\frac{1+z}{(1+z_0)^{\gamma_{\rm z}}}\lambda_{\rm edd}^{\gamma_{\rm e}}, \tag{3.16}$$

where $A = 0.00071$, $\gamma_{\rm z} = 3.47$, $\gamma_{\rm e} = -0.65$, and $z_0 = 0.6$. Thus the specific mass accretion rate in this model has no dependence upon black hole mass, though $\mathrm{d}M_\bullet/\mathrm{dt} \propto M_\bullet$.

In addition to the dependence of $\mathrm{d}M_\bullet/\mathrm{dt}$ on black hole mass and redshift, we use conditional abundance matching (CAM) to introduce correlations between $\mathrm{dlog}M_\bullet/\mathrm{dt}$ and sSFR, the specific star-formation rate of the galaxy. For each galaxy, we calculate the cumulative probability $\mathrm{P}(<\mathrm{sSFR}|\mathrm{M}_\star)$, and use the CAM implementation in Halotools [62] to non-parametrically correlate sSFR and $\lambda_{\rm edd}$, setting the correlation strength to 50%. In cosmoDC2, galaxy sSFR is tightly correlated with broadband color, such that active galaxies have bluer colors; thus our use of CAM in assigning $\lambda_{\rm edd}$ produces synthetic catalogs in

which galaxies with bluer broadband color host black holes that tend to be rapidly accreting mass.

Galaxy Ellipticity

The magnitude of the ellipticity of the galaxy is calculated as the r-band luminosity-weighted average of the disk and bulge ellipticities. These, in turn, are drawn from a Johnson SB probability distribution:

$$f(e, a, b) = \frac{b}{e(1-e)} \phi \left(a + b \log \frac{e}{1-e} \right), \tag{3.17}$$

where e is the ellipticity, ϕ is the normal probability distribution and a and b are model parameters. For the disk ellipticity we use constant values of $a = -0.4$ and $b = 0.7$. For the bulge component, we set $b = 1.0$, and a to a value that depends on the rest frame r-band magnitude as follows: $a = 0.6$ for $M_{\rm r} \leq -21$, a increases linearly with $M_{\rm r}$ for $-21 \leq M_{\rm r} \leq -17$ with slopes such that $a = 1.0$ for $M_{\rm r} = -19$ and $a = 1.6$ for $M_{\rm r} \geq -17$. The values of these parameters have been chosen to match the ellipticity distributions reported in [75], who studied the shapes of galaxy images from the COSMOS survey. The position angles, and thereby the components of the ellipticities, are chosen to correspond to galaxies with random orientations.

3.5.5 Galaxy Catalog Content

The catalog contains ∼2.26 billion galaxies in a 440 deg^2 field that spans $0 < z < 3$. Each galaxy has 551 listed properties. The catalog size is 5.2 TiB and the catalog is subdivided into 393 HDF5 files separated by redshift range and sky pixelization. In order to cover the image-simulation area that was selected for DC2, the catalog is built on the Nside=32 healpixels that are required to cover the area specified by the following (RA, Dec) coordinate

pairs (J2000): (71.46, −27.25), (52.25, −27.25), (73.79, −44.33), (49.42, −44.33).

Briefly, the modeled properties of galaxies in the cosmoDC2 lightcone include lensed and unlensed positions; stellar mass and black hole mass; a range of luminosity information, including broadband flux through both LSST and SDSS bands, as well as coarsely binned SEDs, fluxes supplied with and without dust, in observer and rest frames, with separate fluxes for each galaxy's disk and bulge component; shape information, including ellipticity, shear, magnification, convergence, and size; we also include information about the parent dark matter halo of each galaxy. Note that this extragalactic catalog does not include photometric errors as realistic determinations require survey-specific image simulations.

3.6 Selected Validation Results

In this section, we present selected validation results from cosmoDC2. These results have been chosen to be representative of important aspects of the output catalog and include redshift, magnitude and color properties of the galaxies, and tests of large-scale clustering and lensing distortions. Calibration of the model was driven by the competing demands of DC2 for model complexity and accuracy, codified by the set of validation requirements supplied by DESC science working groups. The results shown here are a subset of the full range of DESCQA validation tests (to be presented elsewhere) which encompass additional tests that were used to validate cosmoDC2.

3.6.1 Cumulative Number Counts as a Function of Magnitude

We compare the number counts of galaxies as a function of apparent magnitude with observational data. These provide an important empirical test of the realism of the model for the redshift-dependent luminosity function.

The dataset that we use for this comparison is the first data release [5] of the Deep layer of the Hyper-SuprimeCam (HSC) survey [4]. The reason for this choice is that the Deep layer

of the HSC survey is the deepest ongoing survey with an area exceeding a few square degrees. Hence, we perform a comparison that involves less extrapolation than with a shallower large-area survey, but without the cosmic variance uncertainty that would come from using a much smaller pencil beam survey. Using this data set, we measure the cumulative number counts of galaxies in the i-band down to 25th magnitude, then fit the result to a power law. This power-law extrapolation is justified by measurements from deep HST surveys, which appear to have power-law number counts down to at least 28th CModel magnitude [e.g., 12]. The results are shown in Fig. 3.12.

While this plot was used for validation, additional checks were made of the number counts against other surveys. For example, Subaru observations in the COSMOS field [28] yield a cumulative number density of 150 arcmin^{-2} or 5.4×10^5 deg^{-2} for $I < 26.5$, which is clearly quite close to the extrapolated HSC curve in Figure 3.12.

The grey-shaded band around the HSC extrapolated curve shows the region of $\pm 40\%$ tolerance in the cumulative number counts, the validation criterion set by the DESC science working groups. The magnitude range for validation region is indicated by the vertical shaded band which covers the range $24 < r < 27.5$. The maximum fractional difference between the cosmoDC2 data and the HSC fit in this region is 0.17 and hence cosmoDC2 passes this important validation test. As discussed in Sec. 3.5.1, the decrease in the number count with increasing i-band magnitude is an effect of the mass resolution of the simulation. Before the inclusion of the ultra-faint galaxies described in Sec. 3.5.1, cosmoDC2 failed this test.

One important caveat in interpreting this test is that galaxy surface-brightness effects have not been included. In order to determine the number of galaxies observed at a given magnitude, a full image simulation is needed to exclude the low surface-brightness galaxies [108]. We note that the size distributions for the cosmoDC2 disk and bulge components are in agreement with observational data, so we do not expect that there will be a system-

atic overestimate in the simulated surface brightnesses. Nevertheless, the results shown in Fig. 3.12 represent an upper limit on the galaxy number densities in cosmoDC2.

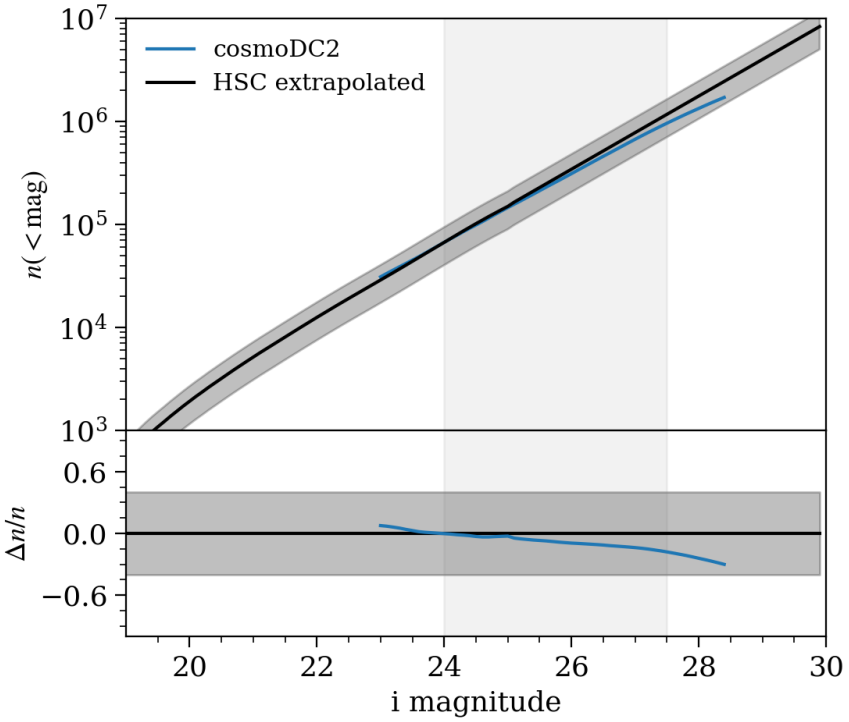

Figure 3.12: Top panel: Observed cumulative i-band number counts per square degree as a function of magnitude from cosmoDC2 (blue) and extrapolated from the HSC survey (black) (see text for more details). The grey shaded band shows a ±40% uncertainty around the HSC extrapolation. The vertical shaded region shows the magnitude range within which the two curves are compared. Bottom panel: Relative difference between the two curves in the top panel.

3.6.2 Redshift Distribution

The distribution of the number of galaxies as a function of redshift is another fundamental test of the realism of the synthetic catalog. This test is complementary to the cumulative-number-density test described above and provides a check on the shape of the redshift-dependent luminosity function. For this test, we compare the probability distribution for the number of galaxies as a function of the cosmological redshift with the observational data

from [106] and [34]. These observations are reported as parameterized fits to the $\mathrm{d}N/\mathrm{d}z$ distributions for a variety of magnitude-limited samples. The selection cuts for these samples were imposed on the CFHT R- and I-band magnitudes. In order to obtain comparisons that are as meaningful as possible, we construct magnitude-limited samples for the catalog data by applying the same cuts to the r- and i-band LSST filter magnitudes in the cosmoDC2 data. The error σ_i for each redshift bin is determined with a jackknife procedure: we estimate the errors due to cosmic variance by excluding subregions of RA and Dec from the catalog footprint with a k-means algorithm as implemented in the scikit-learn package.[8]. The elements of the covariance matrix are then given by

$$\sigma_{ij}^2 = \frac{N_{\mathrm{jack}} - 1}{N_{\mathrm{jack}}} \sum_k (\bar{N}_i - N_i^k) \cdot (\bar{N}_j - N_j^k), \tag{3.18}$$

where N_{jack} denotes the number of jackknife regions (chosen to be 30 for the present work) and $\bar{N}_i$ and N_i^k denote the numbers of galaxies in the i-th redshift bin for the full sample and for the sample with the k-th region excluded, respectively. The score for the test is computed by calculating the average of the reduced χ^2 between the catalog data and observed fit for each of the magnitude-limited samples. In practice, the computation of the covariances can be lengthy and we typically run this test over smaller sky areas of ~ 100 deg^2 at a time. For these smaller areas, the above covariance matrix is often not invertible due to instabilities in the off-diagonal matrix elements, so we use only the diagonal elements in the χ^2 computation. In Figure 3.13, we compare the redshift distributions for cosmoDC2 for three magnitude-limited samples having a sky area of ~ 60 deg^2 with the fits obtained from the DEEP2 data. The redshift distributions are in reasonable agreement with the DEEP2 fits, and we obtain χ^2/d.o.f values of 1.2, 0.75 and 0.73 for magnitude cuts of $r < 22$, $r < 23$ and $r < 24$, respectively. These values of χ^2/d.o.f may be somewhat under-estimated due to

8. https://scikit-learn.org/stable/

the aforementioned problem of obtaining the off-diagonal elements of the covariance matrix. We note that the validation criteria for cosmoDC2 do not specify a quantitative tolerance for χ^2/d.o.f.

3.6.3 Color Distributions

The color distributions of galaxies as a function of redshift and luminosity are critical properties that must be rendered with sufficient realism for many of the DC2 scientific use cases. For example, the accurate characterization of the biases and systematics of photometric redshift algorithms relies on the realism of synthetic color distributions. In Figure 3.14, we compare the color distributions for a magnitude-limited sample of cosmoDC2 galaxies with SDSS r-band < 17.7 and redshifts in the range $0.05 < z < 0.1$ to observational data from the SDSS main galaxy sample in SDSS DR13 [8]. The validation criteria for this test are that the bimodalities, peak locations and luminosity dependencies of the cosmoDC2 color distributions are broadly in agreement with those of the SDSS data.

Two features of the comparisons between cosmoDC2 and the SDSS data are worth noting. First, recall that the empirical model is tuned only for *rest-frame* r-band magnitude and $g-r$ and $r-i$ colors, whereas the quantities shown in the figure are *observer-frame* colors. The level of agreement that has been achieved relies solely on the correlations between the properties available for tuning and the full set of galaxy properties, as discussed in Sec. 3.2. Second, the distributions for cosmoDC2 peak at redder colors for $g-r$ and $i-z$ distributions and a bluer color for $u-g$. There are no empirical parameters available to alter $u-g$ and $i-z$ color distributions independently of the others, so it is difficult to achieve better agreement for the joint distribution of all colors by changing the empirical model. Other constraints, such as the requirement that the catalog have a prominent sample of red-sequence galaxies, impose further restrictions on the joint color distributions, and so the comparison shown here represents a compromise between multiple criteria supplied by the DESC science working

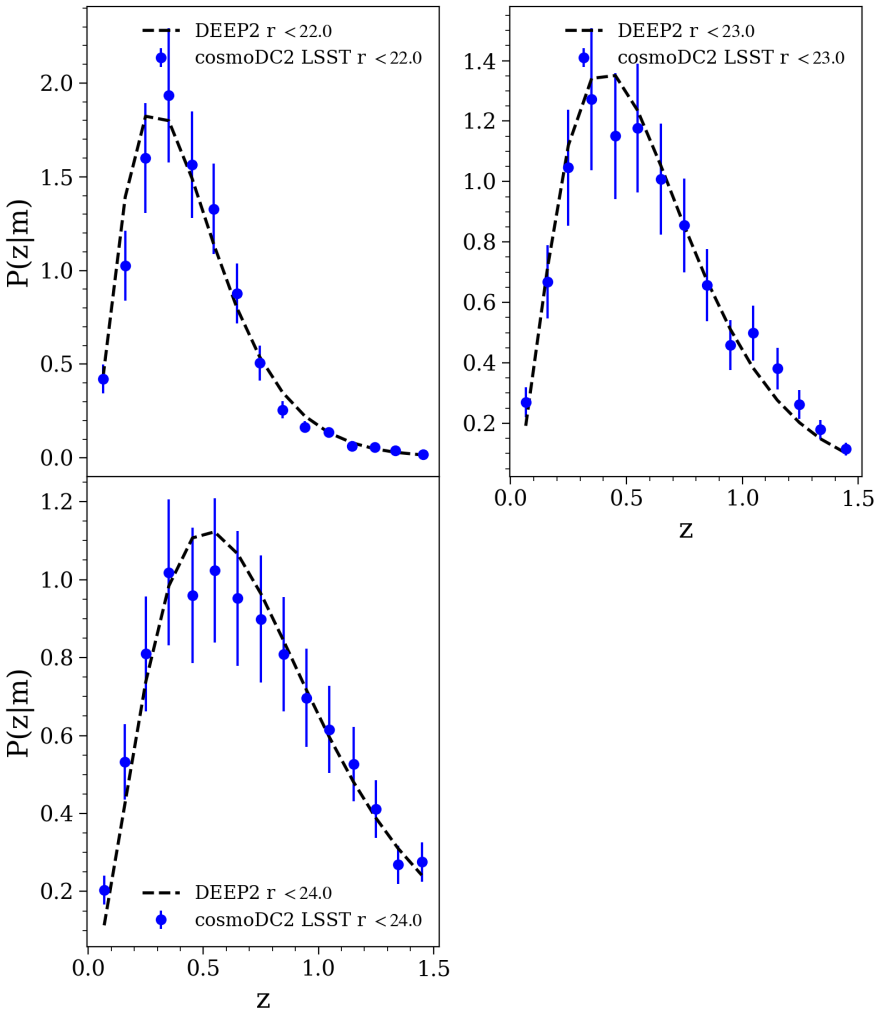

Figure 3.13: Redshift distribution of cosmoDC2 galaxies compared with fits to DEEP2 data for a redshift range of $0 < z < 1.5$ for three magnitude-limited samples. The selection cuts are LSST-$r < 22.0$, $r < 23.0$ and $r < 24.0$ as indicated in the legend.

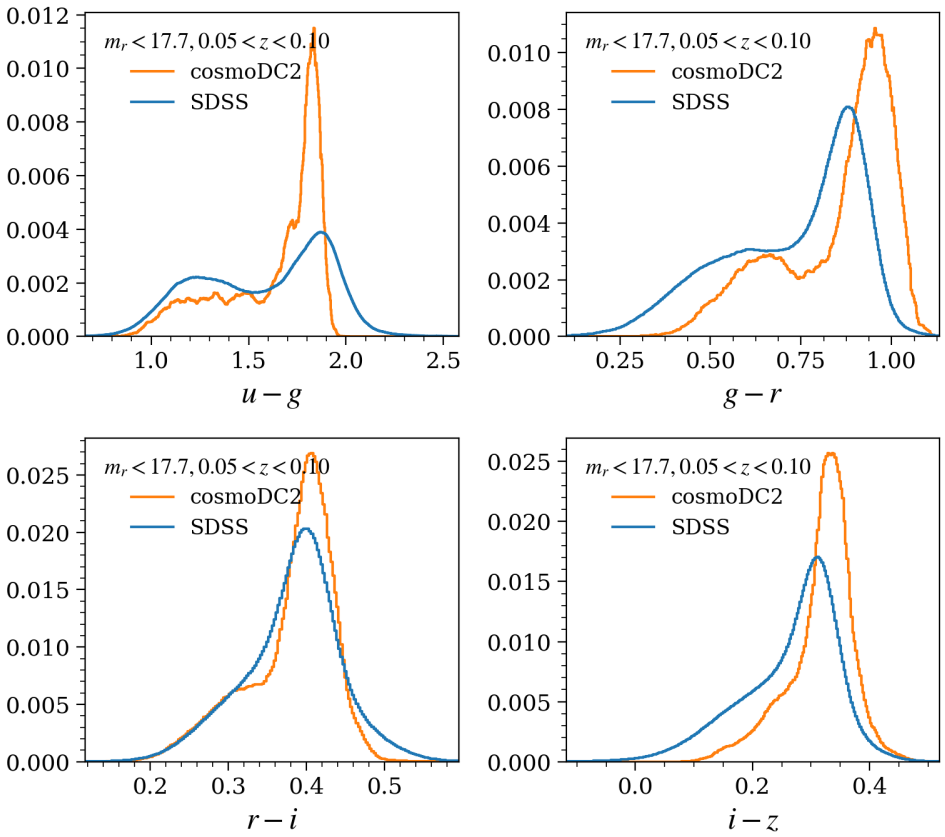

Figure 3.14: SDSS color distributions of cosmoDC2 galaxies compared with those of SDSS data. The cosmoDC2 galaxies comprise a magnitude- and redshift-limited sample with SDSS r-band magnitude < 17.7 and a redshift range of $0.05 < z < 0.1$. The colors shown in the figure are SDSS $u - g$, $g - r$, $r - i$ and $i - z$.

groups.

3.6.4 Two-Point Correlation Function

Galaxy clustering measurements provide a biased, but high signal-to-noise probe of the underlying matter field, and are commonly used both individually and in combination with lensing measurements to constrain the underlying matter field [e.g., 1]. We therefore validate the catalog for such measurements by comparing mock observations of the galaxy two-point correlation function to data. In this highlighted test we use the `TreeCorr` package [74] to mimic measurements from [153] of $w(\theta)$, the over-abundance of galaxy pairs at some angular

separation θ relative to a random distribution, given selected SDSS r-band magnitude cuts. This uses the estimator of [88]

$$w(\theta) = \frac{N_{dd} - 2N_{dr} + N_{rr}}{N_{rr}}, \tag{3.19}$$

where N is a normalized pair count within an angular separation bin centered at θ, and the subscripts d and r refer to the data or randomly generated galaxies used to compute pairs, respectively.

Figure 3.15 shows the shape and amplitude of the angular clustering of different galaxy samples selected by apparent r-band magnitude. The overall trends with scale, amplitude, and luminosity are the same as seen in SDSS; more pronounced deviations are visible in the 1-halo term for brighter galaxy samples, particularly the steeper slope of the clustering in the mock relative to SDSS. To estimate the variance within this patch we have included error bars using the jackknife procedure detailed in Sec. 3.6.2 with $N_{\rm jack} = 20$.

For this test, the validation criteria supplied by the DESC science working groups for the purposes of DC2 amounted to a check that the synthetic catalog and validation data were in reasonable agreement, and that the clustering strength scales with galaxy brightness in the expected fashion. Achieving higher precision agreement in the future would involve a much more costly optimization exercise, as well as a larger catalog area to reduce the effect of cosmic variance, and a more realistic estimation of statistical and systematic errors.

3.6.5 Galaxy-galaxy Lensing

Figure 3.16 displays the results of the galaxy-shear correlation test on the cosmoDC2 catalog. This computes the average tangential shear γ_t of a collection of background source galaxies at a given projected physical distance R from foreground lens galaxies, where

$$\gamma_t = -[\gamma_1 \cos(2\phi_c) + \gamma_2 \sin(2\phi_c)]. \tag{3.20}$$

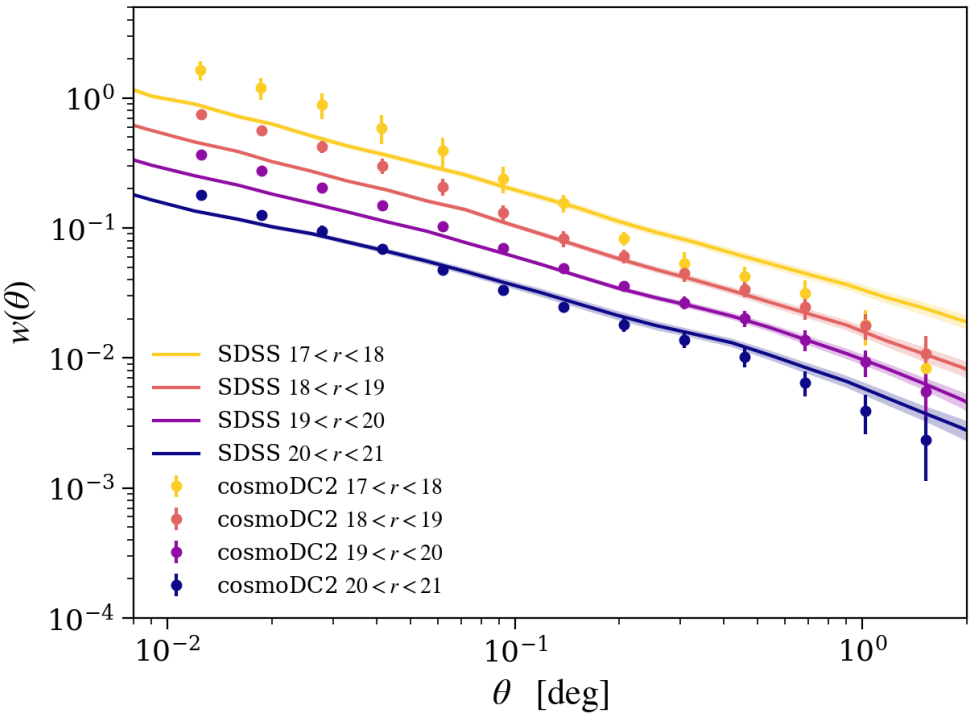

Figure 3.15: Validation test for the two-point correlation function, computed on the 440 sq. deg. cosmoDC2 catalog for cuts in SDSS r-band magnitude as given in the legend. The points correspond to measurements on the cosmoDC2 catalog, with error bars obtained through jackknife resampling, and solid lines to SDSS measurements from Table 2 of [153].

Here ϕ_c is the angle of the vector connecting the projected lens and source galaxies. These values are scaled by the geometry-dependent critical surface density $\Sigma_{\rm crit}$ to give the excess surface mass density $\Delta\Sigma(R)$. This test uses color, magnitude and redshift cuts designed to mimic the SDSS LOWZ measurement of [133], compute the excess surface mass density, and compare it to the observed values.

While no quantitative validation criteria were provided by DESC science working groups for this test, nonetheless the cosmoDC2 results show a qualitatively good fit to the SDSS data at large scales, and a realistic galaxy number density for the LOWZ sample; the falloff of the synthetic lensing signal on small scales is expected due to pixelization and noise in the shear maps as noted in Sec. 3.4. This test shows that the LOWZ-like population in the catalog is very similar to observations, both in number density and correlation with the underlying

tangential shear field, validating an important use case of the catalog, as well as confirming that the weak-lensing quantities and galaxy positions are appropriately correlated.

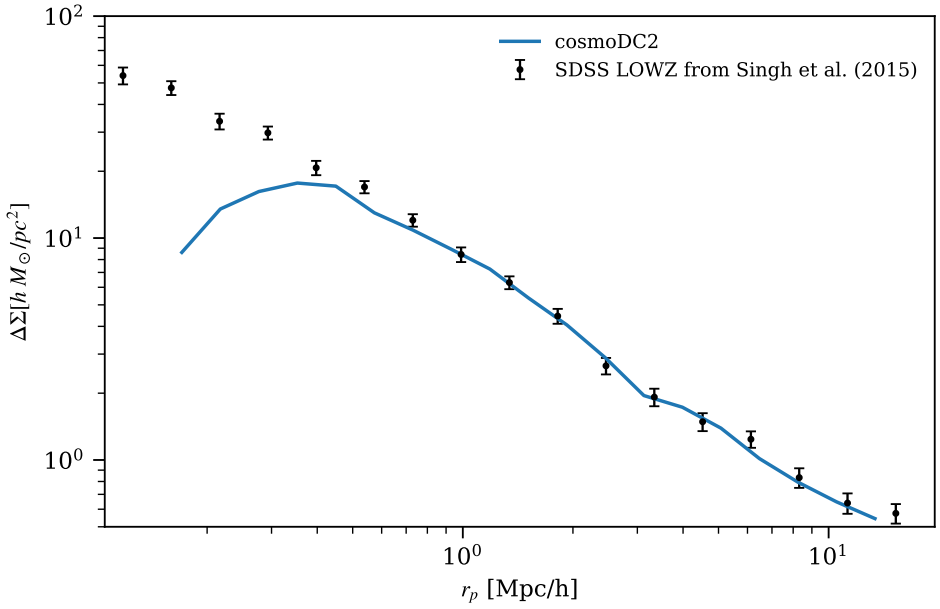

Figure 3.16: Validation test for galaxy-galaxy lensing, computed on the cosmoDC2 image area with cuts chosen to match those of the SDSS LOWZ sample of [133]. The test returns a total SDSS LOWZ galaxy number density of 58.8 per square degree for the cosmoDC2 image area, compared to the observed value of 57.0 per square degree from [119]. The solid blue line in the figure corresponds to the measurement performed on the cosmoDC2 catalog, and the black points to the measurement on the SDSS sample. At small scales the signal is smoothed due to lensing map resolution limits.

3.7 Summary and Future Directions

In this paper, we have described the production of cosmoDC2, a large synthetic sky catalog tailored to the needs of contemporary cosmological surveys. CosmoDC2 serves as the extra-galactic catalog used in the end-to-end image simulation pipeline developed as part of the LSST DESC Data Challenge 2. To produce the synthetic data, we have developed a new methodology for modeling the galaxy-halo connection that combines empirical and semi-analytic models (A. Hearin et al. 2019, in preparation), as well as a new software pipeline

for ray-tracing computations (P. Larsen et al. 2019, in preparation). The cosmoDC2 light-cone spans the redshift range $0 < z < 3$; each galaxy in cosmoDC2 has more than 500 attributes, including broadband flux through LSST filters, stellar mass, gravitational shear, separate coarse-grained SEDs for disk and bulge components of the image, half-light radius, black hole properties, and a range of other properties. The requirements for the statistical distributions of galaxies in cosmoDC2 were designed in close collaboration with the analysis working groups in DESC, using the DESCQA validation software to ensure appropriate realism.

Here, we have given a detailed account of cosmoDC2, the extragalactic catalog used in the DC2 image simulation, which spans 440 $\deg^2$. In the near future, with a straightforward scaling up of our existing production pipeline, we will generate a 5000 $\deg^2$ synthetic sky; this larger catalog will be used in a range of scientific analyses conducted by DESC science working groups.

Our effort to produce high-quality synthetic data for LSST DESC is ongoing, and several specific areas have already been targeted for further improvement. For example, the explicit halo-mass dependence of the SEDs of galaxies in the red sequence in cosmoDC2 galaxies is inconsistent with assumptions made by the redMaGiC LRG selection algorithm used in concert with the redMaPPer cluster finder. This is because the additional observer-frame color matching that is done for cosmoDC2 red-sequence cluster members introduces a difference in scatter between cosmoDC2 red-sequence field and cluster galaxies (whereas redMaGiC and redMaPPer assume that there is no difference). The effect is to lower the efficiency of the redMaGiC selections and is most pronounced at redshifts above $z \sim 0.8$. Recalibrating the color model after eliminating this explicit halo-mass dependence will improve the applicability of the catalog for studies of redMaPPer projection effects.

We also plan to improve the physical realism of the spatial distribution of cluster satellites. The current intra-halo distribution has a truncated Navarro-Frenk-White (NFW) profile [104]

that is inherited from Rockstar subhalos that are populated with UniverseMachine galaxies; in the future, the radial profiles will no longer be truncated and will follow an ellipsoidal NFW profile that is aligned with the large-scale tidal field.

In future releases of the model, we plan to extend our framework to utilize hydrodynamical simulations. For example, gas profiles from high-resolution hydrodynamical simulations can be painted onto group and cluster halos using the Galsampler technique, creating mock catalogs that could be used to study baryonic effects in the environments of massive halos.

Finally, we will soon provide high-resolution cutouts of the density and shear field surrounding cluster-mass halos. A comprehensive and up-to-date list of all planned improvements is available.[9]

Due to the evolving nature of the validation criteria of modern surveys, any method for generating synthetic cosmological data must be flexible enough to accommodate the demands imposed upon it by validation, as well as have sufficient computational efficiency to facilitate repeated iteration. Our hybrid method has a number of features that make it particularly suitable for producing a catalog that meets a series of potentially evolving validation criteria. The workflow is quite flexible in that the empirical models that drive the initial distributions of the limited set of galaxy properties can be easily changed. The pipelines can be run relatively quickly so that it is quite feasible to iterate on the empirical model to improve the agreement between the catalog and the observational data. The most time-consuming part of the modeling, namely running the SAM, need only be done once. If the resulting library spans the range of observed properties, the matching with empirically-modeled galaxies is straightforward. In the future, as more observational data become available from deeper surveys, the validation of synthetic catalogs will become much more demanding. It will be crucial to continue development of methodologies that efficiently and flexibly generate realistically complex synthetic cosmological data.

9. https://github.com/LSSTDESC/cosmodc2/issues

The cosmoDC2 catalog is publicly available from the NERSC website[10] as a collection of HDF5 files. The files are labeled by redshift range and healpixel numbers, which correspond to the Nside = 32 healpixels in the DC2 image-simulation area. The LSST DESC provides a Python package to facilitate user access to cosmoDC2. The cosmoDC2 code is publicly available.[11]

10. See https://portal.nersc.gov/project/lsst/cosmoDC2/_README.html

11. https://github.com/LSSTDESC/cosmodc2

CHAPTER 4

CONCLUSION

We have presented two topics in improving galaxy modeling. First, we introduced a new method of substructure tracking within halos to be used in conjunction with galaxy modeling. In the second half, we discussed the construction of cosmoDC2, which used the novel techniques of resampling a smaller catalog into a larger catalog and combination of empirical and semi-analytical models. The rest of the section is devoted to more detailed discussion of conclusions and future direction of work.

4.1 Halo Core Tracking

4.1.1 Primary Results

In Chapter 2 we introduced the concept of halo core tracking as a method of identifying and tracking substructure within dark matter halos. The goal is to enable better galaxy modeling without relying on subhalos. Subhalos and their merger trees are difficult to robustly define and are computationally expensive to run on cosmological sized simulations. Subhalo based methods also rely on modeling 'orphan' galaxies which are not associated with any subhalo to successfully reconstruct the clustering on small scales [27] which are dominated by the cluster environment. We validate halo core tracking by constructing a simple empirical galaxy model in the highly complicated galaxy cluster environment.

The process of constructing halos cores can be broken into two main steps. The first step identifies halo cores in individual simulation snapshots, either as a post processing step or *in-situ* of HACC simulations. The second step links cores across snaps shots using halo merger tress to construct a continuous view of halo cores evolution with time. We generated cores for a number of Argonne cosmological gravity only simulations, including the flagship Outer Rim simulation [63]. From these cores, we construct an empirical model of galaxies

to reproduce observation data from galaxy clusters.

Using two galaxy cluster catalogs, redMaPPer [127] and SPIDERS [32], to construct stacked profiles of galaxies from the SDSS DR15 [3] main galaxy sample. RedmaPPer clusters are optically selected while SPIDERS are selected by x-ray. We created a simple empirical galaxy model by utilizing properties of cores, chiefly the recorded infall mass of the core and the effective radius of the cores. The model was successful in reproducing the stacked profiles. The modeling the disruption of galaxies was fundamentally important to successfully reconstruct the galaxy profile.

4.1.2 Future Work

There are several directions halo core tracking can progress into. One direction is to improve the empirical model. While we have shown the success of the model to reconstruct the stacked profiles of galaxy clusters. The same model fails to simultaneously reconstruct the abundance of galaxies and the two point clustering of field galaxies to reproduce the two point clustering statistics has proven be more difficult. While the current model works well in clusters, a more complicated model may be required to capture the correct behavior across a wider range of environments.

Another major direction is the use of cores instead of subhalos for already established galaxy models. As mentioned in previous sections, subhalos have a number complications associated with them such as the computation expense and robustness of defining a subhalo. Cores can be substituted instead of subhalos in galaxy modeling codes. Work already has started on using substitute cores for subhalos in the L-Galaxies model [66]. In principle, any subhalo based galaxy model can have subhalos substituted, as long as the cores can correctly to capture the distribution and evolution of subhalos. The L-Galaxies model is tuned for subhalos. While it may be possible to re-tune the entire L-Galaxies to work with cores instead of subhalos, it would require significant effort and computational expense. A

more straight forward approach would be to leave L-Galaxies unmodified and instead use the cores to model subhalos. The core-modeled subhalos would be passed to L-Galaxies without any additional modifications to L-Galaxies.

4.2 Synthetic Galaxy Catalog

4.2.1 Primary Results

In Chapter 3 we present the methods and novel techniques used to construct the synthetic galaxy catalog, named cosmoDC2, for the Large Synoptic Survey Telescope Dark Energy Science Collaboration (LSST-DESC). The catalog spans 440 deg^2 and up to $z = 3$. The catalog is a hybridization of three separate simulations and three separate galaxy models. CosmoDC2 galaxies are founded from the Universe Machine synthetic galaxy catalog [13], which provides the abundance and stellar masses of galaxies within the MultiDark Planck 2 simulation (MDPL2) [82]. Galaxies from Universe Machine catalog are resampled into the Outer Rim [63] light cone by matching halos between Outer Rim and MDPL2, and transplanting galaxies on a halo basis. Careful additional modeling is applied to the galaxies in the Outer Rim light cone to provide luminosity and several key measures of galaxy color. The modeling of the luminosity and colors was done to pass the validations requirements set by multiple working groups with LSST-DESC. These improved galaxies were then matched to a library of galaxies generated by the Galacitcus SAM. The library was generated by running on the much smaller AlphaQ simulation. The matching was done to preserve the careful tuning of the empirical model. The library galaxies contained significantly more information on each galaxies that the earlier empirical model. Weak gravitational lensing shears and displacements for each galaxy are interpolated from ray tracing within the Outer Rim particle light cone. The final catalog is assembled as a set of files and python reader that is served on NERSC.

The entire process of generating and validating the catalog was a highly iterative process, requiring many individual catalogs to be generated quickly and examined. As the requirements evolved and issues within the catalog were discovered, the unique combination of the empirical model and the SAM library enable to rapid generation and iteration on a catalog that is highly complex. Without the hybridization of the empirical and SAM model, it would not have been possible to accomplish this task. Empirical models do not contain enough detail galaxy information for goals of the project and SAM catalogs cannot be quickly tuned and generated in a rapid manner on the scales required.

4.2.2 Future Work

There are several possible direction to improve the catalog. While the 440 deg^2 catalog is large enough for the image simulation for DC2, which is limited by the computation cost of running the image simulations and not by cosmoDC2, there are a number analyses that do not necessarily rely on the simulated images. Analyses, such as two point correlations and cluster finding, benefit greatly from an increase in volume. As the cosmoDC2 generation pipeline is already parallelized by breaking up the lightcone into spatial segments, scaling to larger sizes is trivial. We are planning to construct a 5000 deg^2 catalog which would include additional improvements.

Currently in cosmoDC2, the distribution of galaxies within a halo are completely randomly aligned relative to the shape of the halo. To improve the physical realism of the galaxy positions, we have replaced the spherical Navarro-Frenk-White (NFW) profiles with ellipsoidal NFW profiles that align with the large scale tidal field.

The weak lensing interpolation is not high enough resolution to accurately resolve accurate weak lensing shear for galaxies within clusters. We have started to create high resolution light cone cut outs around cluster sized halos to produce high resolution gravitational lensing maps. With smaller cut out light cones, we would not be computationally limited to run

such high resolution leak lensing ray tracing.

REFERENCES

[1] T. M. C. Abbott, F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen, A. Amara, J. Annis, J. Asorey, S. Avila, D. Bacon, E. Balbinot, M. Banerji, N. Banik, W. Barkhouse, M. Baumer, E. Baxter, K. Bechtol, M. R. Becker, A. Benoit-Lévy, B. A. Benson, G. M. Bernstein, E. Bertin, J. Blazek, S. L. Bridle, D. Brooks, D. Brout, E. Buckley-Geer, D. L. Burke, M. T. Busha, A. Campos, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, N. Chen, M. Childress, A. Choi, C. Conselice, R. Crittenden, M. Crocce, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, R. Das, T. M. Davis, C. Davis, J. De Vicente, D. L. DePoy, J. DeRose, S. Desai, H. T. Diehl, J. P. Dietrich, S. Dodelson, P. Doel, A. Drlica-Wagner, T. F. Eifler, A. E. Elliott, F. Elsner, J. Elvin-Poole, J. Estrada, A. E. Evrard, Y. Fang, E. Fernandez, A. Ferté, D. A. Finley, B. Flaugher, P. Fosalba, O. Friedrich, J. Frieman, J. García-Bellido, M. Garcia-Fernandez, M. Gatti, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, M. S. S. Gill, K. Glazebrook, D. A. Goldstein, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, S. Hamilton, W. G. Hartley, S. R. Hinton, K. Honscheid, B. Hoyle, D. Huterer, B. Jain, D. J. James, M. Jarvis, T. Jeltema, M. D. Johnson, M. W. G. Johnson, T. Kacprzak, S. Kent, A. G. Kim, A. King, D. Kirk, N. Kokron, A. Kovacs, E. Krause, C. Krawiec, A. Kremin, K. Kuehn, S. Kuhlmann, N. Kuropatkin, F. Lacasa, O. Lahav, T. S. Li, A. R. Liddle, C. Lidman, M. Lima, H. Lin, N. MacCrann, M. A. G. Maia, M. Makler, M. Manera, M. March, J. L. Marshall, P. Martini, R. G. McMahon, P. Melchior, F. Menanteau, R. Miquel, V. Miranda, D. Mudd, J. Muir, A. Möller, E. Neilsen, R. C. Nichol, B. Nord, P. Nugent, R. L. C. Ogando, A. Palmese, J. Peacock, H. V. Peiris, J. Peoples, W. J. Percival, D. Petravick, A. A. Plazas, A. Porredon, J. Prat, A. Pujol, M. M. Rau, A. Refregier, P. M. Ricker, N. Roe, R. P. Rollins, A. K. Romer, A. Roodman, R. Rosenfeld, A. J. Ross, E. Rozo, E. S. Rykoff, M. Sako, A. I. Salvador, S. Samuroff, C. Sánchez, E. Sanchez, B. Santiago, V. Scarpine, R. Schindler, D. Scolnic, L. F. Secco, S. Serrano, I. Sevilla-Noarbe, E. Sheldon, R. C. Smith, M. Smith, J. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, G. Tarle, D. Thomas, M. A. Troxel, D. L. Tucker, B. E. Tucker, S. A. Uddin, T. N. Varga, P. Vielzeuf, V. Vikram, A. K. Vivas, A. R. Walker, M. Wang, R. H. Wechsler, J. Weller, W. Wester, R. C. Wolf, B. Yanny, F. Yuan, A. Zenteno, B. Zhang, Y. Zhang, J. Zuntz, and Dark Energy Survey Collaboration. Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. *PhRvD*, 98:043526, Aug 2018.

[2] O. Agertz, A. V. Kravtsov, S. N. Leitner, and N. Y. Gnedin. Toward a Complete Accounting of Energy and Momentum from Stellar Feedback in Galaxy Formation Simulations. *ApJ*, 770:25, June 2013.

[3] D Aguado, Henrik Jönsson, Hu Zou, and et al. The fifteenth data release of the sloan digital sky surveys: First release of manga-derived quantities, data visualization tools, and stellar library. *Astrophysical Journal, Supplement Series*, 240(2), 2019. Export Date: 20 March 2019.

[4] H. Aihara, N. Arimoto, R. Armstrong, S. Arnouts, N. A. Bahcall, S. Bickerton, J. Bosch, K. Bundy, P. L. Capak, J. H. H. Chan, M. Chiba, J. Coupon, E. Egami, M. Enoki, F. Finet, H. Fujimori, S. Fujimoto, H. Furusawa, J. Furusawa, T. Goto, A. Goulding, J. P. Greco, J. E. Greene, J. E. Gunn, T. Hamana, Y. Harikane, Y. Hashimoto, T. Hattori, M. Hayashi, Y. Hayashi, K. G. Hełminiak, R. Higuchi, C. Hikage, P. T. P. Ho, B.-C. Hsieh, K. Huang, S. Huang, H. Ikeda, M. Imanishi, A. K. Inoue, K. Iwasawa, I. Iwata, A. T. Jaelani, H.-Y. Jian, Y. Kamata, H. Karoji, N. Kashikawa, N. Katayama, S. Kawanomoto, I. Kayo, J. Koda, M. Koike, T. Kojima, Y. Komiyama, A. Konno, S. Koshida, Y. Koyama, H. Kusakabe, A. Leauthaud, C.-H. Lee, L. Lin, Y.-T. Lin, R. H. Lupton, R. Mandelbaum, Y. Matsuoka, E. Medezinski, S. Mineo, S. Miyama, H. Miyatake, S. Miyazaki, R. Momose, A. More, S. More, Y. Moritani, T. J. Moriya, T. Morokuma, S. Mukae, R. Murata, H. Murayama, T. Nagao, F. Nakata, M. Niida, H. Niikura, A. J. Nishizawa, Y. Obuchi, M. Oguri, Y. Oishi, N. Okabe, S. Okamoto, Y. Okura, Y. Ono, M. Onodera, M. Onoue, K. Osato, M. Ouchi, P. A. Price, T.-S. Pyo, M. Sako, M. Sawicki, T. Shibuya, K. Shimasaku, A. Shimono, M. Shirasaki, J. D. Silverman, M. Simet, J. Speagle, D. N. Spergel, M. A. Strauss, Y. Sugahara, N. Sugiyama, Y. Suto, S. H. Suyu, N. Suzuki, P. J. Tait, M. Takada, T. Takata, N. Tamura, M. M. Tanaka, M. Tanaka, M. Tanaka, Y. Tanaka, T. Terai, Y. Terashima, Y. Toba, N. Tominaga, J. Toshikawa, E. L. Turner, T. Uchida, H. Uchiyama, K. Umetsu, F. Uraguchi, Y. Urata, T. Usuda, Y. Utsumi, S.-Y. Wang, W.-H. Wang, K. C. Wong, K. Yabe, Y. Yamada, H. Yamanoi, N. Yasuda, S. Yeh, A. Yonehara, and S. Yuma. The Hyper Suprime-Cam SSP Survey: Overview and survey design. *PASJ*, 70:S4, January 2018.

[5] H. Aihara, R. Armstrong, S. Bickerton, J. Bosch, J. Coupon, H. Furusawa, Y. Hayashi, H. Ikeda, Y. Kamata, H. Karoji, S. Kawanomoto, M. Koike, Y. Komiyama, D. Lang, R. H. Lupton, S. Mineo, H. Miyatake, S. Miyazaki, T. Morokuma, Y. Obuchi, Y. Oishi, Y. Okura, P. A. Price, T. Takata, M. M. Tanaka, M. Tanaka, Y. Tanaka, T. Uchida, F. Uraguchi, Y. Utsumi, S.-Y. Wang, Y. Yamada, H. Yamanoi, N. Yasuda, N. Arimoto, M. Chiba, F. Finet, H. Fujimori, S. Fujimoto, J. Furusawa, T. Goto, A. Goulding, J. E. Gunn, Y. Harikane, T. Hattori, M. Hayashi, K. G. Hełminiak, R. Higuchi, C. Hikage, P. T. P. Ho, B.-C. Hsieh, K. Huang, S. Huang, M. Imanishi, I. Iwata, A. T. Jaelani, H.-Y. Jian, N. Kashikawa, N. Katayama, T. Kojima, A. Konno, S. Koshida, H. Kusakabe, A. Leauthaud, C.-H. Lee, L. Lin, Y.-T. Lin, R. Mandelbaum, Y. Matsuoka, E. Medezinski, S. Miyama, R. Momose, A. More, S. More, S. Mukae, R. Murata, H. Murayama, T. Nagao, F. Nakata, M. Niida, H. Niikura, A. J. Nishizawa, M. Oguri, N. Okabe, Y. Ono, M. Onodera, M. Onoue, M. Ouchi, T.-S. Pyo, T. Shibuya, K. Shimasaku, M. Simet, J. Speagle, D. N. Spergel, M. A. Strauss, Y. Sugahara, N. Sugiyama, Y. Suto, N. Suzuki, P. J. Tait, M. Takada, T. Terai, Y. Toba, E. L. Turner, H. Uchiyama, K. Umetsu, Y. Urata, T. Usuda, S. Yeh, and S. Yuma. First data release of the Hyper Suprime-Cam Subaru Strategic Program. *PASJ*, 70:S8, January 2018.

[6] James Aird, Alison L. Coil, John Moustakas, Michael R. Blanton, Scott M. Burles, Richard J. Cool, Daniel J. Eisenstein, M. Stephen M. Smith, Kenneth C. Wong, and

Guangtun Zhu. PRIMUS: The Dependence of AGN Accretion on Host Stellar Mass and Color. *The Astrophysical Journal*, 746(1):90, Feb 2012.

[7] R. Akeson, L. Armus, E. Bachelet, V. Bailey, L. Bartusek, A. Bellini, D. Benford, D. Bennett, A. Bhattacharya, R. Bohlin, M. Boyer, V. Bozza, G. Bryden, S. Calchi Novati, K. Carpenter, S. Casertano, A. Choi, D. Content, P. Dayal, A. Dressler, O. Doré, S. M. Fall, X. Fan, X. Fang, A. Filippenko, S. Finkelstein, R. Foley, S. Furlanetto, J. Kalirai, B. S. Gaudi, K. Gilbert, J. Girard, K. Grady, J. Greene, P. Guhathakurta, C. Heinrich, S. Hemmati, D. Hendel, C. Henderson, T. Henning, C. Hirata, S. Ho, E. Huff, A. Hutter, R. Jansen, S. Jha, S. Johnson, D. Jones, J. Kasdin, P. Kelly, R. Kirshner, A. Koekemoer, J. Kruk, N. Lewis, B. Macintosh, P. Madau, S. Malhotra, K. Mandel, E. Massara, D. Masters, J. McEnery, K. McQuinn, P. Melchior, M. Melton, B. Mennesson, M. Peeples, M. Penny, S. Perlmutter, A. Pisani, A. Plazas, R. Poleski, M. Postman, C. Ranc, B. Rauscher, A. Rest, A. Roberge, B. Robertson, S. Rodney, J. Rhoads, J. Rhodes, R. Ryan, Jr., K. Sahu, D. Sand, D. Scolnic, A. Seth, Y. Shvartzvald, K. Siellez, A. Smith, D. Spergel, K. Stassun, R. Street, L.-G. Strolger, A. Szalay, J. Trauger, M. A. Troxel, M. Turnbull, R. van der Marel, A. von der Linden, Y. Wang, D. Weinberg, B. Williams, R. Windhorst, E. Wollack, H.-Y. Wu, J. Yee, and N. Zimmerman. The Wide Field Infrared Survey Telescope: 100 Hubbles for the 2020s. *arXiv e-prints*, February 2019.

[8] F. D. Albareti, C. Allende Prieto, A. Almeida, F. Anders, S. Anderson, B. H. Andrews, A. Aragón-Salamanca, M. Argudo-Fernández, E. Armengaud, E. Aubourg, and et al. The 13th Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-IV Survey Mapping Nearby Galaxies at Apache Point Observatory. *ApJS*, 233:25, December 2017.

[9] Astropy Collaboration, Thomas P. Robitaille, Erik J. Tollerud, Perry Greenfield, Michael Droettboom, Erik Bray, Tom Aldcroft, Matt Davis, Adam Ginsburg, Adrian M. Price-Whelan, Wolfgang E. Kerzendorf, Alexander Conley, Neil Crighton, Kyle Barbary, Demitri Muna, Henry Ferguson, Frédéric Grollier, Madhura M. Parikh, Prasanth H. Nair, Hans M. Unther, Christoph Deil, Julien Woillez, Simon Conseil, Roban Kramer, James E. H. Turner, Leo Singer, Ryan Fox, Benjamin A. Weaver, Victor Zabalza, Zachary I. Edwards, K. Azalee Bostroem, D. J. Burke, Andrew R. Casey, Steven M. Crawford, Nadia Dencheva, Justin Ely, Tim Jenness, Kathleen Labrie, Pey Lian Lim, Francesco Pierfederici, Andrew Pontzen, Andy Ptak, Brian Refsdal, Mathieu Servillat, and Ole Streicher. Astropy: A community Python package for astronomy. *A&A*, 558:A33, Oct 2013.

[10] Yannick M Bahé, Joop Schaye, David J Barnes, Claudio Dalla Vecchia, Scott T Kay, Richard G Bower, Henk Hoekstra, Sean L McGee, and Tom Theuns. Disruption of satellite galaxies in simulated groups and clusters: the roles of accretion time, baryons, and pre-processing. *Monthly Notices of the Royal Astronomical Society*, 485(2):2287–2311, may 2019.

[11] M. R. Becker. Connecting Galaxies with Halos Across Cosmic Time: Stellar mass assembly distribution modeling of galaxy statistics. *ArXiv e-prints*, July 2015.

[12] S. V. W. Beckwith, M. Stiavelli, A. M. Koekemoer, J. A. R. Caldwell, H. C. Ferguson, R. Hook, R. A. Lucas, L. E. Bergeron, M. Corbin, S. Jogee, N. Panagia, M. Robberto, P. Royle, R. S. Somerville, and M. Sosey. The Hubble Ultra Deep Field. *AJ*, 132:1729–1755, November 2006.

[13] P. Behroozi, R. Wechsler, A. Hearin, and C. Conroy. UniverseMachine: The Correlation between Galaxy Growth and Dark Matter Halo Assembly from z=0-10. *ArXiv:1806.07893*, June 2018.

[14] P. S. Behroozi, C. Conroy, and R. H. Wechsler. A comprehensive analysis of uncertainties affecting the stellar mass-halo mass relation for $0 < z < 4$. *ApJ*, 717:379–403, jul 2010.

[15] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu. The rockstar phase-space temporal halo finder and the velocity offsets of cluster cores. *ApJ*, 762:109, jan 2013.

[16] Peter S. Behroozi, Charlie Conroy, and Risa H. Wechsler. A comprehensive analysis of uncertainties affecting the stellar mass-halo mass relation for $0 < z < 4$. *Astrophysical Journal*, 717(1):379–403, jul 2010.

[17] A. J. Benson. Galacticus: A semi-analytic model of galaxy formation. *NewA*, 17:175–197, feb 2012.

[18] A. A. Berlind, D. H. Weinberg, A. J. Benson, C. M. Baugh, S. Cole, R. Davé, C. S. Frenk, A. Jenkins, N. Katz, and C. G. Lacey. The Halo Occupation Distribution and the Physics of Galaxy Formation. *ApJ*, 593:1–25, August 2003.

[19] Andreas A. Berlind and David H. Weinberg. The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass. *The Astrophysical Journal*, 575(2):587–616, aug 2002.

[20] M. Bernyk, D. J. Croton, C. Tonini, L. Hodkinson, A. H. Hassan, T. Garel, A. R. Duffy, S. J. Mutch, G. B. Poole, and S. Hegarty. The Theoretical Astrophysical Observatory: Cloud-based Mock Galaxy Catalogs. *ApJS*, 223:9, March 2016.

[21] Jérémy Blaizot, Yogesh Wadadekar, Bruno Guiderdoni, Stéphane T. Colombi, Emmanuel Bertin, François R. Bouchet, Julien E. G. Devriendt, and Steve Hatton. MoMaF: the Mock Map Facility. *MNRAS*, 360(1):159–175, Jun 2005.

[22] Chris Blake and Karl Glazebrook. Probing dark energy using baryonic oscillations in the galaxy power spectrum as a cosmological ruler. *The Astrophysical Journal*, 594(2):665–673, sep 2003.

[23] G. R. Blumenthal, S. M. Faber, R. Flores, and J. R. Primack. Contraction of dark matter galactic halos due to baryonic infall. *The Astrophysical Journal*, 301:27, feb 1986.

[24] Richard G. Bower, J. R. Lucey, and Richard S. Ellis. Precision photometry of early-type galaxies in the Coma and Virgo clusters: a test of the universality of the colour–magnitude relation – II. Analysis. *Monthly Notices of the Royal Astronomical Society*, 254(4):601–613, 02 1992.

[25] David Branch. Type ia supernovae and the hubble constant. *Annual Review of Astronomy and Astrophysics*, 36(1):17–55, 1998.

[26] A. M. Brooks and A. Zolotov. Why Baryons Matter: The Kinematics of Dwarf Spheroidal Satellites. *ApJ*, 786:87, May 2014.

[27] Duncan Campbell, Frank C van den Bosch, Nikhil Padmanabhan, Yao-Yuan Mao, Andrew R Zentner, Johannes U Lange, Fangzhou Jiang, and Antonio Villarreal. The galaxy clustering crisis in abundance matching. *Monthly Notices of the Royal Astronomical Society*, 477(1):359–383, 02 2018.

[28] P. Capak, H. Aussel, M. Ajiki, H. J. McCracken, B. Mobasher, N. Scoville, P. Shopbell, Y. Taniguchi, D. Thompson, S. Tribiano, S. Sasaki, A. W. Blain, M. Brusa, C. Carilli, A. Comastri, C. M. Carollo, P. Cassata, J. Colbert, R. S. Ellis, M. Elvis, M. Giavalisco, W. Green, L. Guzzo, G. Hasinger, O. Ilbert, C. Impey, K. Jahnke, J. Kartaltepe, J.-P. Kneib, J. Koda, A. Koekemoer, Y. Komiyama, A. Leauthaud, O. Le Fevre, S. Lilly, C. Liu, R. Massey, S. Miyazaki, T. Murayama, T. Nagao, J. A. Peacock, A. Pickles, C. Porciani, A. Renzini, J. Rhodes, M. Rich, M. Salvato, D. B. Sanders, C. Scarlata, D. Schiminovich, E. Schinnerer, M. Scodeggio, K. Sheth, Y. Shioya, L. A. M. Tasca, J. E. Taylor, L. Yan, and G. Zamorani. The First Release COSMOS Optical and Near-IR Data and Catalog. *ApJS*, 172:99–116, September 2007.

[29] JE Carlstrom, PAR Ade, KA Aird, BA Benson, LE Bleem, S Busetti, CL Chang, E Chauvin, H-M Cho, TM Crawford, et al. The 10 meter south pole telescope. *Publications of the Astronomical Society of the Pacific*, 123(903):568, 2011.

[30] A. Challinor, G. Chon, S. Colombi, E. Hivon, S. Prunet, and I. Szapudi. PolSpice: Spatially Inhomogeneous Correlation Estimator for Temperature and Polarisation. Astrophysics Source Code Library, September 2011.

[31] Hillary L. Child, Salman Habib, Katrin Heitmann, Nicholas Frontiere, Hal Finkel, Adrian Pope, and Vitali Morozov. Halo Profiles and the Concentration–Mass Relation for a ΛCDM Universe. *The Astrophysical Journal*, 859(1):55, may 2018.

[32] N. Clerc, A. Merloni, Y. Y. Zhang, A. Finoguenov, T. Dwelly, K. Nandra, C. Collins, K. Dawson, J. P. Kneib, Eduardo Rozo, E. Rykoff, T. Sadibekova, J. Brownstein, Y. T. Lin, J. Ridl, M. Salvato, A. Schwope, M. Steinmetz, H. J. Seo, and J. Tinker. Spiders:

The spectroscopic follow-up of x-ray-selected clusters of galaxies in sdss-iv. *Monthly Notices of the Royal Astronomical Society*, 463(4):4490–4515, 1 2016.

[33] J. D. Cohn. Some observational tests of a minimal galaxy formation model. *MNRAS*, 466:2718–2735, April 2017.

[34] A. L. Coil, J. A. Newman, N. Kaiser, M. Davis, C.-P. Ma, D. D. Kocevski, and D. C. Koo. Evolution and Color Dependence of the Galaxy Angular Correlation Function: 350,000 Galaxies in 5 Square Degrees. *ApJ*, 617:765–781, December 2004.

[35] C. Conroy, J. E. Gunn, and M. White. The Propagation of Uncertainties in Stellar Population Synthesis Modeling. I. The Relevance of Uncertain Aspects of Stellar Evolution and the Initial Mass Function to the Derived Physical Properties of Galaxies. *ApJ*, 699:486–506, July 2009.

[36] C. Conroy, R. H. Wechsler, and A. V. Kravtsov. Modeling luminosity-dependent galaxy clustering through cosmic time. *ApJ*, 647:201–214, aug 2006.

[37] D. J. Croton, V. Springel, S. D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. F. Navarro, and N. Yoshida. The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *MNRAS*, 365:11–28, January 2006.

[38] Richard H Cyburt, Brian D Fields, and Keith A Olive. Primordial nucleosynthesis in light of wmap. *Physics Letters B*, 567(3-4):227–234, 2003.

[39] S. Das and P. Bode. A Large Sky Simulation of the Gravitational Lensing of the Cosmic Microwave Background. *ApJ*, 682:1–13, July 2008.

[40] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *ApJ*, 292:371–394, may 1985.

[41] J. DeRose, R. H. Wechsler, M. R. Becker, M. T. Busha, E. S. Rykoff, N. MacCrann, B. Erickson, A. E. Evrard, A. Kravtsov, D. Gruen, S. Allam, S. Avila, S. Bridle, D. Brooks, E. Buckley-Geer, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, M. Crocce, L. N. da Costa, C. Davis, J. De Vicente, J. P. Dietrich, P. Doel, A. Drlica-Wagner, P. Fosalba, J. Frieman, J. Garcia-Bellido, G. Gutierrez, W. G. Hartley, D. L. Hollowood, B. Hoyle, D. J. James, E. Krause, K. Kuehn, N. Kuropatkin, M. Lima, M. A. G. Maia, F. Menanteau, C. J. Miller, R. Miquel, R. L. C. Ogando, A. Plazas Malagón, A. K. Romer, E. Sanchez, R. Schindler, S. Serrano, I. Sevilla-Noarbe, M. Smith, E. Suchyta, M. E. C. Swanson, G. Tarle, and V. Vikram. The Buzzard Flock: Dark Energy Survey Synthetic Sky Catalogs. *arXiv e-prints*, January 2019.

[42] DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E. Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger Ballester, Charles Baltay, Lucas Beaufore, Chris Bebek, Timothy C.

Beers, Eric F. Bell, José Luis Bernal, Robert Besuner, Florian Beutler, Chris Blake, Hannes Bleuler, Michael Blomqvist, Robert Blum, Adam S. Bolton, Cesar Briceno, David Brooks, Joel R. Brownstein, Elizabeth Buckley-Geer, Angela Burden, Etienne Burtin, Nicolas G. Busca, Robert N. Cahn, Yan-Chuan Cai, Laia Cardiel-Sas, Raymond G. Carlberg, Pierre-Henri Carton, Ricard Casas, Francisco J. Castand er, Jorge L. Cervantes-Cota, Todd M. Claybaugh, Madeline Close, Carl T. Coker, Shaun Cole, Johan Comparat, Andrew P. Cooper, M. C. Cousinou, Martin Crocce, Jean-Gabriel Cuby, Daniel P. Cunningham, Tamara M. Davis, Kyle S. Dawson, Axel de la Macorra, Juan De Vicente, Timothée Delubac, Mark Derwent, Arjun Dey, Govinda Dhungana, Zhejie Ding, Peter Doel, Yutong T. Duan, Anne Ealet, Jerry Edelstein, Sarah Eftekharzadeh, Daniel J. Eisenstein, Ann Elliott, Stéphanie Escoffier, Matthew Evatt, Parker Fagrelius, Xiaohui Fan, Kevin Fanning, Arya Farahi, Jay Farihi, Ginevra Favole, Yu Feng, Enrique Fernandez, Joseph R. Findlay, Douglas P. Finkbeiner, Michael J. Fitzpatrick, Brenna Flaugher, Samuel Flender, Andreu Font-Ribera, Jaime E. Forero-Romero, Pablo Fosalba, Carlos S. Frenk, Michele Fumagalli, Boris T. Gaensicke, Giuseppe Gallo, Juan Garcia-Bellido, Enrique Gaztanaga, Nicola Pietro Gentile Fusillo, Terry Gerard, Irena Gershkovich, Tommaso Giannantonio, Denis Gillet, Guillermo Gonzalez-de-Rivera, Violeta Gonzalez-Perez, Shelby Gott, Or Graur, Gaston Gutierrez, Julien Guy, Salman Habib, Henry Heetderks, Ian Heetderks, Katrin Heitmann, Wojciech A. Hellwing, David A. Herrera, Shirley Ho, Stephen Holland, Klaus Honscheid, Eric Huff, Timothy A. Hutchinson, Dragan Huterer, Ho Seong Hwang, Joseph Maria Illa Laguna, Yuzo Ishikawa, Dianna Jacobs, Niall Jeffrey, Patrick Jelinsky, Elise Jennings, Linhua Jiang, Jorge Jimenez, Jennifer Johnson, Richard Joyce, Eric Jullo, Stéphanie Juneau, Sami Kama, Armin Karcher, Sonia Karkar, Robert Kehoe, Noble Kennamer, Stephen Kent, Martin Kilbinger, Alex G. Kim, David Kirkby, Theodore Kisner, Ellie Kitanidis, Jean-Paul Kneib, Sergey Koposov, Eve Kovacs, Kazuya Koyama, Anthony Kremin, Richard Kron, Luzius Kronig, Andrea Kueter-Young, Cedric G. Lacey, Robin Lafever, Ofer Lahav, Andrew Lambert, Michael Lampton, Martin Land riau, Dustin Lang, Tod R. Lauer, Jean-Marc Le Goff, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Su-Jeong Lee, Daniela Leitner, Michael Lesser, Michael E. Levi, Benjamin L'Huillier, Baojiu Li, Ming Liang, Huan Lin, Eric Linder, Sarah R. Loebman, Zarija Lukić, Jun Ma, Niall MacCrann, Christophe Magneville, Laleh Makarem, Marc Manera, Christopher J. Manser, Robert Marshall, Paul Martini, Richard Massey, Thomas Matheson, Jeremy McCauley, Patrick McDonald, Ian D. McGreer, Aaron Meisner, Nigel Metcalfe, Timothy N. Miller, Ramon Miquel, John Moustakas, Adam Myers, Milind Naik, Jeffrey A. Newman, Robert C. Nichol, Andrina Nicola, Luiz Nicolati da Costa, Jundan Nie, Gustavo Niz, Peder Norberg, Brian Nord, Dara Norman, Peter Nugent, Thomas O'Brien, Minji Oh, Knut A. G. Olsen, Cristobal Padilla, Hamsa Padmanabhan, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Antonella Palmese, Daniel Pappalardo, Isabelle Pâris, Changbom Park, Anna Patej, John A. Peacock, Hiranya V. Peiris, Xiyan Peng, Will J. Percival, Sandrine Perruchot, Matthew M. Pieri, Richard Pogge, Jennifer E. Pollack, Claire Poppett, Francisco Prada, Abhishek Prakash, Ronald G.

Probst, David Rabinowitz, Anand Raichoor, Chang Hee Ree, Alexandre Refregier, Xavier Regal, Beth Reid, Kevin Reil, Mehdi Rezaie, Constance M. Rockosi, Natalie Roe, Samuel Ronayette, Aaron Roodman, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, Eduardo Rozo, Vanina Ruhlmann-Kleider, Eli S. Rykoff, Cristiano Sabiu, Lado Samushia, Eusebio Sanchez, Javier Sanchez, David J. Schlegel, Michael Schneider, Michael Schubnell, Aurélia Secroun, Uros Seljak, Hee-Jong Seo, Santiago Serrano, Arman Shafieloo, Huanyuan Shan, Ray Sharples, Michael J. Sholl, William V. Shourt, Joseph H. Silber, David R. Silva, Martin M. Sirk, Anze Slosar, Alex Smith, George F. Smoot, Debopam Som, Yong-Seon Song, David Sprayberry, Ryan Staten, Andy Stefanik, Gregory Tarle, Suk Sien Tie, Jeremy L. Tinker, Rita Tojeiro, Francisco Valdes, Octavio Valenzuela, Monica Valluri, Mariana Vargas-Magana, Licia Verde, Alistair R. Walker, Jiali Wang, Yuting Wang, Benjamin A. Weaver, Curtis Weaverdyck, Risa H. Wechsler, David H. Weinberg, Martin White, Qian Yang, Christophe Yeche, Tianmeng Zhang, Gong-Bo Zhao, Yi Zheng, Xu Zhou, Zhimin Zhou, Yaling Zhu, Hu Zou, and Ying Zu. The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv e-prints*, page arXiv:1611.00036, Oct 2016.

[43] K. Dolag. The Magneticum Simulations, from Galaxies to Galaxy Clusters. *IAU General Assembly*, 22:2250156, August 2015.

[44] Persis S. Drell, Thomas J. Loredo, and Ira Wasserman. Type ia supernovae, evolution, and the cosmological constant. *The Astrophysical Journal*, 530(2):593–617, feb 2000.

[45] G. Efstathiou, W. J. Sutherland, and S. J. Maddox. The cosmological constant and cold dark matter. *Nature*, 348:705–707, 1990.

[46] Albert Einstein. Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin*, pages 142–152, Jan 1917.

[47] Daniel J. Eisenstein, Wayne Hu, and Max Tegmark. Cosmic complementarity: [ITAL]h[/ITAL][TINF]0[/TINF] and ω[TINF][CLC][ITAL]m[/ITAL][/CLC][/TINF] from combining cosmic microwave background experiments and redshift surveys. *The Astrophysical Journal*, 504(2):L57–L60, sep 1998.

[48] J. D. Emberson, N. Frontiere, S. Habib, K. Heitmann, P. Larsen, H. Finkel, and A. Pope. The Borg Cube Simulation: Cosmological Hydrodynamics with CRK-SPH. *arXiv e-prints*, November 2018.

[49] G. Erfanianfar, A. Finoguenov, K. Furnell, P. Popesso, A. Biviano, S. Wuyts, C. A. Collins, M. Mirkazemi, J. Comparat, H. Khosroshahi, K. Nandra, R. Capasso, E. Rykoff, D. Wilman, A. Merloni, N. Clerc, M. Salvato, and J. I . Chitham. Stellar mass - halo mass relation for the brightest central galaxies of X-ray clusters since z˜0.65. *Astronomy & Astrophysics*, aug 2019.

[50] A. E. Evrard, T. J. MacFarland, H. M. P. Couchman, J. M. Colberg, N. Yoshida, S. D. M. White, A. Jenkins, C. S. Frenk, F. R. Pearce, J. A. Peacock, and P. A. Thomas. Galaxy Clusters in Hubble Volume Simulations: Cosmological Constraints from Sky Survey Populations. *ApJ*, 573:7–36, July 2002.

[51] G. J. Ferland, M. Chatzikos, F. Guzmán, M. L. Lykins, P. A. M. van Hoof, R. J. R. Williams, N. P. Abel, N. R. Badnell, F. P. Keenan, R. L. Porter, and P. C. Stancil. The 2017 Release Cloudy. *RMxAA*, 53:385–438, October 2017.

[52] A. Ferrara, S. Bianchi, A. Cimatti, and C. Giovanardi. An Atlas of Monte Carlo Models of Dust Extinction in Galaxies for Cosmological Applications. *ApJS*, 123:437–445, August 1999.

[53] P. Fosalba, E. Gaztañaga, F. J. Castander, and M. Crocce. The MICE Grand Challenge light-cone simulation - III. Galaxy lensing mocks from all-sky lensing maps. *MNRAS*, 447:1319–1332, February 2015.

[54] L. Gao, G. De Lucia, S. D.M. White, and A. Jenkins. Galaxies and subhaloes in ACDM galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 352(2):L1–L5, aug 2004.

[55] L. Gao, S. D.M. White, A. Jenkins, F. Stoehr, and V. Springel. The subhalo populations of ΛCDM dark haloes. *Monthly Notices of the Royal Astronomical Society*, 355(3):819–834, dec 2004.

[56] Michael D. Gladders and H. K. C. Yee. The Red-Sequence Cluster Survey. I. The Survey and Cluster Catalogs for Patches RCS 0926+37 and RCS 1327+29. *The Astrophysical Journal Supplement Series*, 157(1):1–29, mar 2005.

[57] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *ApJ*, 622:759–771, April 2005.

[58] Qi Guo, Simon White, Cheng Li, and Michael Boylan-Kolchin. How do galaxies populate dark matter haloes? *Monthly Notices of the Royal Astronomical Society*, 404(3):1111–1120, mar 2010.

[59] S. Habib, A. Pope, H. Finkel, N. Frontiere, K. Heitmann, D. Daniel, P. Fasel, V. Morozov, G. Zagaris, T. Peterka, V. Vishwanath, Z. Lukić, S. Sehrish, and W.-k. Liao. HACC: Simulating sky surveys on state-of-the-art supercomputing architectures. *NewA*, 42:49–65, January 2016.

[60] Salman Habib, Vitali Morozov, Nicholas Frontiere, Hal Finkel, Adrian Pope, and Katrin Heitmann. Hacc: Extreme scaling and performance across diverse architectures. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, SC '13, pages 6:1–6:10, New York, NY, USA, 2013. ACM.

[61] Sarah M. Hansen, Timothy A. McKay, Risa H. Wechsler, James Annis, Erin Scott Sheldon, and Amy Kimball. Measurement of Galaxy Cluster Sizes, Radial Profiles, and Luminosity Functions from SDSS Photometric Data. *The Astrophysical Journal*, 633(1):122–137, oct 2005.

[62] Andrew P. Hearin, Duncan Campbell, Erik Tollerud, Peter Behroozi, Benedikt Diemer, Nathan J. Goldbaum, Elise Jennings, Alexie Leauthaud, Yao-Yuan Mao, Surhud More, John Parejko, Manodeep Sinha, Brigitta Sipöcz, and Andrew Zentner. Forward Modeling of Large-scale Structure: An Open-source Approach with Halotools. *AJ*, 154:190, Nov 2017.

[63] Katrin Heitmann, Hal Finkel, Adrian Pope, Vitali Morozov, Nicholas Frontiere, Salman Habib, Esteban Rangel, Thomas Uram, Danila Korytov, Hillary Child, Samuel Flender, Joe Insley, and Silvio Rizzi. The Outer Rim Simulation: A Path to Many-Core Supercomputers, apr 2019.

[64] Heitmann et al. *arXiv e-prints*, 2019.

[65] B. M. B. Henriques, S. D. M. White, P. A. Thomas, R. Angulo, Q. Guo, G. Lemson, V. Springel, and R. Overzier. Galaxy formation in the Planck cosmology - I. Matching the observed evolution of star formation rates, colours and stellar masses. *MNRAS*, 451:2663–2680, August 2015.

[66] Bruno M. B. Henriques, Simon D. M. White, Peter A. Thomas, Raul Angulo, Qi Guo, Gerard Lemson, Volker Springel, and Roderik Overzier. Galaxy formation in the Planck cosmology – I. Matching the observed evolution of star formation rates, colours and stellar masses. *Monthly Notices of the Royal Astronomical Society*, 451(3):2663–2680, 06 2015.

[67] S. Hilbert, J. Hartlap, S. D. M. White, and P. Schneider. Ray-tracing through the Millennium Simulation: Born corrections and lens-lens coupling in cosmic shear and galaxy-galaxy lensing. *A&A*, 499:31–43, May 2009.

[68] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. NINE-YEARWILKINSON MICROWAVE ANISOTROPY PROBE(WMAP) OBSERVATIONS: COSMOLOGICAL PARAMETER RESULTS. *The Astrophysical Journal Supplement Series*, 208(2):19, sep 2013.

[69] P. F. Hopkins, A. Wetzel, D. Kereš, C.-A. Faucher-Giguère, E. Quataert, M. Boylan-Kolchin, N. Murray, C. C. Hayward, S. Garrison-Kimmel, C. Hummels, R. Feldmann, P. Torrey, X. Ma, D. Anglés-Alcázar, K.-Y. Su, M. Orr, D. Schmitz, I. Escala, R. Sanderson, M. Y. Grudić, Z. Hafen, J.-H. Kim, A. Fitts, J. S. Bullock, C. Wheeler, T. K. Chan, O. D. Elbert, and D. Narayanan. FIRE-2 simulations: physics versus numerics in galaxy formation. *MNRAS*, 480:800–863, October 2018.

[70] Wayne Hu and Zoltan Haiman. Redshifting rings of power. *Phys. Rev. D*, 68:063004, Sep 2003.

[71] S. Huang, A. Leauthaud, A. Hearin, P. Behroozi, C. Bradshaw, F. Ardila, J. Speagle, A. Tenenti, K. Bundy, J. Greene, C. Sifon, and N. Bahcall. Weak Lensing Reveals a Tight Connection Between Dark Matter Halo Mass and the Distribution of Stellar Mass in Massive Galaxies. *arXiv e-prints*, November 2018.

[72] Fabio Iocco, Gianpiero Mangano, Gennaro Miele, Ofelia Pisanti, and Pasquale D Serpico. Primordial nucleosynthesis: from precision cosmology to fundamental physics. *Physics Reports*, 472(1-6):1–76, 2009.

[73] Bhuvnesh Jain and Uroš Seljak. Cosmological model predictions for weak lensing: Linear and nonlinear regimes. *The Astrophysical Journal*, 484(2):560, 1997.

[74] Mike Jarvis. TreeCorr: Two-point correlation functions. Astrophysics Source Code Library, August 2015.

[75] B. Joachimi, E. Semboloni, P. E. Bett, J. Hartlap, S. Hilbert, H. Hoekstra, P. Schneider, and T. Schrabback. Intrinsic galaxy shapes and alignments - I. Measuring and modelling COSMOS intrinsic galaxy ellipticities. *MNRAS*, 431:477–492, May 2013.

[76] David E. Johnston, Erin S. Sheldon, Risa H. Wechsler, Eduardo Rozo, Benjamin P. Koester, Joshua A. Frieman, Timothy A. McKay, August E. Evrard, Matthew R. Becker, and James Annis. Cross-correlation Weak Lensing of SDSS galaxy Clusters II: Cluster Density Profiles and the Mass–Richness Relation. 2007.

[77] G. Kauffmann, J. M. Colberg, A. Diaferio, and S. D. M. White. Clustering of galaxies in a hierarchical universe - I. Methods and results at z=0. *MNRAS*, 303:188–206, February 1999.

[78] G. Kauffmann, S. D. M. White, and B. Guiderdoni. The formation and evolution of galaxies within merging dark matter haloes. *MNRAS*, 264:201, sep 1993.

[79] N. Khandai, T. Di Matteo, R. Croft, S. Wilkins, Y. Feng, E. Tucker, C. DeGraf, and M.-S. Liu. The MassiveBlack-II simulation: the evolution of haloes and galaxies to z $\sim$ 0. *MNRAS*, 450:1349–1374, June 2015.

[80] J.-h. Kim, T. Abel, O. Agertz, G. L. Bryan, D. Ceverino, C. Christensen, C. Conroy, A. Dekel, N. Y. Gnedin, N. J. Goldbaum, J. Guedes, O. Hahn, A. Hobbs, P. F. Hopkins, C. B. Hummels, F. Iannuzzi, D. Keres, A. Klypin, A. V. Kravtsov, M. R. Krumholz, M. Kuhlen, S. N. Leitner, P. Madau, L. Mayer, C. E. Moody, K. Nagamine, M. L. Norman, J. Onorbe, B. W. O'Shea, A. Pillepich, J. R. Primack, T. Quinn, J. I. Read, B. E. Robertson, M. Rocha, D. H. Rudd, S. Shen, B. D. Smith, A. S. Szalay, R. Teyssier, R. Thompson, K. Todoroki, M. J. Turk, J. W. Wadsley, J. H. Wise, A. Zolotov, and t. AGORA Collaboration29. The AGORA High-resolution Galaxy Simulations Comparison Project. *ApJS*, 210:14, January 2014.

[81] Manfred G. Kitzbichler and Simon D. M. White. The high redshift galaxy population in hierarchical galaxy formation models. *Mon. Not. Roy. Astron. Soc.*, 376:2–12, 2007.

[82] A. Klypin, G. Yepes, S. Gottlöber, F. Prada, and S. Heß. MultiDark simulations: the story of dark matter halo concentrations and density profiles. *MNRAS*, 457:4340–4359, April 2016.

[83] Alexander Knebe, Frazer R. Pearce, Hanni Lux, Yago Ascasibar, Peter Behroozi, Javier Casado, Christine Corbett Moran, Juerg Diemand, Klaus Dolag, Rosa Dominguez-Tenreiro, Pascal Elahi, Bridget Falck, Stefan Gottlöber, Jiaxin Han, Anatoly Klypin, Zarija Lukić, Michal Maciejewski, Cameron K. McBride, Manuel E. Merchán, Stuart I. Muldrew, Mark Neyrinck, Julian Onions, Susana Planelles, Doug Potter, Vicent Quilis, Yann Rasera, Paul M. Ricker, Fabrice Roy, Andrés N. Ruiz, Mario A. Sgró, Volker Springel, Joachim Stadel, P. M. Sutter, Dylan Tweed, and Marcel Zemp. Structure finding in cosmological simulations: the state of affairs. *Monthly Notices of the Royal Astronomical Society*, 435(2):1618–1658, oct 2013.

[84] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year wilkinson microwave anisotropy probe (wmap) observations: Cosmological interpretation. *ApJS*, 192:18, feb 2011.

[85] J. Kormendy and L. C. Ho. Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies. *ARA&A*, 51:511–653, August 2013.

[86] A. V. Kravtsov, A. A. Berlind, R. H. Wechsler, A. A. Klypin, S. Gottlöber, B. Allgood, and J. R. Primack. The dark side of the halo occupation distribution. *ApJ*, 609:35–49, jul 2004.

[87] C. G. Lacey, C. M. Baugh, C. S. Frenk, A. J. Benson, R. G. Bower, S. Cole, V. Gonzalez-Perez, J. C. Helly, C. D. P. Lagos, and P. D. Mitchell. A unified multi-wavelength model of galaxy formation. *MNRAS*, 462:3854–3911, November 2016.

[88] S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *ApJ*, 412:64–71, July 1993.

[89] R. Laureijs, J. Amiaux, S. Arduini, J. . Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, and et al. Euclid Definition Study Report. *arXiv e-prints*, October 2011.

[90] E. Lawrence, K. Heitmann, J. Kwan, A. Upadhye, D. Bingham, S. Habib, D. Higdon, A. Pope, H. Finkel, and N. Frontiere. The Mira-Titan Universe. II. Matter Power Spectrum Emulation. *ApJ*, 847:50, September 2017.

[91] A. Lewis. Lensed CMB simulation and parameter estimation. *PhRvD*, 71(8):083008, April 2005.

[92] Antony Lewis and Sarah Bridle. Cosmological parameters from cmb and other data: A monte carlo approach. *Phys. Rev. D*, 66:103511, Nov 2002.

[93] Eric V. Linder. Exploring the expansion history of the universe. *Phys. Rev. Lett.*, 90:091301, Mar 2003.

[94] LSST Dark Energy Science Collaboration. Large Synoptic Survey Telescope: Dark Energy Science Collaboration. *arXiv e-prints*, November 2012.

[95] LSST Science Collaboration, P. A. Abell, J. Allison, S. F. Anderson, J. R. Andrew, J. R. P. Angel, L. Armus, D. Arnett, S. J. Asztalos, T. S. Axelrod, and et al. LSST Science Book, Version 2.0. *arXiv e-prints*, December 2009.

[96] R. Mandelbaum. Weak Lensing for Precision Cosmology. *ARA&A*, 56:393–433, September 2018.

[97] Rachel Mandelbaum, Anže Slosar, Tobias Baldauf, Uroš Seljak, Christopher M Hirata, Reiko Nakajima, Reinabelle Reyes, and Robert E Smith. Cosmological parameter constraints from galaxy–galaxy lensing and galaxy clustering with the sdss dr7. *Monthly Notices of the Royal Astronomical Society*, 432(2):1544–1575, 2013.

[98] Yao-Yuan Mao, Eve Kovacs, Katrin Heitmann, Thomas D. Uram, Andrew J. Benson, Duncan Campbell, Sofía A. Cora, Joseph DeRose, Tiziana Di Matteo, Salman Habib, Andrew P. Hearin, J. Bryce Kalmbach, K. Simon Krughoff, François Lanusse, Zarija Lukić, Rachel Mandelbaum, Jeffrey A. Newman, Nelson Padilla, Enrique Paillas, Adrian Pope, Paul M. Ricker, Andrés N. Ruiz, Ananth Tenneti, Cristian A. Vega-Martínez, Risa H. Wechsler, Rongpu Zhou, Ying Zu, and LSST Dark Energy Science Collaboration. DESCQA: An Automated Validation Framework for Synthetic Sky Catalogs. *ApJS*, 234:36, Feb 2018.

[99] A. I. Merson, C. M. Baugh, J. C. Helly, V. Gonzalez-Perez, S. Cole, R. Bielby, P. Norberg, C. S. Frenk, A. J. Benson, R. G. Bower, C. G. Lacey, and C. d. P. Lagos. Lightcone mock catalogues from semi-analytic models of galaxy formation - I. Construction and application to the BzK colour selection. *MNRAS*, 429:556–578, February 2013.

[100] Alexander Merson, Yun Wang, Andrew Benson, Andreas Faisst, Daniel Masters, Alina Kiessling, and Jason Rhodes. Predicting Hα emission-line galaxy counts for future galaxy redshift surveys. *MNRAS*, 474:177–196, Feb 2018.

[101] B. P. Moster, T. Naab, and S. D. M. White. EMERGE - an empirical model for the formation of galaxies since z $\sim$ 10. *MNRAS*, 477:1822–1852, June 2018.

[102] B. P. Moster, R. S. Somerville, C. Maulbetsch, F. C. van den Bosch, A. V. Macciò, T. Naab, and L. Oser. Constraints on the Relationship between Stellar Mass and Halo Mass at Low and High Redshift. *ApJ*, 710:903–923, February 2010.

[103] Benjamin P. Moster, Rachel S. Somerville, Christian Maulbetsch, Frank C. Van Den Bosch, Andrea V. MacCiò, Thorsten Naab, and Ludwig Oser. Constraints on the relationship between stellar mass and halo mass at low and high redshift. *Astrophysical Journal*, 710(2):903–923, feb 2010.

[104] J. F. Navarro, C. S. Frenk, and S. D. M. White. The Structure of Cold Dark Matter Halos. *ApJ*, 462:563, May 1996.

[105] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. The Structure of Cold Dark Matter Halos. *Astrophysical Journal v.462, p.563*, 462:563, may 1996.

[106] J. A. Newman, M. C. Cooper, M. Davis, S. M. Faber, A. L. Coil, P. Guhathakurta, D. C. Koo, A. C. Phillips, C. Conroy, A. A. Dutton, D. P. Finkbeiner, B. F. Gerke, D. J. Rosario, B. J. Weiner, C. N. A. Willmer, R. Yan, J. J. Harker, S. A. Kassin, N. P. Konidaris, K. Lai, D. S. Madgwick, K. G. Noeske, G. D. Wirth, A. J. Connolly, N. Kaiser, E. N. Kirby, B. C. Lemaux, L. Lin, J. M. Lotz, G. A. Luppino, C. Marinoni, D. J. Matthews, A. Metevier, and R. P. Schiavon. The DEEP2 Galaxy Redshift Survey: Design, Observations, Data Reduction, and Redshifts. *ApJS*, 208:5, September 2013.

[107] Julian Onions, Alexander Knebe, Frazer R. Pearce, Stuart I. Muldrew, Hanni Lux, Steffen R. Knollmann, Yago Ascasibar, Peter Behroozi, Pascal Elahi, Jiaxin Han, Michal Maciejewski, Manuel E. Merchán, Mark Neyrinck, Andrés N. Ruiz, Mario A. Sgró, Volker Springel, and Dylan Tweed. Subhaloes going Notts: the subhalo-finder comparison project. *Monthly Notices of the Royal Astronomical Society*, 423(2):1200–1214, jun 2012.

[108] R. Overzier, G. Lemson, R. E. Angulo, E. Bertin, J. Blaizot, B. M. B. Henriques, G. D. Marleau, and S. D. M. White. The Millennium Run Observatory: first light. *MNRAS*, 428(1):778–803, Jan 2013.

[109] P. Panuzzo, A. Bressan, G. L. Granato, L. Silva, and L. Danese. Dust and nebular emission. I. Models for normal galaxies. *A&A*, 409:99–114, October 2003.

[110] J. A. Peacock and R. E. Smith. Halo occupation numbers and galaxy bias. *Monthly Notices of the Royal Astronomical Society*, 318(4):1144–1156, nov 2000.

[111] Laurence Perotto, Julien Lesgourgues, Steen Hannestad, Huitzu Tu, and Yvonne Y Y Wong. Probing cosmological parameters with the CMB: forecasts from monte carlo simulations. *Journal of Cosmology and Astroparticle Physics*, 2006(10):013–013, oct 2006.

[112] Planck Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L.

Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. jul 2018.

[113] Douglas Potter, Joachim Stadel, and Romain Teyssier. PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys. *Computational Astrophysics and Cosmology*, 4(1):2, May 2017.

[114] William H. Press and Paul Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *The Astrophysical Journal*, 187:425, feb 1974.

[115] E. Rangel, N. Li, S. Habib, T. Peterka, A. Agrawal, W. Liao, and A. Choudhary. Parallel dtfe surface density field reconstruction. In *2016 IEEE International Conference on Cluster Computing (CLUSTER)*, pages 30–39, Sept 2016.

[116] Esteban Rangel, Nicholas Frontiere, Salman Habib, Katrin Heitmann, Wei-keng Liao, Ankit Agrawal, and Alok Choudhary. Building halo merger trees from the q continuum simulation. In *2017 IEEE 24th International Conference on High Performance Computing (HiPC)*, pages 398–407. IEEE, 2017.

[117] Esteban Rangel, Nicholas Frontiere, Salman Habib, Katrin Heitmann, Wei Keng Liao, Ankit Agrawal, and Alok Choudhary. Building halo merger trees from the Q continuum

simulation. In *Proceedings - 24th IEEE International Conference on High Performance Computing, HiPC 2017*, volume 2017-Decem, pages 398–407. IEEE, dec 2018.

[118] R. M. Reddick, R. H. Wechsler, J. L. Tinker, and P. S. Behroozi. The connection between galaxies and dark matter structures in the local universe. *ApJ*, 771:30, jul 2013.

[119] B. Reid, S. Ho, N. Padmanabhan, W. J. Percival, J. Tinker, R. Tojeiro, M. White, D. J. Eisenstein, C. Maraston, A. J. Ross, A. G. Sánchez, D. Schlegel, E. Sheldon, M. A. Strauss, D. Thomas, D. Wake, F. Beutler, D. Bizyaev, A. S. Bolton, J. R. Brownstein, C.-H. Chuang, K. Dawson, P. Harding, F.-S. Kitaura, A. Leauthaud, K. Masters, C. K. McBride, S. More, M. D. Olmstead, D. Oravetz, S. E. Nuza, K. Pan, J. Parejko, J. Pforr, F. Prada, S. Rodríguez-Torres, S. Salazar-Albornoz, L. Samushia, D. P. Schneider, C. G. Scóccola, A. Simmons, and M. Vargas-Magana. SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues. *MNRAS*, 455:1553–1573, January 2016.

[120] Reinabelle Reyes, Rachel Mandelbaum, Uros Seljak, Tobias Baldauf, James E Gunn, Lucas Lombriser, and Robert E Smith. Confirmation of general relativity on large scales from weak lensing and galaxy velocities. *Nature*, 464(7286):256, 2010.

[121] Adam G. Riess, Louis-Gregory Strolger, John Tonry, Stefano Casertano, Henry C. Ferguson, Bahram Mobasher, Peter Challis, Alexei V. Filippenko, Saurabh Jha, Weidong Li, Ryan Chornock, Robert P. Kirshner, Bruno Leibundgut, Mark Dickinson, Mario Livio, Mauro Giavalisco, Charles C. Steidel, Txitxo Benitez, and Zlatan Tsvetanov. Type ia supernova discoveries at z<1 from the hubble space telescope: Evidence for past deceleration and constraints on dark energy evolution. *The Astrophysical Journal*, 607(2):665–687, jun 2004.

[122] A. Rodriguez-Puebla, J. R. Primack, V. Avila-Reese, and S. M. Faber. Constraining the galaxy-halo connection over the last 13.3 Gyr: star formation histories, galaxy mergers and structural properties. *MNRAS*, 470:651–687, September 2017.

[123] E. Rozo, E. S. Rykoff, M. Becker, R. M. Reddick, and R. H. Wechsler. redMaPPer - IV. Photometric membership identification of red cluster galaxies with 1 per cent precision. *MNRAS*, 453:38–52, October 2015.

[124] A. N. Ruiz, S. A. Cora, N. D. Padilla, M. J. Domínguez, C. A. Vega-Martínez, T. E. Tecce, Á. Orsi, Y. Yaryura, D. García Lambas, I. D. Gargiulo, and A. M. Muñoz Arancibia. Calibration of semi-analytic models of galaxy formation using particle swarm optimization. *ApJ*, 801:139, mar 2015.

[125] E. Rykoff et al. redMaPPer. I. Algorithm and SDSS DR8 Catalog. *MNRAS*, 785:104, April 2014.

[126] E. Rykoff et al. The RedMaPPer Galaxy Cluster Catalog From DES Science Verification Data. *ApJS*, 224:1, May 2016.

[127] E. S. Rykoff, E. Rozo, M. T. Busha, C. E. Cunha, A. Finoguenov, A. Evrard, J. Hao, B. P. Koester, A. Leauthaud, B. Nord, M. Pierre, R. Reddick, T. Sadibekova, E. S. Sheldon, and R. H. Wechsler. redMaPPer. i. ALGORITHM AND SDSS DR8 CATALOG. *The Astrophysical Journal*, 785(2):104, apr 2014.

[128] J. Schaye, R. A. Crain, R. G. Bower, M. Furlong, M. Schaller, T. Theuns, C. Dalla Vecchia, C. S. Frenk, I. G. McCarthy, J. C. Helly, A. Jenkins, Y. M. Rosas-Guevara, S. D. M. White, M. Baes, C. M. Booth, P. Camps, J. F. Navarro, Y. Qu, A. Rahmati, T. Sawala, P. A. Thomas, and J. Trayford. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *MNRAS*, 446:521–554, January 2015.

[129] Hee-Jong Seo and Daniel J. Eisenstein. Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys. *The Astrophysical Journal*, 598(2):720–740, dec 2003.

[130] Hee-Jong Seo and Daniel J. Eisenstein. Probing dark energy with baryonic acoustic oscillations from future large galaxy redshift surveys. *The Astrophysical Journal*, 598:720–740, December 1, 2003 2003. n/a.

[131] F. Shankar, A. Lapi, P. Salucci, G. De Zotti, and L. Danese. New relationships between galaxy properties and host halo mass, and the role of feedbacks in galaxy formation. *The Astrophysical Journal*, 643(1):14–25, may 2006.

[132] Melanie Simet, Tom McClintock, Rachel Mandelbaum, Eduardo Rozo, Eli Rykoff, Erin Sheldon, and Risa H. Wechsler. Weak lensing measurement of the mass–richness relation of SDSS redMaPPer clusters. *Mon. Not. Roy. Astron. Soc.*, 466(3):3103–3118, 2017.

[133] S. Singh, R. Mandelbaum, and S. More. Intrinsic alignments of SDSS-III BOSS LOWZ sample galaxies. *MNRAS*, 450:2195–2216, June 2015.

[134] Alex Smith, Shaun Cole, Carlton Baugh, Zheng Zheng, Raul Angulo, Peder Norberg, and Idit Zehavi. A Lightcone Catalogue from the Millennium-XXL Simulation. *Mon. Not. Roy. Astron. Soc.*, 470(4):4646–4661, 2017.

[135] R. S. Somerville and R. Davé. Physical Models of Galaxy Formation in a Cosmological Framework. *ARA&A*, 53:51–113, August 2015.

[136] R. S. Somerville and J. R. Primack. Semi-analytic modelling of galaxy formation: the local universe. *MNRAS*, 310:1087–1110, dec 1999.

[137] V. Springel, R. Pakmor, A. Pillepich, R. Weinberger, D. Nelson, L. Hernquist, M. Vogelsberger, S. Genel, P. Torrey, F. Marinacci, and J. Naiman. First results from the IllustrisTNG simulations: matter and galaxy clustering. *MNRAS*, 475:676–698, March 2018.

[138] Volker Springel, Simon D. M. White, Giuseppe Tormen, and Guinevere Kauffmann. Populating a cluster of galaxies – I. Results at z = 0. *Monthly Notices of the Royal Astronomical Society*, 328(3):726–750, 12 2001.

[139] Gary Steigman. Primordial nucleosynthesis in the precision cosmology era. *Annu. Rev. Nucl. Part. Sci.*, 57:463–491, 2007.

[140] N Suzuki, D Rubin, C Lidman, G Aldering, R Amanullah, K Barbary, LF Barrientos, J Botyanszki, M Brodwin, N Connolly, et al. The hubble space telescope cluster supernova survey. v. improving the dark-energy constraints above z> 1 and building an early-type-hosted supernova sample. *The Astrophysical Journal*, 746(1):85, 2012.

[141] Ariel G. Sánchez, C. G. Scóccola, A. J. Ross, W. Percival, M. Manera, F. Montesano, X. Mazzalay, A. J. Cuesta, D. J. Eisenstein, E. Kazin, C. K. McBride, K. Mehta, A. D. Montero-Dorta, N. Padmanabhan, F. Prada, J. A. Rubiño-Martín, R. Tojeiro, X. Xu, M. Vargas Magaña, E. Aubourg, N. A. Bahcall, S. Bailey, D. Bizyaev, A. S. Bolton, H. Brewington, J. Brinkmann, J. R. Brownstein, J. Richard Gott, J. C. Hamilton, S. Ho, K. Honscheid, A. Labatie, E. Malanushenko, V. Malanushenko, C. Maraston, D. Muna, R. C. Nichol, D. Oravetz, K. Pan, N. P. Ross, N. A. Roe, B. A. Reid, D. J. Schlegel, A. Shelden, D. P. Schneider, A. Simmons, R. Skibba, S. Snedden, D. Thomas, J. Tinker, D. A. Wake, B. A. Weaver, David H. Weinberg, Martin White, I. Zehavi, and G. Zhao. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological implications of the large-scale two-point correlation function. *Monthly Notices of the Royal Astronomical Society*, 425(1):415–437, 09 2012.

[142] M. Takada, R. S. Ellis, M. Chiba, J. E. Greene, H. Aihara, N. Arimoto, K. Bundy, J. Cohen, O. Doré, G. Graves, J. E. Gunn, T. Heckman, C. M. Hirata, P. Ho, J.-P. Kneib, O. Le Fèvre, L. Lin, S. More, H. Murayama, T. Nagao, M. Ouchi, M. Seiffert, J. D. Silverman, L. Sodré, D. N. Spergel, M. A. Strauss, H. Sugai, Y. Suto, H. Takami, and R. Wyse. Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph. *PASJ*, 66:R1, February 2014.

[143] R. Takahashi, T. Hamana, M. Shirasaki, T. Namikawa, T. Nishimichi, K. Osato, and K. Shiroyama. Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments. *ApJ*, 850:24, November 2017.

[144] Max Tegmark, Daniel J. Eisenstein, Michael A. Strauss, David H. Weinberg, Michael R. Blanton, Joshua A. Frieman, Masataka Fukugita, James E. Gunn, Andrew J. S. Hamilton, Gillian R. Knapp, Robert C. Nichol, Jeremiah P. Ostriker, Nikhil Padmanabhan, Will J. Percival, David J. Schlegel, Donald P. Schneider, Roman Scoccimarro, Uro š Seljak, Hee-Jong Seo, Molly Swanson, Alexander S. Szalay, Michael S. Vogeley, Jaiyul Yoo, Idit Zehavi, Kevork Abazajian, Scott F. Anderson, James Annis, Neta A. Bahcall, Bruce Bassett, Andreas Berlind, Jon Brinkmann, Tamás Budavari, Francisco Castander, Andrew Connolly, Istvan Csabai, Mamoru Doi, Douglas P. Finkbeiner, Bruce Gillespie, Karl Glazebrook, Gregory S. Hennessy, David W. Hogg, Željko Ivezić, Bhuvnesh Jain, David Johnston, Stephen Kent, Donald Q. Lamb, Brian C. Lee, Huan Lin,

Jon Loveday, Robert H. Lupton, Jeffrey A. Munn, Kaike Pan, Changbom Park, John Peoples, Jeffrey R. Pier, Adrian Pope, Michael Richmond, Constance Rockosi, Ryan Scranton, Ravi K. Sheth, Albert Stebbins, Christopher Stoughton, István Szapudi, Douglas L. Tucker, Daniel E. Vanden Berk, Brian Yanny, and Donald G. York. Cosmological constraints from the sdss luminous red galaxies. *Phys. Rev. D*, 74:123507, Dec 2006.

[145] Max Tegmark, Michael A. Strauss, Michael R. Blanton, Kevork Abazajian, Scott Dodelson, Havard Sandvik, Xiaomin Wang, David H. Weinberg, Idit Zehavi, Neta A. Bahcall, Fiona Hoyle, David Schlegel, Roman Scoccimarro, Michael S. Vogeley, Andreas Berlind, Tamás Budavari, Andrew Connolly, Daniel J. Eisenstein, Douglas Finkbeiner, Joshua A. Frieman, James E. Gunn, Lam Hui, Bhuvnesh Jain, David Johnston, Stephen Kent, Huan Lin, Reiko Nakajima, Robert C. Nichol, Jeremiah P. Ostriker, Adrian Pope, Ryan Scranton, Uro š Seljak, Ravi K. Sheth, Albert Stebbins, Alexander S. Szalay, István Szapudi, Yongzhong Xu, James Annis, J. Brinkmann, Scott Burles, Francisco J. Castander, Istvan Csabai, Jon Loveday, Mamoru Doi, Masataka Fukugita, Bruce Gillespie, Greg Hennessy, David W. Hogg, Željko Ivezić, Gillian R. Knapp, Don Q. Lamb, Brian C. Lee, Robert H. Lupton, Timothy A. McKay, Peter Kunszt, Jeffrey A. Munn, Liam O'Connell, John Peoples, Jeffrey R. Pier, Michael Richmond, Constance Rockosi, Donald P. Schneider, Christopher Stoughton, Douglas L. Tucker, Daniel E. Vanden Berk, Brian Yanny, and Donald G. York. Cosmological parameters from sdss and wmap. *Phys. Rev. D*, 69:103501, May 2004.

[146] A. Vale and J. P. Ostriker. Linking halo mass to galaxy luminosity. *MNRAS*, 353:189–200, sep 2004.

[147] M. P. van Daalen, B. M. B. Henriques, R. E. Angulo, and S. D. M. White. The galaxy correlation function as a constraint on galaxy formation physics. *MNRAS*, 458:934–949, May 2016.

[148] Frank C. van den Bosch. Dissecting the evolution of dark matter subhaloes in the bolshoi simulation. *Monthly Notices of the Royal Astronomical Society*, 468(1):885–909, jun 2017.

[149] Frank C van den Bosch, Go Ogiya, Oliver Hahn, and Andreas Burkert. Disruption of dark matter substructure: fact or fiction? *Monthly Notices of the Royal Astronomical Society*, 474(3):3043–3066, mar 2018.

[150] Terry P Walker, Gary Steigman, David N Schramm, Keith A Olive, and Ho-Shik Kang. Primordial nucleosynthesis redux. *The Big Bang And Other Explosions In Nuclear And Particle Astrophysics*, pages 43–61, 1991.

[151] L. Wang, A. A. Dutton, G. S. Stinson, A. V. Macciò, C. Penzo, X. Kang, B. W. Keller, and J. Wadsley. NIHAO project - I. Reproducing the inefficiency of galaxy formation across cosmic time with a large sample of cosmological hydrodynamical simulations. *MNRAS*, 454:83–94, November 2015.

[152] Limin Wang and Paul J Steinhardt. Cluster abundance constraints for cosmological models with a time-varying, spatially inhomogeneous energy component with negative pressure. *The Astrophysical Journal*, 508(2):483, 1998.

[153] Y. Wang, R. J. Brunner, and J. C. Dolence. The SDSS galaxy angular two-point correlation function. *MNRAS*, 432:1961–1979, Jul 2013.

[154] R. H. Wechsler and J. L. Tinker. The Connection Between Galaxies and Their Dark Matter Halos. *ARA&A*, 56:435–487, September 2018.

[155] S. D. M. White and M. J. Rees. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. *Monthly Notices of the Royal Astronomical Society*, 183(3):341–358, jul 1978.

[156] You-Cai Zhang and Xiao-Hu Yang. Size distribution of galaxies in SDSS DR7: weak dependence on halo environment. *Research in Astronomy and Astrophysics*, 19:006, Jan 2019.

[157] Z. Zheng, A. L. Coil, and I. Zehavi. Galaxy Evolution from Halo Occupation Distribution Modeling of DEEP2 and SDSS Galaxy Clustering. *ApJ*, 667:760–779, October 2007.

[158] Zheng Zheng, Andreas A. Berlind, David H. Weinberg, Andrew J. Benson, Carlton M. Baugh, Shaun Cole, Romeel Dave, Carlos S. Frenk, Neal Katz, and Cedric G. Lacey. Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies. *The Astrophysical Journal*, 633(2):791–809, nov 2005.