Cyberinfrastructure for Cosmology and Line-of-Sight Projection in Optical Galaxy Clusters

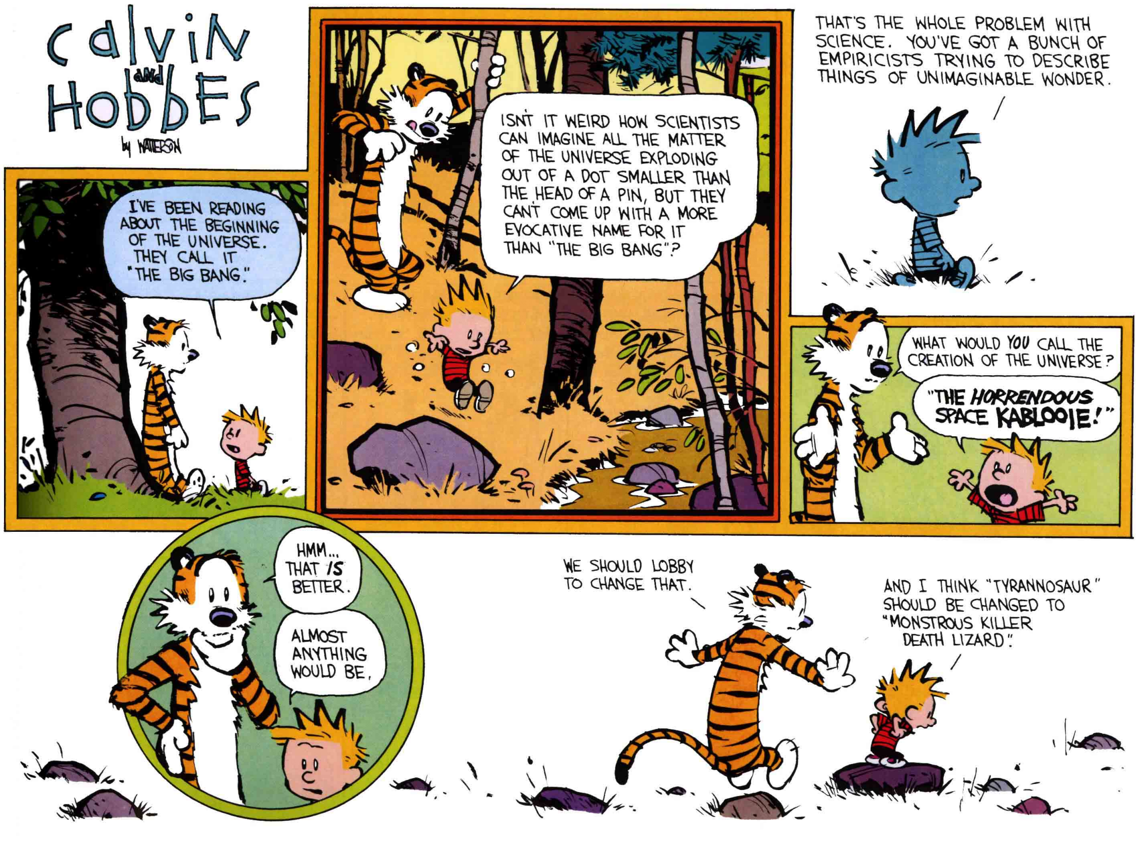
by

Brandon M. S. Erickson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in the University of Michigan
2013

Doctoral Committee:

Professor August Evrard, Chair
Professor Dante Eric Amidei
Associate Professor Dragan Huterer
Assistant Professor Christopher John Miller
Professor Gregory Tarlé


Calvin and Hobbes
by WATTERSON
I'VE BEEN READING ABOUT THE BEGINNING OF THE UNIVERSE. THEY CALL IT "THE BIG BANG."
ISN'T IT WEIRD HOW SCIENTISTS CAN IMAGINE ALL THE MATTER OF THE UNIVERSE EXPLODING OUT OF A DOT SMALLER THAN THE HEAD OF A PIN, BUT THEY CAN'T COME UP WITH A MORE EVOCATIVE NAME FOR IT THAN "THE BIG BANG"?
THAT'S THE WHOLE PROBLEM WITH SCIENCE. YOU'VE GOT A BUNCH OF EMPIRICISTS TRYING TO DESCRIBE THINGS OF UNIMAGINABLE WONDER.
WHAT WOULD YOU CALL THE CREATION OF THE UNIVERSE?
"THE HORRENDOUS SPACE KABLOOIE!"
HMM... THAT IS BETTER.
ALMOST ANYTHING WOULD BE.
WE SHOULD LOBBY TO CHANGE THAT.
AND I THINK "TYRANNOSAUR" SHOULD BE CHANGED TO "MONSTROUS KILLER DEATH LIZARD".

21 June 1992

ACKNOWLEDGMENTS

ecquid iis videretur mimum vitae commode transegisse?
εἰ δέ τι
ἔχοι καλῶς τὸ παίγνιον, κρότον δότε
καὶ πάντες ἡμᾶς μετὰ χαρᾶς προπέμψατε.
–Augustus

I have spent the last third of my life performing scientific research, and though completing a PhD is often the beginning of a scientific career, this is most likely the end of mine. As I move onto a new life, I would like to thank all of the people who inspired and made this work possible.

I would like to thank the people who shaped me before graduate school: family, friends and teachers.

First, my parents, who shepherded me through the first two-thirds of life, for without the job they did, nothing useful would have followed. Bill Watterson taught me at a very young age that science is fun, especially when it goes *boink*. Robert Henderson and Alan Albrandt started me down a true scientific path, without their charismatic high school teaching, I certainly wouldn't have majored in a pure science in college. At Clemson University, Gary Lickfield offered helpful advice and guidance that eventually led me to Physics and the University of Colorado. Ariel Paul, Manuel Castellanos-Beltran and Richard Sandberg offered a view of grad school that made me consider my choice to come to the University of Michigan very carefully. Their words and stories offered a perspective with significant value to me as I had my own experiences at Michigan. Thanks to Konrad Lehnert, my undergraduate research advisor who taught me what it takes to succeed in research. I have seen many people without such guidance fail to meet expectations in their research, and I surely would have followed down the same path if Konrad had not been such a great advisor. I would also like to thank Tom DeGrand, for holding his students to an ideal rather than to a curve in class. Not many survived, but those of us who did are far better people for it. He also provided the most useful philosophical statement about life that I have ever heard, and it is helpful to remember that statement when life seems unjust. Mike Ayers spent many an evening watching sports with me, which was most relaxing and restorative. Jim Holtsnider, lab partner extraordinaire, put me on the path when I might otherwise have been playing too many videogames.

I would like to thank the people who helped me at the University of Michigan, my friends, colleagues, advisor and committee.

The list of people who have made this task possible while in graduate school is quite long. I hope I do not leave anybody out who helped me along the way, for it is as Voltaire says;

> *Il en est des livres comme du feu de nos foyers; on va prendre ce feu chez son voisin, on lallume chez soi, on le communique à dautres, et il appartient à tous.*

Thanks to Sarah, Casey and Elizabeth for convincing me to come to Michigan; especially Sarah for being a basketball watching and Friday night dinner friend, until she moved on. Jim Allen was my first research advisor at Michigan, and he was very gracious when I decided I wanted to switch to a field that offered windows and sunlight in the workspace.

Particular thanks should go to my advisor Gus, because he was willing to take me on, even after I explained to him that because of health issues I would not be able to work the kinds of hours that are expected of many grad students. He also allowed me to explore my interests outside of research, spending time studying public policy. And he really reached when I told him I was going to move to Buckeye country because my wife had found a new job in Columbus, Ohio. I imagine many a less kind advisor would not tolerate their student living 200 miles away, and essentially never being seen in person.

I would also like to thank the committee who I hope will approve this document. When I was a novice graduate student, Dragan provided a lot of useful explanation about various equations, calculations and methods. Greg and Dan were both the faculty in charge of the undergraduate labs when I was lead GSI, and I enjoyed my time with them, especially in that they essentially left me unfettered to run the labs. Chris and I watched Tim Tebow lead the Broncos by the Steelers in the playoffs from Austin, Texas; something I will never forget, though he might like to.

Carlos Cunha really got me started working hard in this field, his no nonsense attitude while on the clock was highly instructive. Matt Becker provided a lot of help coding, though he probably doesn't know it, and at least seemed to take notice of my occasionally dry humor, even if he didn't appreciate it. Brian Nord was an excellent office mate in that he could provide help, and was always willing to discuss philosophy, economics and politics without popping a blood vessel whenever we couldn't seem to focus on the work we should have been doing. Jake Ketchum was also an excellent office mate, providing the soundtrack to which we all worked. Anbo bailed me out when I needed some temporary lodging, and for that I am grateful. Devin Harper, Dan Stuart and I spent a lot of time doing homework together, where many a yarn was spun when we just couldn't make progress. Tomasz Biesiadzinski and his need for coffee at regular intervals probably saved me from herniating a disk from being bent over the computer screen for too many consecutive hours. Jeeseon Song for knowing about astronomy, and for surviving. Ryo Saotome was like a collection of real life wikipedia links, I don't think we ever had a conversation where I did not learn something interesting and yet unrelated to anything I would have thought about otherwise. Cameron Gibelyou was always ready to toss a baseball. Tim Cohen taught me everything I know about particle physics. Andrew Morgan taught me to program. At various times, Chris, Chrissy, Stephanie and Kellie were tethers to a reality that exists outside a physics department. They also helped keep the paycheck going, and provided a lot of help when I was lead GSI. Also while a GSI, Dennis Allen was amazing. He makes it clear that some of the most interesting people on campus are outside the faculty. Eli and Eduardo helped quite a bit with the redMaPPer portion of this dissertation, and I especially appreciate Eli's humor. Sometimes scientists seem to take themselves too seriously, but this kind of work should be fun, and these guys keep it that way. Having all these friends floating on this ship of graduate school certainly kept cabin fever from setting in.

I would like to thank the people who helped me through life while in graduate school from outside the University.

Special thanks go to a few people I have never met. Mike & Mike got me going every morning while it seemed the rest of the world was still asleep, while Arthur, Douglas, Eoin and Joanne cleared my mind and put me to sleep every night.

Matt and Moria have my deepest gratitude for allowing me to sleep in their house when I needed to make a journey to Ann Arbor over the last year. My brother has been an excellent resource for the last several years. From his ability to explain why one might use a fast multipole method for a calculation, to rants about Bronco play-calling. We conversed on many special topics that were exclusive to us.

Lastly, and probably most importantly, my wife Kate. She has had to put up with being married to a graduate student for almost 5 years now. Work has many times pulled me away from things we might have done together, and she has never complained. She has offered much love and encouragement, and I don't know how I could have finished this without her, despite her assertion that "[she] didn't have any part in the writing or research, I just made sure you were fed."

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

2dFGRS Two-degree-Field Galaxy Redshift Survey

2LPTic Second-order Langrangian Perturbation Theory Initial Conditions

2MASS Two Micron All-Sky Survey

ACT Atacama Cosmology Telescope

ADDGALS Adding Density Determined GAlaxies to Lightcone Simulations

ALMA Atacama Large Millimeter Array

API Application Programming Interface

APM Automatic Plate Mover

BAO Baryon Acoustic Oscillation

BCC Blind Cosmology Challenge

CALCLENS Curved-sky grAvitational Lensing for Cosmological Light conE simulatioNS

CAMB Code for Anisotropies in the Microwave Background

CfA Harvard–Smithsonian Center for Astrophysics

CFCP Cluster Finder Comparison Project

CMB Cosmic Microwave Background

COBE *COsmic Background Explorer*

CTIO Cerro Tololo Inter-American Observatory

CXC *Chandra X-Ray Observatory*

DES Dark Energy Survey

DoE Department of Energy

ECSS Extended Collaborative Support Service

EIS ESO Imaging Survey

ESA European Space Agency

ESO European Southern Observatory

flops FLoating-point Operations Per Second

FoF Friends-of-Friends

FLRW Freidman–Lemaître–Robertson–Walker

GFac Generic Application Factory

HIPASS HI Parkes All Sky Survey

HOD Halo Occupation Distribution

HPC High Performance Computing

HST *Hubble Space Telescope*

KDK kick-drift-kick

LSS Large-Scale Structure

ΛCDM Lambda Cold Dark Matter

LRIS Low Resolution Imaging Spectrometer

LSS Large Scale Structure

LSST Large Synoptic Survey Telescope

NASA National Aeronautics and Space Administration

NSF National Science Foundation

NGC New General Catalogue

NGS-POSS National Geographic Society–Palomar Observatory Sky Survey

OGCE Open Gateways Computing Environment

Pan-STARRS Panoramic Survey Telescope and Rapid Response System

redMaPPer red-sequence Matched-filter Probabilistic Percolation

ROCKSTAR Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement

SDSS Sloan Digital Sky Survey

SHAM Sub-Halo Abundance Matching

SKA Square Kilometer Array

SLAC SLAC National Accelerator Laboratory, formerly Stanford Linear Accelerator Center

SNe Supernovae

SpARCS *Spitzer* Adaptation of the Red-sequence Cluster Survey

SPT South Pole Telescope

SED Spectral Energy Distribution

STScI Space Telescope Science Institute

SU Service Unit

SZ Sunyaev–Zel'dovich

TACC Texas Advanced Computing Center

VISTA Visible and Infrared Survey Telescope for Astronomy

WFI Wide Field Imager

WMAP *Wilkinson Microwave Anisotropy Probe*

XMM–Newton *X-ray Multi-Mirror Mission–Newton*

XSEDE eXtreme Science and Engineering Discovery Environment

ZA Zel'dovich Approximation

ABSTRACT

Cyberinfrastructure for Cosmology and Line-of-Sight Projection in Optical Galaxy Clusters

by

Brandon M. S. Erickson

Chair: August E. Evrard

Upcoming wide-area sky surveys such as the Dark Energy Survey (DES) offer the power to test the source of cosmic acceleration by placing extremely precise constraints on existing cosmological model parameters. These observational surveys will employ multiple tests based on statistical signatures of galaxies and larger-scale structures such as clusters of galaxies. Simulations of large-scale structure provide the means to maximize the power of sky survey tests by characterizing key sources of systematic uncertainties. This dissertation explores two subjects motivated by these facts. First, it explores how grid-aware cyberinfrastructure needs to be utilized in current and upcoming simulation campaigns that support large-area sky surveys. Second, it shows how line-of-sight projection plays into cosmological analysis based on galaxy cluster counts in the same wide-area sky surveys. In the first part, an Apache Airavata-enabled grid-aware application workflow for managing simulations is described. Results pertaining to efficiency in producing N-body simulations are reported. In the second part, bias in cosmological parameter estimates caused by incorrectly assuming a Gaussian (projection-free) mass–observable relation when the true relation is non-Gaussian due to projection is explored. Projection tends to skew the mass–observable relation of galaxy clusters by creating a small fraction of severely blended systems, those for which the measured observable property of a cluster is strongly boosted relative to the value of its primary host halo. A model motivated by optical cluster-finding applied to the Millennium Simulation is introduced for projection and Fisher information matrix parameter bias forecasts are produced for a DES-like sky survey. The model predicts significant biases in the dark energy density and equation of state parameters. The model additional predicts an increase in uncertainties in dark energy parameters to a factor of about two larger than forecast uncertainties. Additionally, new parameters used to characterize the model degrade uncertainties in the dark energy parameters. Motivated by this result, this dissertation also contains preliminary results for a new projection model meant to reduce bias in cluster analysis based on redMaPPer identified clusters for the DES.

CHAPTER 1

Introduction

Through the early- and mid- twentieth century, it was clear that the Universe is a staggeringly complex place, but cosmology was a data-starved science, where each newer and larger instrument heralded a leap forward in understanding. Then, in the late 1980s, desktop computers became powerful and ubiquitous, and a few important space telescopes were launched. By 1991, desktop computers had processor speeds measured in MegaHertz and storage space measured in MegaBytes. The *COsmic Background Explorer* (*COBE*) and *Hubble Space Telescope* (*HST*) were in orbit, and cosmology entered its golden age. High precision measurements combined with heavy numerical calculations combined to pick out particular parameter sets in Big Bang cosmology, and by the year 2000 cosmologists had zeroed in on the Lambda Cold Dark Matter (ΛCDM) cosmological model.

In the mid to late 1990s a large number of space telescopes were launched or in development, and digital ground telescopes were being built. Cluster computing centers could provide Teraflops computing systems to run cosmological simulations. With these computers, rigorous, predictive theoretical frameworks could be established and compared with precision observations. The *Wilkinson Microwave Anisotropy Probe* (*WMAP*) satellite and Sloan Digital Sky Survey (SDSS)—along with other observations—combined to tightly constrain parameters of the ΛCDM cosmology. By the middle of the first decade of the twenty-first century, cosmologists knew the three main components of the Universe—baryonic matter, dark matter and dark energy—and their proportions at the percent level. But, we understand very little about two of the three components.

An enormous number of NASA missions will end in the next few years, and cosmologists will be left with a smaller number of telescopes, but those instruments will produce enormous data sets. The data provided will be so precise that to continue to build knowledge of the Universe, precision modeling and detailed understanding of potential systemic errors will be necessary. Perhaps cosmology is entering a platinum age, where cosmologists will be data-rich and insight driven, as opposed to data-starved and driven by the need to build data sets. To fully harness the power of these data sets, theory and modeling will need to support from large simulation campaigns, so large that the traditional model of running a simulation on a cluster computer and then moving data around for analysis will no longer be viable. Cosmologists will need to utilize federated computing resources to manage large data and strengthen collaborative study.

This dissertation touches on each of the major issues facing cosmologists in this transition. I have made an attempt to using grid computing resources to improve simulation efficiency and data sharing. I have also explored one potential systematic effect in galaxy-cluster–based cosmological studies: line-of-sight projection.

Chapter 2 traces the history of cosmology, from paleolithic times to modern wide-area sky surveys. Unfortunately, the history is abbreviated and focuses only on developments in *Western* culture, which is not meant to be a slight to great *Eastern* astronomers, or downplay the importance of historical details, but due to the limited time and space for this document. By the end of Chapter 2, we should have an appreciation for how current and next-generation sky-surveys require precision modeling to fully exploit their large data sets. Chapter 3 presents the standard cosmological model (ΛCDM) and how we use current computing methods to generate artificial Universes on our desktops—or at a computing center—and how we are entering a phase in computational cosmology where grid computing methods are needed to manage the rising deluge of simulated data.

The next three chapters cover published work, or work that will be submitted for publication shortly after presenting this dissertation. The first research oriented chapter describes efforts to build a grid-aware cyberinfrastructure for cosmological simulations. Chapter 4 is an amalgamation of two published works—Erickson et al. (2012) and Erickson et al. (2013)—as well as some new results. This chapter describes my efforts to develop a workflow environment for efficiently running cosmological simulations, which could ideally one day serve as, or inform, a backend for a cosmological simulations gateway, as described in Chapter 3. The next two research chapters are related to precision modeling for wide-area sky surveys. Chapter 5 is a study of how line-of-sight projection in optical galaxy cluster observations can bias extracted cosmological parameters, published in Erickson, Cunha, and Evrard (2011). Chapter 6, partially motivated by the work found in Chapter 5, presents a method for accounting for line-of-sight projection with the optical cluster finder red-sequence Matched-filter Probabilistic Percolation (redMaPPer). This work, with sufficient expansion and improvement, will be submitted as Erickson, Rykoff, and Rozo (2013). Chapter 7 summarizes the results of the dissertation, and offers some thoughts for future work.

CHAPTER 2

From Naked Eye Astronomy to Precision Cosmology

Astronomy is by far the oldest of the natural sciences, representing the growth of humanity's concept of the world and its place in it. While other sciences have developed relatively recently in human history, astronomy has had at least an implicit place in humanity's mind from the earliest times. The lives of early humans were driven by the rising Sun, which brings light and warmth, and also panic and confusion from an occasional eclipse. Attempts to understand these daily events would have driven the earliest astronomers. Early astronomy was an amalgamation of religion, philosophy, astrology and astronomy; but, by the Neolithic Era, astronomy was truly scientific, evidenced by surviving megaliths such as Stonehenge (Challener, 1999; Hawkins, 1963, 1985; Hoyle, 1966a,b; Lockyer, 1906; Newham, 1966) that illustrate astronomical alignments. Humanity has continued to study the sky since Neolithic times, progressing from megaliths marking the solstice to precision cosmology, aiming to describe the history and structure of the Universe with detailed observational data and mathematical models.

In this chapter, I will provide an overview of the history of occidental astronomy and cosmology from Mesopotamian times up to today. With that historical context in place, I will give examples of observations that led to general relativity and the Friedmann–Lemaître world model. Next, I will chronicle efforts to investigate that model with improved technology led to the standard ΛCDM cosmology. Finally, I will describe modern cosmological surveys, culminating with a description of the Dark Energy Survey (DES), which the work described in this dissertation supports.

2.1 Foundations

Ancient astronomy is minimally documented, but we can imagine paleolithic humans looking out from the Lascaux caves at night wondering about their place in the cosmos. Megalithic structures show astronomical alignments, and so surely ancient civilizations were able to make some measurements of the night sky, though no written records survive to describe their conception of the Universe. Thanks to near invincible clay tablets, we know the Babylonians believed in a six-level Universe, with two heavens and two underworlds, and that Earth was a disk floating in the sea.

But, we have much more details about astronomy in classical times. The word cosmology comes to us from ancient greek—κόσμος (world) + -λογία (treating of)—by way of medieval latin—cosmōlogia—carrying with it thousands of years of history. Much of Greek cosmology fixated on solving a problem set by Plato (b. 427 BC). Plato believed heavenly bodies moved exclusively in circles, and he posed this problem to mathematicians: what circular motions, uniform and perfectly regular, are the essence behind the apparent irregular wanderings of the planets (Pannekoek, 1961)? His student Eudoxius (b. 408 BC) devised a system of 27 homocentric rotating spheres, with one assigned to the background stars, and the rest divided among the moon, known planets and the Sun. Eudoxius's contemporary Callippus (b. 370 BC) added seven more spheres to account for additional observations. Another of Plato's students, Aristotle (b. 384 BC), devised another system of the cosmos, influenced by Callippus. Aristotle required 55 crystalline spheres, with 22 of them counter-rotating to explain the apparent motions of planets and stars (Pannekoek, 1961). These are the Greek explanations of the cosmos that we normally remember, but shortly after Aristotle, another Greek made a profound leap in logic to produce another model of the Universe, one that would not be repeated for quite some time.

Aristarchus of Samos (b. 310 BC) made an effort to calculate the relative distances to the Sun and Moon, and then computed their relative volumes to the Earth—by using the arc of Earth's shadow during eclipses to relate the size of the Moon to the Earth. He found that the sun had a volume between 254 and 368 times that of Earth (the modern value is $\approx 1.3 \times 10^6$). Perhaps because he found it absurd that such a large object would move about the earth he adopted a heliocentric model for the cosmos. Unfortunately, the only surviving text from Aristarchus is the one in which he computes the relative distances and sizes of the Sun and Moon. We know of his heliocentric theory today because Archimedes references Aristarchus in his works, and Archimedes is *nearly* contemporary with Aristarchus (Heath and Aristarchus of Samos, 1913). However, the heliocentric theory did not gain sufficient popularity to compete with Aristotle's model. Astronomy continued to develop through antiquity, building on Aristotle's model, culminating with Ptolemy's (b. AD 90) theory of the Universe.

Ptolemy put forth a much more complete and useful theory of the cosmos than his Greek predecessors, combining much of the work from the preceding 500 years—especially that of Hipparchus (b. 190 BC)—in one useful set of tables and mathematical relations. Ptolemy's opus, *Almagest*, presents a geocentric Universe—solar system plus background stars—where a number of phenomena are explained mathematically, by nested circular motions and epicycles, illustrated in Figure 2.1. Ptolemy presents an explanation of the motion of the planets, and their apparent retrograde motion against the zodiac, precession of the equinoxes, eclipses and more (Hetherington, 1993; Pannekoek, 1961); the work was so powerful that it remained the basis for cosmological studies for nearly 1500 years.

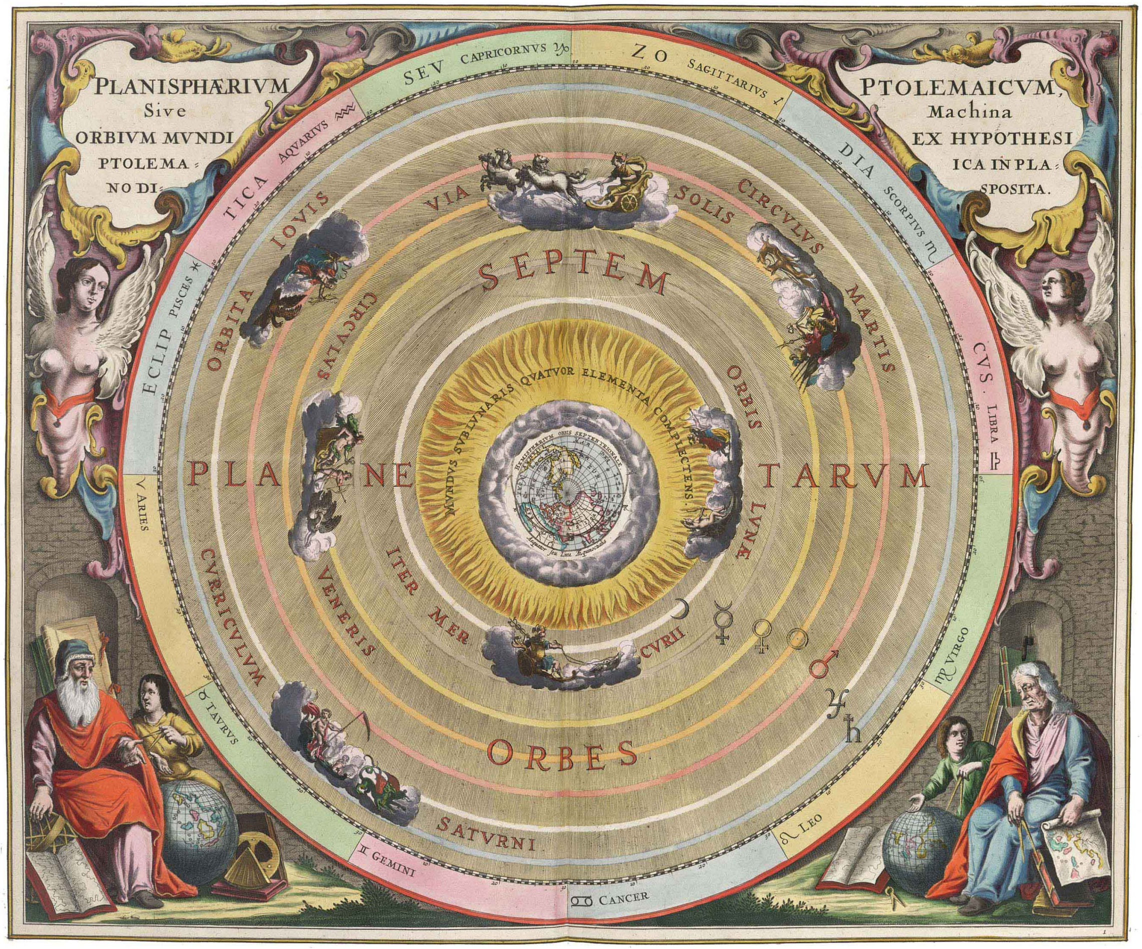


Figure 2.1: Ptolemy's Universe as illustrated in Cellarius (1661).

Between the final dissolution of the Roman Empire (476) and the beginning of the scientific revolution (1543), astronomy was a subject limited to study within the Catholic Church, due to its need to correctly reckon religious holidays. In the thirteenth century Thomas Aquinas (b. 1225) united the Aristotelean cosmology with the doctrine of the Church. Much Greek knowledge disappeared during the dark ages, but many classical manuscripts were stored at libraries in the Middle East. During the renaissance, classical knowledge flowed with commercial trade back to western Europe from the Middle East. The Ptolemaic system again became the accepted system, but it needed to be improved. It did not provide sufficient accuracy for religious events to occur on the same days as they had been in historical times, and as trade routes expanded in the renaissance, more accurate navigational tools were required. The need for a better astronomical system for the Church and for commerce provided incentive to improve astronomical observations and calculations.

Modern scientific analysis of the cosmos began in earnest in 1543 when Copernicus placed the Sun at the center of the Solar System and set Earth spinning on an axis (Copernicus, 1543). With this system, and a number of recent observations, Copernicus provided a simpler explanation of the cosmos—in that there were fewer motions that explained more observable phenomenon—that could make better predictions than the Ptolemaic system. The Copernican Universe is illustrated in Figure 2.2. The year 1543 is considered the start of the scientific revolution, and advances in astronomical observations and theories would come much more frequently thereafter.

Even with the new heliocentric worldview in place, background stars were thought constant and unchanging, but in 1572 a supernova appeared in the constellation Cassiopeia. Noticing this new star, Tycho Brahe was struck by the idea that amassing observational data would be the best way to improve astronomical theories. From 1572, observation would be his life's work. He was meticulous in his observations, demanding precision beyond any previous observer—to the point that when concerned about atmospheric refraction affecting his measurements at the arcminute level, he would send assistants to different latitudes to double check measurements. Late in his career, Brahe hired Johannes Kepler, who would inherit and analyze Brahe's data, arriving at his eponymous laws of planetary motion. These laws were invaluable to Isaac Newton as support for his law of gravitation.

Building on Galileo's work on linear motion, and Huygens's on *centrifugal force*, Newton was presented a scientific world where it was well established that a force of attraction directed toward the center of an orbit acts upon the planets and moon. The question Newton addressed was how that force varied with distance. Huygens had estimated in 1673 (Bell, 1941; Huygens, 1673) that the centrifugal force on a

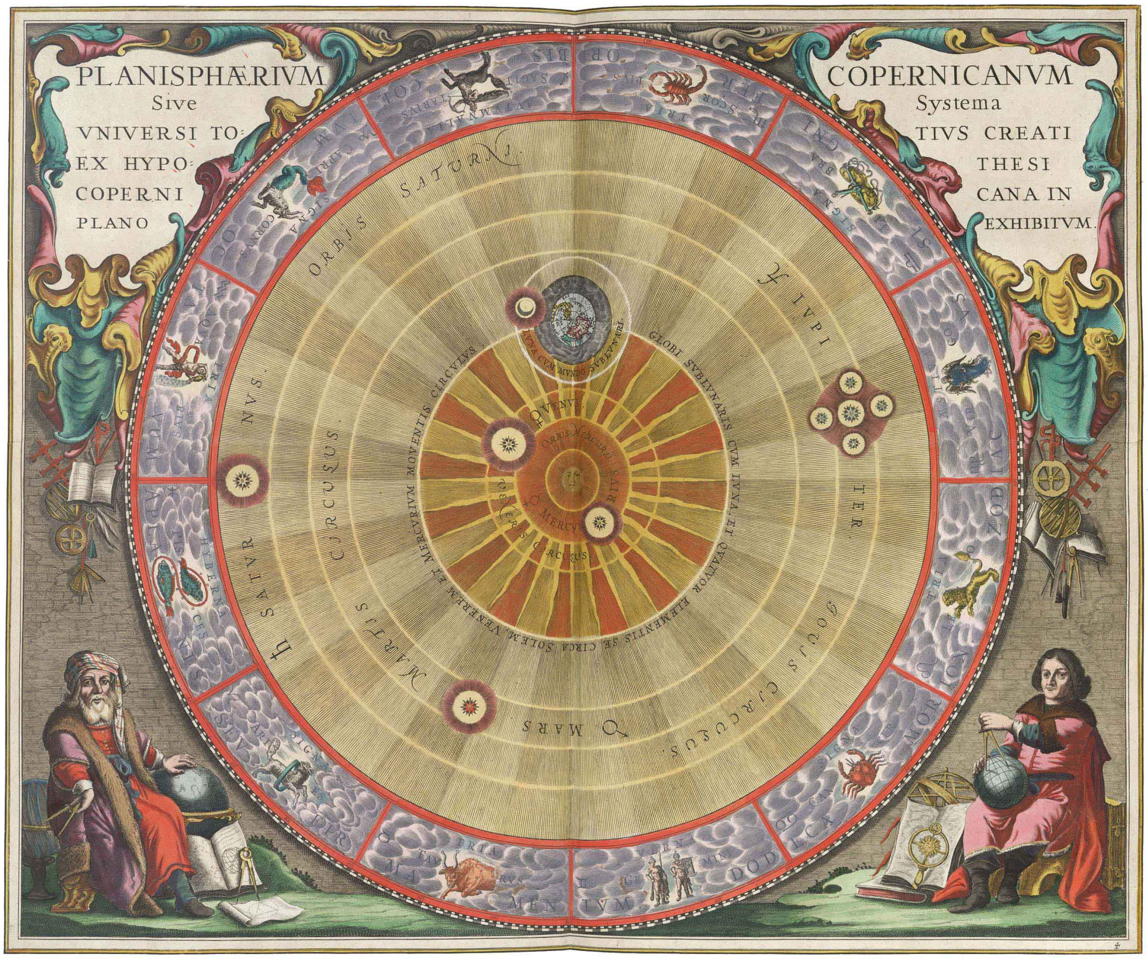
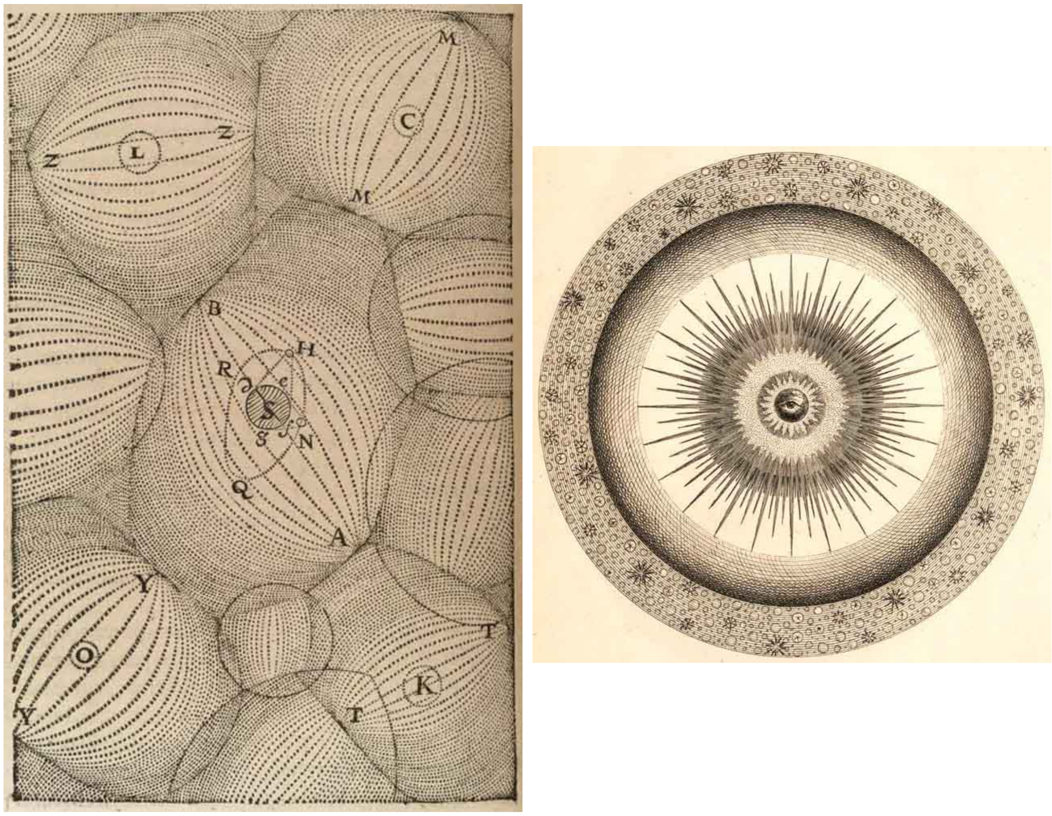


Figure 2.2: The Copernican Universe as illustrated in Cellarius (1661).

body is proportional to both the square of the its velocity and to the inverse of the its radius of motion. With Kepler's third law—the square of the orbital period is proportional to the cube of the semi-major axis—Newton was able to explain how centrifugal force varied with distance. In particular, a four times larger circular orbit has an eight times larger period, hence a two times smaller velocity. Then, using Huygens relationship, that orbit experiences a sixteen times smaller centrifugal force. If Gravity compensates in the same ratio, then such a planetary system obeys an inverse square law for the gravitational force. His theory was completed in 1684–1685, presented to the royal society in 1686, and published in 1687, along with his mathematical theory, calculus (Newton, 1687).

Newton's theory reinforced the heliocentric model for the Solar System but, scientists were becoming aware that the Universe extends beyond the Solar System, and started to address the nature of the Milky Way and stellar background. The first cosmological models extending beyond the Solar System were speculative conjectures, but they are interesting nonetheless. René Descartes put forward a model of interlocking Solar Systems (Descartes, 1677) as shown in Figure 2.3a. In 1750, Thomas Wright attempted to put the Sun and the other stars in orbit around a divine center (Figure 2.3b, and noted that an observer looking out from a location on the spherical shell orbiting the center would observe a bright band of stars exactly like the Milky Way (Wright, 1750). In 1755, Immanual Kant proposed a hierarchical structure for the Universe—from the level of a galaxy down to a solar system—and that celestial objects are flattened by rotation (Kant, 1755).

William Herschel mapped the stars in the Milky Way in 1785 (W. Herschel, 1785), making cosmology a more staid subject. Herschel argued that the observable distant nebulae were actually galaxies like our own, and set about the task of resolving stars in those distant *nebulae*. His endeavors led him to doubt his own theory, as he discovered cases where there were regions of luminous gas that could not be resolved into stars. He lost faith in his model and started to believe the galaxy had no edge. However, his study led to the first large catalog of *interesting nebulae* (W. Herschel, 1786). This and a contemporary catalog of objects by Charles Messier (Messier, 1781) were the earliest extensive galaxy catalogs, and we still use Messier identification numbers today (*e.g.*, M51 in Figure 2.6). William Herschel's son, John, continued cataloging objects visually, releasing the General Catalogue of Nebulae in 1846 (J. Herschel, 1864), which provided a large fraction of the objects for John Dreyer's New General Catalogue (NGC) of Nebulae and Clusters of Stars (Dreyer, 1888) and two Index Catalog supplements (Dreyer, 1895, 1908). All of these catalogs are relevant to astronomy today, because many observable objects are identified by the number



(a) René Descartes (b) Thomas Wright

Figure 2.3: The Universe according to Descartes and Wright. Descartes imagined a world of interlocking Solar Systems (*left*). In this image, S represents the sun, while C, L, K and O are other stars in the center of their own regions of circular motions. Wright imagined a shell of stars circulating about a divine center (*right*). Wright adopted this model, because looking out in the tangential direction from the shell, an observer would see a band of stars across the sky, which could explain the Milky Way.

that they were assigned in these early visual catalogs (*e.g.*, NGC 1365 in Figure 2.6). The visual cataloging of objects was done by 1908, as photographic methods became viable.

After Newton introduced his laws of motion, he and other scientists appreciated that the nature of this gravitational law had important implications for the large-scale structure of the Universe. In particular, if all objects are gravitationally attracted to each other, how can the Universe be stable as it appears to be? One way to resolve this problem is to allow space to extend infinitely in all directions, with a

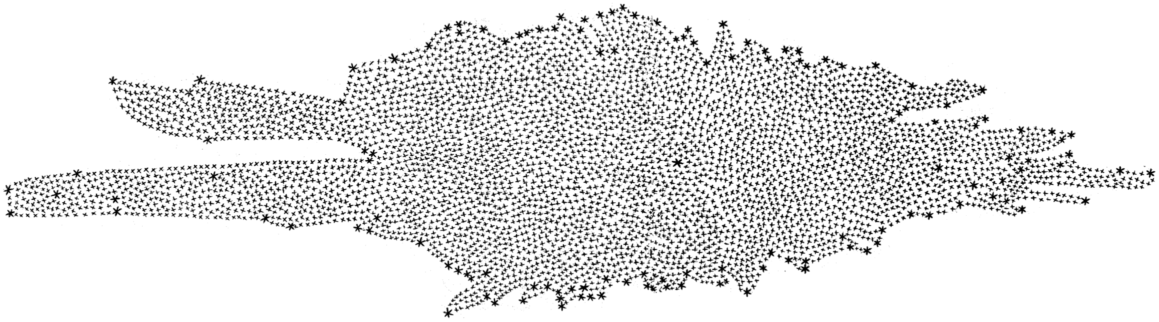


Figure 2.4: William Herschel's measurement of star positions; the Milky Way Galaxy (W. Herschel, 1785).

uniform distribution of stars. Early observations suggested that there was an even distribution of stars in all directions, although the Herschel map showed otherwise. This uniform distribution theory had problems, one of which was the idea that if a star were perturbed from its equilibrium position, it would be attracted gravitationally to another star. Nonetheless, with no better theory in place, the idea of a static and infinite–as well as spatially flat–Universe, was the prevalent theory, until the late 19th century. Several observations and theoretical developments in the early 20th century created doubt about the static Universe model.

2.2 Early Twentieth Century Observations

There were a number of observations and predictions made in the early 20th century that are consistent with, and even point to the modern standard cosmological model, though at the time they were made, they may have seemed confusing and unrelated.

Though not a true 20th century observation, an interesting question to ask about the sky is, why is it dark at night? A dark night sky fits neatly with the medieval view that the Universe was created for man to live in, with Earth at its center, and a finite number of celestial spheres (Harrison, 1987; P. Schneider, 2006), but does not fit as well in the post-scientific resolution and enlightenment world. A dark sky does follow from an infinite Universe with a uniform distribution of stars, which would make the sky bright. This question is commonly known as Olbers's paradox, though the problem was posed by others before Heinrich Olbers posed the problem in 1823 (Harrison, 1987). The dark sky implies that the Universe is not infinite, or the distribution of stars is not uniform, or some effect causes distant starlight to be diminished on its path to Earth. There is another possibility, one that would have made many of the early natural philosophers uncomfortable but that gained traction in the 20th century: the Universe is not static. Perhaps stars evolved from some factor that occurred before stars could exist, or perhaps some property of the Universe evolves in time that influences starlight.

In 1927 Georges Lemaître proposed that the Universe is expanding while investigating the equations of Einstein's general relativity (Lemaître, 1927, 1931). In 1928 Edwin Hubble observed that the doppler shift of spectral lines from distant galaxies showed that the recessional velocity of those galaxies increases linearly with their distance (Hubble, 1929) as shown in Figure 2.5;

$$v = H_0 D, \tag{2.1}$$

where H_0 is the Hubble constant. This finding offered support for Lemaître's prediction, but comes with an odd implication. If everything in the observable Universe is moving apart, then if we were to run time in reverse, at some point, everything could come together. If we could wind the clock back far enough, all matter and energy would be so densely packed that the laws of physics as we know them would not apply. We call the early hot and dense state from which the Universe started the Big Bang.

Aside from all these large-scale observations, there were some smaller scale investigations of individual galaxies and clusters of galaxies around the same time. Studying

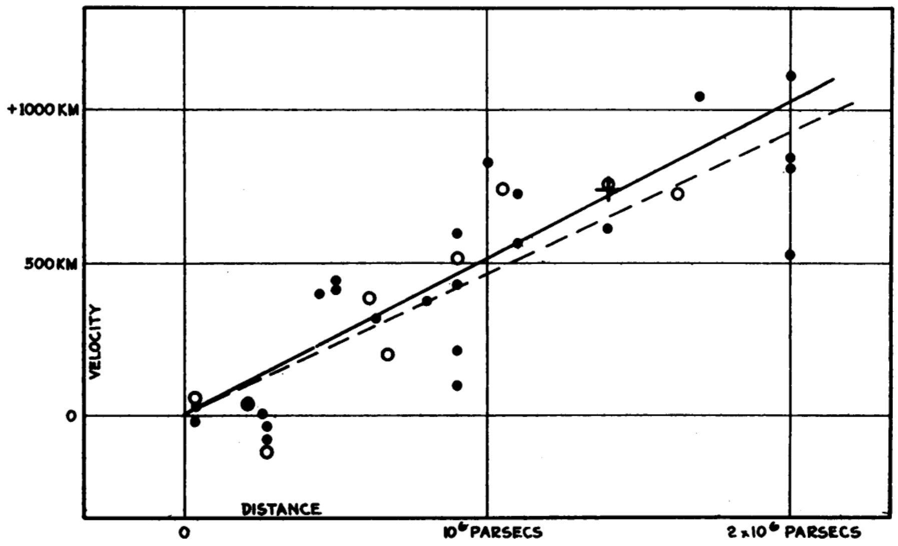


Figure 2.5: Edwin Hubble's original galaxy recession diagram. The horizontal axis is the distance to the galaxy, while the vertical axis is their measured recessional velocity. Filled circles and solid line are computed using galaxies individually, while the open circles and dotted line result from combining galaxies into groups. The cross is the mean velocity and mean distance corresponding to 22 galaxies that could not be estimated individually (Hubble, 1929).

spectra of stars in the Milky Way, in 1932 Jan Oort found that the stars tended to move faster than expected if moving only under the influence of mass associated with visible matter (stars, gas, dust) (Oort, 1932). In fact, the stars had sufficient velocity to escape the gravitational pull of the other visible matter in the galaxy, but Oort knew this could not be the case, and he postulated the existence of some unobservable matter. Near the same time, Fritz Zwicky was studying the Coma Cluster of Galaxies, and applied the virial theorem—derived by Claussius (1870), applied to star clusters by Eddington (1916)—to their measured velocities and found evidence for hundreds of times more mass in the Coma Cluster than would be expected based on the luminosity of the galaxies (Zwicky, 1933, 1937). These observations by Oort and Zwicky were the first evidence for the dark sector of the Universe, a sector which will play an more important role later in the 20th century.

2.3 Classical Cosmology

As World War II approached, astronomers had come around to the Friedmann–Lemaître world models, and there were several questions to address: is space curved; what is the value of Hubble's parameter; how fast is the Universe's expansion decelerating (or so they thought...); and what is the mean density of the Universe? A number of technological advances in the early- and mid-twentieth century would help answer these questions, though determining some of the answers would take much longer than anybody in the early-twentieth century expected.

To address these questions and examine the large-scale structure of the Universe, the visual sky catalogs described in section 2.1 were useful, but not sufficient. Information about the sky would need to be acquired and cataloged much more quickly and reliably, and observations would need to be more precise. In the late nineteenth century, photographic plates were developed for use in astronomy. Early photographic materials were not very sensitive, and telescopes were not sufficiently stable to produce sharp images for the length of required exposures. In 1900, the Lick Observatory commissioned a new telescope designed for astrophotography, and opened by photographing M51, shown in Figure 2.6a. Two large telescopes that were designed to be capable of photography are interesting to note. A 100-inch telescope was commissioned for the Mount Wilson Observatory in California in 1906 and saw first light in 1917. Hubble used the Mount Wilson telescope for several of his important observations. A 48-inch aperture Schmidt camera—a new wide-field design (B. Schmidt, 1938)—was built at the Mount Palomar Observatory in California between 1939 and 1949, and then used for a nine-year, all-sky photographic survey, with 14 inch photographic plates in two colors. The survey covered the entire sky in the northern hemisphere (Minkowski and Abell, 1963). The NGS-POSS was an indispensable research tool for years, providing images for the first Abell catalog of rich galaxy clusters (Abell, 1958) (*e.g.,* Abell 2151 in Figure 2.6), as well as the later

(a) 1900: M51 (b) 1949: M33 (c) 2000: NGC 2403 (d) 2012: NGC 1365

(e) Lick Observatory: Abell 2151 (f) SDSS: Abell 2151

Figure 2.6

Figure 2.6: A series of images to illustrate improving telescope technology. The *top* row shows images of a series of individual galaxies from a variety of telescopes spanning more than a century. The first image shows detail of a commissioning photograph from the Lick Observatory in 1900 containing M51, the *Whirlpool Galaxy* (2.6a). The second image shows M33, the *Triangulum Galaxy*, on a red sensitive photographic plate from the NGS-POSS (2.6b). The next image shows NGC 2403 from the SDSS (2.6c). The final image shows a recent image of NGC 1365 from the DES (2.6d). This set of images illustrates how much more detail is available to astronomers today than in the early and middle twentieth century, as recent images have more resolution, color information and can detect fainter objects. The *bottom* row shows two images of the same region of space with two different telescopes separated by half a century. They each show Abell 2151 from the NGS-POSS (*left*, 2.6e) and SDSS (*right*, 2.6f) respectively. This pair of images is meant to illustrate how new digital surveys with color information that find dimmer objects make identifying clusters faster and simpler, especially when paired with computer analysis, which was not available to Abell when he first identified Abell 2151.

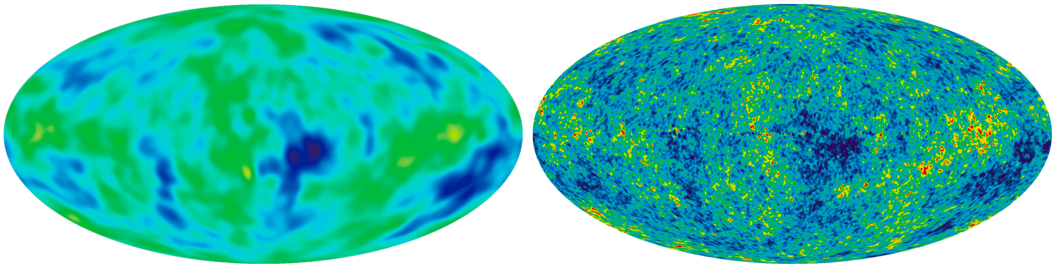
Zwicky and Abell catalogs (Abell, Corwin, and Olowin, 1989; Zwicky et al., 1961–1968). The World War II era not only saw telescopes with controls precise enough for photographic surveys, but also technological development allowing observation at frequencies outside the visible spectrum, which led to discoveries beyond what simple visible observations would provide. Figure 2.6 shows how sky surveys have improved from the first photograph-enabled telescopes to the most recent CCD telescopes, described later in the chapter.

2.4 Recent Observations

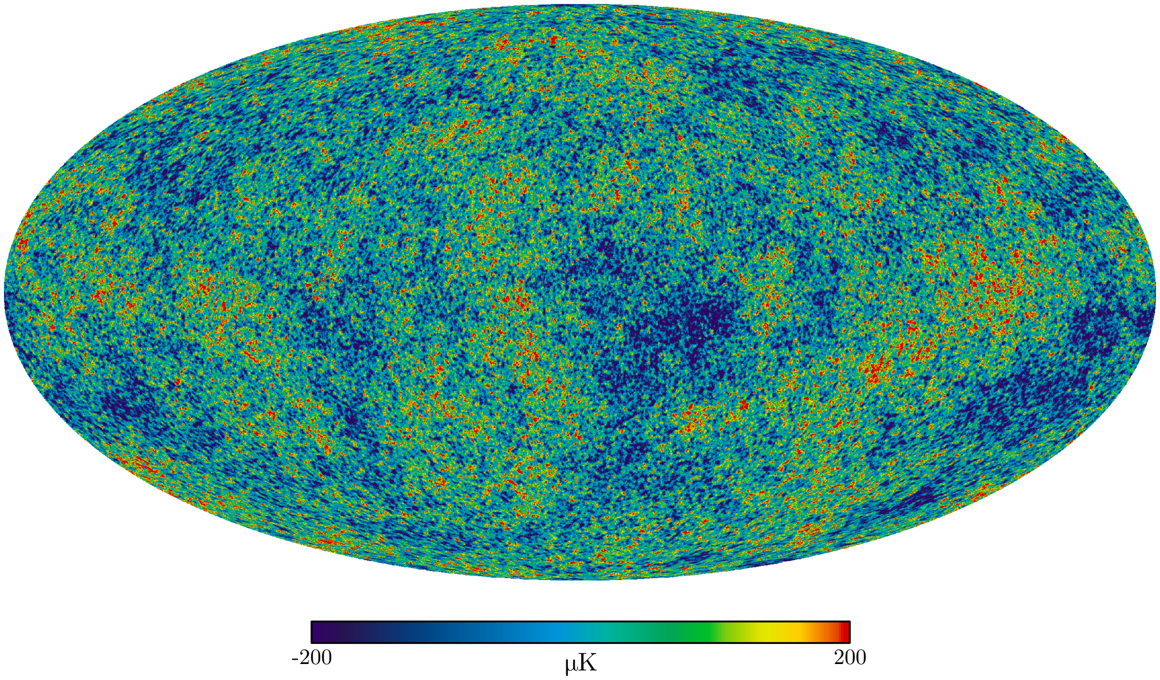
Early observations were consistent with world models developed by Friedmann and Lemaître. The recessional velocity measurements suggested some early hot dense period in the early Universe, emergence from which was dubbed the *Big Bang* (or *horrendous space kablooie*, see front matter). However, some astronomers held out for some other ideas, most notably Fred Hoyle with the steady state theory (Hoyle, 1948; Hoyle, Burbidge, and Narlikar, 1993). But, the post World War II years came with new observations that bolstered Big Bang cosmology, in particular the ΛCDM paradigm that will be explained in Chapter 3.

Accepting the Big Bang, it is possible to make predictions about some properties of the Universe and compare them to observations to see if we are forming a consistent picture. Imagining the Universe as a hot and dense matter fluid, and combining thermodynamics and Universal expansion, it is possible to estimate the relative abundance of light elements that should be observed in the Universe. Alpher first did this estimation in 1948 (Alpher, Bethe, and Gamow, 1948) and found rough agreement with the light element abundances observed at the time, though he missed badly on heavy element abundances. More detailed calculations, taking into account more thermodynamic processes is possible (Coc et al., 2004; Kolb and M. Turner, 1990), and these calculations agree with modern measurements.

Another implication of the description of space-time that Lemaître—and others—used when predicting that the Universe is expanding is that space is isotropic and homogeneous on large scales (de Sitter, 1917; Friedmann, 1922, 1924; Lanczos, 1922; Lemaître, 1927, 1931; Robertson, 1935; Walker, 1937). Penzias and Wilson (1965) measured an excess diffuse background radiation while calibrating a sensitive maser receiver system. They found that the radiation intensity was uniform over the sky in the wavelength range 1 cm $< \lambda <$ 1 m, corresponding to a black-body temperature of about 3 Kelvin. The uniform background radiation was a key hot Big Bang measurement, and Penzias and Wilson were awarded the 1978 Nobel Prize in Physics. In the distant past, matter was so hot that it existed only in ionized form, and a measurable low temperature radiation background is a relict of the early ionization. Theorists predicted that the radiation could be measured, though Penzias and Wilson did not know it at the time (Doroshkevich and Novikov, 1964; Penzias, 1979)! From the moment when the Universe cooled enough to allow for atoms to form, photons have been able to stream freely, carrying information about the state of the Universe at that



(a) *COBE*: launched 1989 (b) *WMAP*: launched 2001



(c) *Planck*: launched 2009

Figure 2.7: The CMB as measured by *COBE*, *WMAP* and *Planck*. The uniform component of the CMB has been removed, as well as the dipole component due to Earth's velocity through the background radiation. Additionally, the images have been corrected to remove radiation from the plane of the galaxy, leaving a map of the fluctuations in the microwave background. Each plot is on the same scale, in units of microKelvin. These anisotropies give hints about the geometry of the Universe, as well as the seeds of structure formation. Notice how the more recent maps have much finer detail. The *COBE* instrument had an angular resolution of 7 degrees, while *WMAP* had 13.5 arcminutes, and *Planck* 5 arcminutes, improvements by factors of 30 and 2.5 respectively. Additionally, *WMAP*'s instruments were 45 times more sensitive than *COBE*'s, and *Planck*'s 10 times more sensitive still. This shows how much more precise measurements have become in the early 21st century. Data for *COBE* and *WMAP* from NASA, and for *Planck* from ESA and the Planck collaboration (Planck Collaboration (2013) and related).

time. The background radiation is called the Cosmic Microwave Background (CMB). Given that we have structure in the Universe today, one would expect to see very small fluctuations in the CMB that would be amplified by gravitational infall to give the structure we see today (Kolb and M. Turner, 1990). Many experiments have gathered data on the CMB, space-based telescopes are especially notable: the *COBE*, starting in 1989; the *WMAP* starting in 2001; and the *Planck* satellite, launched in 2009. *COBE* verified that the CMB is a perfect source of blackbody radiation with a temperature of 2.7 K, and measured that there were anisotropies in the CMB, but was not capable of exploring those anisotropies in detail (Mather et al., 1994; Smoot et al., 1992). For discovering the anisotropies and verifying the black body spectrum of the CMB, the *COBE* team was awarded the Nobel Prize in 2006. *WMAP* carried much more sensitive instruments and was able to measure anisotropies in detail (Komatsu et al., 2011; Spergel et al., 2003). Figure 2.7 shows maps of the CMB as measured by *COBE* and *WMAP*. However, those fluctuations are smaller than would be expected for a Universe in which baryons are the only source of matter, giving support to the idea of Dark Matter, in accordance with Oort's and Zwicky's observations. The CMB also favors a Universe in the Lemaître model with no curvature (see Chapter 3), requiring a specific amount of total energy in the Universe, but the amplitude of the temperature fluctuations suggested that matter (luminous or dark) can only account for ∼30% of that required energy.

In 1998 and 1999, two groups performing experiments similar to Hubble's recessional velocity measurements but using Supernovae (SNe), found evidence supporting an expanding Universe, and that the expansion is accelerating (Perlmutter et al., 1997, 1999; Riess et al., 1998). They measured the *redshift* of the SNe, which is equivalent to Hubble's original recessional velocity measurements, and applied standard candle assumptions and measurements to measure the distance to those objects. So velocity/redshift is on x-axis in both cases, and distance on y-axis. Their results are

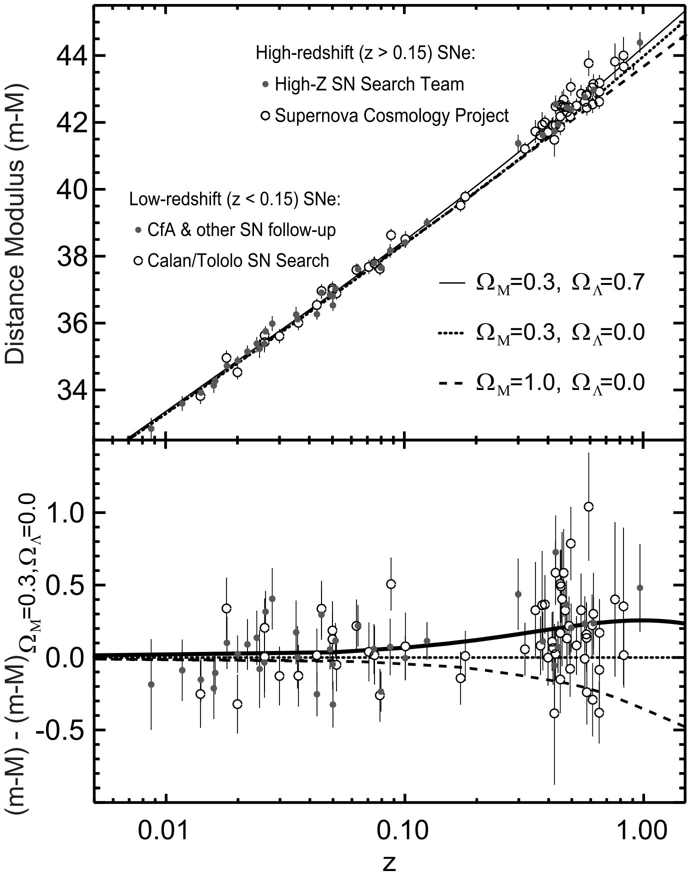


Figure 2.8: The Hubble diagram for Type Ia SNe from the High-Z Supernova Search (Riess et al., 1998) and the Supernova Cosmology Project (Perlmutter et al., 1999). The bottom panel shows residuals for the cosmological fits given in the upper panel legend. Figure from Perlmutter and B. P. Schmidt (2003).

shown in Figure 2.8. Accelerating expansion is a clear indicator that the missing energy density required to give the Universe its flat geometry exists in a form with negative pressure. We call the energy driving the accelerating expansion *dark energy*, and there are many models to describe it, one of which is a cosmological constant Λ, described in detail in Chapter 3. There are also more exotic explanations for dark energy which involve dynamical non-canonical scalar fields (see *e.g.*, Saotome (2013)), but those lie outside the scope of this work.

In addition to previously mentioned virial theorem and CMB measurements implying the existence of non-luminous matter, careful analysis of a cluster commonly called the *Bullet Cluster* shows strong evidence for the existence of dark matter. Galaxy clusters are the observational counterpart to dark matter halos, and are the largest graitationally bound objects in the Universe. Density fluctuations in the early Universe grow by gravitational collapse into large dark matter structures that we

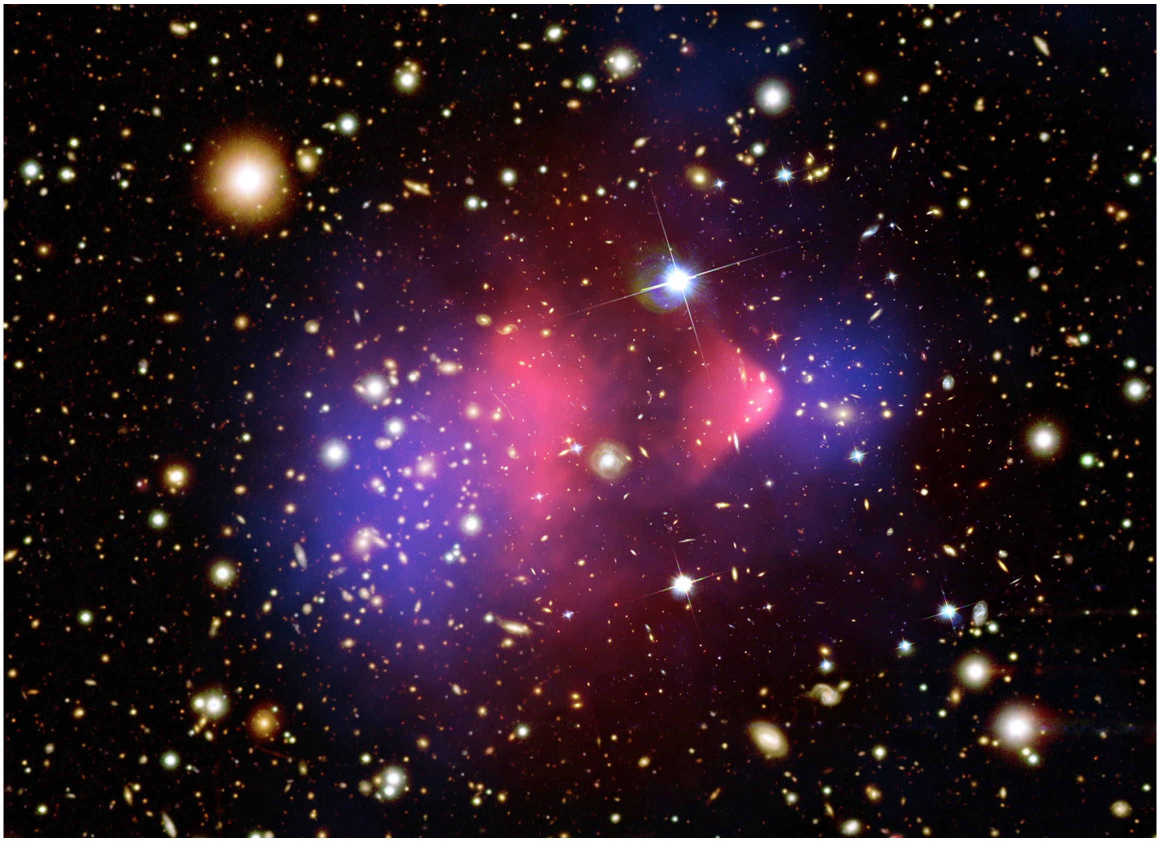


Figure 2.9: A Composite X-ray, dark matter and optical wavelength image of galaxy cluster 1E 0657-56, the *Bullet Cluster*, with stars shown as bright points, the X-ray gas temperature shown in pink and the weak-lensing mass distribution shown in blue. This cluster provides strong evidence for the existence of dark matter. Credits: X-ray: NASA/*CXC*/CfA/Markevitch (2006); Lensing Map: NASA/STScI; ESO WFI; Magellan/U.Arizona/Clowe et al. (2006) Optical: NASA/STScI; Magellan/U.Arizona/Clowe et al. (2006).

call halos. These halos also contain luminous matter, such as stars, galaxies and hot intergalactic—or intracluster—gas. X-ray observations of the spatial gas distribution about the Bullet Cluster cluster suggests that it is a high-velocity merger event of two different clusters, with the interacting gas lagging behind the motion of the collisionless galaxies. Additionally, a weak gravitational lensing study of the mass distribution shows that the majority of the overall cluster mass corresponds to the center of the galaxy distribution. This suggests that the dark matter has continued on its trajectory along with the luminous matter, implying that it is a non- or weakly-interacting form of matter, while the diffuse gas from the two clusters has collided and slowed.

The measurements are combined into one false color image overlaid on the cluster's galaxy distribution in Figure 2.9. The clear separation of mass and gas is strong evidence for dark matter of the same type that would match other observations (Clowe et al., 2006; Clowe, Gonzalez, and Markevitch, 2004; Markevitch, 2006; Markevitch et al., 2004).

2.5 Theoretical Developments

Some of the observations described in the previous section create more problems to be solved. Observing small anisotropies in the CMB creates three problems (Dicke and Peebles, 1979). First, the horizon problem: why are parts of the sky that should have been causally separated in the early Universe appear to have been in thermal equilibrium? Second, the flatness problem: the equations of generally relativity show that if at any point in the history of the Universe, it had not been flat, the spatial curvature would have rapidly evolved to be less flat (see 3.1 for a discussion of flatness). To observe the level of flatness we see today, the early Universe would have had to be flat to a part in $\sim 10^{16}$, which is an unsatisfying fine-tuning problem. Third, many theories predict that we should observe magnetic monopoles. Guth (1981) proposed an idea that could solve all of these problems simultaneously. He proposed that in the very early Universe there was a rapid, exponential expansion of space. This means that patches of the sky that *appear* to not be causally connected actually were, since the correspond to much smaller patches before inflation. Also, if there were a region of space that had a complicated curved structure, the rapid expansion could drive the region to become so large that locally, it appears to be flat. Finally, inflation reduces the monopole density to such low levels that we should not necessarily expect to observe them.

Computer simulation of physical systems became sufficiently sophisticated in the second half of the twentieth century to address some of the questions raised by the observations. At first, simulators had difficulty producing the correct distribution of matter on both large and small scales in a variety of cosmological models—*e.g.*, *Standard Cold Dark Matter* (Davis et al., 1985; Dodelson, Gates, and M. Turner, 1996; Maddox et al., 1990a). Early models favored matter densities that produce flat spatial curvature (see section 3.1), but the distribution of matter could more readily be matched with simulations where the matter density of the Universe is only ∼30% of that required to give a flat spatial curvature (Efstathiou, Sutherland, and Maddox, 1990; Evrard et al., 1993). Also, observed X-ray emission from galaxy clusters when compared with simulations suggest that the ratio of baryonic matter to dark matter in clusters does not differ from the cosmic value, and that observations are with the matter density being only ∼30% of the total required for a flat Universe (Evrard, 1997; S. White et al., 1993). However, these simulations were independent of the presence of dark energy, so a flat spatial curvature—favored by CMB measurements—could be achieved by a combination of low matter density and large dark energy density, ∼30% and ∼70% respectively (Carroll, Press, and E. Turner, 1992; Ostriker and Steinhardt, 1995).

The 1980s saw inflation and computer simulation combine with ever higher precision observations to buttress the conclusion that the Universe not only has luminous matter, but dark matter and dark energy, described by Einstein's general relativity.

2.6 Modern Cosmology

Modern cosmology is built around measurement in many wavebands, but I focus on optical galaxy catalogs, because that is where I have focused my effort in DES. The early galaxy cluster catalogs of Abell (1958) and Abell, Corwin, and Olowin (1989) were determined by visual inspection of photographic plates. Visual inspection is highly prone to systematic errors (Dekel et al., 1989). Other visually identified galaxy and galaxy cluster catalogs (*e.g.,* Shane and Wirtanen (1967); Zwicky et al. (1961–1968)) also had significant systematic effects that generated interesting discussions about how reliably clusters could be identified, whether the galaxy counts (or richness) for clusters were reliable and if clusters are prone to projection effects alon a line-of-sight (de Lapparent, Kurtz, and Geller, 1986; Groth and Peebles, 1977, 1986a,b). Not only did visual inspection introduce systematic errors, but larger sky surveys produced so much data that the only way to extract useful information was to develop high speed methods for identifying galaxy clusters. In the years after the Lick Survey (Shane and Wirtanen, 1967), computational and photographic power advanced significantly, and automatic scanning machines allowed for digitization of sky surveys. The Automatic Plate Mover (APM) galaxy survey eventually found 3 million galaxies using automated photographic plate scanners (Dalton et al., 1997; Maddox, Efstathiou, and Sutherland, 1990; Maddox et al., 1990b), which was a significant improvement compared to the 30,000 galaxies in Zwicky et al. (1961–1968). The next round of galaxy surveys added the ability to determine the distance to galaxies by measuring their distance—or redshift, see section 3.1—with multi-object spectrometers. With the distance information, the Harvard–Smithsonian Center for Astrophysics (CfA) redshift survey (Huchra et al., 1983) and the impressive Two-degree-Field Galaxy Redshift Survey (2dFGRS) were able to probe the large-scale structure of the Universe with their statistical power and three dimensional position measurements, discovering the *Great Wall* (Geller and Huchra, 1989) and finding Baryon Acoustic Oscillation (BAO) (Cole et al., 2005). The distribution of galaxies observed by the CfA survey and 2dFGRS are shown in Figure 2.10, along with more recent observations from the SDSS described below, and corresponding galaxies from a recent large-scale structure simulation.

The SDSS was fully a fully end-to-end digital survey, which allowed data to be

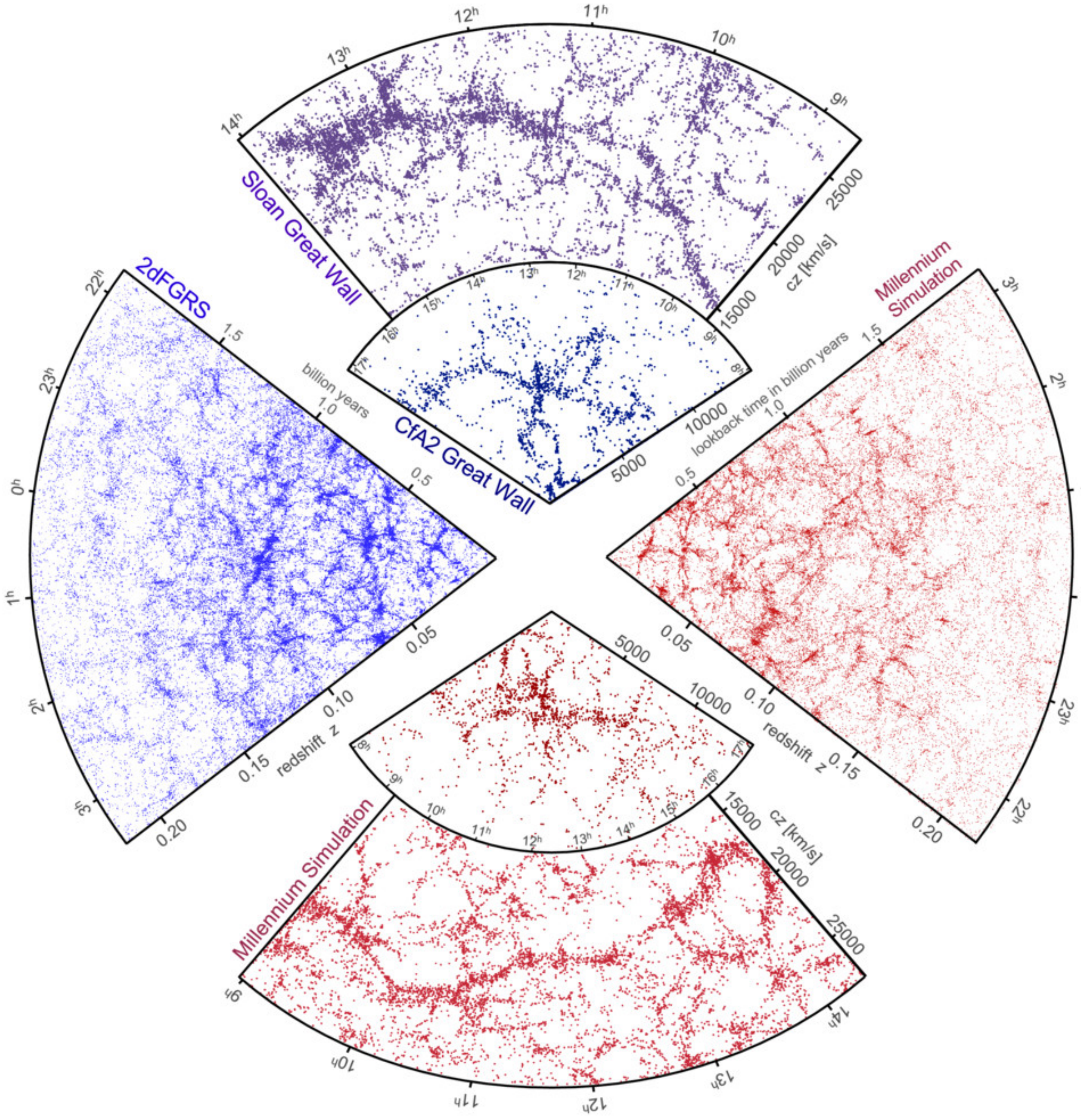


Figure 2.10: Observed galaxy distributions from three galaxy redshift surveys, and corresponding simulated galaxy distribution from the Millennium SImulation (Springel et al., 2005). The *inner top* wedge shows galaxies from the Harvard–Smithsonian Center for Astrophysics redshift survey, while the *outer top* wedge shows SDSS galaxies, and the *left* wedge shows the distribution measured by 2dFGRS. The remaining wedges show corresponding sections of the Millennium SImulation (image from Springel, Frenk, and S. White (2006)).

produced even faster than with the automatic plate-scanning surveys. SDSS used a wide field of view (3 square degrees), 120 Megapixel CCD camera and covered nearly 1/4 of the sky during operation. The camera was equipped with 5 waveband filters, improving on earlier surveys, which usually used 2 color photographic plates, for precise determination of galaxy positions in three dimensions. The telescope operated in time-delay integrate mode (D. Schneider, M. Schmidt, and Gunn, 1994), allowing for long exposures which could be used to find quasars and galaxies at high redshift. The SDSS is the largest sky survey completed to date. The ninth data release (Ahn et al., 2012) contains over 900 million celestial objects (stars, galaxies, etc.), and has spectra for over 2 million of them.

While these surveys were combing the visible wavelengths of light, other large surveys have explored other wavelengths to investigate properties of the Universe beyond the visible mass distribution. The previously mentioned *WMAP* experiment has observed the sky at microwave, and its successor experiment, the Planck satellite will improve the *WMAP* measurements (Planck Collaboration, 2011a). The Two Micron All-Sky Survey (2MASS) has covered the sky with infrared observations (Skrutskie et al., 2006), and the HI Parkes All Sky Survey (HIPASS) covered the southern sky with radio frequency observations (D. Barnes et al., 2001). The *X-ray Multi-Mirror Mission–Newton* (*XMM–Newton*) and *Chandra X-Ray Observatory* satellites provides X-ray coverage of the sky (Cappelluti et al., 2009) and the *Fermi Gamma-ray Space Telescope* searches the sky for gamma ray emission (Abdo et al., 2010).

All of these deep and wide-field sky surveys have wrought a fundamental change in cosmological studies. No longer is there only a race to build collections of objects. Finding objects is bordering on banal, but understanding measurement systematics and theoretical uncertainties is becoming a priority for extracting more detailed cosmological properties from the plethora of surveys available. With the next generation of experiments, including the DES, Large Synoptic Survey Telescope (LSST) (LSST

Science Collaboration et al., 2009), Atacama Large Millimeter Array (ALMA) (Wootten, 2003) and Square Kilometer Array (SKA) (Rawlings and Schilizzi, 2011), there will need to be advances made in modeling to exploit the full observational power of the survey. I will now focus on particulars of the DES, which is the survey for which the work described here supports.

2.7 The Dark Energy Survey

Building on the success of the SDSS, the DES is an optical survey using a new camera mounted on the 4-meter Blanco telescope at Cerro Tololo Inter-American Observatory (CTIO) in Chile (Annis, Bridle, et al., 2005; Flaugher et al., 2012). The survey will observe 5,000 square degrees of the southern sky in five different optical bands: g, r, i, z, and Y (Kuehn et al., 2012), out to redshift ~1.3 using a ~3 square degree field of view, ~570 megapixel CCD camera. The resulting data set should contain about 200 million galaxies and about 2000 Type Ia SNe (The Dark Energy Survey Collaboration, 2005). DES aims to greatly improve constraints on the dark energy equation of state parameters w_0 and w_a(described in Chapter 3) and test general relativity. Currently, w_0 is constrained at the ~10% level, and w_a is only weakly constrained (*e.g.*, Sullivan et al. (2011)), while some tests of models that modify general relativity are inconclusive (Reyes et al., 2010). DES will tighten these constraints and tests, by improving and combining measurements of four cosmological probes: galaxy clusters, weak gravitational lensing, large-scale structure and Type Ia SNe.

2.7.1 Galaxy Clusters

Galaxy clusters are associated with large dark matter halos, the largest gravitationally bound objects in the Universe. Galaxy clusters are exceptionally interesting objects, they were early indicators of dark matter and provided evidence that the matter content of the Universe was less than that of the density that corresponds to a flat spatial geometry (Zwicky, 1933, 1937). For a review of how cluster properties can inform cosmology, see the review by Allen, Evrard, and Mantz (2011). Clusters can be observed in optical, X-ray, infrared, radio or microwave bands, and cross-correlating the signals can increase the statistical power of surveys to constrain cosmology (Cunha, 2009; Cunha, Huterer, and Frieman, 2009). Galaxy clusters should be tightly correlated with the large scale dark matter distribution in the Universe, and the amplitude of the dark matter structure is set by the cosmological parameters σ_8 and $\Omega_{\rm m}$ (described in Chapter 3). Observing galaxy cluster counts and relating their observational signatures back to a theoretical underlying dark matter halo distribution—dark matter halos host galaxy clusters—allows cluster counts to constrain cosmology (Voit, 2005).

2.7.2 Large-Scale Structure

Baryon Acoustic Oscillation (BAO) are fluctuations in the angular correlation of the visible baryon density caused by acoustic waves in the early Universe. In the early Universe, photons and baryons were tightly coupled. If we consider a dense region of dark matter, baryons and photons in the early Universe, the pressure of the baryons and photons would create an outward traveling spherical sound wave. As the Universe expands and cools, the photons and baryons decouple; the photons continue on their free streaming path and reduce some of the pressure in the system. A shell of baryonic matter will remain at a characteristic radius—the sound horizon—away from the original dense region. As the Universe evolves, gravitational collapse causes the dense regions and the shells of baryonic matter to grow, leaving a pattern on the sky that allows the sound horizon to be measured as a function of redshift. This allows the expansion history of the Universe to be characterized, and the time evolution of dark energy to be constrained. BAO can be seen in measurements of the CMB, or in the angular galaxy-galaxy correlation function, with the latter being the technique that will be employed in DES data. For an extraordinarily detailed discussion of BAO, see Hu (1995).

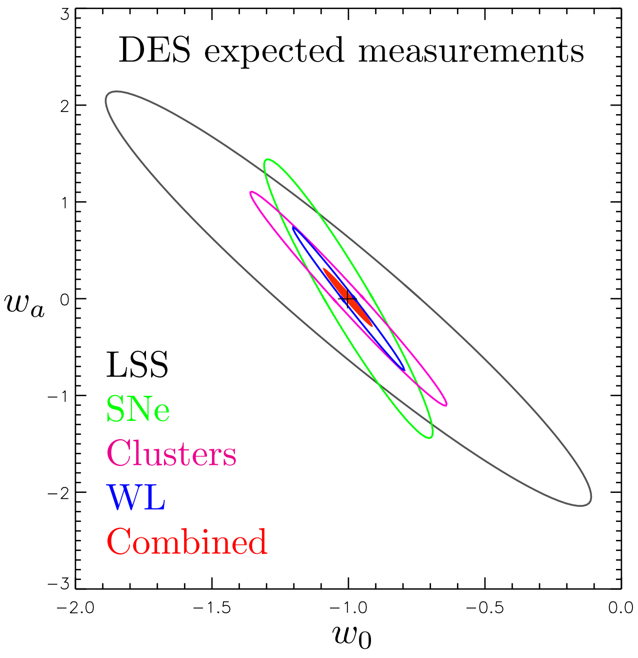
2.7.3 Weak Gravitational Lensing

Light from distant galaxies passes massive structures on its journey to Earth. In some cases, the light is deflected so significantly as to create multiple images of the background galaxy. This effect is called strong gravitational lensing. A far more common effect is for the light to be slightly deflected, creating distortions to the size and shape—shear—of the observable signature of the background galaxy. This is called weak gravitational lensing. Since one does not know the properties of the unlensed galaxy, signals can only be analyzed in large statistical samples. The evolution of the weak gravitational lensing shear power spectrum, and the cross correlation between foreground galaxies and background shear are both sensitive to the matter content and expansion history of the Universe (Huterer, 2002), and can be used to constrain cosmological parameters.

2.7.4 Type Ia Supernovae

Type Ia SNe start as white dwarf stars, which then accrete sufficient mass from a companion star to pass the Chandrasekhar limit (Hillebrandt and Niemeyer, 2000) and explode. The luminosity of the supernova rises and falls as a function of time, with a specific shape when plotted as a function of time and is called a light-curve. This combination of peak brightness and shape of the light-curves allows the distance to a supernova to be precisely calculated (Frieman, M. Turner, and Huterer, 2008). The extreme brightness of SNe means they can be seen at great distances. With extensive and precise distance measurements, a collection of Type Ia SNe provide a probe of cosmic acceleration and the expansion history of the Universe, as illustrated

Figure 2.11: Forecasted 1-σ constraints on w_0 and w_a for DES. The parameters w_0 and w_a describe a particular model for time evolution of the dark energy equation of state, and are described in section 3.1. The fiducial model is taken to be $w_0 = -1, w_a = 0$. Each individual constraint includes *Planck* priors. From Soares-Santos and the DES Collaboration (2012).



in Figure 2.8. Type Ia SNe are a powerful probe. As mentioned earlier, they are the observable that convinced the cosmology community that the Universe's expansion is accelerating, *i.e.*, that dark energy exists.

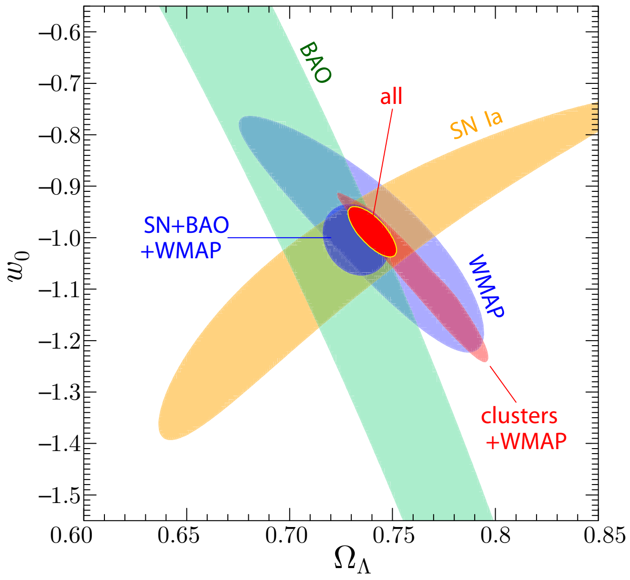
2.7.5 Combined Probes

While DES will have power to constrain cosmology in each of these probes individually, by combining probes the constraints on parameters can be dramatically improved. DES will be the first experiment designed to use all four probes from one data set, and Figure 2.11 shows the power of combining probes. With all four probes, DES aims to constrain the dark energy equation of state parameters $\{w_0, w_a\}$ at the 5% and 30% level respectively (Soares-Santos and the DES Collaboration, 2012).

2.8 Conclusion

A large fraction of the sky is now covered with large surveys at every useful wavelength in the electromagnetic spectrum. Long gone are the days of looking through a tele-

Figure 2.12: Dark energy constraints from combining all cosmological datasets as described in Vikhlinin et al. (2009). Here, we can see that all of the different cosmological probes are in general agreement in constraining the dark energy parameters w_0 and Ω_Λ discussed in Chapter 3. They find $w_0 = 0.991 \pm 0.045(\pm 0.04 \text{ systematic})$ and $\Omega_\Lambda = 0.740 \pm 0.012$.



scope, attempting to identify *nebulae* visually. From the early days when the nature of the Universe was completely unknown, we are now in an era where tight constraints on specific models are achieved by wide-field all sky digital surveys. Cosmology has become a true precision science, with a standard cosmological model covered in the next Chapter, and all observations from every method and wavelength agreeing with the model predictions spectacularly well. Most of the parameters of the model have been identified to a few-percent level precision, and all of the cosmological probes agree, as shown in Figure 2.12. In this era, to gain more precision or to probe slightly modified models, such as time evolution of dark energy, small systematic effects and properties of detection and selection need to be identified and characterized. In this dissertation, I explore one particular problem in optical observations of galaxy clusters, and explore how significant the effect can be in estimating model parameters, as well as how to produce data to aid the modeling. In Chapter 4 we will look at how to produce simulated data to improve modeling, and then in Chapters 5 and Chapter 6 perform some analysis and modeling of the simulated data in support of DES science operations.

CHAPTER 3

Modeling the Universe

Current and next generation sky-surveys will be able to measure cosmological parameters at the percent level, so potential systematic measurement errors must be understood in detail, and implications of underlying theoretical models must be clear. To support analysis for Dark Energy Survey (DES), large simulations for a variety of cosmological models are needed to make detailed predictions about the dark matter distribution, and to vet the scientific analysis process. These simulations are run using N-body simulations that trace the growth of structure over cosmic time in an expanding background. Simulations supporting DES are large and computationally intensive, requiring large national computing infrastructure. To expedite production, these simulations can be managed with software developed to maximize efficiency on that infrastructure, and some of this dissertation is dedicated to documenting efforts to use that software by mapping our N-body problem into the appropriate form. Overall, by building large simulations efficiently, DES as well as future surveys such as the Large Synoptic Survey Telescope (LSST) should be able to realize their full power.

This chapter outlines the standard ΛCDM cosmological model, then describes the development of N-body simulations and how simulations for DES produce synthetic sky surveys. Then, it lays out why exploiting cyberinfrastructure is useful for scientific collaboration, and how it should be helpful for DES and future sky surveys.

3.1 The Standard Cosmological Model

The observations described in Chapter 2 show that the Universe started with a hot Big Bang, then experienced a period of rapid inflation. After inflation, we were left with a spatially flat Universe that is accelerating due to dark energy. The energy required to make the Universe flat comes in three major components:

1. baryonic matter, making up about 4% of the energy;
2. dark (non-luminous) matter (for which there are many potential explanations, see *e.g.*, Harper (2013)), making up about 23%;
3. dark energy, with negative pressure, making up the remaining 73%.

These facts can be combined with Einstein's theory of general relativity to give the ΛCDM —or standard—cosmological model of the Universe.

The standard cosmological model arises using general relativity, combining the energy-momentum properties of a perfect fluid and spacetime geometry or gravity with a specific choice of spacetime metric. We will start our study of standard cosmology with Einstein's equation from general relativity, closely following Hobson, Efstathiou, and Lasenby (2006). I will not repeat the long and necessary derivations to arrive at Einstein's equation, but start with Einstein's equation and find properties of space-time with observable implications—with details in Appendix A. In particular, I will investigate the Universe's kinematic and geometric properties.

Einstein's equation encompasses all of the differential equations relating the curvature of space-time and the energy density in the Universe;

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} - \Lambda g_{\mu\nu} = -\frac{8\pi G}{c^4}T_{\mu\nu} \tag{3.1}$$

where $g_{\mu\nu}$ is the metric tensor, $R_{\mu\nu}$ and R are the Ricci tensor and scalar, G is Newton's gravitational constant and Λ is the cosmological constant.

Careful inspection of each term in the equation, using the Freidman–Lemaître–Robertson–Walker (FLRW) metric (see Appendix A.1), and checking all of the indices $\mu, \nu = 0, 1, 2, 3$ and doing the associated algebra gives us two independent equations:

$$3\frac{\ddot{R}}{R} = -\frac{1}{2}\frac{8\pi G}{c^4}\left(\rho c^2 + 3p\right)c^2 + \Lambda c^2 \tag{3.2}$$

$$R\ddot{R} + 2\dot{R} + 2c^2 k = R\left[\frac{1}{2}\frac{8\pi G}{c^4}\left(\rho c^2 - p\right)c^2 + \Lambda c^2\right] \tag{3.3}$$

Combining these two equations to eliminate $\ddot{R}(t)$ from the second equation, we get

$$\ddot{R} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right)R + \frac{1}{3}\Lambda c^2 R \tag{3.4}$$

$$\dot{R}^2 = \frac{8\pi G}{3}\rho R^2 + \frac{1}{3}\Lambda c^2 R^2 - c^2 k, \tag{3.5}$$

which are known as the Friedmann–Lemaître equations.

Before moving forward with the Friedmann–Lemaître equations, it is convenient to recast the density term into it's components and recast them as dimensionless densities relative to the critical density of the Universe. There are three significant components to the density from a cosmological perspective: matter, radiation and dark energy, Λ, so that the total energy density can be written

$$\rho(t) = \rho_{\mathrm{m}}(t) + \rho_{\mathrm{r}}(t) + \rho_{\Lambda}(t). \tag{3.6}$$

The energy density required to give a flat Universe is

$$\rho_{\mathrm{crit}} = \frac{3H^2}{8\pi G}, \tag{3.7}$$

and thus the densities of each component can be written

$$\Omega_{i=\{\mathrm{m},\mathrm{r},\Lambda\}}(t) \equiv \frac{8\pi G}{3H^2(t)}\rho_i(t), \qquad \Omega_k(t) = -\frac{c^2 k}{H^2(t)R^2(t)}. \tag{3.8}$$

If we consider our three components, and use the Friedmann–Lemaître equations to determine how they evolve with time (See Appendix A) , we find

$$\rho_{\mathrm{m}}(t) = \rho_{\mathrm{r},0}(1+z)^3 \tag{3.9}$$

$$\rho_{\mathrm{r}}(t) = \rho_{\mathrm{r},0}(1+z)^4 \tag{3.10}$$

$$\rho_{\Lambda}(t) = \rho_{\Lambda,0} = \frac{\Lambda c^2}{8\pi G} \tag{3.11}$$

where we have introduced the cosmological redshift z,

$$1+z = \frac{R_0}{R(t)}. \tag{3.12}$$

We will from here forward also used the normalized scale factor,

$$a(t) = \frac{R(t)}{R_0} \tag{3.13}$$

which is convenient because $a_0 = 1$ by definition. Then the redshift z and scale factor are related by

$$a(t) = \frac{1}{1+z}. \tag{3.14}$$

Rewriting equation 3.5 in terms of the dimensionless densities gives the simple equation

$$\Omega_{\mathrm{m}} + \Omega_{\mathrm{r}} + \Omega_{\Lambda} = 1 - \Omega_k, \tag{3.15}$$

where all the variables have time dependance. There are three types of Universes that can be specified by the parameters on the left hand side of this equation:

$$\Omega_{\mathrm{m}} + \Omega_{\mathrm{r}} + \Omega_{\Lambda} < 1 \Leftrightarrow k = -1, \mathrm{open} \tag{3.16}$$

$$\Omega_{\mathrm{m}} + \Omega_{\mathrm{r}} + \Omega_{\Lambda} = 1 \Leftrightarrow k = 0, \mathrm{flat} \tag{3.17}$$

$$\Omega_{\mathrm{m}} + \Omega_{\mathrm{r}} + \Omega_{\Lambda} > 1 \Leftrightarrow k = 1, \mathrm{closed}. \tag{3.18}$$

CMB measurements have essentially limited the parameters to be those that give a flat Universe, where k=0. In this case, $\Omega = 1$ and total energy density is equal to the critical density.

Returning to the Friedmann-Lemaître equations, in particular, equation 3.5, recalling that $H = \dot{R}/R$, and using equations (3.8–3.11) we have an equation for the dynamical behavior of the Universe.

$$H^2 = \left(\frac{da}{dt}\right)^2 = H_0^2\left(\Omega_{\mathrm{m},0}a^{-3} + \Omega_{\mathrm{r},0}a^{-4} + \Omega_{\Lambda,0} + \Omega_{k,0}a^{-2}\right). \tag{3.19}$$

Up until this point, we have only considered *cosmological constant* dark energy, but there are a variety of other more general possibilities for dark energy. In this work, we are especially interested in simple models for time evolution of the dark energy density, especially given by the form

$$\Omega_{\mathrm{DE}}(a) = \frac{\Omega_{\Lambda,0}}{a^{3(1+w(a))}}. \tag{3.20}$$

The most common time evolution model discussed is that from Linder (2003) where

$$w(a) = w_0 + w_a(1-a) \tag{3.21}$$

which introduces additional interesting time evolution into the right hand side of equation 3.19.

3.2 Structure Formation

The previous section describing the smooth expansion of the Universe ignores that the Universe is not perfectly smooth. The existence of stars, galaxies and other dense regions in the Universe today must have been present as small perturbations in an almost perfectly smooth early Universe, that have since grown via gravitational collapse. At some early time, there must have been regions in the Universe that must have differed in their density from the mean density of the Universe $\overline{\rho}$:

$$\delta_x(\vec{x}) = \frac{\rho(\vec{x}) - \langle\rho\rangle}{\langle\rho\rangle}, \tag{3.22}$$

where $\langle\rangle$ denotes a mean value. Birkhoff's theorem (Birkhoff and Langer, 1923; Peebles, 1993) tells us that for small perturbations, we can essentially treat each of them independently and in the Newtonian limit. As such, we can turn to Newtonian mechanics to describe the growth of perturbations. If we write the three standard Newtonian equations in comoving cosmological coordinates, and use equation 3.22 for the density (See Appendix A.2), we arrive at three equations governing the time evolution of density perturbations:

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot\vec{v} = 0 \tag{3.23}$$

$$\nabla^2\tilde{\varphi} = 4\pi G\langle\rho\rangle a^2\delta \tag{3.24}$$

$$\frac{\partial\vec{v}}{\partial t} + H(t)\vec{v} + \frac{1}{a}\nabla\tilde{\varphi} = 0. \tag{3.25}$$

We can eliminate the peculiar velocity by subtracting the time derivative of the first equation from the divergence of the third. Combined with the Poisson equation for $\nabla^2\tilde{\varphi}$, we are left with an equation for the time evolution of density perturbations

$$\frac{\partial^2\delta}{\partial t^2} + 2H(t)\frac{\partial\delta}{\partial t} = 4\pi G\langle\rho\rangle\delta \tag{3.26}$$

For a flat Universe with a cosmological constant ($w = -1$ for dark energy) equation 3.26 has a solution for $\delta(z) = D(z)\delta_0$ given by

$$D(z) = \frac{H(z)}{H_0}\int_z^\infty \frac{dz'(1+z')}{H^3(z')} \tag{3.27}$$

and can be normalized ($D = 1$ at $z = 0$) by dividing by the same integral over all z. This expression is useful for evolving the power spectrum that describes the early density fluctuations to different epochs. However, if the equation of state for dark energy is time-dependent, the solution given by equation 3.27 does not apply. See equation A.75 for the full time-dependent dark energy expression.

As useful as this solution is for investigating the power spectrum, galaxies and clusters of galaxies arise from highly non-linear evolution of the initial density perturbations ($\delta \gg 1$). Using the Press–Schechter formalism, covered in Appendix A, the spatial mass distribution of matter can be treated approximately for $\delta \gtrsim 1$, but to evaluate the full non-linear growth of structure, a more extensive calculation method is necessary. For this task, N-body simulations are useful.

3.3 Cosmological Simulations

Analytic treatments are only able to evaluate cosmic structure growth in the linear regime, but interesting features such as galaxies and clusters of galaxies arise in the highly non-linear regime. As such, numerical techniques are needed to calculate how these structures grow and evolve in time. There are many scales at which these types of simulations are done, for various purposes, and include different physical treatments. The simulation backbone for DES is built on cosmological volume simulations of collisionless massive particles, *i.e.*, N-body dark matter simulations. Cosmological volume simulations sometimes include hydrodynamics to model the baryonic component of the Universe, and include various treatments heating and star formation and other special cases, but lie outside the scope of this work. For details on the effects of gas physics, see Stanek (2009). For reviews of simulations, see Bertschinger (1998); Dolag et al. (2008); Hockney and Eastwood (1988) and Borgani and Kravtsov (2009).

3.3.1 The N-body Problem

The N-body problem is an initial value problem, where, given a set of initial positions $\vec{x}_i$ and velocities $\dot{\vec{x}}_i$ (the dot denoting a time derivative) of a collection of N particles, the system can be described by the Hamiltonian

$$H = \sum_i \frac{\vec{p}_i^{\,2}}{2m_i a(t)^2} + \frac{1}{2}\sum_{ij} \frac{m_i m_j \varphi(\vec{x}_i - \vec{x}_j)}{a(t)}, \tag{3.28}$$

where H is a function of the positions, $\vec{x}$ and momenta, $\vec{p}$. The $\vec{x}_i$ are vectors in comoving coordinates, and the corresponding momentum is $\vec{p}_i = a^2 m_i \dot{\vec{x}}_i$. Rather than calculate the acceleration of each particle by direct summation, simulations often solve the first order formulation of Newton's laws in comoving coordinates (Bertschinger, 1998; Peebles, 1993):

$$\frac{d\vec{x}}{dt} = \frac{1}{a(t)}\vec{v}; \tag{3.29}$$

$$\frac{d\vec{v}}{dt} + H\vec{v} = \vec{g}; \tag{3.30}$$

$$\nabla \cdot \vec{g} = -4\pi G a(t)\left(\rho(\vec{x}, t) - \bar{\rho}(t)\right). \tag{3.31}$$

3.3.2 History

The earliest N-body simulations were done by Holmberg (1941) to study tidal capture of Galaxies by close encounters. He used 37 lightbulbs to represent point masses and photocells to measure the amount of light reaching specific points. He exploited the similarity between gravitational potential and the intensity of light ($\propto 1/r^2$) to map the magnitude and direction of the gravitational force on particles in various configurations. It was then possible to map particle trajectories by graphical integration. This simulation of course, was done without the aid of electronic digital computers. The earliest simulations to use computers were performed by von Hoerner (1960) and

Aarseth (1963) They studied small clusters of point masses, their moment of inertia and velocity dispersions, but were limited to less than 100 particles.

Through the 1970s, the number of particles in the simulations increased, as well as the complexity of what was measured. Early cosmological simulations included Peebles (1970) and S. White (1976). Peebles (1970) used 300 particles and a direct summation technique to find that gravitational interactions over the age of the Universe could form a compact cluster of galaxies like the Coma Cluster. S. White (1976) extended Peebles's calculation to 700 particles to show that large clusters form by amalgamation of smaller clusters. There were additional simulations in the late 1970s to study structure formation, hierarchical clustering and the two-point galaxy correlation function (see Bertschinger (1998) for an extensive list of references). Up to this point, the simulations used a direct summation technique, evaluating all forces pairwise, an $\mathcal{O}(N^2)$ operation.

Then, in the 1980s, interest in simulations exploded, as new computational methods made performing simulations with large particle numbers ($\sim 10^5$) possible. Efstathiou and Eastwood (1981) introduced the Particle-Particle/Particle-Mesh ($\mathrm{P}^3\mathrm{M}$) method to cosmological simulations The $\mathrm{P}^3\mathrm{M}$ algorithm uses direct summation over pair of particles on short ranges, but cuts out direct summation over longer ranges. A mesh grid is constructed, and mass is assigned to each grid point based on the mass within the grid cell. The gravitational potential due to the distant particles can be computed quickly in Fourier space, and varies slowly with time. The algorithm scales as $\mathcal{O}(N \log N)$, and allowed Efstathiou and Eastwood (1981) to use 20,000 particles, and Efstathiou et al. (1985) to show that P^3M is desirable over previous methods to study hierarchical clustering. The resulting data sets show that the virial theorem can provide a good mass estimator for galaxy clusters, and that the same analysis on real data would show $\Omega_{\mathrm{m}} < 1$ (Evrard, 1986; Evrard and Yahil, 1985a,b).

J. Barnes and Hut (1986) introduced a hierarchical tree structure for force cal-

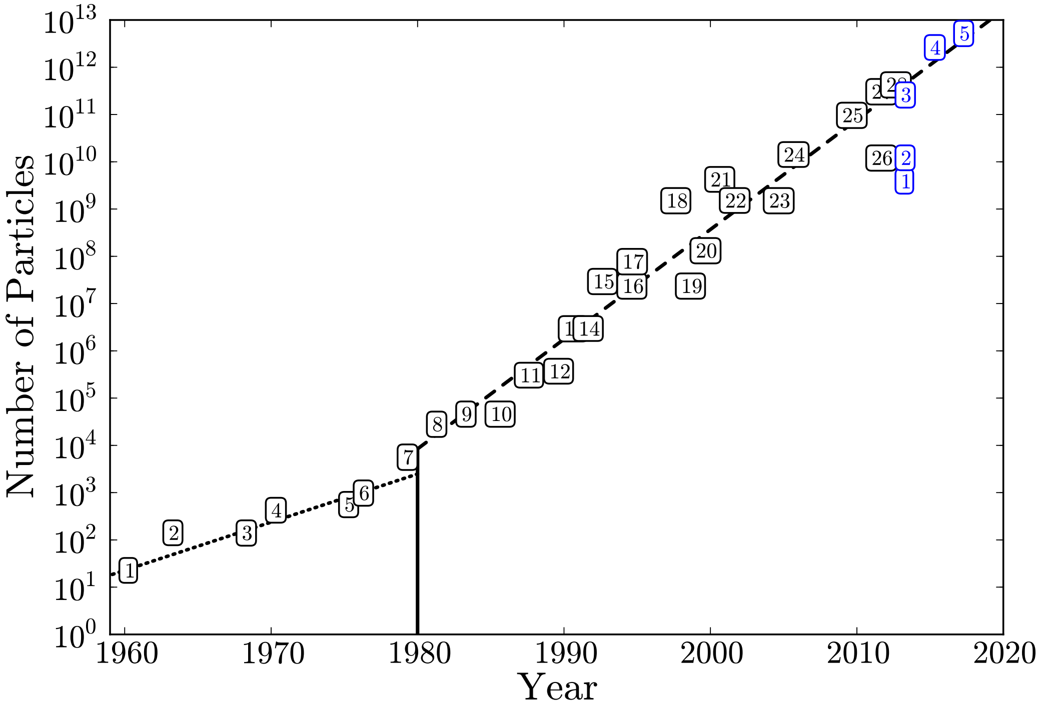


Figure 3.1: Plot showing the historical development of simulation sizes. The vertical line at 1980 represents the time when algorithms more efficient that direct summation started to be developed. Notice the *dotted* line shows how fast simulations were growing with direct summation techniques, while the *dashed* line shows how they have grown with advanced techniques. *Black* points represent simulations by other authors, and *blue* are simulations for this work. The *black* points are (1) von Hoerner (1960); (2) Aarseth (1963); (3) Standish (1968); (4) Peebles (1970); (5) Miyoshi and Kihara (1975); (6) S. White (1976); (7) Aarseth, E. Turner, and Gott (1979); (8) Efstathiou and Eastwood (1981); (9) Melott (1983); (10) Davis et al. (1985); (11) S. White et al. (1987); (12) Carlberg and Couchman (1989); (13) Park (1990); (14) Suto and Suginohara (1991); (15) Warren et al. (1992); (16) Gelb and Bertschinger (1994); (17) Park et al. (1994); (18) Park (1997); (19) Jenkins et al. (1998); (20) Governato et al. (1999); (21) Colberg et al. (2000); (22) Bode et al. (2001); (23) Wambsganss, Bode, and Ostriker (2004); (24) Springel et al. (2005); (25) Teyssier et al. (2009); (26) Klypin, Trujillo-Gomez, and Primack (2011); (27) Kim et al. (2011); (28) Angulo et al. (2012). The *blue* points are DES BCC simulations. Points (1) and (2) represent individual BCC simulations, using 1400^3 and 2048^3 particles respectively. Point (3) is the total number of particles simulated for the BCC so far, five complete cosmologies with a 1400^3 box and four 2048^3 boxes each. Points (4) and (5) represent the two year and four year goals for the BCC which is to run 50 and 100 cosmologies respectively. The comparison between *blue* points (3), (4) and (5) and other simulations may not be fair, because earlier simulations were for single models, while the BCC will sample many cosmological models.

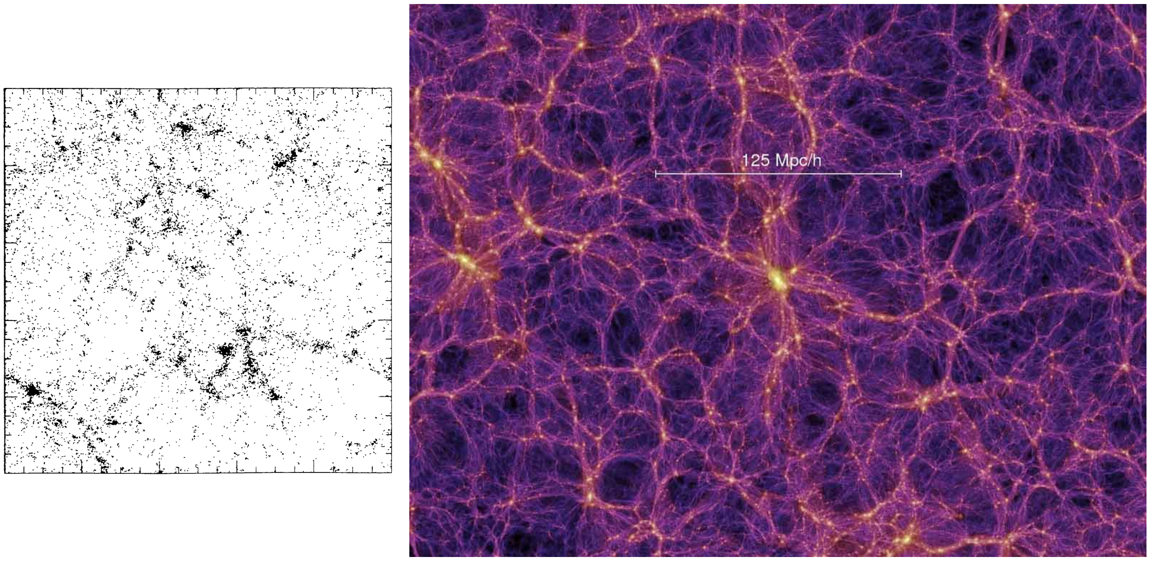


Figure 3.2: Projections from two dark matter simulations. The *left* panel comes from Efstathiou et al. (1985) and shows a two dimensional projection of a 32^3 particle simulation. The *right* panel shows a projection from the Millennium Simulation (Springel et al., 2005) which traced 2160^3 dark matter particles.

culations. Here, at each timestep, the simulation volume is divided into subcells recursively until each cell contains only one particle. Contributions to the force from nearby particles can be calculated by direct summation, while distant particles are grouped into large cells, and the force contribution is calculated from the center of mass of the large cell. The overall algorithm scales as $\mathcal{O}(N \log N)$, but has an advantage over P^3M in that this method does not require any special tuning to the geometry of the problem (*e.g.*, spherical harmonics of Fourier transforms. J. Barnes and Hut (1986)).

Xu (1995) combined the advantages of P^3M for computing long range forces quickly, with the advantages of a tree code for computing short range forces. This eliminates the problems that particle mesh-based codes have in dealing with highly clustered regions. Xu (1995) also added variable timesteps for clustered regions, allowing greater accuracy, and fully parallelized the code for modern computing clusters. The resulting TreePM algorithm is the basis for much of recent N-body simulation work.

For the N-body calculations done for this work, a modified version of the cosmological simulation code Gadget-2 (Springel, 2005). Gadget-2 is a publicly available TreePM + Smooth Particle Hydrodynamics code for cosmological simulations. However, we performed N-body only simulations, and used a modified version of the code dubbed "LGadget-2." LGadget-2 has had all of the gas properties removed in order to reduce the per-particle memory footprint, allowing for more particles per processor and more efficient code scaling than Gadget-2. LGadget-2 was created for the Millennium Simulation (Springel et al., 2005). For details on LGadget-2 and producing N-body simulations in general, see Appendix B, where all of the steps for producing our N-body simulations are discussed in detail.

3.4 Simulations for the Dark Energy Survey

The Dark Energy Survey (DES), introduced in section 2.7, will be the first sky survey to examine all four probes of dark energy—Type Ia Supernovae (SNe), Baryon Acoustic Oscillation (BAO), Clusters and Gravitational Lensing—in a single data set which will cover 5,000 square degrees of the southern sky. Three of these probes—Clusters, BAO and Weak Lensing—are collectively referred to as Large Scale Structure (LSS). Large Scale Structure formation through cosmic time is a highly non-linear process, as described in section 3.2. To study LSS in detail requires N-body simulations to trace non-linear structure growth. An essential feature of all LSS studies is the nature of the input data, which consists of property catalogs of discrete objects that lie along the past lightcone of observers. To calibrate science analysis for a survey like DES, a collection of DES-like synthetic data LSS realizations along the past lightcone of a virtual observer allows real-world observers to test science code, understand systematic errors, and if blinded, test for confirmation bias.

This work describes the methods by which we provide the DES collaboration with full DES-sized synthetic catalogs of varying cosmologies in the parameter space

$$\vec{\theta} = \{\Omega_m, \Omega_b, \sigma_8, n_s, h, \Omega_\Lambda, w_0, w_a, f_{\rm NL}\}\,, \tag{3.32}$$

where most parameters have been described so far, but $h = H_0/(100\ \mathrm{km/s/Mpc})$ and $f_{\rm NL}$ characterizes non-Gaussian initial conditions (A. Becker (2012) discusses $f_{\rm NL}$ extensively, but details of $f_{\rm NL}$ will not be discussed here).

Figure 3.4 shows the generic process for producing a DES catalog, but it omits some details in the N-body simulations. In order to accomplish the DES science goals, it is necessary to accurately model the dark matter distribution on a lightcone over the full volume and dynamic range probed by the DES. This requirement puts competing constraints on the underlying N-body simulation:

1. it must be large enough to probe ~5,000 square degrees to a distance of $\sim 6\ h^{-1}$ Gpc, the survey area and depth of DES;
2. it must have sufficient mass resolution to model the host galaxies brighter than the DES survey sensitivity limit.

The former requirement fixes the length of the periodic simulation cube while the latter specifies the particle mass resolution. Achieving a particle mass of $\sim 3 \times 10^{10}\ h^{-1} M_\odot$ (sufficient to resolve small, low-redshift galaxies) within a single ~6 Gpc volume would require more than 5×10^{11} particles. Such a large simulation is technically feasible at present (Angulo et al., 2012; Kim et al., 2011), but producing an ensemble of such simulations is impractical and unnecessary. We have adopted a more efficient approach using multiple, nested volumes to construct a deep lightcone out of segments of progressively coarser mass and spatial resolution.

Table 3.1 lists the sidelength, L, of the four main simulation volumes used to produce a synthetic DES BCC survey. A fifth, high-resolution simulation is needed to train the Adding Density Determined GAlaxies to Lightcone Simulations (ADDGALS)

Table 3.1: Simulation set for a single cosmology

L (h^{-1} Gpc)	$N_{\rm part}$	$M_{\rm part}(10^{10} h^{-1}\ M_\odot)$	kSU	Storage (TB)
1.05	1400^3	3	45	2.7
2.60	2048^3	15	125	8.4
4.00	2048^3	60	115	8.4
6.00	2048^3	200	115	8.4
0.40	2048^3	0.05	300	28
All	—	—	700	55.9

method. The first (1.05 h^{-1} Gpc) volume resolves nearby galaxies to faint absolute magnitudes. At redshift $z = 0.34$, we switch to the 2.6 h^{-1} Gpc simulation, with mass resolution sufficient to model galaxies above the DES magnitude limit at these redshifts. This transition is repeated two more times, at $z = 0.9$ and $z = 2.0$ to extend the edge of the lightcone to $z = 6$. A schematic of the stitching is shown in Figure 3.3.

The process for producing a synthetic DES optical galaxy catalog is shown in Figure 3.4. This figure also illustrates the process by which multi-wavelength data can be produced from the same underlying data, enabling collaboration across experiments that cover different wavebands, *e.g.*, South Pole Telescope (SPT) for Sunyaev–Zel'dovich (SZ) signals and *X-ray Multi-Mirror Mission–Newton* (*XMM–Newton*) for X-ray signals. The first step, N-body production is performed in three steps. First, we compute the matter power spectrum for the initial redshift of a simulation using the

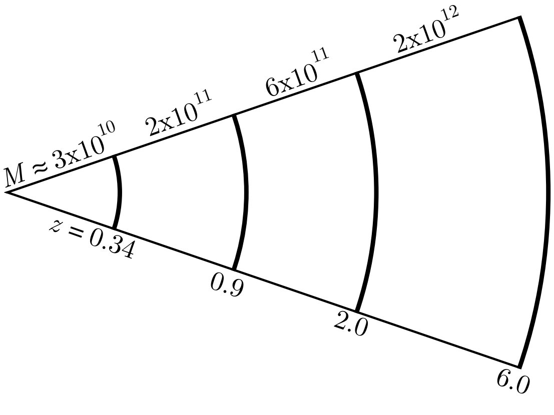


Figure 3.3: Simulation lightcones are stitched together to produce a DES depth catalog. The particle masses for each box size are shown, as well as the redshift at which they are stitched together.

Code for Anisotropies in the Microwave Background (CAMB) (Challinor and Lewis, 2005; Lewis, Challinor, and Lasenby, 2000). CAMB solves the coupled Einstein and Boltzmann equations to calculate the expected clustering of the dark matter—in both Fourier and harmonic space. Then, to calculate initial particle kinematics, we employ the robustly tested Second-order Langrangian Perturbation Theory Initial Conditions (2LPTic) code (Crocce, Pueblas, and Scoccimarro, 2006). This code, as the name suggests, uses second-order perturbation theory to generate initial conditions. With the initial conditions in place, the particles are evolved forward in time using the code LGadget-2, which is a streamlined version of the public code Gadget-2, developed by Volker Springel for the Millennium Simulation (Springel, 2005). Gadget is a hybrid Tree-PM code designed to scale to thousands of compute cores for dark matter simulations (Gadget is described in detail in Appendix B). LGadget produces snapshots—the positions and velocities of all the particles in the simulation at a fixed time—and lightcones—the dark matter along the past lightcone of a virtual observer in the simulation box. We archive all the raw snapshot and particle data to a storage system, and perform a number of analysis steps on the data.

For each simulation snapshot, we compute the distance to the nth nearest neighbor for each particle, with n depending on the simulation resolution. The distance provides a local density estimate for each particle in each output snapshot. We also calculate the distance to the nth nearest neighbor for lightcone output for each simulation. For the nearest-neighbor calculation, we use CALCRNN, a code developed by Matthew Becker specifically for our purposes (details can be found in Appendix C.1). We also perform dark matter halo finding and compute halo properties on the output snapshots and lightcones. For halo finding, we use the Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement (ROCKSTAR) halo-finder (Behroozi, Wechsler, and Wu (2013); described in Appendix C.2). We also build halo merger trees using CTREES (Behroozi et al., 2013), which provides additional infor-

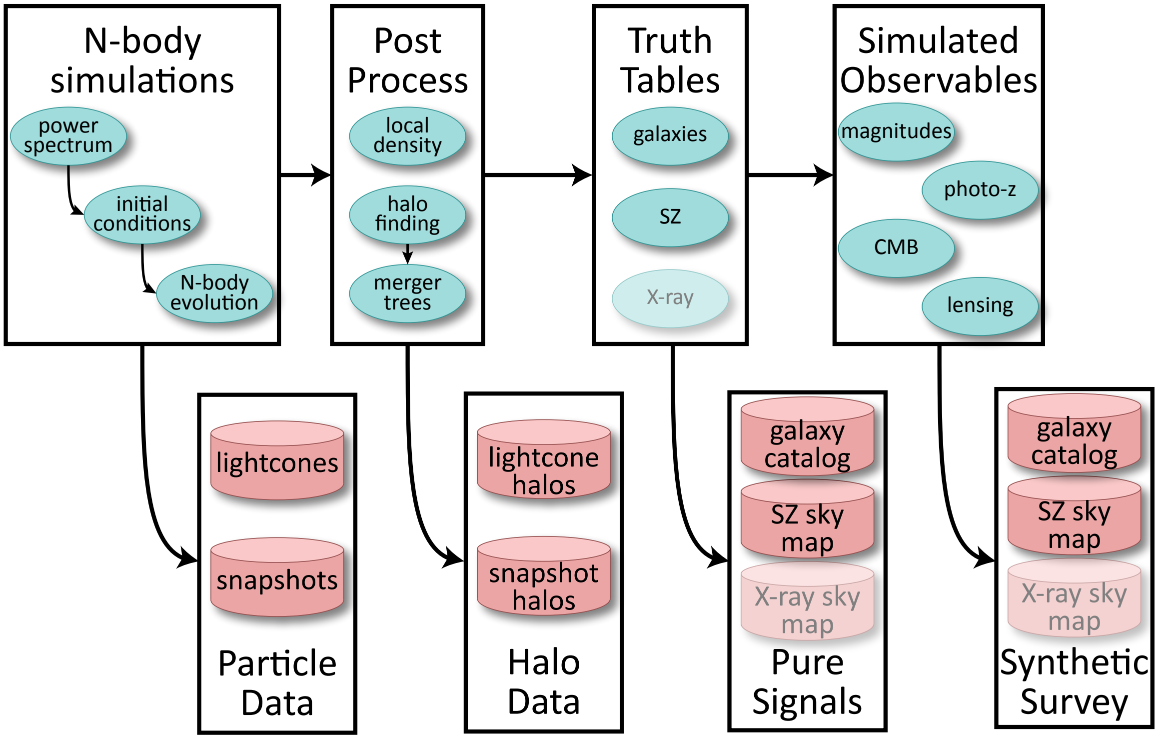


Figure 3.4: This diagram shows a schematic of the full process for producing a multi-wavelength synthetic sky. Each large block in the *upper* row shows a major computational step, and the inner ovals show the sub-steps for each major step. The *lower* row shows the data products that are produced by each computational block. See text for details on each step.

mation about the dark matter halos and informs later galaxy assignment. Then, we compute the distance to the nth nearest particle neighbor for the dark matter halo centers in all of the cases where we performed halo finding. With this suite of density estimates, dark matter halos and associated properties, we proceed to assign galaxies to the matter distribution.

We dress the dark matter distribution with galaxies brighter than the DES limiting magnitudes in each passband using the ADDGALS algorithm (Busha and Wechsler, 2013). The galaxy catalogs contain all of the *true* information about the galaxies and their parent dark matter halos from the simulation. From those truth values, we then proceed to create observational catalogs, which contain values that correspond to the actual quantities that DES will observe.

To calculate the effect of weak gravitational lensing, and in turn the *observed* galaxy properties, we use Curved-sky grAvitational Lensing for Cosmological Light conE simulatioNS (CALCLENS) (M. Becker (2012); described in Appendix C.5). We also collaborate with Tomasz Biesiadzinski to produce SZ maps matching observational properties of the SPT. Since our simulations are dark matter only, the SZ halo profiles are created by laying an Arnaud gas profile (M. Arnaud et al., 2010) down at the halo location. At the time of writing, we have not developed the ability to produce X-ray flux maps based on the halo catalogs, illustrated by the faded appearance of the X-ray products in Figure 3.4. Producing these catalogs would be an excellent future project to expand the scope of these synthetic skies. Note that for DES-only science, neither the SZ nor X-ray skies are necessary, though they do allow for collaboration with surveys in the SZ and X-ray wavebands that have survey are overlapping DES. The final *observed* galaxy properties included are the magnitudes in each DES survey filter band ($grizY$), as well as weak gravitational lensing properties. We also calculate photometric redshift probability distributions, $P(z)$, for each galaxy using the zCarlos code (based on Cunha et al. (2009)), and provide photometric redshift values for galaxies from the ARBORz (Gerdes et al., 2010) and ANNz (Collister and Lahav, 2004) algorithms.

The final product of this entire process is a synthetic DES sky survey in the form of an *observed* galaxy catalog, as well as underlying *truth* about the dark matter and galaxies, with corresponding SZ signals. These catalogs can then be tested and analyzed by the full DES collaboration for code validation and quality control. Cosmological parameters derived from DES science analysis codes can then be compared against the original model and parameters used to generate the synthetic survey, to characterize the fidelity of cosmological analysis in the collaboration.

A collection of simulations with the parameters unknown by the science analysis teams would form the basis of a Blind Cosmology Challenge (BCC).

3.4.1 The Blind Cosmology Challenge

The Dark Energy Survey is embarking on a BCC to ensure reliable cosmological analysis on real science data. The BCC is built on the synthetic observational surveys described in the previous section. The general idea is to sample cosmological parameter space given in equation 3.32 in such a way as to test science analysis code for a variety of models. We want to make sure that the codes can return a number of possible models, avoiding confirmation bias, and that the analysis returns the same cosmology used to generate the synthetic surveys.

Toward a BCC we have already produced four different N-body models, of which, two have been fully transformed into observational catalogs. However, in the near term, we would like to expand the number of cosmologies to at least ten, and in the longer term, on the order of one hundred. Looking at the steps laid out to produce a synthetic catalog in the previous section, we can see that there are a number of heterogenous computational tasks that need to be managed. Producing one hundred simulations could require as many as 500 separate computing jobs to be managed at the computing centers used to run the simulations. To expedite these jobs, and remove human management, we look to grid-based cyberinfrastructure.

3.5 Cyberinfrastructure

Producing an entire simulation set requires significant time and effort to manage job submissions. Every compute cluster available for running DES simulations has different sets of queue policies, limiting the number of jobs that can be run in parallel, and limiting the amount of time that each job can run. For example, the Texas Advanced Computing Center (TACC) Ranger system, where we performed most of our early production simulations, limits a 1024 cpu core job to no more than 48 hours of continuous run-time. Given the total runtime of our simulations, that means a single job requires three submissions to the queue. Combined with the power spectrum calculation, and the initial conditions calculation, that makes five submissions per simulation box, plus 7–9 more for the high resolution tuning box (bottom row in Table 3.1). Adding the post-processing steps, nearest neighbor calculations and halo finding, a single simulation box may require twenty separate job scripts and submissions, and a full cosmology may need close to one hundred. Plus, each cosmology generates nearly 60 TB of data that need to be managed, documented and stored for later access. For the several cosmological models we intend to run, the task of managing data and jobs becomes error prone and inefficient for an individual using a standard terminal window and command line.

An automated solution would reduce errors in job submission, increase efficiency and reduce the learning curve for running BCC simulations. Part of this thesis documents development errors and results of automating the BCC production process. Ideally, this solution will provide a basis for other surveys that will need simulation support to improve their simulation quality and efficiency.

Recent cosmological simulations have been performed on high-performance computing clusters. The standard procedure is to consume a large number of CPU-hours to generate simulated data, then, move the data to local disk and perform scientific analysis for publication. However, as algorithms and compute resources have advanced, data products from simulations have grown in size and complexity so that the standard production mechanism should be reexamined. Dark matter simulations and their associated data products for DES generate $\sim$50 TB of information per cosmological model. Moving 50 TB of data via the internet can be extremely time consuming—almost 5 days at 1 Gbps and 48 days at 100 Mbps. If computational research groups were to regularly generate this much data, university systems would quickly be saturated with data and data transfer. Additionally, data output from a single simulation is sufficient to support many analysis projects. When taken together, these points suggest that computational cosmology will pass a critical point, from a data-limited state to an insight-limited state if computational power is not efficiently harnessed.

For the past 30 years, generating ever larger simulations was the driving factor in computational cosmological research (see Figure 3.1). Adding more particles, or increasing resolution, while adjusting cosmological parameters was sufficient to generate new and useful results. The important difficulties to address were how to improve algorithms, or how to access larger computing resources. Now, computing power has scaled to a level that accessing resources and producing enormous data sets is not hard, but making use of the large data sets is more so. There needs to be a shift in research processes to harness the data we now generate, and increase the output of scientifically rich insights. Computational cosmology is not the only field entering the domain of data rich science, many others are facing a similar transition. For an overview, see Hey, Tansley, and Tolle (2009). Some groups and fields have worked to develop infrastructure to more efficiently use data, often built around the concept of grid computing. A shift to this data-intensive style of research can be enabled by IT specialists by providing cyberinfrastructure and grid based computing solutions. Grid computing is usually taken to be federated computer resources that can be used for on-demand computing and on-demand storage, where sharing and transferring data are optimized across the federation for simplified collaboration when compared to trying to work across various institutional networks. The motivation behind developing grid computing environments is to reduce the amount of computer science work that domain scientists need to perform and to maximize the amount of time they spend on data analysis. One example of a computing grid is eXtreme Science and Engineering Discovery Environment (XSEDE)

3.5.1 XSEDE

XSEDE is an NSF supported grid computing environment that follows the TeraGrid (Berman, 2001; Towns, 2011) and is, in the words of the National Science Foundation (NSF):

> the most advanced, powerful, and robust collection of integrated digital resources and services in the world. It is a single virtual computing system that scientists can use to interactively share resources, data and expertise.

XSEDE is funded by a $121 million, five-year NSF grant and is designed to integrate resources and services, making them easier to use. XSEDE, led by the University of Illinois's National Center for Supercomputing Applications, supports 16 supercomputers, visualization, data analysis and data storage facilities across the country. It is designed to streamline research by offering support for developing and optimizing applications for the systems supported.

DES has made use of supercomputers under the XSEDE umbrella for running BCC simulations, but running simulations alone does not use the grid nature of XSEDE to its full extent. XSEDE encourages researchers to create *gateways*, which make it much easier to generate and share data collaboratively. A science gateway is a collection of tools, applications and data that are integrated via a common interface. Often a gateway is a web service that allows users associated with the gateway to access national resources to further their research goals with little application development required. DES has engaged with XSEDE staff in an effort to build a science gateway for cosmological simulations, first for internal DES use, but ideally one which could be exposed to the larger research community, allowing for much more efficient simulation production and analysis than can be accomplished by individual researchers working on their own data products.

3.5.2 Toward Data Rich N-body Production

Not so long ago, exploring the possibility of adding features to codes was possible by highly-motivated and hard-working individuals on a reasonable time scale (*e.g.*, Evrard (1988)). New simulation programs are massively parallel, sometimes requiring functions built specifically to match the physical layout of a particular computer to optimize performance (Habib et al., 2012). Long gone are the days when an individual could produce code for a useful end-to-end simulation and analysis in a reasonable timescale. If domain scientists spend 90% of their time developing code to produce or reduce data for analysis, and only 10% performing scientific analysis of the tasks there is an enormous loss of potential scientific insight. With this in mind, it is clear that an online collaborative environment, where a small number of scientists specialize in developing optimized code, and others specialize in tools for analyzing data, is desirable. In such and environment, scientists could invert the time split; 10% computer science work, and 90% domain science work. This would vastly improve efficiency in producing scientific results per unit simulated (or experimental) data.

For this work, I have been developing software for a gateway for cosmological simulations. Ideally, there would be a web interface where a user could select cosmological models, parameters, simulation size, resolution, and output options. After selecting the options, the gateway would be capable of selecting the best grid computing resource for the simulation and launch the job. When data is ready, the user requesting the simulation should be notified, so the data could be further analyzed, or served to the user if they wanted to move it to local resources. Also, the idea environment would have a database of completed simulations and data products that could be shared with the research community, so that researchers would not repeat simulations if one similar or identical had already been produced. We would like to have a gateway that could provide synthetic surveys on-the-fly for the DES collaboration/

Our first task is to collect and import codes that we can use immediately with a purpose and start to wrap them in workflow middleware that launches compute jobs and then build toward a gateway service. The BCC for DES has provided an opportunity to build such a service. Here, our goal is to produce an interface where a DES simulator need only select a cosmological model and parameters, and a full BCC simulation will be run with some data reduction, and the simulator being notified when the task is complete. I detail progress on this task in Chapter 4.

3.6 Conclusion

To produce cosmological simulations is no small task. The N-body problem must be solved in an expanding background, and simulations need to have sufficiently high resolution to model survey observables. To efficiently produce these simulations requires harnessing significant computing resources, and for maximum efficiency, those resources should be managed by automated middleware. With a fully automated cosmology portal, DES will be able to quickly produce synthetic surveys, which will allow science analysis teams to test their code for a variety of cosmological models. This sort of testing should improve science output from the survey when it produces real data.

CHAPTER 4

An Environment to Efficiently Produce Synthetic Skies

The Dark Energy Survey (DES) (Annis, Bridle, et al., 2005; Annis, Castander, et al., 2005; The Dark Energy Survey Collaboration, 2005) is a Stage III[1] dark energy project jointly sponsored by Department of Energy (DoE) and National Science Foundation (NSF) that will soon begin survey operations. The project has deployed a new 560 Megapixel panoramic camera on the Blanco 4-m telescope at the Cerro Tololo Inter-American Observatory (CTIO) in Chile. This instrument will be used to image ≈5000 square degrees of the sky in the South galactic cap in four optical bands, and to carry out repeat imaging over a smaller area to identify distant type Ia Supernovae (SNe) and measure their distances. The camera began science verification in fall 2012—discovering four new SNe (Abbott, Abdalla, Achitouv, et al., 2012)—and survey operations will begin in fall 2013.

The main imaging area of the DES overlaps the South Pole Telescope (SPT) (Carlstrom et al., 2011) sub-mm survey that will identify galaxy clusters via the Sunyaev–Zel'dovich (SZ) effect (Sunyaev and Zel'dovich, 1972) as well as the Visible and Infrared Survey Telescope for Astronomy (VISTA) infrared survey of galaxies, which will provide additional information on galaxy photometric redshifts and on the properties of galaxy clusters at large cosmological redshift, $z > 1$. Roughly three

---

[1] In the language of the Dark Energy Task Force, see Albrecht et al. (2006)

hundred scientists across nearly thirty institutions comprise the DES collaboration.

The DES will be the first project to combine four different methods to probe the properties of the dark sector (dark matter and dark energy) and test general relativity gravity via evolution of the Hubble expansion parameter and the linear growth rate of structure. The methods—baryon acoustic oscillations in the matter power spectrum, the abundance and spatial distribution of galaxy groups and clusters, weak gravitational lensing by large-scale structure, and type Ia supernovae—are quasi-independent. Each has sources of systematic error associated with it, some of which are unique to the method but some of which are shared. Examples of the latter are the accuracy of photometric redshift estimates (Cunha et al., 2012), line-of-sight contamination of foreground galaxies (Erickson, Cunha, and Evrard, 2011), the form of the non-linear matter clustering power spectrum, and shape measurement errors for galaxy images that affect cosmic shear and galaxy cluster mass estimates. DES will thus be the first survey to address joint systematics in multiple methods probing accelerating expansion of the Universe. N-body simulations provide key support for the analysis of systematics in the three methods associated with cosmic large-scale structure (all but supernovae in the above list). To validate science analysis codes, the DES Simulations Working Group is coordinating a Blind Cosmology Challenge (BCC) process, in which a variety of sky realizations in different cosmologies will be analyzed, in a blind manner, by DES science teams. The BCC process will enable a variety of DES science, as well as collaborative science, particularly with the SPT, detailed in the next section. I have described the BCC in some detail in Chapter 3, so I will give a brief reminder here, and then describe our efforts to automate BCC production using eXtreme Science and Engineering Discovery Environment (XSEDE) software and resources.

4.1 The Blind Cosmology Challenge

The BCC process will require generating multiple galaxy catalogs to the full photometric depth across the full 5000 square degrees of the DES survey.

The dark matter structure from N-body simulations of a given cosmology forms the basis for galaxy catalog expectations. Generating a large cosmological BCC simulation has two principal steps. The first is to realize an appropriate set of initial conditions within a chosen cosmological model. We generate an initial matter-power spectrum for a given cosmological model using Code for Anisotropies in the Microwave Background (CAMB) (Lewis, Challinor, and Lasenby, 2000). To produce accurate initial particle kinematics, we use Second-order Langrangian Perturbation Theory Initial Conditions (2LPTic) (Crocce, Pueblas, and Scoccimarro, 2006). We evolve the N-body particle set using a streamlined version of the Gadget-2 code developed by Volker Springel for the Millennium Simulation (Springel et al., 2005). The code scales well for large-volume simulations for which the tree calculations are mainly computed within a single node. We store particle configurations, $\{x_i(t), v_i(t)\}$, in two forms: *snapshots* of the positions and velocities of all particles i in the simulation volume at a fixed time t, and *lightcones* (Evrard et al., 2002) that hold kinematic information for particles lying on the past lightcone of a virtual observer located at a fixed position, x_{cen}, in the computational volume.

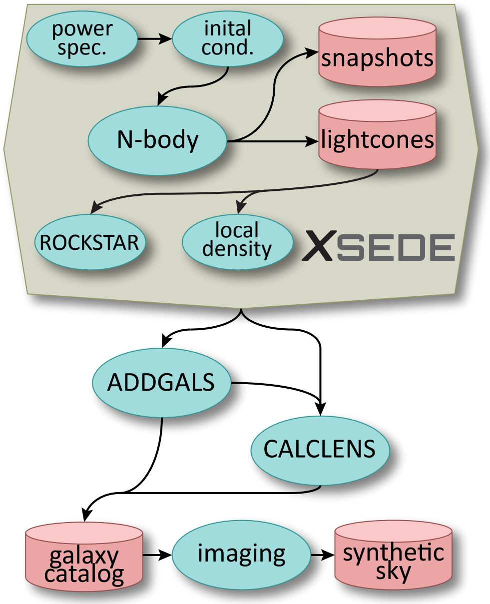
Dark matter halos, bound systems that host galaxies and clusters of galaxies, are identified in these outputs and their properties can be used to determine their central galaxy characteristics. Halos, as well as a local density estimate are used by the Adding Density Determined GAlaxies to Lightcone Simulations (ADDGALS) algorithm to assign galaxy properties to suitably selected dark matter particles. The matter along the past lightcone also sets the gravitational lensing shear signal applied to these galaxies, and we use the Curved-sky grAvitational Lensing for Cosmological Light conE simulatioNS (CALCLENS) algorithm to apply weak lensing signals to the galaxy catalog.

4.2 Workflow Abstractions

The simulation codes discussed in section 3.4 are executed on large scale XSEDE resources where heterogeneity and complexity in interfacing with batch resources managers slow scientists in harnessing the vast amount of available computing power. Scientific workflows are a prominent abstraction tool that allow scientists to carry out their scientific discovery and experimentation without having to completely understand underlying systems that are constantly evolving (Belhajjame et al., 2008; Curcin and Ghanem, 2008; Goecks, Nekrutenko, and J. Taylor, 2010; Ludäscher et al., 2006; Marru et al., 2011). These abstractions lower the entry costs and learning curves to using computing resources. The also help reduce human inefficiency and errors in running jobs, and can provide detailed provenance.

To build our cosmological workflow, we leverage the experience and software developed by the Open Gateways Computing Environments (OGCEs) project (Pierce et al., 2010) facilitated by the XSEDE Extended Collaborative Support Services. The workflow infrastructure is based upon the Apache Airavata (Marru et al., 2011) framework. The Airavata workflow system is primarily targeted to support long running scientific applications on computational resources. Airavata's XBaya is a graphical workflow tool, allows composition, execution and monitoring of tasks. The Airavata workflow engine, dubbed "XBaya," requires these applications to be conform to a specific Application Programming Interface (API) that allows XBaya to execute the applications using a standard protocol. The Airavata Generic Application Factory (GFac) component bridges this gap between applications and the workflow systems by providing a network accessible web service interface to the scientific application.

Figure 4.1: Processing steps to build a synthetic galaxy catalog are illustrated here and described in the text. The Apache Airavata suite currently controls the elements shown in the *tan* polygon which consists of methods to sample a cosmological power spectrum, generate an initial set of particles, evolve the particles forward in time with LGadget (N-body) and perform basic post processing—dark matter halo finding and local density estimation. The remaining methods are run manually on other resources, mostly at SLAC.



Once the simulation codes are deployed on XSEDE computational resources, we register descriptions of these applications with the Apache Airavata registry service. These descriptions are used by the Airavata GFac component to generate the artifacts required to expose the application as a service. The workflow developer can access these wrapped application services and construct workflows and orchestrate executions on target compute resources.

4.2.1 Implementation

To illustrate the construction of a cosmological workflow, we will describe developing the N-body simulation workflow illustrated in Figure 4.1. First, the nature of the applications, their execution characteristics, and their input and output data are analyzed. The application meta information, including the executable location, its nature (*e.g.*, serial, MPI, TCP/IP), inputs and outputs, are described and registered with Airavata registry. This process was followed for the following four applications:

BCC Parameter Maker

This initial setup code is written as a PYTHON script and prepares necessary configurations and parameter files for the workflow execution. This simple script is forked on the XSEDE Ranger job management nodes.

CAMB

The Code for Anisotropies in the Microwave Background application computes the power spectrum of dark matter, which is necessary for generating the simulation initial conditions. This application is a serial FORTRAN code. The output files are small ASCII files describing the log-binned power spectrum.

2LPTic

The Second-order Langrangian Perturbation Theory Initial Conditions code is an MPI C code that computes the initial conditions for the simulation from the input power spectrum generated by CAMB. The output of this application are a set of binary files that vary in size from $\sim$80–250 GB depending on the simulation resolution, scaling with the number of particles.

LGadget

The LGadget simulation code is MPI based C code that uses a TreePM algorithm to evolve a gravitational N-body system. The outputs of this step are system state snapshot files, as well as lightcone files, and some properties of the matter distribution, including diagnostics such as total system energies and momenta. The total output from LGadget depends on resolution and the number of system snapshots stored, and approaches 10 TB for DES simulation volumes.

ROCKSTAR The ROCKSTAR application is a C code parallelized with a specially written TCP/IP communication library for efficient communication. This code identifies halos by first running a percolation algorithm on the simulation particles to determine how to split particles between processors. The code performs a hierarchical phase space refinement to find structures in the particle distribution. The output is a collection of binary or ASCII files containing information that describes identified dark matter halos, including positions, velocities, angular momentum and several mass measurements—*e.g.*, $M_{200c}, M_{500c}, M_{vir}$.

CALCRNN This is a hybrid MPI/OPENMP algorithm written in C that uses a simple k-d tree to perform nearest neighbor searches. The novel part of the code is that it reduces interprocessor communication overhead by using the Peano-Hilbert space filling curve domain decomposition from Gadget to identify which parts of the Gadget output each processor needs to read to perform the nearest neighbor search. The output is a set of binary files with particle ids and distance to the nth nearest neighbor.

After all the above applications are registered, the BCC workflow is constructed using Airavata XBaya. With the workflow in place, all of the jobs can be executed with appropriate dependency on an XSEDE resource by setting a master parameter file and beginning execution with a single click. Users can be notified by email as steps complete, and when all of the computations have completed. The nature of the ADDGALS algorithm requires it to be run at SLAC, so the final step in our current workflow process is a gridFTP data transfer of necessary data products to SLAC for final processing, as well as an archive or all data products to the Texas Advanced Computing Center (TACC) Ranch archive. An example workflow graph is shown in Figure 4.2. This figure is somewhat simplified from a workflow for a full set of cosmology boxes, in that it only shows two simulation box sizes. We can follow the data and computational flow by reading from left to right across the figure. In the figure, solid lines show data flow, and dotted lines indicate dependency. The leftmost boxes are the basic run parameters, the directories containing the python script to set-up the simulation boxes and the output directories for the various applications. The first large box is the python set-up scripts, which accepts the initial input parameters and outputs locations of parameter files and executables for every other task. The first task to execute is CAMB, followed by 2LPTic for each of the simulation boxes.

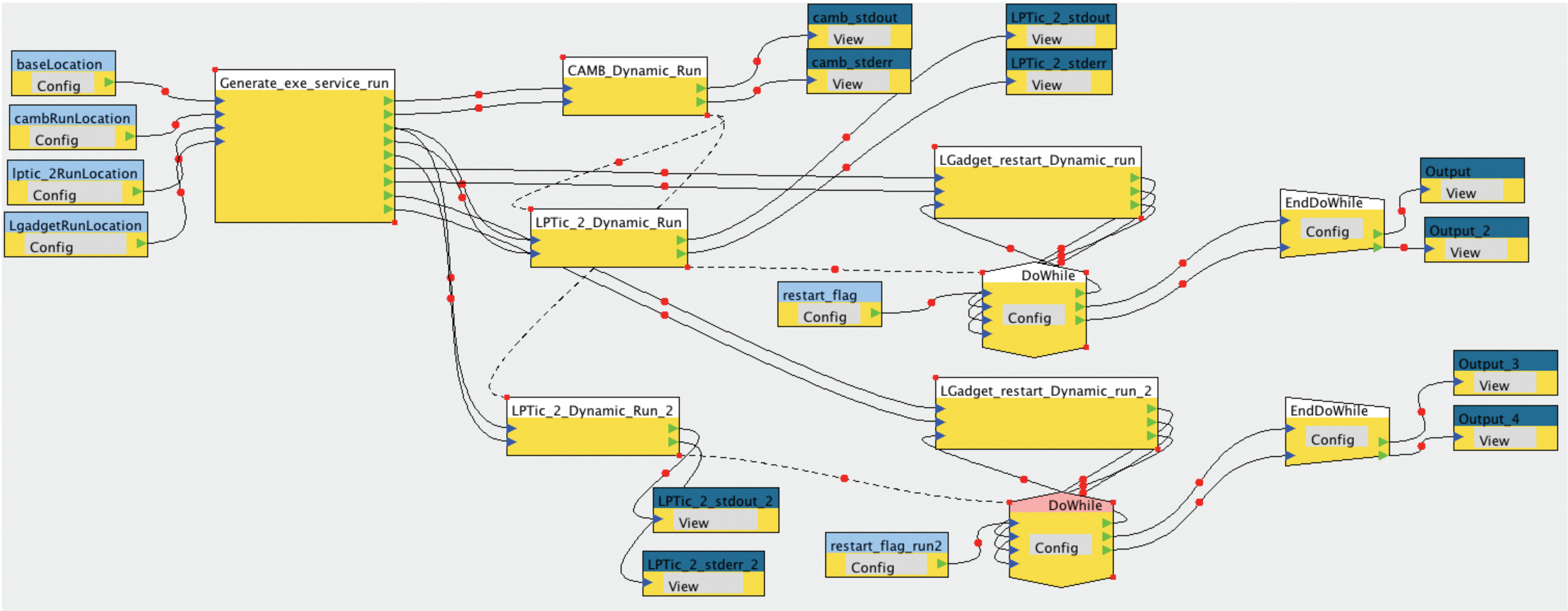


Figure 4.2: A screenshot of the XBaya graphical workflow composer. In this example, scripts for running two simulation boxes are generated, then initial conditions are generated and then LGadget is launched. *Light blue* headers are input parameters, *dark blue* headers are output and *white* headers are applications. Solid lines show where output is passed to the next application, and dotted lines indicate job dependency. The LGadget step includes a new do-while construct that automatically resubmits a job to start from a checkpoint if the computation is not completed in the 48 hours wallclock limit.

Finally comes LGadget, which is wrapped with a DoWhile construct, which ensures that LGadget executes at least once, with the possibility of executing many times if necessary to reach the end of the simulation.

4.2.2 Workflow System Enhancements

Apache Airavata is being constantly updated to reflect the needs of its science use cases on XSEDE resources. Executing the DES BCC workflow required new features to be added to the Airavata structure by the Airavata team, in particular by our collaborator Raminderjeet Singh.

Iterative execution support for long running applications. The N-Body simulation requires multiple days of execution, but the XSEDE Ranger and Stampede clusters limit maximum wall time to 48 hours. To mitigate this limitation, the workflow infrastructure has to allow iterative support so the job can be broken down into multiple increments of 48 hour jobs harnessing the check-point restart capabilities within the application. These capabilities required sophistication beyond the blind restarts, in order to account for application execution patterns and exception handling. To add these capabilities, we have added a formal Do-While construct semantic characteristic of workflow engines. The Do-While construct requires that an application execute at least once, but it may execute an indefinite number of times depending on a condition. In our case, the condition to stop execution is LGadget reaching the final simulation time, if it hasn't, LGadget is submitted again.

Output Transfers. The workflow executions tend to produce terabytes of data residing on the cluster scratch file systems to long persistence archival systems. The data movement to archival systems like TACC Ranch for long term storage have to be provided. The large file data movement is non-trivial process. Even though advancements have been made in this area, seamless reliable data transfers are still challenging. The emerging solutions like Globus Online (`https://www.`

`globusonline.org`) GridFTP client API (`http://www.globus.org/toolkit/docs/latest-stable/gridftp/rn/#gridftpRN`) and bbcp (`http://www.slac.stanford.edu/~abh/bbcp`) are potentially viable options. Globus Online would be the best of the three options based on testing, however, the command line interface to Globus Online is not sufficiently stable to implement in the workflow. In the meantime, we have implemented GridFTP in a test workflow, but have been using the web interface for Globus Online in most cases.

4.3 Results

We have run a number of cosmological models on XSEDE resources using the Airavata XBaya workflow engine, and a few individual boxes *by hand*. For a given cosmology, we generate five N-body simulations in nested volumes, consisting of three large-volume realizations with 2048^3 particles, one small volume realization with 2048^3 particles, and one intermediate volume of 1400^3 particles. This approach allows a better match to halo mass selection imposed by the magnitude-limited nature of the DES galaxy sample. The mass resolution varies by nearly a factor of 100 from our smallest to largest volumes. A halo resolved by a minimum of 100 particles ranges from a mass of $3 \times 10^{12}\ h^{-1}\ M_\odot$ in the near-field simulation to $2 \times 10^{14}\ h^{-1}\ M_\odot$ in the far-field. We nest the simulations as mentioned in Chapter 3 and illustrated in

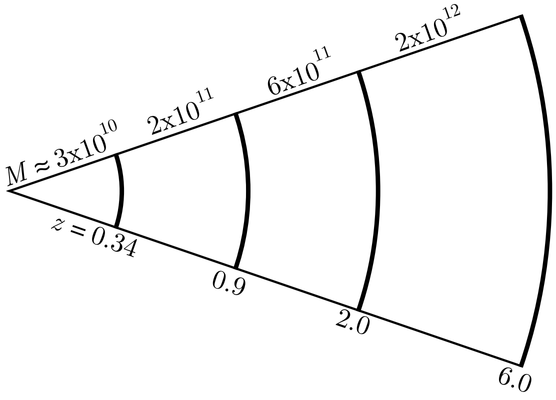


Figure 4.3: Simulation lightcones are stitched together to produce a DES depth catalog. The particle masses for each box size are shown, as well as the redshift at which they are stitched together. Described in more detail in section 3.4.1.

Table 4.1: Kilo-SUs (1 SU = 1 cpu-hour) consumed producing BCC simulations on XSEDE resources in 2012–2013. For these simulations, we have consumed nearly 1.5 million SUs and produced approximately 200 TB of data to be reduced to final galaxy catalogs.

	Size: L (h^{-1} Gpc)						data
Sim	0.4	1.05	2.60	4.00	6.00	total	in TB
Aardvark-2	—	25.2	67.7	57.4	53.8	204.1	30
Beluga	205.8	23.3	65.3	55.7	52.8	402.8	60
Chinchilla-1	239.5	38.0	100.8	94.7	—	473.0	50
Chinchilla-2	—	14.1	71.7	59.2	54.5	199.5	30
Chinchilla-3	—	25.7	72.9	60.5	56.2	215.3	30
total						1,494.6	200

Figure 4.3 to produce DES depth lightcones. The former is roughly the mass of our Milky Way galaxy's halo while the latter corresponds to the mass scale of clusters of galaxies. The computational time and data produced from those simulations are summarized in Table 4.1.

Each simulation produces lightcone outputs centered on each of the eight corners of the computational volume. By employing the periodic boundary conditions of the computational domain, we can stitch these octants of sky into a single 4π representation of the full past lightcone of a hypothetical observer placed at the origin of the simulation. A map of the resultant structure in a thin radial slice of synthetic sky is shown in Figure 4.4.

4.3.1 Efficiency Gains(?) with XBaya

The XBaya workflow shown in Figure 4.2 was tested and refined using smaller simulations over the period October 2011 to June 2012. From June 2012 to mid-August 2012, some production simulations were run, but with a combination of manual and XBaya management due to some bugs being discovered that had not been vetted during the testing period. From late-August 2012 to December 2012, we had a burst

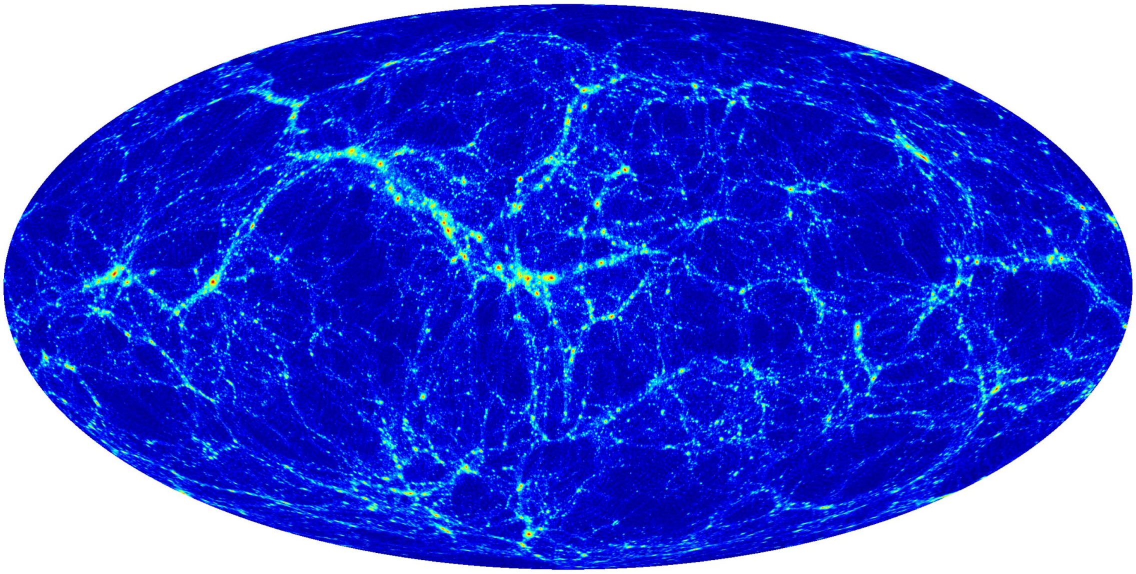


Figure 4.4: Full-sky image of the dark matter density in a thin radial slice (50–75 h^{-1} Mpc distance from observer) taken from one of our 1050 h^{-1} Mpc ΛCDM simulation. Color maps the local matter density relative to the mean value on a logarithmic scale ranging from -1 (blue) to roughly 500 (red). Figure courtesy of Matthew Becker (University of Chicago).

of simulation production, as we used XBaya to manage several jobs, and at the same time ran a few additional jobs by hand. The production period in the last quarter of 2012 provides an interesting set of run times to investigate if, and to what extent, the XBaya workflow was able to improve our production efficiency. Jobs were submitted to the long queue at TACC Ranger, which has maximum resource limits of 1024 processors and 48 hour runtime.

The first two columns of Table 4.2 give the simulation run, and the third column gives the method used to manage the job. The next two columns give: Total Time, the real world wallclock time interval for the entire production process (dd:hh:mm:ss); and the CPU Time—the sum of the run times of the required jobs. Times reflect the full N-body production process, from generating initial conditions all the way through to completing of the final N-body timestep. Efficiency is the ratio of CPU to total time, with 100% representing the ideal scenario of running without interruption and with no queue wait time.

Table 4.2: Comparison of Manual and Workflow-enabled production times (dd:hh:mm:ss) for production runs between September and December 2012.

Simulation	Size	Control	Total Time	CPU Time	Efficiency
Beluga-xbao	1050	Manual	4:05:39:07	2:17:43:43	64.8%
Beluga-xrnsd	2600	XBaya	3:01:56:41	2:17:35:50	88.4%
Beluga-xbao	2600	Manual	8:15:33:05	4:07:24:10	50.0%
Beluga-xbao	4000	XBaya	2:09:53:23	2:05:08:59	92.4%
Beluga	1050	XBaya	2:22:02:41	2:00:37:16	69.4%
Beluga	2600	XBaya	4:09:22:21	2:19:09:50	63.7%
Beluga	4000	XBaya	4:10:59:04	2:07:16:45	51.7%
Beluga	6000	XBaya	3:12:35:25	2:04:30:52	62.1%
Chinchilla-1	1050	Manual	4:09:46:18	3:02:49:49	70.7%
Chinchilla-1	2600	Manual	4:12:45:04	4:03:19:00	91.3%
Chinchilla-1	4000	Manual	5:05:07:26	3:22:14:43	75.3%
Chinchilla-2	1050	XBaya	1:20:39:18	1:07:51:23	71.3%
Chinchilla-2	2600	XBaya	10:13:16:03	3:08:14:43	31.7%
Chinchilla-2	4000	XBaya	5:07:11:26	2:13:23:15	48.3%
Chinchilla-2	6000	XBaya	6:16:51:06	2:12:14:13	37.4%
Chinchilla-3	1050	XBaya	2:04:42:56	2:02:36:42	96.0%
Chinchilla-3	2600	XBaya	4:05:45:11	3:02:44:09	73.4%
Chinchilla-3	4000	XBaya	3:04:18:18	2:13:12:00	80.2%
Chinchilla-3	6000	XBaya	4:14:54:22	2:09:00:24	51.4%
Total		XBaya	59:10:28:15	34:03:55:31	57.5%
Total		Manual	27:00:51:00	18:05:55:18	67.5%

At a first glance, this table presents a counterintuitive result that an automated job management system can be less efficient that manually managing those jobs. However, that is not the correct conclusion to draw. Consider the factors that contribute to the wall time to run a job

$$t_{\text{job}} = t_{\text{cpu}} + t_{\text{q}}, \tag{4.1}$$

where t_{cpu} is the amount of cpu time required to run the job, and t_{q} is the amount of time a job spends waiting in the queue to begin execution. If a task requires more than one job to be run, *i.e.*, multiple queue submissions, then the total run time is

$$t_{\text{tot}} = \sum_{\text{jobs}} t_{\text{job},i} + \sum_{\Delta\text{jobs}} t_{\text{lag},i}, \tag{4.2}$$

where t_{lag} is the time between job (i) completing, and job $(i+1)$ being submitted to the queue. For these efficiency measurements, the idea is to eliminate human inefficiency in resubmitting to the queue. For example, if job (i) were to finish in the middle of the night, and a user did not become aware that the job was complete until morning, t_{lag} would be the time while the user was asleep, before checking on the progress of job (i). If a job completes at a convenient time, a user may submit the next job almost immediately, reducing t_{lag}. An automated job management system that monitors the job status will submit job $(i+1)$ immediately after job i completes, taking $t_{\text{lag}} \to 0$. In this case, we would expect that if t_{cpu} and t_{q} are the same for a given job, then an automated system should always do at least as well as manual job management, and often better.

Let us keep the assumption that t_{cpu} is the same for similar jobs, but examine the effect of t_{q} on efficiency. Given that we are defining efficiency as the ratio or cpu run time to job time, the only source of inefficiency in an automatically managed job is the queue time. If t_{q} is often much larger than t_{lag}, we would expect that an automated system would only be marginally more efficient than a user. Also, if t_{q} varies widely, then a set of user managed jobs could have a serendipitously small $\sum t_{\text{q}}$, while automated jobs could have an unfortunately large $\sum t_{\text{q}}$, and the automated system would appear to be less efficient compared to the user jobs.

To see this effect in our table, we can consider the simulation *Chinchilla-2* and when it occurred in the XSEDE allocation cycle. XSEDE allocates SUs for a calendar year, accepting applications every quarter. Generally, as the end of an allocation period approaches, the number of jobs submitted on machines rises, as users attempt to use all of their hours before their allocation period ends. The deluge of extra jobs generally increases the queue time for jobs, as the computing resources are much busier. Chinchilla-2 was run in the middle of December of 2012, while all other runs were earlier in the allocation cycle. We can imagine that in addition to the usual end-of-quarter run on the computing system, there may have been an extra boost because it was the end of the year, and users may have been trying to use their time before leaving campuses for winter breaks.

We can excise Chinchilla-2 from the efficiency calculation and compare with manual run jobs, as shown in Table 4.3. Here, we can see that if we exclude a time that had spuriously large t_q from the efficiency calculation, then the automated system is indeed marginally more efficient than manual user management.

Table 4.3: New XBaya total value excluding runs near the end of the allocation quarter.

Simulation	Control	Total Time	CPU Time	Efficiency
Total	XBaya	59:10:28:15	34:03:55:31	57.5%
Excl. mid-Dec.	XBaya	35:00:30:22	24:10:11:57	69.7%
Total	Manual	27:00:51:00	18:05:55:18	67.5%

The difference between these two efficiencies ia driven by t_lag. But even for manual job management, it is possible to drive $t_\mathrm{lag} \to 0$ if the jobs being run are well understood. On all of the XSEDE batch computing resources used for this work, it is possible to tag dependencies between compute jobs. So, if job $(i+1)$ depends on the output from job (i), the two jobs can be submitted at the same time by the user, with job $(i+1)$ marked in such a way that it will wait to begin execution until job (i) has completed successfully. I am sufficiently familiar with these simulations that I usually know beforehand how long an LGadget job will take, and as such, how many times a simulation will need to be submitted to run to completion. For most of the manual jobs in Table 4.2, I was able to submit a set of jobs with appropriate dependencies to take $t_\mathrm{lag} \to 0$, essentially performing the same task as XBaya using the batch system. In this case, I would expect that XBaya and manual management should have essentially the same efficiency, and that the efficiency could be characterized by the average queue time on the systems over the quarter that the jobs were submitted.

However, even though XBaya might not have a significant advantage by this measure of efficiency, there were still some efficiency gains to be had. When I would submit jobs with dependencies, I would have to put considerable effort in to plan all of the jobs beforehand, and interact several times with the batch system, at least once per job to be submitted, as well as write the necessary job submission scripts with appropriate parameters for walltime, number of processors, account to be charged for SUs, etc.. On the other hand, with XBaya, execution of all jobs can be started with a single click, and a previously programmed set of rules for each job is automatically applied. So, XBaya can give the same efficiency by this measure with a smaller time commitment to manage jobs by a user, which is advantageous.

4.3.2 Other XBaya Advantages

XBaya has advantages beyond saving the user time in job management. The workflow process creates parameter files for an entire set of simulations from a master file, which creates two distinct advantages over manual execution.

First, an entire suite of simulations can be set from a single parameter file, and a number of parameters are dynamically set correctly by the workflow process. For example, CAMB takes an $\Omega_{\mathrm{dm}} = \Omega_{\mathrm{m}} + \Omega_{\mathrm{b}}$ parameter and an Ω_{b} parameter, while 2LPTic and LGadget take Ω_{m} and Ω_{b} seperately. This can create confusion when setting parameter files manually. The workflow reduces the learning curve, and eliminates potential errors from confusing which values belong in which configuration file. Also, the configuration step runs almost instantaneously, significantly reducing the set-up time for each simulation.

Also, since all of the simulation parameters for all boxes in a given cosmology are set by a script, they are guaranteed to be consistent. In the early testing period, we ran a set of XBaya simulations with identical parameters to an earlier manual run. Upon completion, we found that not all outputs from the automated simulation agreed with the manual simulation. Upon careful investigation, we found that for one of the jobs in the manual case, a parameter in one configuration file had been mistakenly set to an incorrect value.

On the other hand, this configuration feature is included in the workflow as a python script that can be executed separately, in conjunction with manual runs, so the advantages from automated configuration can be realized in a manually managed job. But, all of the submission scripts still need to be addressed in a fully manual sense, leaving XBaya with a still significant advantage over manual management.

4.3.3 Some XBaya disadvantages

The Apache Airavata framework is a software project supported by a team of software engineers and designed to be an easy environment to import code, but at this stage of development, some help is needed to embed code in its XBaya environment. Our Airavata use was attached to an XSEDE award that come with support from those developers in the form of Extended Collaborative Support Service (ECSS) time, which ensured that our code could be integrated in the XBaya workflow manager. Needing to interact with a developer each time that we need to make changes or add features adds a layer of communication and inefficiency in our development process. If that developer is working on many other projects simultaneously, then making changes to our workflow could incur delays.

Another challenge that we faced was a mismatch between the computer application paradigm in which scientists often operate, and the paradigm that computer scientists expect. While a scientist often develops code, it is often to get a particular task done in a minimalist fashion. Scientists often do not include significant error handling, as there is an implicit assumption that users will know how to use the code, and rarely are unit tests written. Computer scientists with experience working on large software projects expect that a significant amount of time is dedicated to these kinds of tasks. This created specific problems for us trying to develop our workflow. Consider

LGadget as an example.

LGadget returns few error messages that are useful. LGadget will return a useful error message if the user supplies contradictory memory bounds and run parameters in the configuration file, relating to allocation performed within LGadget. However, several packages that LGadget calls do not provide similar memory bounds checking, so it is possible that LGadget can overrun available memory without providing a useful error message. The job will then crash, with little or no useful information about what caused the crash. One feature we would have liked to include in our workflow is the ability to resubmit jobs that experience runtime errors. However, without useful error messages and return codes built into LGadget, the XBaya is unable to determine if the cause of the crash is something that needs to be addressed or can be ignored before resubmission. This means that every time there is a runtime error, a user needs to manually inspect all of the outputs to determine what kind of error caused the crash before the job can be resubmitted, just as in a manually managed case.

Another LGadget example is that the code needs to be compiled differently depending on the type of cosmological model and parameters, *e.g.*, enabling $w \neq -1$ models is a compile time option for LGadget. Each time a new application is compiled it needs to be added to the XBaya application registry, so every time we wanted to run a new model, we would need our developer to register a new version of LGadget and construct a new workflow, which added time between BCC runs. To alleviate these sorts of issues and adjust our codes to match the standard computer science paradigm, error handling in particular, would require a significant time investment. I estimate that to make the appropriate changes to LGadget would require at least 6 months of effort.

4.3.4 Enabling DES Projects

With the N-body simulations in place, additional processing proceeds at SLAC to produce galaxy catalogs from the N-body data. This section explores some of the work done with the data products. Much of the work here is done by others, but I will include a summary of some of the work for completeness, and to show concrete examples of DES scientific results that follow from these N-body simulations.

4.3.4.1. Projection in Optical Galaxy Clusters

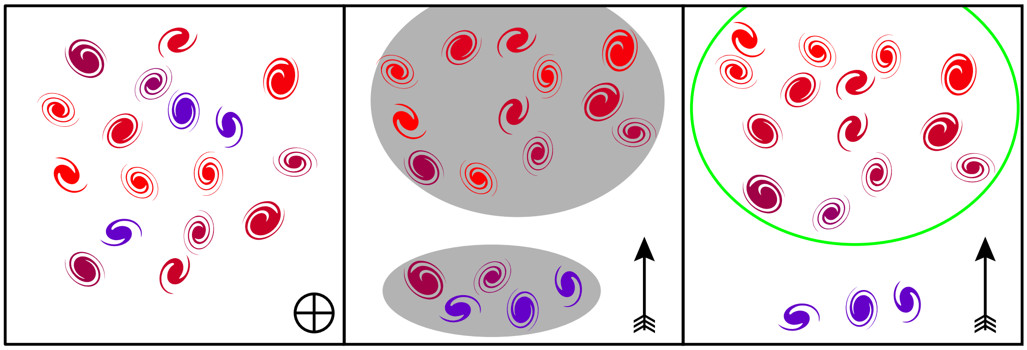


Figure 4.5: A cartoon showing how a line-of-sight projection can influence a cluster finder's mass estimation. The *left* panel represents a patch of sky that might be observed on the sky. The symbol in the bottom right represents an arrow facing away, indicating that we are looking along the line-of-sight. The *middle* panel shows true galaxy positions and underlying dark matter halos, indicated by the shaded area. The *right* panel shows inferred galaxy distances based on their colors The arrows in these two panels show the line-of-sight direction. A cluster finder might identify one cluster in this data (circled in *green*). This results in a richness estimate biased high relative to either halo.

The first project described is related to optical projection effects. This project is covered in detail in Chapter 6, so I will omit most of the details, but include the introduction to the project published in Erickson et al. (2013). Galaxy cluster surveys aim to measure the space density of dark matter halos by locating and counting groups and clusters of galaxies as a function of their size and distance from the Milky Way (equivalent to cosmic epoch, z). Galaxies will tend to be clustered in regions with massive underlying dark matter halos. A signal that can be used to identify the number of galaxies in a halo is the measure λ, known as richness (Rozo, Rykoff, Koester, McKay, et al., 2009; Rykoff et al., 2012). In a pure and complete sample, the expected number of galaxy clusters, N, in a richness bin λ_i and redshift bin z_j lying within a solid angle $\Delta\Omega_k$ can be written as a convolution

$$N_{ijk} = \frac{\Delta\Omega_k}{4\pi} \int_{z_j}^{z_{j+1}} \frac{dV}{dz}\, dz \int_{\lambda_i}^{\lambda_{i+1}} \int_0^\infty \frac{dn(M,z)}{dM} P(\lambda|M,z)\, dM\, d\lambda, \qquad (4.3)$$

where M is the halo mass, $dn(M,z)/dM$ is the halo space density (also called the mass function), dV/dz is the cosmological volume element, and $P(\lambda|M,z)$ is the probability of observing a particular λ given an underlying halo mass and redshift. Dark matter, dark energy and other cosmological parameters enter this expression through $dn(M,z)/dm$ and dV/dz, while astrophysics and cluster finding algorithm properties can be characterized by the probability, $P(\lambda|M,z)$, the mass–observable relation. Relating observed galaxy cluster counts to a theoretical underlying distribution of dark matter halos allows cosmological parameters to be extracted.

Our synthetic survey catalogs can be used to characterize the mass-observable likelihood, $P(\lambda|M,z)$, in a variety of cosmological models and to study the nature of noise arising from projection of unrelated halos along the line-of-sight. Such studies inform analysis of the real data, where the true underlying cosmology is unknown, since constraints on cosmological models depends on how well characterized is the mass–observable relation.

The mass–observable relation is often assumed to be log-normal about a power-law mean. That is, a signal can be expected to be related to the true mass correctly on average, but has some scatter about the true mass value. However, there are a variety of reasons why the observed relation may not be log-normal. One possible systematic effect could come from galaxies along the line of sight to a target halo. Dark matter halos usually lie at the confluence of dark matter filaments, and if an observer happens to be looking along such a filament, there will be excess galaxies along the line of sight. Halos projected along the line-of-sight of a given target boost its signal in a stochastic manner, resulting in a $P(\lambda|M, z)$ that grows a tail to high values.

Figure 4.5 shows a cartoon version of how line of sight projection can boost the mass estimate of a large halo. A cluster finder would locate an apparent galaxy overdensity on the two dimensional visible sky, represented by the *left* panel, but given that the Universe has three dimensions, there can be some confusion about the galaxy positions in the third axis, which is illustrated in the *middle* and *right* panels—the line of sight is shown by the direction of the arrows in the lower right hand corner of each box. DES will use photometric redshift estimators to determine the distance to galaxies in the line-of-sight direction. That means that the color of the galaxy as measured a couple of filter bands will determine its redshift. The *right* panel shows the positions of galaxies that might be measured by photometric redshifts, while the *middle* panel shows the true positions. A cluster finder in this cartoon example would assign two galaxies from a small foreground halo to the larger background halo. Since optical cluster finders use the number of galaxies measured in a cluster as a mass proxy, this would result in the mass of the background halo being overestimated. There is a much larger number of small dark matter halos in the Universe than massive ones. These small halos do not contain enough galaxies to be identified as clusters but can bias mass estimates via projection to the galaxy counts for larger halos.

If a log-normal assumption is used in cosmological analysis, when there is in fact line-of-sight projection, cosmological parameter estimates will be biased, see Chapter 5. However, we can account for this projection effect in red-sequence Matched-filter Probabilistic Percolation (redMaPPer) identified clusters, as described in Chapter 6, most clearly illustrated by Figure 6.5 and the supporting text.

Projects done by other authors have also been able to use the XSEDE-produced data described here, and I will now illustrate some of those.

4.3.4.2. CFCP

Cluster cosmology is built on making a link between a theoretical prediction for the distribution of dark matter halos and observed galaxy clusters. Different cluster finding algorithms and cluster definitions will lead to different cluster identifications. While identifying clusters differently is not a problem in itself, extracting cosmological information from sky surveys depends on the ability to characterize the properties of each method and relate their results to the underlying dark matter distribution. A galaxy cluster finder is more powerful when, compared to others, it can trace the underlying dark matter halos more efficiently and accurately.

In our synthetic galaxy surveys, scientists have an advantage in that they know the true underlying dark matter distribution and dark matter halo locations. Detected clusters can thus be compared directly to the underlying dark matter halo population. A variety of competing cluster-finding algorithms are available, and the synthetic surveys allow an assessment of their relative performance.

Figure 4.6 shows in two dimensional cartoon form how two different cluster finding algorithms might perform. Galaxies are represented by the *black* spiral shapes, and dark matter halos by *gray* circles. The *green* and *red* circles show clusters identified by two different algorithms. The “best” cluster finding algorithm would be one that most reliably matches the underlying halo distribution, and can most robustly relate the observed galaxy distribution to the underlying dark matter halo mass.

Looking at this cartoon, we can quickly identify some strengths and weaknesses of each cluster finder. If we look at the three dark matter halos on the left, we see that the *green* cluster finder identifies galaxy clusters that are highly correlated with the underlying dark matter halos, while the *red* finder merges halos together.

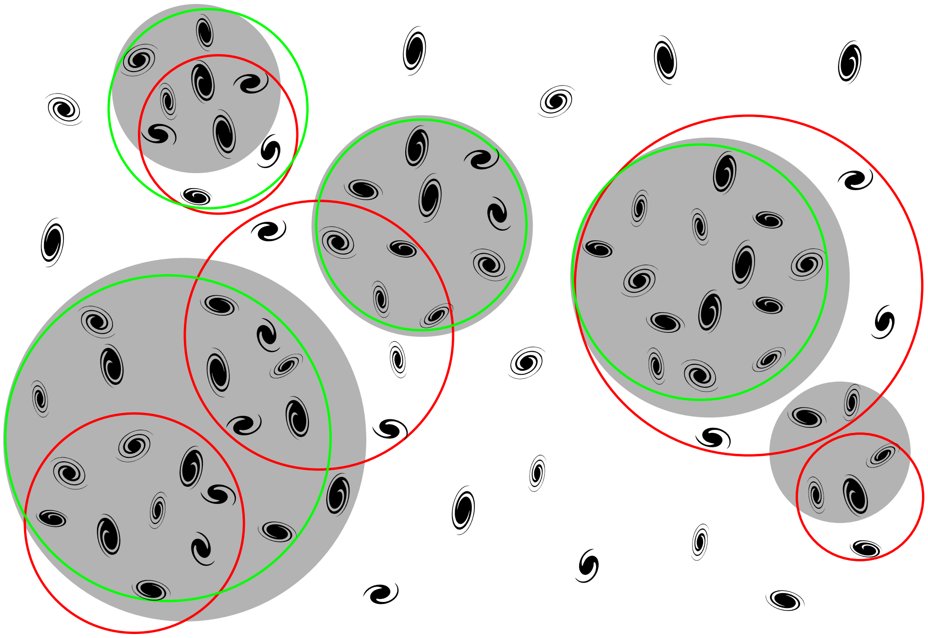


Figure 4.6: A gedanken cluster finder comparison. The it grey circles represent true underlying dark matter halos, in which there is an overdensity of galaxies relative to the diffuse galaxy background. The *red* and *green* circles represent clusters identified by two galaxy cluster finders. Which more reliably reproduces the underlying dark matter halos? The *green* finder more closely matches the halos where it identifies a cluster, but the *red* finder identified the same number of clusters as underlying halos. Note that in real galaxy clusters, most galaxies are elliptical type, rather than the spiral type illustrated here.

Such confusion could arise because of projection effects arising in the third dimension as described in section 4.3.4.1. In this region, it seems that the *green* algorithm is clearly superior to the *red*. Looking at the two dark matter halos on the right, we again see that the *green* finder, where it identified a cluster, is highly correlated with the underlying halo. But, the *green* finder only identifies one cluster where there are two halos, while the *red* finder, though being poorly correlated with the underlying halos, has identified two clusters. From this small cartoon picture it is hard to say if *red* or *green* would be a better cluster finding algorithm to use for a large data set.

The DES Clusters Working group has embarked on a Cluster Finder Comparison Project (CFCP) (Song et al., 2013). There are six cluster finding algorithms available to DES: C4 (Miller et al., 2005), GMBCG (Hao et al., 2010), redMaPPer (Rozo and

Rykoff, 2013; Rykoff et al., 2013), VTRS (Barkhouse et al., 2006), WAzP (Benoist et al., 2013), and zVT (Soares-Santos et al., 2011) and each could be used to generate cluster catalogs from DES data. Using the synthetic skies, these catalogs can be analyzed and include descriptive information about their performance in tracing the underlying dark matter.

In order to test how the cluster finders trace the dark matter, it is necessary to map clusters to halos. The CFCP defines two methods to match a galaxy cluster to an underlying dark matter halo, *proximity matching* and *membership matching*. *Proximity matching* attempts to match each cluster center with the nearest dark matter halo center. This process can be complicated by several factors; *e.g.*, how to deal with a situation where a cluster finder identifies two clusters that exist in one true underlying halo. To overcome this, CFCP proximity matching defines a search volume and matches clusters with more galaxies to halos with larger masses. *Membership matching* assigns clusters to halos based on the overlap of the member galaxies. That is, a cluster matches a halo if the plurality of the clusters galaxies overlap a single halo and vice versa.

Once clusters can be matched to halos, we can define statistics that describe the performance of the cluster finders, the completeness, C and purity, P:

$$C = \frac{N_{\text{matches}}}{N_{\text{total halos}}} \quad \text{and} \quad P = \frac{N_{\text{matches}}}{N_{\text{total clusters}}} \tag{4.4}$$

where C characterizes how well the cluster finder is at finding clusters that actually correspond to underlying dark matter halos, and P characterizes what fraction of the clusters that were found matched halos. The completeness and purity can be checked in redshift bins. For a concrete example, let us examine the *red* and *green* cluster finders of Figure 4.6. The *green* algorithm identifies 4 clusters, of which all 4 are well matched to a dark matter halo. Since there are 5 halos, this finder has

$C = 80\%$ and $P = 100\%$. The *red* algorithm identifies 5 clusters, but only 3 are well matched to underlying halos. Again since there are 5 halos, this finder has $C = 100\%$ and $P = 60\%$. Figure 4.7 shows preliminary results for purity and completeness of a proximity matching based DES CFCP.

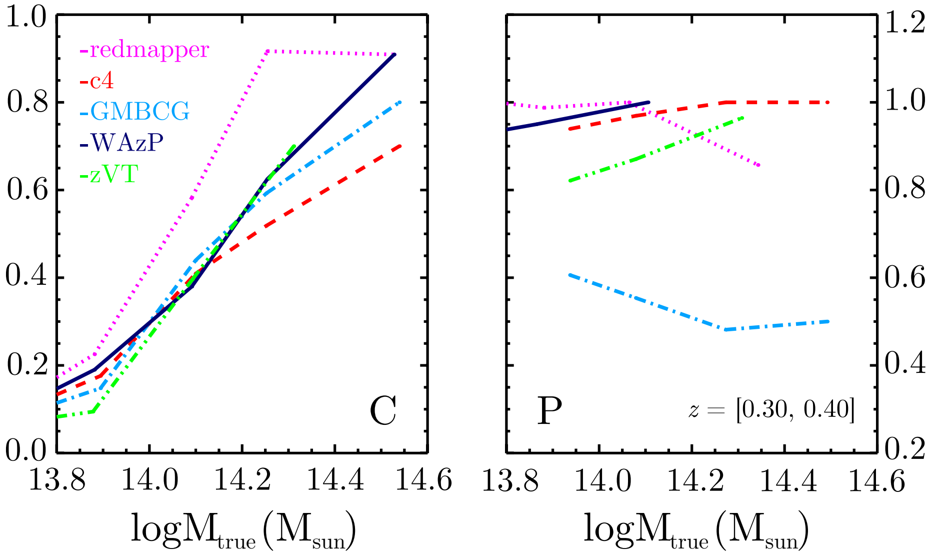
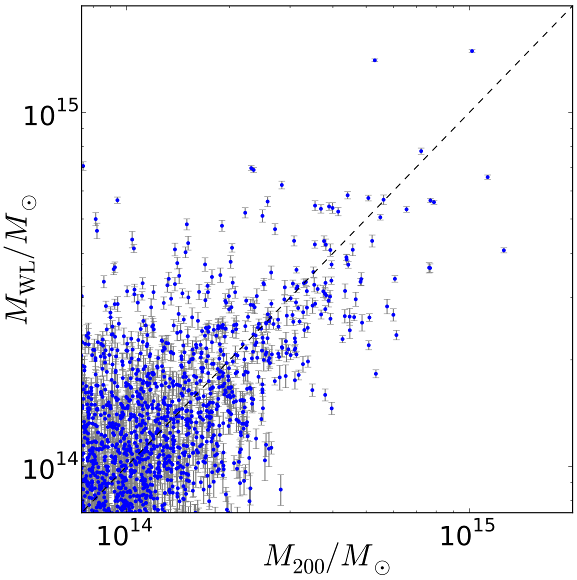


Figure 4.7: Preliminary results of a cluster finding comparison on a BCC synthetic galaxy catalog in a $z = 0.3$–0.4 redshift shell. The plots show completeness (*left*) and purity (*right*) for (note the different vertical scales) for 5 cluster finders available to DES. The horizontal axis is the mass associated with proximity matched halos for each finder. Figure courtesy of Jeeseon Song (University of Michigan).

4.3.4.3. Weak Lensing Mass Calibration

Members of the Weak Lensing Working Group have been investigating how best to estimate the mass of an observed cluster based on the weak lensing distortion of background galaxies. Using a maximum likelihood estimator of P. Schneider, King, and Erben (2000) to fit the synthetic observations, and comparing to the true halo masses from the simulation, the group can understand how best to estimate masses and also provide feedback on the model used to generate the synthetic observations. A recent calibration plot comparing estimated and true masses is shown in Figure 4.8. The dotted line shows an identity relation. A best fit to the data returns a mean mass

Figure 4.8: Comparison of the mass inferred from weak lensing shear analysis (y-axis) to the true halo mass (x-axis) for several thousand galaxy clusters identified in the first synthetic DES sky survey. The dotted line is the identity relation. Figure courtesy of Jörg Dietrich (Universitäts-Sternwarte München).



that is on average, biased by -8.7%. The scatter is about 85%. This is an early test case, and results should improve.

4.3.4.4. Joint Cluster Finsing with the SZ Effect

The N-Body simulations and halo catalogs we produce also provide a platform for generating simulated SZ observations for multi-wavelength cosmological studies. DES was designed to have a significant overlap with the SPT survey of the SZ effect in order to take advantage of joint cosmological analysis. Theoretical work shows that this multi-wavelength approach can improve the characterization of dark energy and other cosmological parameters by DES (Cunha, 2009; Wu, Zentner, and Wechsler, 2010). In order to prepare for this joint analysis computational framework needs to be built and tested on synthetic observations. Therefore, analogously to the ADDGALS observable catalog creation for the DES optical data, tools have been developed in order to create realistic simulated observations for SPT and other instruments (Biesiadzinski et al., 2013). Just as optically identified galaxy cluster counts are related to an underlying halo distribution by a mass–observable relation, so are SZ clusters, where the observable is the SZ decrement. The SZ effect is a spectral distortion in the observed Cosmic Microwave Background (CMB). It is caused by CMB scattering off of with electrons in the hot gas bound to dark matter halos. The change in observed intensity is characterized by the SZ decrement, y, multiplied by the frequency dependent scaling $f(\nu)$,

$$\frac{\Delta T_{\mathrm{SZE}}}{T_{\mathrm{CMB}}} = f(\nu)y, \tag{4.5}$$

where the SZ decrement is governed by the temperature of the gas, T_{g}, electron density, n_{e} and Thomson electron interaction cross-section, σ_{T},

$$y = \int \frac{k_{\mathrm{B}} T_{\mathrm{g}}}{m_{\mathrm{e}} c^2} \sigma_{\mathrm{T}} n_{\mathrm{e}} dl \quad \text{where} \quad \sigma_{\mathrm{T}} = \frac{8\pi}{3} \left(\frac{e^2}{m_{\mathrm{e}} c^2} \right)^2, \tag{4.6}$$

and integration is over the line of sight. The DES BCC simulations do not include gas dynamics, so the method for generating SZ maps is to embed gas pressure profiles within the dark matter halos (M. Arnaud et al., 2010) and treat that gas as the source of the decrement. Figure 4.9 shows a patch of sky for which an SZ map has been generated from BCC halos.

Maps generated using this framework have already been utilized in the study of SZ–mass scaling of clusters of galaxies observed by the *Planck* satellite (Biesiadzinski et al., 2012) and to measure the correlation between optical and infra-red detected galaxies and CMB lensing (Bleem et al., 2012). In addition, cluster finding algorithms run on these simulated maps are also being combined in a rigorous fashion with cluster finding data obtained from the equivalent optical simulations—based on the same underlying N-Body simulations—in order to explore whether the joint galaxy cluster catalogs can offer improved purity and completeness as well as reduce the scatter in optical observables (Nord et al., 2013).

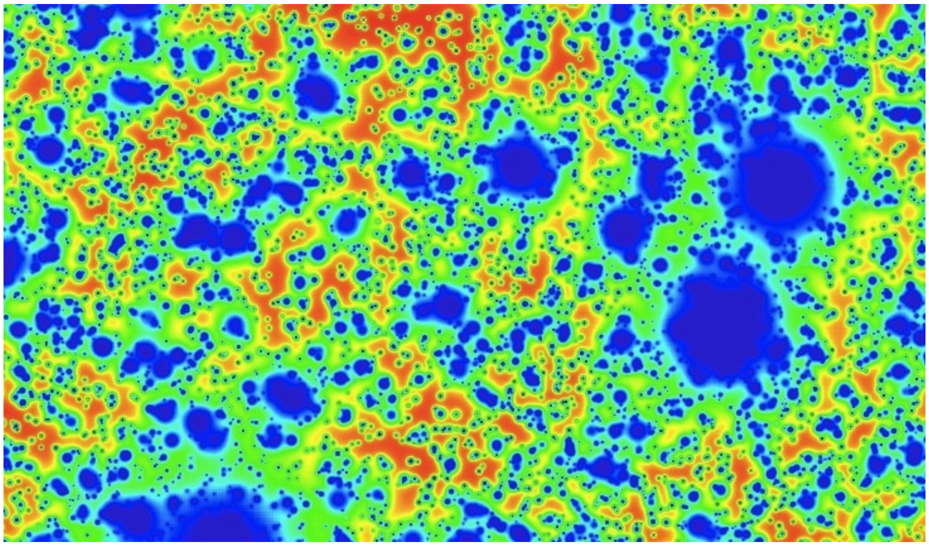


Figure 4.9: Synthetic SZ sky map patch from BCC dark matter halos identified with ROCKSTAR. This map does not include the overall cosmic microwave background, only sources of the SZ effect. Red regions indicate that there are no sources, while blue indicated a strong source. The large blue circles represent nearby massive dark matter halos that have a large SZ footprint.

4.4 Summary

Implementing the Airavata workflow for this project has entailed significant overhead. Scripts that set up the input parameter files needed to be developed, and new features were added to the existing codebase so that our applications would integrate more effectively with the workflow framework. Interaction between the co-authors of this document—domain scientists along with XSEDE ECSS team members—was essential to achieving a production-level service. The effort invested has been worthwhile in that we have been able to produce a large number of N-body simulations for DES with much less effort than would have been necessary to produce them manually.

In past two years we have automated codes in a workflow infrastructure which improves simulation production efficiency, by reducing effort needed by a user to set-up a simulation, and then using wallclock time with ∼70% efficiency. Eventually, this service could serve as the backend for a web-based gateway. The gateway could be used by astrophysicists who may want to generate a synthetic sky in a particular model without needing to build their own production system and locate sufficient computing resources.

The workflow system can capture provenance, including information such as when the data set was created, by whom, where, and with what application version and which input parameters. Currently this information is distributed across the High Performance Computing (HPC) filesystems where runs are performed, and is not easily accessible. We would like to modify how Airavata handles provenance tracking, *e.g.*, by centralized reportng in a single, web-accessible database. Improved provenance can enable broader sharing, reuse, and long-term preservation of our simulations and the resultant galaxy catalogs.

In the longer term, we could also expand our scope, generalizing our galaxy catalog construction process into a science gateway that would support broader classes of astrophysical studies. The optical catalogs we create could be augmented by synthetic surveys at other wavelengths, from radio to X-ray, and our focus on galaxies could be expanded to include quasars, galactic stars, and other astrophysical objects.

Now, we turn away from long discussions of simulation and computer science issues, and turn to some scientific analysis. First, in the next Chapter, I will look into an observational systematic, line-of-sight projection, and how it can affect cosmological analysis. Then, in Chapter 6, I will use some of the DES simulations discussed here to show how to account for line-of-sight projection in DES analysis.

CHAPTER 5

Influence of Projection in Optical Cluster Cosmology Studies

We examine the bias in cosmological parameter estimates caused by incorrectly assuming a Gaussian (projection-free) mass–observable relation when the true relation is non-Gaussian due to projection. Projection tends to skew the mass–observable relation of galaxy clusters by creating a small fraction of severely blended systems, those for which the measured observable property of a cluster is strongly boosted relative to the value of its primary host halo. We introduce a mixture model for projection and explore Fisher forecasts for a survey of 5,000 square degrees to $z = 1.1$ and an equivalent mass threshold of $10^{13.7}\,h^{-1}M_{\odot}$. Using a blended fraction motivated by optical cluster-finding applied to the Millennium Simulation, and applying *Planck* and otherwise weak priors, we find that the biases in $\Omega_{\rm DE}$ and w are significant and that new parameters describing the model increase uncertainty by factors of 2.8 and 2.4, respectively, above previous forecast uncertainties. Incorporating eight new degrees of freedom to describe cluster selection with projection increases the forecast uncertainty in $\Omega_{\rm DE}$ and w by similar factors. Knowledge of these additional parameters at the 5% level limits degradation in dark energy constraints to $\lesssim$10% relative to projection-free forecasts. We discuss strategies for using simulations and complementary observations to characterize the fraction of blended clusters and their mass selection properties.

5.1 Introduction

Galaxy clusters offer tests of large-scale gravity and cosmology, as their space density is exponentially sensitive to the time-dependent amplitude of the matter power spectrum and the cosmic expansion history (see Allen, Evrard, and Mantz (2011) and Voit (2005) for recent reviews). Because their counts and clustering probe the gravitational growth of structure, clusters provide information beyond that provided by Cosmic Microwave Background (CMB) and cosmic distance measurements, such as Supernovae (SNe) and Baryon Acoustic Oscillation (BAO). Joining cluster data with such measurements significantly improves cosmological parameter constraints (Cunha, Huterer, and Frieman, 2009).

While the potential for clusters to constrain parameters such as the dark energy equation of state, the energy densities of cosmic components, and the amplitude of matter density fluctuations has long been known (Bahcall and Fan, 1998; Henry and K. Arnaud, 1991; Lima and Hu, 2004; Majumdar and Mohr, 2004; Weller, Battye, and Kneissl, 2002), early work also emphasized the importance of understanding systematic errors associated with survey modeling (*e.g.,* Evrard et al., 2002; Levine, Schulz, and M. White, 2002). The agreement in cosmological parameters derived recently from independent samples selected at optical (Rozo et al., 2010) and X-ray (Henry et al., 2009; Mantz et al., 2010; Vikhlinin et al., 2009) wavelengths indicates progress in addressing systematic errors. However, early *Planck* analysis of the thermal Sunyaev–Zel'dovich (SZ) effect in the optically-selected `maxBCG` sample (Koester et al., 2007) suggests there is more work to be done, *e.g.,* the mean gas thermal energies inferred from *Planck* measurements lie a factor of two below simple model expectations (Planck Collaboration, 2011d). The origins of this disagreement are due to differences in mass calibration (Rozo, Bartlett, et al., 2012; Rozo, Evrard, et al., 2012; Rozo, Rykoff, Bartlett, et al., 2012; Rozo, Vikhlinin, and More, 2012), the effective offset in total mass of $\sim$40% is in $\sim 2.5\sigma$ conflict with the mass calibration errors quoted in the above cosmological studies.

The *Planck* optical results highlight the importance of a key functional element of cluster cosmology from counts, namely the likelihood, $p(M_{\rm obs}|M, z)$, that a halo of mass M at redshift z has an observed property $M_{\rm obs}$. For SZ observations, $M_{\rm obs}$ is the total thermal energy of the hot gas inferred from the spectral distortion in the cosmic microwave background. For the case of optical studies we consider $M_{\rm obs}$ to be the optical richness, defined by the number of red galaxies in the cluster above a given magnitude limit. Sky counts expected within a particular model are calculated by a convolution of this mass–observable function with the space density of halos. The latter has been calibrated to high precision by N-body simulations (Crocce et al., 2010; Evrard et al., 2002; Jenkins et al., 2001; Tinker et al., 2008; Warren et al., 2006).

Since the scaling of most observables with mass are power-laws, and therefore linear in the logarithm, the convolution kernel is typically assumed to be log-normally distributed about a power-law mean. The log-normal assumption for deviations in hot gas properties about the mean is supported by X-ray observations of core-excised luminosity and temperature in clusters (Pratt et al., 2009) and from a range of hot gas properties in simulated halo samples (Stanek et al., 2010; Yang, Ricker, and Sutter, 2009). However, some degree of departure from log-normality should be expected intrinsically, potentially driven by different formation histories (Yang, Bhattacharya, and Ricker, 2010) and by major merging events (Poole et al., 2006).

A given intrinsic likelihood for halo observables will be modified when $M_{\rm obs}$ is projected onto the sky (Cohn et al., 2007; Hallman et al., 2007; M. White, Hernquist, and Springel, 2002). Halos projected along the line-of-sight of a given target boost its signal in a stochastic manner, resulting in a $P(M_{\rm obs}|M, z)$ that grows a tail to high values. Optical richness is an attractive property to measure because it only requires broad-band photometry; however, it is also sensitive to line of sight projections. Richness scales roughly linearly with mass (Johnston et al., 2007; M. White et al., 2011; Zheng et al., 2009), while X-ray and SZ signals scale more steeply, $\sim M^{1.6}$ (Planck Collaboration, 2011c; Reiprich and Böhringer, 2002; Rykoff et al., 2008; Stanek et al., 2006), making these observables less susceptible to contamination from (spatially more abundant) lower mass halos. Indeed, the Abell catalog (Abell, Corwin, and Olowin, 1989) contains cautionary notes about projected confusion. Spectroscopic studies of optically selected clusters occasionally reveal multiple peaks and complex structures in velocity space (A1689 (Łokas et al., 2006), CL1604 (Gal et al., 2008), A85 (Boué et al., 2008) and EIS clusters (Grove, da Costa, and Benoist, 2008)), and simulation studies are beginning to explore these issues in detail (Noh and Cohn, 2011; M. White, Cohn, and Smit, 2010). Joint X-ray and optical studies of three nearby clusters show thermal signatures anticipated by gas dynamic simulations for actively merging systems (Maurogordato et al., 2011).

The statistical ingredients (the space density, spatial clustering, and galactic content of halos) needed to calculate projected confusion are coming into focus, and a generic expectation is that most massive halos suffer little contamination while a modest percentage are strongly affected by projection (Cohn et al., 2007; Rozo et al., 2011). These studies motivate a Gaussian mixture model for projection that we explore in this paper. The mixture represents a dominant component of clusters whose sight-lines are largely clean, along with a minority of clusters whose signal is strongly boosted. The latter category we refer to as blended systems, or *blends*, and in these objects the $M_{\rm obs}$ signal is not dominated by a single halo. Our treatment here is intended to be illustrative, but model parameters could be tuned using sky simulations tailored to specific surveys (Cai et al., 2009).

An earlier study of projection used an Edgeworth expansion to model cluster counts including non-zero skewness and kurtosis in $p(M_{\rm obs}|M, z)$. Shaw, Holder, and Dudley (2010) find that the detailed shape becomes important when the product of the scatter in the mass–observable relation and the logarithmic slope of the mass function at the limiting mass is greater than one. Our study differs from that work in two ways: our Gaussian mixture approach, which includes eight new degrees of freedom, is more general than their expansion. Also, we use a Fisher matrix approach to explicitly calculate both the bias that projection induces in a projection-free (single component Gaussian) analysis, as well as the additional variance that is incurred when the extra degrees of freedom are included. We explore the latter under a variety of prior constraints.

The discussion is organized as follows. In section 5.2, we briefly recount the procedure to extract dark energy constraints via cluster counts and variance in counts, and we present our parameterization of the mass–observable relation that includes line-of-sight projection. In section 5.3, we present our results and, in section 5.4, a discussion of the results.

5.2 Cluster Selection Model with Projection

When Abell published the first homogeneous cluster catalog from photographic plate imaging, he employed the count of galaxies within a fixed-metric aperture and a scaled magnitude range as a measure of *galactic richness*, a proxy for cluster mass (Abell, 1958). The development of multi-band imaging cameras in the late 1990s (Boulade et al., 2003; Gunn et al., 1998) enabled cluster samples to be selected using color selection techniques, whereby counts within a joint magnitude and color (or photometric redshift) range are employed as a mass proxy (Adami et al., 2010; Gladders and Yee, 2005; Hao et al., 2010; Koester et al., 2007; Szabo et al., 2011; Wen, Han, and Liu, 2009). These samples contain up to 69,000 clusters extending to $z \sim 1$ across nearly 8,000 square degrees of sky. The next generation of optical and near-infrared surveys—the Dark Energy Survey (DES), the Visible and Infrared Survey

Telescope for Astronomy (VISTA) surveys, Panoramic Survey Telescope and Rapid Response System (Pan-STARRS), with Large Synoptic Survey Telescope (LSST) and *Euclid* to follow—will identify hundreds of thousands of clusters.

Modern, color-based–cluster-finders rely on the 4000 Å break, a spectral feature of old stellar populations (Bower, Lucey, and Ellis, 1992). Observations show, and stellar population models expect, that the mean color in a fixed observed band straddling 4000 Å will vary with redshift. A single color can therefore be used as a simple photometric redshift estimator (Gladders and Yee, 2000). The redshift accuracy is limited by the finite color width of the red galaxy population at a given epoch. The finite width of the color filter employed for cluster finding in turn corresponds to a comoving length scale of order hundreds of Megaparsecs (Cohn et al., 2007). Red galaxies in spatially distinct halos that fall within a cylinder of this length aligned toward an observer will be catalogued as a single cluster (Milkeraitis et al., 2010), illustrated in cartoon form in Figure 4.5. We generically refer to this process as *blending*. While all clusters suffer some degree of projected blending, we are particularly interested in extreme cases, and so adopt a specific definition for classifying clusters. A cluster of observed richness $M_{\rm obs}$ will be referred to as a *blended cluster* if no single halo contributes $M_{\rm obs}/2$ or more to the richness. Conversely, a cluster for which a single halo does contribute $\geq M_{\rm obs}/2$ of the richness is referred to as *clean* (We assume here that the radial scale for observed and intrinsic measures are aligned.).

The massive halos that host clusters tend to be embedded in filaments and/or supercluster regions. Viewpaths that traverse such structures will have a locally boosted background in the color-magnitude sub-space used for cluster detection. Empirical studies of a new red-sequence matched filter method applied to Sloan Digital Sky Survey (SDSS) `maxBCG` data (Rozo et al., 2011; Rykoff et al., 2012) indicate that such boosts generate a blended fraction of ∼10% in the cluster population.

Previous Fisher matrix forecasts for the cosmological parameter yield from upcoming surveys (Cunha, 2009; Lima and Hu, 2004; Majumdar and Mohr, 2004; Weller, Battye, and Kneissl, 2002) have assumed a log-normal distribution for the observable likelihood, $p(M_{\rm obs}|M,z)$. While the log-normal form may reflect the intrinsic (*e.g.*, spherically averaged) dispersion in the halo ensemble, blended clusters will have $p(M_{\rm obs}|M,z)$ strongly boosted at high $M_{\rm obs}$. We use a Gaussian mixture model in log-mass, described in section 5.2.2, to model a bimodal cluster population consisting of clean and blended systems.

While our model is general, we tune default parameters using the results of Cohn et al. (2007). That study applies a red sequence–based algorithm to projected galaxy maps from the Millennium Simulation (Croton et al., 2006; Springel et al., 2005). They use a single $R-z$ color applied in narrow redshifts sliced centered at $z = 0.4$, 0.7 and 1.0. Matching halos to clusters by galactic membership, they identify a blended subset of clusters whose mass–observable relation is shifted to higher $M_{\rm obs}$ values and whose variance is larger than that of clean clusters. At higher redshifts, the mean color in the old stellar population varies more weakly with z, and the color width of the red sequence traces out an increasingly longer comoving cylinder, reaching $\sim$500 h^{-1} Mpc at $z = 1$. The longer search cylinder drives an increase in the blended fraction of clusters, from 11% at $z = 0.4$ to 22% at $z = 1$. Note that the fraction of halos at fixed true mass that are blended will be lower than this, as convolution with a steeply falling mass function increases the fraction of blended clusters at fixed $M_{\rm obs}$ (Rozo et al., 2011).

5.2.1 Reference Model Survey

Our reference model survey, based on DES+VISTA, is assumed to cover 5,000 square degrees and extend to a limiting redshift of $z_{\rm max} = 1.1$. Our choice of maximum redshift is somewhat conservative since with the addition of IR filters from VISTA, the combined surveys should have accurate redshifts for field galaxies up to $z \sim 1.5$.

We assume that DES+VISTA will detect clusters above an observational threshold, $M_{\rm obs} \geq M_{\rm th}$, with $M_{\rm th} = 10^{13.7} h^{-1} M_\odot$, comparable to what is achieved by low redshift surveys (Johnston et al., 2007; Koester et al., 2007). Based on the `maxBCG` $N_{\rm gal}$ richness measure, the zero-redshift variance in the mass–observable relation is taken to be $\sigma_0^2 = 0.25$ (Rozo, Rykoff, Evrard, et al., 2009).

We subdivide the sky into 500 bins of 10 square degrees each, and calculate the counts and sample variance using richness bins of width $\Delta_{\ln M_{\rm obs}} = 0.2$ with the exception of the highest mass bin, which we extend to infinity. We set the width of our redshift bins to $\Delta z = 0.1$. These bin sizes imply 11 redshift bins and 10 mass bins. We assume fiducial cosmological parameters based on the fifth year data release of the *Wilkinson Microwave Anisotropy Probe* (*WMAP*) (WMAP5, Komatsu et al. (2009)). Thus, we set the baryon density, $\Omega_b h^2 = 0.0227$, the dark matter density, $\Omega_m h^2 = 0.1326$, the normalization of the power spectrum at $k = 0.05 \text{Mpc}^{-1}$, $\delta_\zeta = 4.625 \times 10^{-5}$, the tilt, $n = 0.963$, the optical depth to reionization, $\tau = 0.087$, the dark energy density, $\Omega_{\rm DE} = 0.742$, and the dark energy equation of state, $w = -1$. In this cosmology, $\sigma_8 = 0.796$. With the exception of w, the cosmological parameters we use have been determined to an accuracy of a few percent. We apply *Planck* priors[1] to all cosmological parameters. We use CMBfast (Seljak and Zaldarriaga, 1996), version 4.5.1, to calculate the transfer functions.

5.2.2 The Mass–Observable Relation for Clean Clusters

We assume that the majority of clusters are clean systems whose selection properties are described by a single log-normal form. Following the notation of Cunha and Evrard (2010), we write the probability of observing a cluster with observable mass

---

[1] Planck Fisher matrix courtesy of Wayne Hu.

proxy, $M_{\rm obs}$ given a true mass M, as

$$p(M_{\rm obs}|M,z) = \frac{1}{\sqrt{2\pi\sigma^2_{\ln M}(M,z)}} \exp\left[-x^2\right] \tag{5.1}$$

with

$$x \equiv \frac{\ln M_{\rm obs} - \ln M - \ln M_{\rm bias}(M,z)}{\sqrt{2\sigma^2_{\ln M}(M,z)}}. \tag{5.2}$$

The model allows for systematic error in the observable by allowing redshift-dependent bias and variance

$$\ln M_{\rm bias}(M,z) = B_0 + B_1 \ln(1+z), \tag{5.3}$$

$$\sigma^2_{\ln M}(M,z) = \sigma_0^2 + \sum_{i=1}^{3} s_i z^i. \tag{5.4}$$

We set the fiducial values of B_0, B_1 and the s_i to zero throughout this Chapter. The baseline mass scatter, σ_0, is taken to be 0.5, a value consistent with `maxBCG` findings for that survey's original $N_{\rm gal}$ richness estimator (Rozo, Rykoff, Evrard, et al., 2009). Recently, Rykoff et al. (2012) proposed an improved mass-estimator for MaxBCG, with scatter expected to be 0.2 to 0.3, making our assumption about the scatter conservative.

Below, we apply this single Gaussian model to fit a set of data that are described by our extended, two Gaussian case. For that fit, σ_0 has a slightly different value, and B_0, B_1 and the s_i elements will be non-zero, as described in section 5.3.1.

5.2.3 Selection with Projection: Blended Clusters

To model selection with projection, we use a Gaussian mixture form for $M_{\rm obs}$ that combines clean and blended sub-populatons,

$$p(M_{\rm obs}|M,z) = (1-\gamma(z))\, G_{\rm clean} + \gamma(z)\, G_{\rm blend} \tag{5.5}$$

where $G_{\rm clean}$ and $G_{\rm blend}$ are log-normal distributions of the form given by equation (5.1), and the blend factor, $\gamma(z)$, controls the fraction of blended clusters.

For the component representing blends, we introduce a set of parameters for the bias and scatter different than that of the clean component,

$$\ln M_{\rm bias,b}(z) = \mu_0 + \alpha \ln(1+z) + \beta\,(\ln M - \ln M_{\rm th}), \tag{5.6}$$

$$\sigma^2_{\ln M,\rm b}(M,z) = \sigma^2_{0,\rm b} + s_z z + s_M(\ln M - \ln M_{\rm th}). \tag{5.7}$$

We highlight below the role of the mass bias terms, especially the constant offset, μ_0, and its logarithmic redshift gradient, α. The parameter β allows for a mass-dependent bias. For the scatter of the blended component, we focus on a pessimistic scenario where $\sigma^2_{0,\rm b} = 2\sigma^2_0$. This is consistent with results derived from Millennium Simulation analysis Cohn et al., 2007. The more optimistic case of $\sigma^2_{0,\rm b} = \sigma^2_0$ yields qualitatively similar results. The variance is allowed to evolve linearly with redshift and log-mass.

Default parameter values for the blended component model are $\beta = s_z = s_m = 0$, and $\sigma^2_{0,\rm b} = 2\sigma^2_0$. We consider three specific combinations of μ_0 and α that reflect different scenarios of redshift evolution in the bias of the blended component: none ($\alpha = 0,\ \mu_0 = 0.75$); weak ($\alpha = 0.5,\ \mu_0 = 0.5$); and strong ($\alpha = 1,\ \mu_0 = 0.25$). In all cases, the log-mean halo mass of blended clusters is biased low, by $\mu_0 + \alpha\ln(1+z)$ relative to that of the clean component.

The blend factor controls the overall fraction of blended clusters, and we write its evolution as

$$\gamma(z) = \gamma_0 + \gamma_1 \ln(1+z)e^{-z}, \tag{5.8}$$

where the exponential damping is added only to regularize γ at high redshift. The blend factor grows with redshift to $z = 0.77$, then flattens and decreases weakly toward the $z = 1.1$ redshift limit.

We choose this parameterization because it allows sufficient freedom to roughly match the blending fraction as a function of redshift found in Cohn et al. (2007). We calculate the blended fraction of clusters as a function of their observable mass proxy, $M_{\rm obs}$, via convolution with the mass function, as described below. In Figure 5.1 we show the resulting fraction of blended clusters above the survey threshold,

$$f_{\rm blend} = \frac{\overline{N}_{\rm blend}}{\overline{N}_{\rm blend} + \overline{N}_{\rm clean}}. \tag{5.9}$$

as a function of redshift bin for $\gamma_0 = 0$ and $\gamma_1 = 0.05$, 0.15 and 0.25. The mean counts, $\overline{N}$, are given by equation (5.11) below, where the clean and blended components are calculated using the associated components of equation (5.5). For each γ_1, the three lines show results for the three choices of $\{\mu_0,\ \alpha\}$ pairs discussed above. The results of Cohn et al. (2007), shown as the three black dots in the figure, are roughly matched by the choice of $\gamma_1 = 0.15$.

5.2.4 Cluster Counts and Clustering

The subject of deriving cosmological constraints from cluster number counts and clustering of clusters has been treated extensively in the literature (see, *e.g.*, Cunha (2009); Lima and Hu (2004, 2005, 2007)). We give a brief summary in this section, following the approach described in Cunha (2009), and leave other details to Appendix D.3.

The number density of clusters at a given redshift z with observable in the range $M_a \leq M_{\rm obs} \leq M_{a+1}$ is given by

$$\overline{n}_a(z) \equiv \int_{M_a}^{M_{a+1}} \frac{dM_{\rm obs}}{M_{\rm obs}} \int \frac{dM}{M} \frac{dn}{d{\rm ln}M}\, p(M_{\rm obs}|M, z). \tag{5.10}$$

where $\frac{dn}{d{\rm ln}M}$ is the mean halos space density, also called the mass function. We use the Tinker et al. (2008) parameterization for the mass function, and ignore errors in redshift estimates. The mean cluster number counts, and sample covariance, in

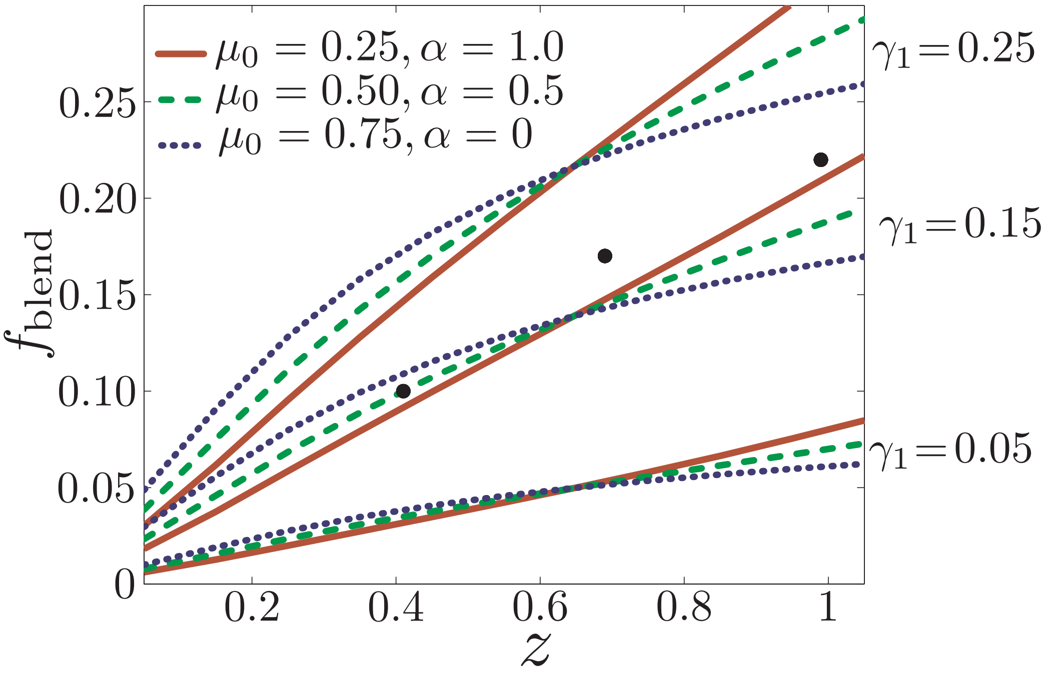


Figure 5.1: The fraction of blended clusters above the survey $M_{\rm obs}$ threshold is plotted for $\gamma_0 = 0$ and $\gamma_1 = 0.05, 0.15, 0.25$ (bottom to top). The three black dots are the values found from the Millennium Simulation study of Cohn et al. (2007). Color-styles correspond to three different redshift dependence forms (none, weak, strong) for the bias of the blending amplitude: $\mu_0 = 0.75, \alpha = 0.0$ (dotted, blue); $\mu_0 = 0.5, \alpha = 0.5$ (dashed, green); $\mu_0 = 0.25, \alpha = 1$ (solid, red). The models are tuned to coincide near the median cluster sample redshift of $z = 0.65$.

bins labeled by $i = \{a, b, c\}$, where a denotes mass proxy, b redshift, and c angular coordinate, are given by

$$\overline{N}_i = \int_{z_b}^{z_{b+1}} dz \, \frac{dV}{dz} \, \overline{n}_a \, W_c^{\rm th}(\Omega) \tag{5.11}$$

$$S_{ij} = \langle (N_i - \overline{N}_i)(N_j - \overline{N}_j) \rangle. \tag{5.12}$$

where ${W_c}^{\rm th}(\Omega)$ is an angular top-hat window function.

Define the covariance matrix of halo counts

$$C_{ij} = S_{ij} + \overline{N}_i \delta_{ij} \tag{5.13}$$

where $\overline{N}_i$ is the vector of mean counts defined in equation (5.11) and S_{ij} is the sample covariance defined in equation (5.12). The indices i and j refer to observable, redshift and angular coordinate bins. Assuming Poisson noise and sample variance are the only sources of noise, the Fisher matrix is, (Holder, Haiman, and Mohr, 2001; Hu and Cohn, 2006; Lima and Hu, 2004)

$$F_{\alpha\beta} = \overline{N}^t_{,\alpha}\mathbf{C}^{-1}\overline{N}_{,\beta} + \frac{1}{2}\mathrm{Tr}\left\{\mathbf{C}^{-1}\mathbf{S}_{,\alpha}\mathbf{C}^{-1}\mathbf{S}_{,\beta}\right\}, \tag{5.14}$$

where commas denote derivatives with respect to the model parameters. The first term on the right-hand side contains the information from the mean counts, $\overline{N}$. The S_{ij} matrix can be thought of as contributing noise to this term, and hence only reduces the information content from counts alone. The second term contains the information from the sample covariance.

The marginalized error in a parameter is given by $\sigma(\theta_\alpha) = [(F^{-1})_{\alpha\alpha}]^{1/2}$. Priors are easily included in the Fisher matrix. If parameter θ_i has a prior uncertainty of $\sigma(\theta_i)$, we add $\sigma^{-2}(\theta_i)$ to the F_{ii} entry of the Fisher matrix before inverting.

5.3 Results

Our model with projection differs from previous models that assume an entirely clean (single log-normal) population. Applying a clean-only model to a sky with projection will generally introduce a bias into derived cosmological parameters. We first address the magnitude of this bias, then turn to the impact that introducing extra degrees of freedom to represent blends has on marginalized constraints of dark energy parameters.

5.3.1 Parameter Bias

To estimate the bias in cosmological constraints that would result if cluster samples with projection are analyzed using a model with no projection, we follow a linearized approach used in previous studies (Huterer and Takada, 2005; Wu, Rozo, and Wechsler, 2008). Our *true* sky counts are based on the bimodal mass–observable relation, equation 5.5, applied using the three redshift evolution cases for the mass bias of the blended component (none, weak, strong) discussed above. The redshift growth rate of the blending factor, γ_1, is a controlling degree of freedom.

If the true sky is analyzed assuming no projection, meaning using a unimodal mass–observable relation equivalent to a $\gamma(z) = 0$ assumption in equation 5.5, then the resultant projection-free counts, $\overline{N}_1$, and sample covariance $\mathbf{S_1}$ may differ from the true values of $\overline{N}$ and $\mathbf{S}$, respectively. The set of model—cosmological and mass–observable—parameters, $\vec{\theta}$, recovered will generally differ from that of the true model. The bias in the model parameters (used in Wu, Rozo, and Wechsler (2008)) is derived in detail in Appendix D.2 and given by

$$\delta\theta_\alpha = \sum_\beta \left(F_1^{-1}\right)_{\alpha\beta} \left[(\overline{N}-\overline{N}_1)^t \mathbf{C_1}^{-1} \overline{N}_{1,\beta} + \frac{1}{2}\mathrm{Tr}\left\{\mathbf{C_1}^{-1}\mathbf{S_1}_{,\beta}\,\mathbf{C_1}^{-1}\left(\mathbf{S}-\mathbf{S_1}\right)\right\}\right]. \quad (5.15)$$

The covariance and Fisher matrix in the above expression are evaluated for the projection-free model using parameter values determined by fitting the redshift behavior of the first two moments of the mass-observable relation with projection. In our optical type survey described in §5.2.1, and for a specific choice of true model parameters μ_0, α, γ_1, (and fixing $\gamma_0, \beta, s_z, s_M = 0$ and $\sigma^2_{0,\mathrm{b}} = 2\sigma^2_0$), we compute the mean mass and variance in redshift bins of width 0.1 and fit these to determine the terms B_0, B_1, s_1, s_2, s_3 of the unimodal model, equations (5.6) and (5.7). Values for the case of $\gamma_1 = 0.15$ are given in Table 5.1.

For $\gamma_1 = 0.15$, the shifts in the mean mass are below one percent at $z = 0$ but

Table 5.1: Projection-free mass–observable parameters fit to the case with projection for $\gamma_1 = 0.15$.

μ_0	α	B_0	B_1	σ_0^2	s_1	s_2	s_3
0.75	0.00	0.0076	0.0389	0.2503	0.1110	-0.1299	0.0491
0.50	0.50	0.0040	0.0470	0.2500	0.0760	-0.0497	0.0097
0.25	1.00	0.0004	0.0551	0.2499	0.0475	0.0111	-0.0163

grow to 3.8% at $z = 1$ for the strong blending evolution case ($\alpha = 1$). The mass bias fit, constrained by the form of equation (5.3) with only two free parameters, can differ from the true bias in the projection model by up to 0.007 at $z = 1$ when the fit is the worst (in the $\alpha = 0$) case, but only by 0.002 for the best fit ($\alpha = 1$) case. The redshift behavior of the variance, with four free parameters of equation (5.4), matches the values of the projection case quite well, with deviations less than 3×10^{-4} in the worst case. The variance at $z = 1$ is larger for larger values of α, with $\sigma^2 = 0.293$ for the $\alpha = 1$ case.

For smaller γ values, the fits deviate less from the true bias in projection. For comparison, the bias at $z = 1$ for a fir with $\gamma_1 = 0.05$, differs by 0.002 in the worst case and the mass bias is about 0.03. The variance also grows more slowly with z, with $\sigma^2 = 0.264$ at $z = 1$ for the $\alpha = 1$ case.

We compute survey expectations for counts ($\overline{N}$) in mass, angle and redshift bins and their covariance ($\mathbf{S}$) for the range $0 \leq \gamma_1 \leq 0.3$. We then calculate the counts ($\overline{N}_1$), sample covariance ($\mathbf{S_1}$), full covariance ($\mathbf{C_1}^{-1}$), and Fisher matrix (F_1) for the respective projection-free case using the best-fit parameters described above. As mentioned in section 5.2.1, we add unbiased *Planck* priors to the Fisher matrix, so that $F_1 \rightarrow F_1 + F_{\text{Planck}}$. The resultant values are used to compute bias in model parameters according to equation (5.15).

Figure 5.2 shows the resulting biases in w and Ω_{DE}. For the cases shown, we assume a *Planck* prior on the cosmological parameters but all other model parameters

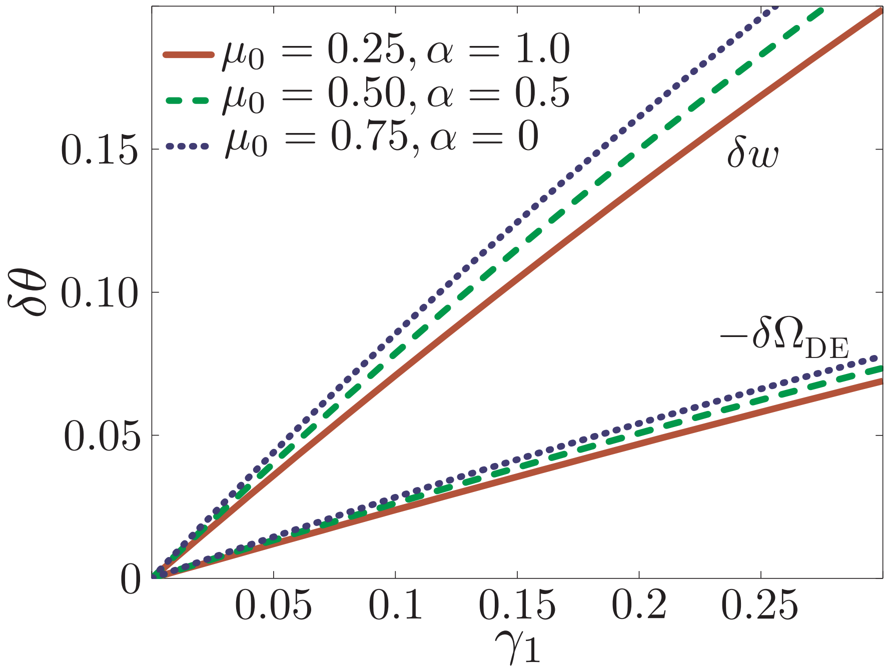


Figure 5.2: Shifts in the cosmological parameters w (upper lines) and $-\Omega_{\rm DE}$ (lower lines) as a function of the blending evolution rate, γ_1. Line-styles correspond to three different redshift dependence forms (none, weak, strong) shown in Fig. 5.1.

are free. When $\gamma_1 = 0$ there is no blended component and therefore no parameter bias (note we assume $\gamma_0 = 0$). The bias in cosmological parameter estimates grows approximately linearly with γ_1, with slopes that are weakly dependent on the assumed redshift evolution of the mass bias in the blended component. For the strong redshift evolution case ($\mu_0 = 0.25, \alpha = 1.0$) with $\gamma_1 = 0.17$, which provides a close match to the simulation results of Cohn et al. (2007), we find a significant biases in the dark energy equation of state, $\delta w = 0.12$, and in the dark energy density, $\delta\Omega_{\rm DE} = -0.04$.

These shifts may be considered pessimistic, in the sense that we have assumed a large scatter in the blended component. For the case of $\sigma^2_{\rm blend} = \sigma^2_0$, the slopes of the equivalent lines in Figure 5.2 are reduced by ∼50%, so that the strong redshift evolution case with $\gamma_1 = 0.17$ produces $\delta w = 0.08$ and $\delta\Omega_{\rm DE} = -0.03$. Reducing the assumed $\sigma_0 = 0.5$ scatter in the clean component would also lead to smaller biases in cosmological parameters.

Table 5.2: Cosmological parameter shifts, $\delta\theta$, for strong redshift evolution ($\mu_0 = 0, \alpha = 1.0$) and $\gamma_1 = 0.17$.

Parameter	$\theta_{\rm true}$	$\delta\theta$
$\Omega_b h^2$	0.0227	-0.0001
$\Omega_m h^2$	0.1326	0.0009
$\Omega_{\rm DE}$	0.742	-0.0401
w	-1.0	0.1178
$\delta_\zeta \times 10^5$	4.625	0.0222
n	0.963	-0.0015
τ	0.087	1.0×10^{-7}

Table 5.2 shows the bias in all cosmological parameters for strong redshift evolution for $\gamma_1 = 0.17$, the case that best matches Cohn et al. (2007). The bias for parameters other than $\Omega_{\rm DE}$ and w is less than 1% of the fiducial value. However, comparing to the fiducial uncertainties from the Fisher matrix with unbiased *Planck* priors show that the shifts can approach a 1-σ level for $\Omega_m h^2$ and δ_ζ.

Figure 5.3 offers insight into the magnitude of the change in cluster counts arising from projection. As a fiducial measure, we use counts, $\overline{N}_{\rm fid}$, for the projection-free (unimodal) case with default parameters (zero bias and redshift-independent variance). The solid lines in Figure 5.3 show the fractional shifts in counts, relative to the fiducial, as a function of redshift for the projected (bimodal) cases with $\mu_0 = 0.25, \alpha = 1.0$. For $\gamma_1 \gtrsim 0.1$, projection boosts counts on the order of a few tens of percent at high redshift. The dotted lines show projection-free expectations when the mass–observable parameters are shifted to the values given in Table 5.1, but the cosmology is held fixed. The dashed lines give projection-free expectations when both cosmological and mass–observable parameters are adjusted according to equation (5.15).

The counts of the projection-free model with fully shifted parameters provide a good match to the counts with projection. The adjustment of the mass–observable parameters alone offers a good match at low redshifts, but at high redshift, a unimodal

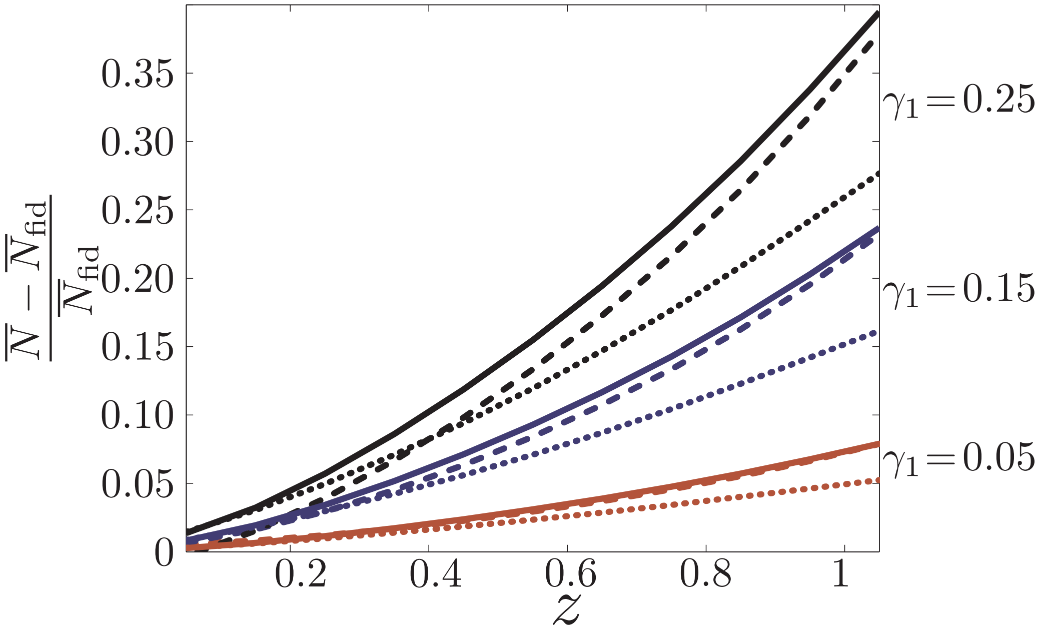


Figure 5.3: Fractional change in counts in the strong redshift evolution case ($\mu_0 = 0.25, \alpha = 1$), relative to a projection-free model, are shown for three values of the blending evolution parameter, γ_1. Solid lines give the case with projection while dotted lines show the projection-free model with parameters tuned to match the mass bias and variance of the projection model, but with cosmology fixed at the fiducial WMAP5 values. Dashed lines show the projection-free case after shifting all parameters (cosmological and mass-observable) according to Eq. (5.15).

fit to the bimodal form of the projected $p(M_{\rm obs}|M, z)$ becomes decreasingly accurate. Adjustments in cosmological parameters shift the amplitude and shape of the mass function as a function of redshift, providing a degree of compensation for deficiencies introduced by a unimodal $p(M_{\rm obs}|M, z)$ assumption. While the quality decreases for higher values of γ_1, the fits are still acceptable in a χ^2 sense.

Note that as γ_1 grows and the associated shifts in parameters grow, the linear approximation for the bias given by equation (5.15) begins to break down. For $\gamma_1 = 0.05$, agreement between the shifted single-Gaussian case and the two-Gaussian case is quite good, while at $\gamma_1 = 0.25$, the divergence is much larger.

Finally, we note that equation (5.15) calculates shifts using the Fisher matrix of the projection-free model. We have verified that we obtain the same results if we employ the projection model matrix with sharp priors added to the eight parameters describing the blended component. This is expected because, for the same free parameters in the two models, the linearized equation should be symmetric under their exchange.

5.3.2 Dark Energy Parameter Constraints

While introducing additional parameters to describe selection with projected blending may eliminate bias in cosmological parameters, that benefit comes with the risk of degrading cosmological parameter constraints. The amount of degradation depends on assumptions about priors on model parameters.

Table 5.3 summarizes results using the projection model that corresponds to our best match of Cohn et al. (2007) ($\mu_0 = 0.25,\ \alpha = 1.00,\ \beta, s_z, s_M = 0,\ \sigma^2_{0,\mathrm{b}} = 2\sigma^2_0,\ \gamma_0 = 0,\ \gamma_1 = 0.17$). In all cases, *Planck* priors are added to the cosmological parameters, and we consider priors on the clean and blended cluster components separately. Given an assumed prior error, σ_i, on the i^{th} parameter, we add to the Fisher matrix

$$F^{ii}_{\mathrm{prior}} = \left(\frac{1}{\sigma_i}\right)^2 . \tag{5.16}$$

We consider sharp priors as being numerically larger than other entries in the Fisher matrix, generally $F_{\mathrm{sharp}} \approx 10^6$, and flat priors are given by $F_{\mathrm{flat}} = 0$.

Table 5.3 shows permutations of three basic cases: a flat prior on model parameters, a prior of $\sigma = 0.1$ added to model parameters as well as a 10% prior added to σ^2_0 or $\sigma^2_{0,\mathrm{b}}$, or sharp priors on model parameters. The last two columns give the marginalized uncertainty in w and Ω_{DE}.

The first three rows compare extremal cases. Sharp knowledge of all mass–observable parameters produces the best constraints possible, ± 0.002 in Ω_{DE} and ± 0.011 in w. The projection–free case with no prior knowledge of the six parameters

Table 5.3: Forecasts for w and $\Omega_{\rm DE}$ constraints using *Planck* priors.

Priors		Uncertainty	
Clean	Blended	$\sigma(\Omega_{\rm DE})$	$\sigma(w)$
sharp	sharp	0.002	0.011
flat	sharp	0.014	0.046
flat	flat	0.034	0.109
0.1	sharp	0.010	0.030
0.1	0.1	0.010	0.030
0.1	flat	0.013	0.042

of the clean component, shown in the second row, produces constraints of ±0.014 in $\Omega_{\rm DE}$ and ±0.046 in w. These errors are worse by factors of 7 and 4, respectively, than the case of perfect knowledge. Introducing eight new degrees of freedom to represent the blended component further degrades the errors by somewhat more than a factor of two, to ±0.034 in $\Omega_{\rm DE}$ and ±0.11 in w.

Targeted follow-up and complementary survey information, from mm or X-ray observations for example, may enable moderate priors to be placed on the bias and variance of the mass–observable relation. These cases are explored in the lower three rows of Table 5.3. Knowledge of the clean component parameters at the level of ±0.1 provides substantial improvement over the flat case. Even with no prior knowledge of the blended component, the errors of ±0.013 in $\Omega_{\rm DE}$ and ±0.042 in w represent improvements over the projection–free case with no prior knowledge (second row). When 0.1 priors are brought to bear on the projected blends, the constraints improve to ±0.010 in $\Omega_{\rm DE}$ and ±0.030 in w. Stronger priors on the blended component do not improve these constraints.

5.3.3 Discussion

Achieving constraints on w and $\Omega_{\rm DE}$ at the few percent level is a goal of next-generation cluster surveys. Our analysis shows that avoiding biases at this level requires projection to be incorporated into the likelihood modeling of optical-IR surveys.

Prior knowledge of the blended component behavior can limit parameter bias. In Figure 5.4, we illustrate the rate at which the forecast uncertainty in w changes with prior uncertainty on the mass–observable parameters of the blended component. The behavior for $\Omega_{\rm DE}$ is similar, mainly because *Planck* priors effectively fix many of the correlations between cosmological parameters. The solid line shows the effect of applying priors to all eight parameters while dashed lines show the behavior when priors are applied only to parameters controlling the blending amplitude ($\gamma_0,\ \gamma_1$), mass bias terms ($\mu_0,\ \alpha,\ \beta$) and mass variance ($\sigma_{\ln M,\rm b},\ s_z,\ s_M$). In all cases, flat priors are imposed to the remaining mass–observable parameters.

Applying priors to only the parameter subsets, we see that all three sets have comparable effects on w uncertainties. Because of covariance, the effect of applying strong priors to all parameters is much stronger than for any of the isolated sets. The error in w is somewhat more sensitive to the blending amplitude $\gamma(z)$ than to the bias and variance, but all parameters need to be known at the level of 0.1 in order to avoid significant degradation.

For red sequence or photometric redshift cluster finding methods, the fraction of blended clusters is not likely to dominate the population, suggesting that the current level of uncertainty is near 0.1. By testing the performance of cluster finding algorithms on sophisticated simulations, and by calibrating mass selection based on multi-wavelength follow-up campaigns of existing deep cluster catalogs, the error may be reduced to $\sim$0.05 as part of next generation survey analysis. In the right-hand panel of Figure 5.4, we show the error in w for cases in which just the amplitude or both the amplitude and mass bias of the blending component are known at the $\sim$0.05 level. These cases limit the degradation of the w constraint, to factors of 2.0 and 1.5, respectively, compared to 2.4 for the case of all parameters free.

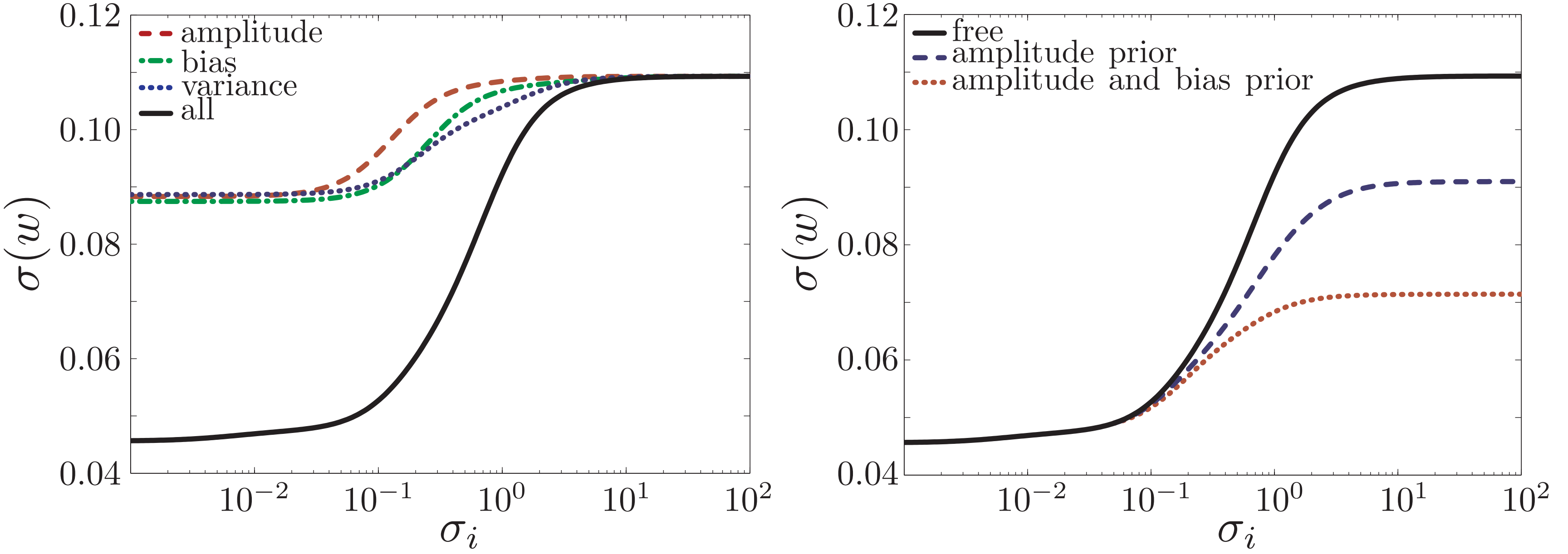


Figure 5.4: Error in w is shown as a function of priors added to the parameters describing projected blends. *Planck* priors are applied to cosmological parameters throughout. The *left* panel shows the error in w as a function of prior σ_i applied to all eight blending parameters (black solid), and separately to only the parameters describing $\gamma(z)$ (red dashed), or only the parameters describing $\ln M_{\rm bias,b}$ (green dot-dashed), or only the parameters for $\sigma^2_{\ln M,\rm b}$ (blue dotted). The *solid black* line is reproduced in the *right* panel, which also shows the resulting error in w when a 0.05 prior is applied to just the blending amplitude, $\gamma(z)$ (blue dashed), or when 0.05 priors are applied to both $\gamma(z)$ and $\ln M_{\rm bias,b}$ (red dotted). In the latter two cases, the results are shown as a function of prior σ_i applied to the remaining blending parameters.

Our assumed value of $\sigma_0 = 0.5$ may be pessimistic, in that future cluster finders may achieve better mass selection. An improved matched filter method Rykoff et al. (2012) applied to the `maxBCG` catalog indicates a mass scatter closer to 0.3 for low redshift clusters for a sample with mass threshold close to the value assumed here. While achieving this level of mass selection at $z > 0.5$ has not yet been demonstrated, the variety of cluster multi-color detection algorithms under active development (Adami et al., 2010; Dong et al., 2008; Koester et al., 2007; Milkeraitis et al., 2010; Miller et al., 2005; Soares-Santos et al., 2011; Song et al., 2012; Wen, Han, and Liu, 2009) offer the potential of future gains.

Spectroscopic observations of cluster fields provide valuable empirical tests of blending. For example, a follow-up of 58 EIS cluster candidates, selected only with I-band imaging and so without the benefit of color-based redshift filtering, found multiple redshift-space structures in a majority of fields (Grove, da Costa, and Benoist, 2008). Studies at high redshift using infrared color selection, which are just beginning, fare better but are not entirely clean. Six rich clusters from the SpARCS sample, which uses z'–3.6μ color from ground-based and *Spitzer* observations, have been followed up with Keck/LRIS spectroscopy. Two cases appear to be strongly blended, with dynamical mass estimates derived from velocity dispersions lower by a factor $\sim$6 than mass estimates based on their galactic richness. Continued follow-up of this and other IR-selected deep cluster samples should be followed vigorously as a means to characterize the amplitude and mass scale of projected blends.

Simulations of large-scale structure provide an effective tool for understanding projection. Work is underway within the DES collaboration to test a variety of cluster finding algorithms against simulated expectations for the multi-band galaxy catalog, as described in Chapters 4 and 6. Using either galaxy membership or redshift-space location as a way to match clusters and halos, the simulations offer the means to test the sensitivity of blending to algorithm choice and to choice of parameters within a fixed algorithm (Gerke et al., 2005). Such studies should produce improved, algorithm-specific characterizations of blends that can be coupled to empirically-derived studies to serve as prior information for cluster likelihood analysis.

As algorithms improve in terms of mass selection, characterization of projection effects will inevitably become apparent through the full shape of the mass–observable relation, $p(M_{\rm obs}|M, z)$ or its inverse, the mass–selection function, $p(M|M_{\rm obs}, z)$. Ultimately, cluster surveys have the potential to achieve the best possible constraints on dark energy parameters given by the first row of Table 5.3. Extracting a one percent constraint on w poses the challenge of precisely characterizing selection.

While we focus our analysis on optical-IR studies, the issue of blending is generic to all cluster finding methods. The blending factor $\gamma(z)$ should be minimal for X-ray selection, due to the compactness of the surface brightness image as well as the strong scaling of luminosity with mass. SZ selected samples are likely to incur blending at a level below that of optical-IR surveys (Cohn and M. White, 2009). However, for X-ray and SZ, angular resolution is also an important factor. The *Planck* satellite has only moderate resolution of 5–$10'$, depending on frequency. Of the 21 new cluster candidates identified in the *Planck* Early SZ sample, four are known to be double or triple systems from *X-ray Multi-Mirror Mission–Newton* (*XMM–Newton*) follow-up imaging (Planck Collaboration, 2011b). Follow-up studies of these and other SZ-selected sources from South Pole Telescope (SPT) (Carlstrom et al., 2011) and Atacama Cosmology Telescope (ACT) (Swetz et al., 2011) is needed to characterize the mass selection of these methods.

5.4 Summary

Cluster counts used in cosmological studies have typically been modeled with log-normal deviations about power-law forms for the mass–observable relation. While a log-normal expectation may reasonably reflect intrinsic scatter, projection will generically boost a minority of systems to higher observed signal. This blending of halo properties is particularly true for optical-IR surveys that use color or photometric redshifts as a distance estimator. We extend previous Fisher matrix studies by introducing a Gaussian mixture model for the mass–observable relation. The model employs eight new parameters to describe a redshift-dependent amplitude and shape of the blended component, in addition to the six parameters of the dominant, non-blended cluster population.

The presence of a minority of strongly blended clusters influences cosmological parameter estimation. For the case of blending parameters tuned to Millennium Simulation analysis (Cohn et al., 2007) (Fig. 5.1), we find that survey analysis using a projection–free (single Gaussian) analysis model introduces biases of 0.1 in w and -0.04 in $\Omega_{\rm DE}$. Comparing their Fisher forecast errors with *Planck* , these shifts are comparable to uncertainties expected using flat priors on mass–observable parameters, but are an order of magnitude larger than the uncertainties possible under precise mass–observable knowledge. Explicit modeling of projection is therefore required to avoid significant bias in next generation cosmological studies using cluster counts and clustering. Optical studies at low redshift, where the blended fraction is below ten percent, or studies using cleaner detection methods, such as X-ray selection, are less susceptible to bias from projection.

Constraints on w and $\Omega_{\rm DE}$ with *Planck* priors degrade by about a factor of 2.4 when new parameters to describe the Gaussian mixture distribution are included. Our analysis indicates that 5% prior knowledge of the blending amplitude and mass bias limit the degradation to a factor of 1.5.

Improved knowledge of blending will come from complementary approaches employing follow-up observations, simulations, and joint analysis of overlapping multi-wavelength surveys. Follow-up campaigns will provide mass estimates based on hydrostatic, virial and lensing masses. Simulations of the galaxy distribution will grow in fidelity, benefitting from empirical studies of the relation between halo mass stellar content to $z \sim 1$ (Leauthaud et al., 2011, 2012). Optical cluster finders applied to such simulated sky expectations will inform prior constraints on projection effects. Over the next decade, the ability to cross-match large cluster samples from mm to X-ray wavelengths will offer a new window into the nature of the relationship between clusters and the massive halos that host them.

CHAPTER 6

Projection in redMaPPer Identified Galaxy Clusters

Chapter 5 shows that line-of-sight projection is potentially a significant source of bias when extracting cosmological information from cluster surveys if not properly taken into account. Chapter 5 also shows that if a large number of parameters are used to describe the projection effects, those parameters need to be constrained to better than 10% level to avoid introducing large uncertainties to cosmological parameters. The Dark Energy Survey (DES) will use cluster counts to probe cosmology, and it has several cluster finding algorithms available to choose from (see Cluster Finder Comparison Project (CFCP) in section 4.3.4). The most recent is red-sequence Matched-filter Probabilistic Percolation (redMaPPer) (Rykoff et al., 2013), which has been developed with DES needs in mind.

Compared to the method used in Cohn et al. (2007) as redMaPPer has improved background subtraction via an iterative red-sequence model for assigning galaxies to clusters (see E.1). This Chapter outlines a model that can be calibrated to account for line-of-sight projection in redMaPPer identified clusters using only a few new parameters. The model treats a redMaPPer richness measurement as a sum of a target halo richness and a projected richness, where the projected richness is assumed to be dominated from the most massive halo along the line-of-sight. This assumption, combined with an assumed form for the richness–mass relation for target halos in clean sight-lines allows us to make a prediction for the richness–mass relation observed from all sight-lines.

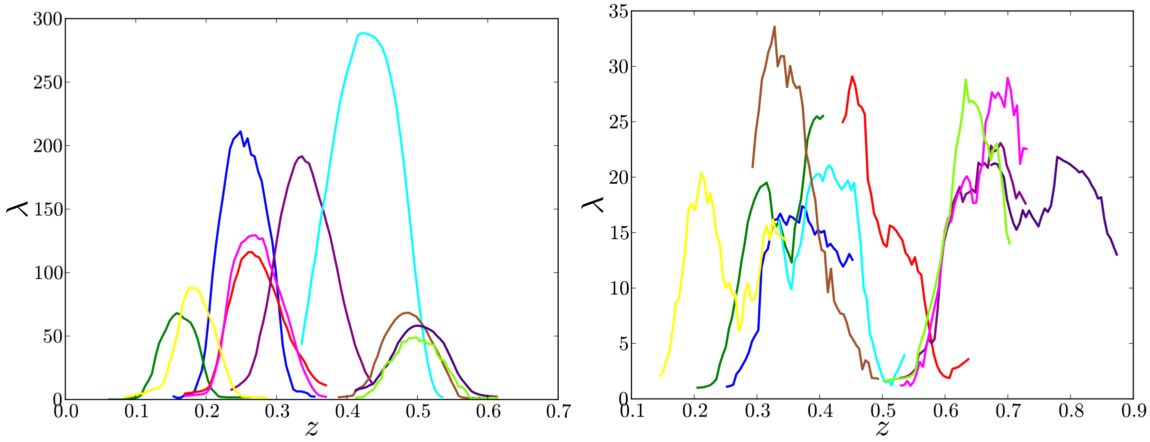
6.1 Projection Effects

We begin our investigation of line-of-sight projection in redMaPPer clusters by looking at dark matter halos in the synthetic DES galaxy catalogs. The catalogs are produced by assigning galaxies to dark matter halos as described in earlier chapters, and the resulting catalogs reproduce low-order spatial statistics of known galaxy populations. RedMaPPer is a photometric, red-sequence based cluster finder. It uses a red-sequence model based on a small spectroscopic training set and returns probable cluster centers, galaxy membership probabilities, and a redshift probability distribution as well as a richness estimate, λ, for each cluster. The richness estimator, λ, is the number of red-galxies in a cluster above $0.2L_*$ in the Schechter luminosity function. For more details, see Appendix E.1. We have added a feature to the redMaPPer algorithm that short-circuits the centering part of the algorithm, and performs just the richness estimation for a fixed redshift. We made this modification seeking to investigate how many halos contribute significantly to a cluster richness estimate due to proximity along the line of sight. The total richness returned by redMaPPer for a cluster will be a sum of the intrinsic richness of a target halo and the richness convolved with that halo due to nearby halos along the line of sight

$$\lambda = \sum_i \lambda_i \tag{6.1}$$

where we take $i = 0$ to be the target halo, and the remaining richnesses for nearby halos to be rank ordered by value, *i.e.*, $\lambda_{i+1} \leq \lambda_i$.

We then examined random sight-lines in a DES synthetic survey. The richness estimates returned along a variety of sight-lines containing at least one halo are shown



(a) Clean sight-lines

(b) Non-clean sight-lines

Figure 6.1: Curves showing the measured λ about halos for clean sight-lines, *left* and projected sight-lines, *right*. The curves extend to ± 0.1 about z_{halo}. The clean sight-lines have Gaussian $\lambda(z)$ curves, while the sight-lines that are likely to have projection effects show more complicated structure.

in Figure 6.1. This figure shows two types of $\lambda(z)$ measurements, some from sight-lines with only one halo, and some from sight-lines with several halos. For the *clean* sight-lines—with only one halo—the λ measurement rises as the redshift used for the measurement approaches the halo redshift, and then falls as the measurement redshift moves away from the halo redshift, with the resulting $\lambda(z)$ curve being Gaussian. For the *projected* sight-lines—with more than one halo—the story is more complicated. One feature that can be seen in the projected sight-lines, however, is that many of the curves appear to be bi-modal, suggesting that two underlying halos contribute significantly to the richness. With this thought in mind, we decided to explore a two-halo projection model, that is

$$\lambda = \lambda_0 + \lambda_1. \tag{6.2}$$

This takes the usual implication that a single halo hosts all of the galaxies identified with a cluster to the next order.

As a check to see how the two-halo model might perform, we randomly selected a

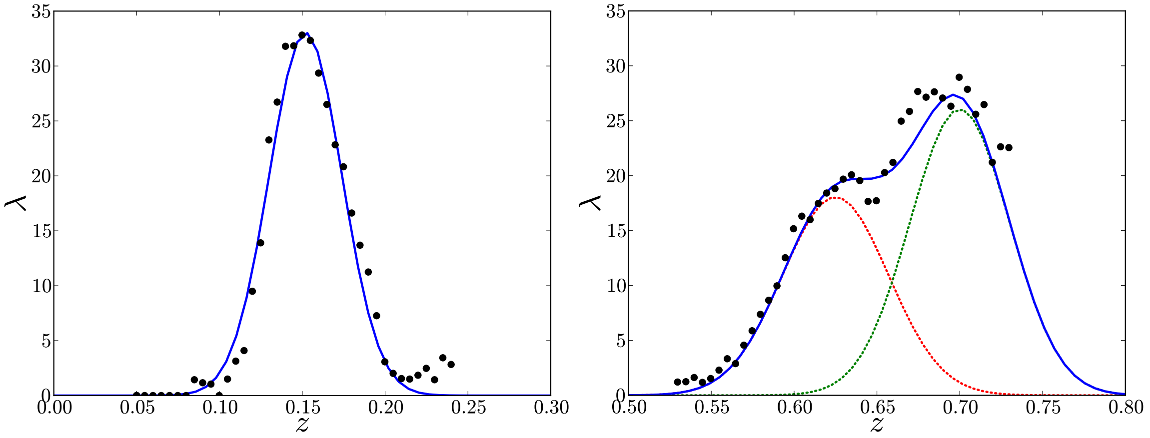


Figure 6.2: A measurement of redMaPPer richness, λ, about a dark matter halo position in a DES synthetic catalog. The measurements are given by *black* markers. Here, there are halos at $z = 0.63$ and $z = 0.70$ along the line-of-sight. Two Gaussian curves, representing how redMaPPer would measure the richness if each halo were isolated along the line-of-sight at the halo positions, sown by the *dotted red* and *green* lines. The two Gaussians to give the *solid blue* line, which closely tracks the redMaPPer richness measurement.

sight-line for a $\lambda(z)$ measurement, and then compared that measurement to a prediction for the measurement when considering the two largest line of sight halos. The result is shown in Figure 6.2. We overlay two Gaussian curves, motivated by the clean sight-lines shown in Figure 6.1, at the two halo positions shown by the dotted *red* and *green* lines. Their sum is shown by the solid *blue* line, while the richness measurements are shown by the points. This illustrates the point that richness measures for clusters identified along this line of sight would have contributions from the two separate host halos, and that those two halos account for most of the richness measure.

6.1.1 The Most Massive Projected Halo

Seeing that the randomly selected sight-lines suggest a two-halo model is important to include in richness measurements, we should check to see if we expect a two-halo model to be sufficient from additional arguments. A useful measure is the mass scale, M_1, that characterizes the typical mass of the most massive projected halo.

Specifically, given a target halo of mass $M_{\rm t}$, we define M_1 as the mass such that one expects, on average, 1 halo of mass $M \in [M_1, M_{\rm t}]$ within the projection volume $2A_{\rm t}\Delta\chi$. Numerically, we set M_1 by the condition

$$1 = 2A_{\rm t}\Delta\chi \int_{M_1}^{M_{\rm t}} \frac{dN}{dM} dM. \tag{6.3}$$

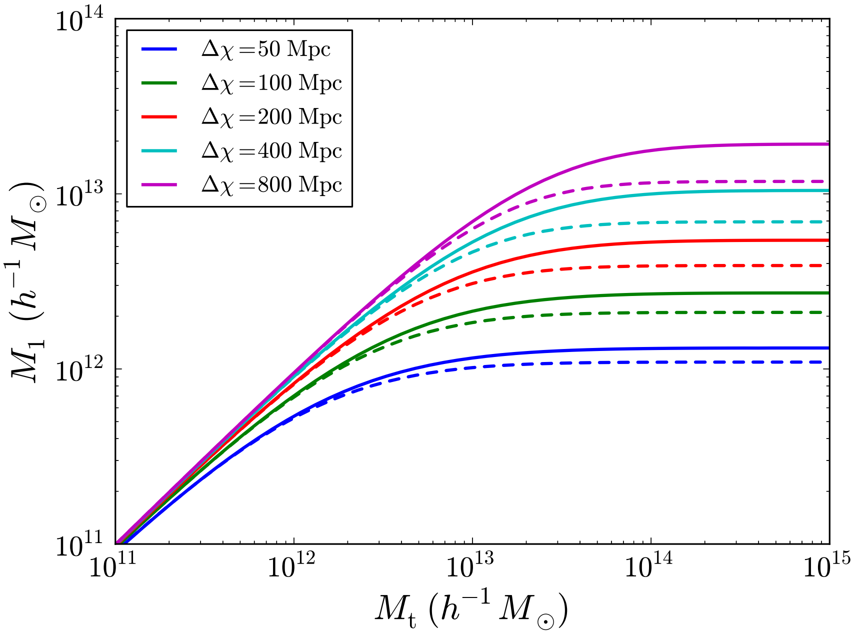


Figure 6.3: The characteristic mass M_1 of the most massive projected halo, as a function of the target halo mass $M_{\rm t}$ at $z = 0.2$, *solid*, and $z = 0.8$, *dashed*. The mass M_1 is defined such that one expects, on average, one halo of mass $M \in [M_1, M_{\rm t}]$ within the volume $2A_{\rm t}\Delta\chi$, where $A_{\rm t} = \pi$ Mpc^2 is the cross-sectional area of the target halo, and $\Delta\chi$ is the length of the cylinder over which projections are effective. For high mass halos that host clusters of galaxies, the most-massive projected neighbor is nearly independent of target halo mass.

Figure 6.3 shows the mass M_1 at $z = 0.2$ and $z = 0.8$. for a variety of cylinder lengths $\Delta\chi$. We have kept the cross section of the cylinder $A_{\rm t} = \pi R^2$ with physical $R = 1$ Mpc fixed, and we have ignored the evolution of the halo mass function along the cylinder. The solid and dashed lines show the mass M_1 evaluated using the $z = 0.2$ and $z = 0.8$ mass functions respectively.

As we can see, evolution of the halo mass function has only a modest impact on the mass M_1, which is nearly independent of the mass $M_{\rm t}$ for high mass halos. The mass of the most massive projector in the $z = 0.2$ and $z = 0.8$ case for $M_{\rm t} = 1.0 \times 10^{15}\ h^{-1} M_\odot$ vary with cylinder length according to:

$$\log_{10} M_1 = 0.97 \log_{10} \Delta\chi + 10.5; z = 0.2; \tag{6.4}$$

$$\log_{10} M_1 = 0.86 \log_{10} \Delta\chi + 10.6; z = 0.8. \tag{6.5}$$

6.1.2 Projection in the Optical

We focus now on projections in the optical. We check to see if a simple Halo Occupation Distribution (HOD) prediction agrees with our expectation that a two-halo model can describe projection effects. If the most massive projecting halo found in the previous section is the dominant factor in projection, then a two-halo model should be able to characterize projection in redMaPPer clusters. We use a simply HOD model for richness where the mean number of galaxies N in a halo of mass M is given by

$$\langle N|M\rangle = \theta(M - M_{\rm min}) \left[1 + \left(\frac{M - M_{\rm min}}{M_{\rm sat}}\right)^{\alpha}\right]. \tag{6.6}$$

Note that in the high mass limit, this expression takes the asymptotic form

$$\langle N|M\rangle = \left(\frac{M}{M_{\rm sat}}\right)^{\alpha}, \tag{6.7}$$

so $M_{\rm sat}$ can be characterized by the number of galaxies in high mass halos. As a fiducial model, we set $M_{\rm sat}$ such that a halo of mass $10^{15}\ h^{-1} M_\odot$ has on average 120 satellite galaxies. We set $M_{\rm min}$ via $M_{\rm sat} = 10 M_{\rm min}$, and the slope $\alpha = 1.0$ in accordance with results from SDSS (Zehavi et al., 2011).

Figure 6.4 shows the expected number of projected galaxies in a cylinder of radius

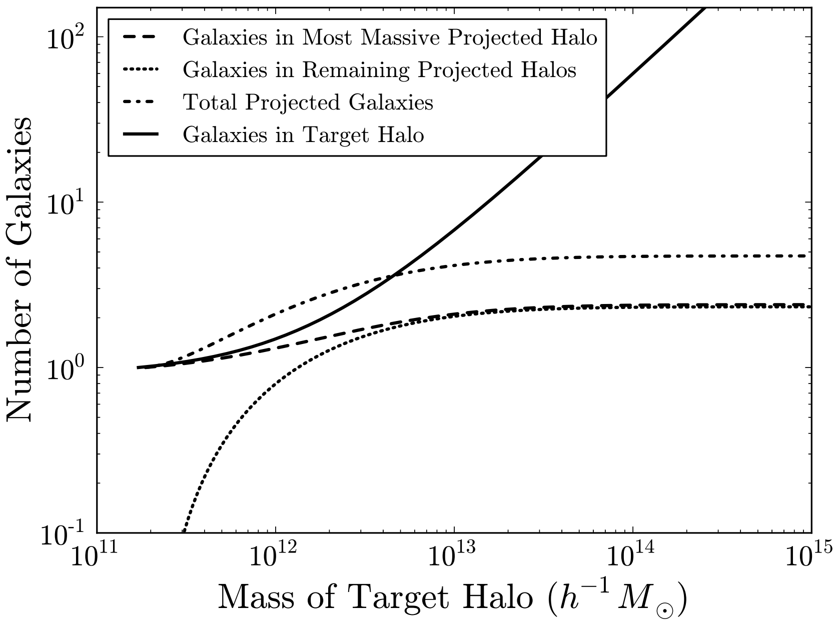


Figure 6.4: Number of galaxies within a cylinder centered on a target halo, as a function of the mass of the target halo. This calculation is done at $z = 0.2$ with $\Delta\chi = 122$ Mpc. Around massive target halos, the most massive projected halo hosts only ≈2 galaxies, with the remaining projected halos hosting no more than one galaxy each. This suggests that the most massive projected halo is a more significant source of contamination than many small mass halos along the line-of-sight.

$R = 1$ Mpc and length $\Delta\chi = \sqrt{3}cH_0^{-1}\Delta z$ with $\Delta z = 0.015(1 + z)$, assuming the cylinder is centered on a halo of mass $M_{\rm t}$. The width of the red-sequence in redshift is approximately $\Delta z \approx 0.015$, while the factor of $\sqrt{3}$ comes from matching the standard deviation of a top-hat and Gaussian distributions. The color distribution of red-sequence galaxies is roughly Gaussian in color, and therefore redshift. Since we are approximating projections using a top-hat selection, one needs to select the width of the top-hat to match the standard deviation of the red-sequence. At $z = 0.2$ (0.8), this corresponds to $\Delta\chi = 122$ Mpc (131 Mpc). The figure is made assuming $z = 0.2$, but the results at $z = 0.8$ are similar. Also shown are the expected number of galaxies in the target halo. We have also split the total projected galaxies into galaxies that reside in halos of mass M_1 or above, and galaxies that reside in halos of masses M_1

and below.

Comparing the dashed and dotted lines in Figure 6.4, we see that the most massive projected halo contributes about the same number of galaxies as the remaining projected halos combined. For massive halos, the most massive projected halo typically contributes a little more than two galaxies (dashed line). Since the remaining halos are all less massive than M_1, we typically expect for only the most massive projected halo to contribute more than one galaxy to the total richness. From this, we expect that a two-halo model should be able to account for most of the projection in redMaPPer identified clusters.

Finally, looking at Figure 6.4 we see that the expected number of projected galaxies in the cylinder spanned by the red-sequence is $\approx$5. For a more detailed study of line-of-sight optical projection using an HOD, see Chen (2013), which covers variance in projection by realizing many random sight-lines and compares optical signals with those of other wavelengths, but these factors are beyond the simple argument presented here. Additionally, for our choice of fiducial HOD parameters, we expect that robust optical identification of halos with $\lesssim$5 galaxies is not possible. This simple argument suggests that a most massive projected halo is the dominant factor in optical line-of-sight projection, and that a two-halo model will capture most projection effects. In our model, we will ignore additional nearby line-of-sight halos, which will effectively contribute additional scatter to the richness–mass relation.

6.2 Projection Model

Consider a halo of mass M at redshift z, and let λ be the richness of the halo. We refer to this halo as the target halo. We assume λ_1 is the largest contribution to λ from all line of sight halos, and λ_0 is the intrinsic richness of the target halo so that the total λ measured for a cluster will be

$$\lambda = \lambda_0 + \lambda_1. \tag{6.8}$$

We say a target halo suffers from projection effects if

$$\lambda_1 \geq \lambda_{\rm c} \tag{6.9}$$

where $\lambda_{\rm c}$ is a constant. In other words, a halo suffers from projection effects if the contribution λ_1 to the total richness is larger than some critical value $\lambda_{\rm c}$. For these preliminary calculations, we fix $\lambda_{\rm c} = 5$ based on arguments in the previous section, but more extended work should consider $\lambda_{\rm c} = cf(M, z)$, where c is a constant, and f is some function describing how clusters of different mass and redshift may be more or less prone to projection. We define $P_0(\lambda_0|M, z)$ as the probability distribution of λ_0. We assume P_0 is a Gaussian of mean $\bar{\lambda}_0(M, z)$ and standard deviation $\sigma_0(M, z) = \bar{\lambda}_0^{1/2}(M, z)$.

6.2.1 The λ_1 Distribution

We are interested in the probability distribution for λ, which depends on the probability distribution for λ_0 and λ_1. Having already assumed a distribution for λ_0, we need to address the distribution for λ_1. We need to consider two factors, the probability that a halo of mass M' at redshift z' overlaps the target halo, and a probability that such a halo contributes richness λ_1. We can address the first term by introducing the function, S given by,

$$S(M', z'|M, z) = \frac{dn}{dM'}\frac{dV}{dz'}\left(1 + b(M)b(M')\xi(r)\right), \tag{6.10}$$

which is built around the halo mass function, dn/dM', the cylindrical comoving volume dV/dz' and the halo bias function $b(M)$ and two-point function $\xi(r)$. Expressions for all of these functions can be found in Appendix E.2. Then, the probability of a halo of mass M' at redshift z' overlaps the target halo is $SdM'dz'$. We assume that λ_1 scales with λ_0', where $\lambda_0' = \lambda_0(M', z')$, so that

$$\lambda_1 = \lambda_0' W(z, z'). \tag{6.11}$$

We require $W(z, z) = 1$, and $W = 0$ for $|z - z'| \to \infty$. Putting these two ideas together, we arrive at an expression for $P_1(\lambda|M, z)$, with details in Appendix E.2:

$$P_1(\lambda|M, z) = \int_0^M dM' \int_0^\infty dz' S(M', z'|M, z) P_0' \left(\frac{\lambda_1}{W}\right). \tag{6.12}$$

where $P_0'(x) = P_0(x|M', z')$. As with λ_0' and P_0', from here on, a prime will indicate $|M', z')$, and no prime notation will indicate $|M, z)$.

6.2.2 The Richness–Mass Relation

We now turn our attention towards the richness–mass relation of galaxy counts. Let then $\widetilde{P}_1$ denote the unity-normalized ($\int d\lambda_1 \widetilde{P}_1(\lambda_1) = 1$) probability distribution for λ_1. Since λ_1 is either identically zero or larger than λ_c, we have then

$$\widetilde{P}_1(\lambda_1) = c(M)\delta(\lambda_1) + P_1(\lambda_1)\theta(\lambda_1 - \lambda_c) \tag{6.13}$$

where the θ term accounts for the fact that λ_1 must be greater than λ_c. The factor $c(M)$ is the *completeness*, or probability that a halo does not suffer from projection effects, note that this definition of completeness similar, but differs from the definition in Chapter 4. In Chapter 4 completeness was the ratio of we were interested in describing how well a cluster finder would recover underlying halos, where failures could come from any number of problems in relating identified clusters to underlying halos. Here, we are treating halos and clusters as equivalent objects, and implying that if a halo is projected onto a target, it will not be independently identified as a cluster. As such, the number of clusters will be less than the number of halos, except if $\int d\lambda_1 P_1 = 0$, in which case there is no projection. In this case, all halos will be identified as clusters, and the sample will be 100% complete. To calculate the completeness, we subtract the probability that a halo does suffer from projection effects from unity. The probability that a halo does suffer from projection effects is the sum over all possible λ_1s weighted by the appropriate probability from the last section:

$$c(M) = 1 - \int_{\lambda_c}^{\infty} d\lambda_1 P_1(\lambda_1). \tag{6.14}$$

Now, to find an expression for $P(\lambda|M, z)$, we need to sum over all possible λ_0 and λ_1 subject to the constraint that $\lambda = \lambda_0 + \lambda_1$,

$$P(\lambda|M, z) = \int d\lambda_0 d\lambda_1 P_0(\lambda_0) \widetilde{P}_1(\lambda_1) \delta(\lambda - \lambda_0 - \lambda_1) \tag{6.15}$$

$$= c(M) P_0(\lambda) + \int d\lambda_0 dM' dz' S P_0(\lambda_0) P_0' \left(\frac{\lambda - \lambda_0}{W} \right) \theta(\lambda - \lambda_0 - \lambda_c). \tag{6.16}$$

If we look at equation 6.16 we can check a couple of cases to make sure that this expression meets basic expectations. First, if a halo does not suffer from projection, $\lambda = \lambda_0$ and therefore λ is distributed according to $P_0(\lambda)$. On the other hand, if there is some probability of projection—encompassed by S—then there will be additional contributions to the richness from a halo at (M', z'). We can make another useful sanity check, though not necessarily an obvious one. The term $P_0((\lambda - \lambda_0)/W)$ is sharply peaked around $\bar{\lambda}_0' W$, so the integral picks out a projector halo that contributes approximately $\bar{\lambda}_0' W$ to the richness. Now we can check our initial effort to asses if this model will be able to address projection in redMaPPer richness in synthetic catalogs.

6.3 Model Calibration

We have started looking into this model using the DES synthetic galaxy catalog version 3.05, which is an unpublished testing synthetic galaxy catalog produced following the steps described in Chapter 3. The Halo catalog contains information about the halos and their ADDGALS assigned observable properties, including mass measures (*e.g.*, M_{200c}, M_{500c}, etc.), radius measures (*e.g.*, R_{200c}, etc.) and the number of galaxies with brighter than a number of magnitude cuts (*e.g.*, -19,-20,-21, etc.), position in right ascention and declination, as well as a variety of other properties, see (DES Simulations Working Group, 2013) for a full ist of properties. The v3.05 catalog covers a small sky area, which limits our ability to test the model, but here follows some preliminary results.

6.3.1 Redshift Weighting

First, we attempt to calibrate the window function introduced in equation 6.11. In terms of the $\lambda(z)$ curves we can measure, W is given by

$$W(z, z') = \frac{\lambda(z')}{\lambda(z)} \tag{6.17}$$

where $\lambda(z)$ is the richness of a halo at redshift z, and $\lambda(z')$ is the richness we would assign to the halo if we assumed the halo was at redshift z'. Consequently, to estimate W all we need to do is take all clusters within some redshift bin $\Delta z = z \pm \delta z$, and compute the above ratio along a grid of redshifts z'.

We isolate clean sight-lines in the v3.05 catalog to investigate $W(z, z')$. We select our sight-lines by limiting analysis to halos that satisfy the following conditions:

1. $M_{\rm halo} > 1.0 \times 10^{14}\ h^{-1} M_{\odot}$;

2. $\lambda > 20$ for redMaPPer richness evaluated at $z_{\rm halo}$;

3. $z_{\rm halo} < 0.8$;

4. $\rm N19_{halo} > \sqrt{N19}_{line\ of\ sight\ halos}$.

The first two conditions limit us to massive halos that can be robustly identified, and are likely to return a non-zero $\lambda(z')$ as z' becomes large. The redshift cut is to limit us to a region of the simulation that is well calibrated in the Adding Density Determined GAlaxies to Lightcone Simulations (ADDGALS) framework. The final condition is to limit us to very clean sight-lines, that is those that have very few potential contaminant galaxies along the line of sight (in a range $z_{\rm halo} \pm 0.2$; here, N19 is the number of galaxies associated with a halo above the 19th magnitude in the synthetic DES catalog). Figure 6.1 shows $\lambda(z)$ for a variety of halos identified in the DES v3.05 synthetic catalog, both clean halos (*left*) that survive all of our cuts for clean sight-lines as well as several that do not (*right*). We can see that halos from clean sight-lines exhibit the Gaussian shape while the curves from non-clean sight-lines have a more complicated structure, illustrated in Figures 6.1 and 6.2. These curves were binned in $\lambda_{\rm peak}$ and z, then averaged together and fit with a Gaussian distribution as motivated by the earlier findings. The resulting fits to extract W were found to have an average standard deviation of 0.04 in redshift. The amplitude and mean of the fits were determined by the bins. We then used the mean amplitude and width to calculate $P(\lambda|M, z)$ according to equation 6.16, as shown in the following section.

6.4 Results

Using the width for $W(z, z')$ we were able to calculate $P(\lambda|M, z)$ for a few target halo masses. The results are shown in Figure 6.5. To produce this plot, we made an additional assumption, that the mean of the richness–mass relation for clean sight-

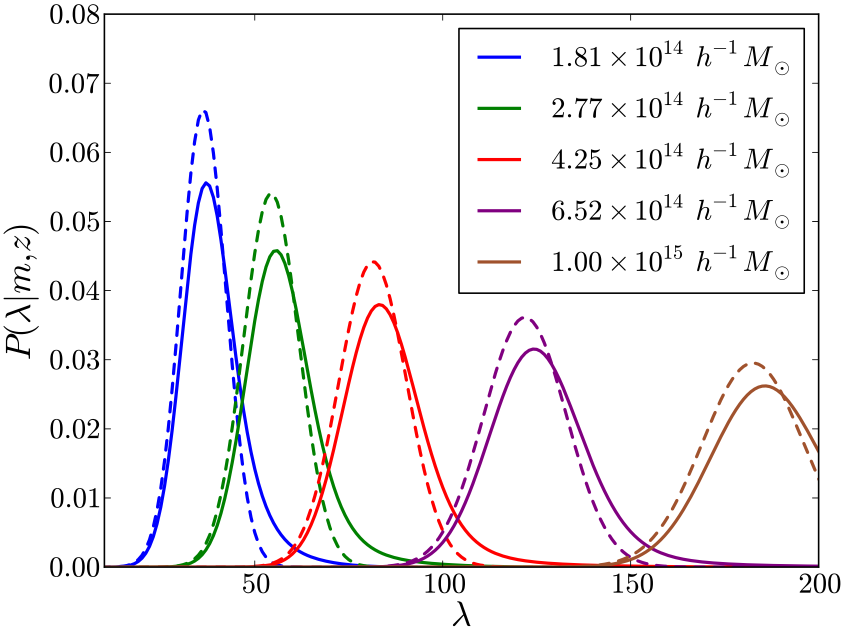


Figure 6.5: The richness–mass relation for λ, $P(\lambda|M,z)$. The *dashed* line shows the intrinsic relation, $P_0(\lambda|M,z)$, in the absence of projection effects, while the *solid* lines include projection, given by equation 6.16. Lines are shown for five target halo masses, M, at redshift $z = 0.3$. There is a boosted probability to larger λ from projection effects.

lines follows the same value as Rykoff et al. (2012):

$$\bar{\lambda}_0(M,z) = 14.85\left(\frac{M}{10^{14}h_{70}^{-1}M_\odot}\right)^{0.94}. \tag{6.18}$$

This relation was determined by Rykoff et al. (2012) by evaluating λ for SDSS `maxBCG` clusters with *ROSAT* (*e.g.,* Voges et al., 1999) X-ray follow-up data as an additional mass proxy. In this figure, the assumed distribution for clean sight-lines, $P_0(\lambda|M,z)$, is shown with a dotted line, and the distribution with projection calculated according to equation 6.16 is shown as the solid line. There is a higher probability of observing a large richness system due to the line-of-sight projection when compared to the original assumption, as expected.

The next step in this work will be to compare the two cases illustrated in Figure

6.5 with the distribution of halos in a large simulation. Since the v3.05 catalog covers only ~200 square degrees of sky, there is not sufficient statistical power to characterize the distributions. Without more statistical power, it is difficult to say how well the model works, or will work in real data. To address these questions, halos from larger area synthetic catalogs will need to be analyzed. The DES Blind Cosmology Challenge (BCC) Aardvark, Beluga and Chinchilla catalogs will provide an opportunity to test the model with larger survey areas.

6.5 Summary

This Chapter presented a work in progress to analyze projection in redMaPPer catalogs. Motivated by the results of Chapter 5 we looked into if redMaPPer clusters show the same kind of contamination found in Cohn et al. (2007). By examining how redMaPPer measures richness as a function of redshift about halo centers, we found that there are cases where projection effects interfere with richness estimates. We then used a simple HOD model to argue that most projection should come from a most massive projector halo, motivating a two-halo projection model. We gave a short overview of the model, and showed how the model predicts projection should appear in the richness–mass relation. The next steps are to more finely calibrate and test this model in large DES synthetic catalogs.

CHAPTER 7

Conclusion

This is a transitional time for cosmology. For much of the twentieth century, discoveries and knowledge were advanced with newer and larger instruments because the field was starved for data. At the end of the century however, digital computers and digital cameras created an opportunity to advance quickly; we settled on the ΛCDM model for the Universe, and all observational campaigns agreed on that model and its parameters at the percent level. Moving into the twenty-first century, we have digital ground-based and space-based telescopes that will produce enormous high-quality data sets. In this era, we will need to be able to extract precise information from these data sets which will require new analysis paradigms: both in computational modeling and data analysis will need to move from desktop to distributed computing resources and analysis will need to focus on previously insignificant systematic sources of error. In this dissertation, I attempted to create an environment for the former, and make a contribution to the latter.

7.1 Workflow

We worked to develop a workflow environment for running cosmological simulations in a grid-aware context. Ideally, the final product would have been a web interface where a user could select a set of cosmological parameters, and the Apache Airavata suite would have launched an appropriate compute job on eXtreme Science and Engineering Discovery Environment (XSEDE) resources. This interface would allow us to rapidly produce synthetic skies for the Dark Energy Survey (DES) Blind Cosmology Challenge (BCC). After two years of effort, we have working solution for producing BCC-like simulations with low user overhead to produce N-body simulations that uses wall time with ∼70% efficiency, though we do not have a web front end. The workflow produces modest gains in efficiency for producing our simulations, and an easier interface for setting up the simulations, and in that sense, our workflow project is a success. However, I believe that other grid-aware computing solutions should be explored for future simulation capaigns.

7.2 Projection

In Chapter 5 we showed that line-of-sight projection in optical galaxy clusters is a source of systematic error that should not be ignored in current and future cluster surveys. Ignoring projection can cause significant bias in recovered cosmological parameters. With this in mind, I worked to model and characterize line-of-sight projection for the red-sequence Matched-filter Probabilistic Percolation (redMaPPer) cluster finding algorithm. Though the model described generates a qualitatively correct richness–mass relationship, I was not able to completely validate the model because of sample size limitations.

7.3 Future directions

There is additional work to be done on both fronts explored in this dissertation. Grid-based computing solutions are helping other fields improve their rate of discovery, exemplified by nanoHUB, and cosmology will surely need to join those fields in maximizing discovery by harnessing computing resources as efficiently as possible. Exploring how to use a hubZERO (`http://hubzero.org/`) interface could help computational cosmologists collaborate. Additionally, other groups are building cyberinfrastructure for cosmology that could yet blossom into a collaborative environment that fully harnesses grid-computing resources, *e.g.*, the yt project (`http://yt-project.org/`) and the Galaxy workflow engine as implemented at Argonne National Laboratory (Heitmann et al., 2013). In any case, all possibilities should be explored for improving how cosmologists harness computational resources, by reducing the effort scientists must spend on infrastructure and computer science issues. Line-of-sight projection for redMaPPer can be explored in more detail as soon as the newest redMaPPer version (v5) can produce the necessary $\lambda(z)$ information. Large synthetic catalogs already exist, which can improve the statistical issues described in Chapter 6. Working to improve simulations and precision cluster cosmology analysis will allow us to address some fundamental questions about dark energy—such as if the equation of state evolves with time—and explaining dark energy is why we play the game.

APPENDICES

APPENDIX A

Details of the FLRW metric

A.1 FLRW

We start with the Einstein equation, relating the curvature of spacetime and the energy density therein:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} - \Lambda g_{\mu\nu} = -\frac{8\pi G}{c^4}T_{\mu\nu}. \tag{A.1}$$

where $g_{\mu\nu}$ is the metric tensor, $R_{\mu\nu}$ and R are the Ricci tensor and scalar, G is Newton's gravitational constant and Λ is the cosmological constant.

We shall investigate these terms one at a time. First, the metric tensor $g_{\mu\nu}$. As argued in section 2.1, we live in a homogenous and isotropic universe, for which the appropriate metric the Freidman–Lemaître–Robertson–Walker (FLRW) metric, given by the space-time interval

$$ds = -c^2dt^2 + R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \tag{A.2}$$

which gives the associated metric tensor components:

$$g_{00} = c^2; \tag{A.3}$$

$$g_{11} = -\frac{R^2(t)}{1-kr^2}; \tag{A.4}$$

$$g_{22} = -r^2 R^2(t); \tag{A.5}$$

$$g_{33} = -R^2(t) r^2 \sin^2\theta. \tag{A.6}$$

For a perfect fluid, where there are no forces between particles and no heat conduction or viscosity, then all of the off-diagonal terms of the energy-momentum tensor will be 0. The T^{00} component is the total energy density, and each T^{ii} is the pressure in the i-direction:

$$T_{00} = \rho c^2; \tag{A.7}$$

$$T_{11} = p; \tag{A.8}$$

$$T_{22} = p; \tag{A.9}$$

$$T_{33} = p. \tag{A.10}$$

$$\tag{A.11}$$

This leaves us the Ricci tensor and scalar. The Ricci scalar can be written $R = 8\pi G/c^4 T^{\mu}_{\mu}$, with $T^{\mu}_{\mu} = \rho c^2 - 3p$, and the Ricci tensor is an awful mess. The Ricci tensor can be written in terms of the affine connection of the metric (also known as a Christoffel symbol) $\Gamma^{\sigma}_{\mu\nu}$,

$$R_{\mu\nu} = \partial_\nu \Gamma^{\sigma}_{\mu\sigma} - \partial_\sigma \Gamma^{\sigma}_{\mu\nu} + \Gamma^{\rho}_{\mu\nu}\Gamma^{\sigma}_{\rho\nu} - \Gamma^{\rho}_{\mu\nu}\Gamma^{\sigma}_{\rho\sigma}, \tag{A.12}$$

where the Christoffel symbols can be written in terms of the metric tensor as

$$\Gamma^{\sigma}_{\mu\nu} = \frac{1}{2} g^{\sigma\rho} \left(\partial_{\nu} g_{\rho\mu} + \partial_{\mu} g_{\rho\nu} - \partial_{\rho} g_{\mu\nu} \right). \tag{A.13}$$

If we were to follow through on all of the algebra for all of these myriad terms, we would find that most of the $\Gamma^{\sigma}_{\mu\nu}$ are zero, and we would be left with the following non-zero components:

$$\Gamma^{0}_{11} = \frac{R\dot{R}}{c^2(1-kr^2)} \qquad \Gamma^{0}_{22} = \frac{R\dot{R}r^2}{c^2} \qquad \Gamma^{0}_{33} = \frac{R\dot{R}r^2 \sin^2\theta}{c^2} \tag{A.14}$$

$$\Gamma^{1}_{01} = \frac{R}{\dot{R}} \qquad \Gamma^{1}_{11} = \frac{kr}{1-kr^2} \qquad \Gamma^{1}_{22} = -r(1-kr^2) \tag{A.15}$$

$$\Gamma^{1}_{33} = -r(1-kr^2)\sin^2\theta \tag{A.16}$$

$$\Gamma^{2}_{02} = \frac{\dot{R}}{R} \qquad \Gamma^{2}_{12} = \frac{1}{r} \qquad \Gamma^{2}_{33} = -\sin\theta\cos\theta \tag{A.17}$$

$$\Gamma^{3}_{02} = \frac{\dot{R}}{R} \qquad \Gamma^{3}_{13} = \frac{1}{r} \qquad \Gamma^{3}_{23} = \cot\theta. \tag{A.18}$$

Then combining the non-zero terms and doing yet more algebra, we would find the following set of components:

$$R_{00} = 3\frac{\ddot{R}}{R}; \tag{A.19}$$

$$R_{11} = -\frac{\left(R\ddot{R} + 2\dot{R} + 2c^2k\right)}{c^2\left(1-kr^2\right)}; \tag{A.20}$$

$$R_{22} = -\frac{\left(R\ddot{R} + 2\dot{R} + 2c^2k\right) r^2}{c^2}; \tag{A.21}$$

$$R_{33} = -\frac{\left(R\ddot{R} + 2\dot{R} + 2c^2k\right) r^2 \sin^2\theta}{c^2}. \tag{A.22}$$

Now we can substitute our expressions into (A.1) to arrive at two independent equations:

$$3\frac{\ddot{R}}{R} = -\frac{1}{2}\frac{8\pi G}{c^4}\left(\rho c^2 + 3p\right)c^2 + \Lambda c^2 \tag{A.23}$$

$$R\ddot{R} + 2\dot{R} + 2c^2k = R\left[\frac{1}{2}\frac{8\pi G}{c^4}\left(\rho c^2 - p\right)c^2 + \Lambda c^2\right] \tag{A.24}$$

Combining these two equations to eliminate $\ddot{R}(t)$ from the second equation, we get

$$\ddot{R} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right)R + \frac{1}{3}\Lambda c^2 R \tag{A.25}$$

$$\dot{R}^2 = \frac{8\pi G}{3}\rho R^2 + \frac{1}{3}\Lambda c^2 R^2 - c^2 k, \tag{A.26}$$

which are known as the Friedmann-Lemaître equations.

A.1.1 Components

Before moving forward with the Friedmann–Lemaître equations, it is convenient to think about the possible components of the density and pressure terms in equations A.25 and A.26. There are three significant components to the density from a cosmological perspective: matter, radiation and dark energy, Λ, so that the total energy density can be written

$$\rho(t) = \rho_{\mathrm{m}}(t) + \rho_{\mathrm{r}}(t) + \rho_{\Lambda}(t), \tag{A.27}$$

where, for practical purposes we assume that the fluids do not interact (not true for matter and radiation), and treat each as an ideal fluid with an equation of state (relationship between state variables p and ρ) given by

$$p_i = w_i \rho_i c^2 \tag{A.28}$$

where w is called the equation-of-state parameter. For matter, or pressureless "dust," $w = 0$. For radiation, $w = \frac{1}{3}$. For vacuum energy, or cosmoloigcal constant dark energy in a ΛCDM model, $w = -1$. Note that matter can come in two forms, baryonic or dark matter, and that dark matter itself can come in two varieties, hot dark matter or cold dark matter. For our purposes here, the differences between the types of matter are not important.

If we differentiate equation A.26 with respect to time and eliminate $\ddot{a}(t)$ in equation A.25, we arrive at a continuity equation for energy,

$$\dot{\rho}(t) + \frac{3\dot{R}}{R}\left(\rho + \frac{p}{c^2}\right) = 0. \tag{A.29}$$

This equation is recognizable as a statement of the first law of thermodynamics for a reversible adiabatic process at fixed number density—that is $dN, dS = 0$ in $dE = -pdV + TdS + \mu dN$. We can rewrite this equation in a suggestive way,

$$\frac{d}{dt}\rho R^3 = -\frac{3p\dot{R}R^2}{c^2} \tag{A.30}$$

and then move $\dot{a}(t)$ to the left hand side, trading a time derivative for one with respect to $a(t)$, we get

$$\frac{d}{dR}\rho R^3 = -\frac{3pR^2}{c^2}. \tag{A.31}$$

If we plug our expression for the equation-of-state of the fluid components into equation A.31, we get

$$\frac{d}{da}\rho R^3 = -3w\rho R^2 \tag{A.32}$$

which gives

$$\rho \propto R^{-3(1+w)} \tag{A.33}$$

If we consider our three components, and plug their equation of state into equation

A.33, we find

$$\rho_{\mathrm{m}}(t) = \rho_{\mathrm{r},0}(1+z)^3 \tag{A.34}$$

$$\rho_{\mathrm{r}}(t) = \rho_{\mathrm{r},0}(1+z)^4 \tag{A.35}$$

$$\rho_{\Lambda}(t) = \rho_{\Lambda,0} = \frac{\Lambda c^2}{8\pi G} \tag{A.36}$$

where we have introduced the cosmological redshift z,

$$1 + z = \frac{R_0}{R(t)}. \tag{A.37}$$

We will from here forward also used the normalized scale factor,

$$a(t) = \frac{R(t)}{R_0} \tag{A.38}$$

which is convenient because $a_0 = 1$ by definition. Then the redshift z and scale factor are related by

$$a(t) = \frac{1}{1+z}. \tag{A.39}$$

Looking at equations A.34–A.36, we see that when $R(t)$ is small, radiation was the dominant source of energy. As the Universe expands, matter becomes the dominant factor, and as the Universe continues to expand, the vacuum or dark energy dominates the energy density.

When discussing cosmology, working with dimensionless parameters is more common and convenient than working with dimensional quantities. For the case of densities, the useful parameters are

$$\Omega_i(t) \equiv \frac{8\pi G}{3H^2(t)}\rho_i(t) \tag{A.40}$$

where $H(t)$ is the Hubble parameter

$$H(t) = \frac{\dot{R}(t)}{R(t)} \tag{A.41}$$

and $i = \{\mathrm{m}, \mathrm{r}, \Lambda\}$. Thus, a particular ΛCDM model can be specified with the parameters

$$H_0, \Omega_{\mathrm{m},0}, \Omega_{\mathrm{r},0}, \Omega_{\Lambda,0} \tag{A.42}$$

Rewriting equation A.26 in terms of the dimensionless densities, and introducing an additional density parameter that describes the curvature

$$\Omega_k(t) = -\frac{c^2 k}{H^2(t) R^2(t)}, \tag{A.43}$$

gives the simple equation

$$\Omega_{\mathrm{m}} + \Omega_{\mathrm{r}} + \Omega_\Lambda = 1 - \Omega_k, \tag{A.44}$$

where all the variables have time dependance.

There are three types of Universes that can be specified by the parameters on the left hand side of this equation:

$$\Omega_{\mathrm{m}} + \Omega_{\mathrm{r}} + \Omega_\Lambda < 1 \Leftrightarrow k = -1, \mathrm{open} \tag{A.45}$$

$$\Omega_{\mathrm{m}} + \Omega_{\mathrm{r}} + \Omega_\Lambda = 1 \Leftrightarrow k = 0, \mathrm{flat} \tag{A.46}$$

$$\Omega_{\mathrm{m}} + \Omega_{\mathrm{r}} + \Omega_\Lambda > 1 \Leftrightarrow k = 1, \mathrm{closed.} \tag{A.47}$$

CMB measurements have essentially limited the parameters to be those that give a flat Universe, where k=0. In this case, $\Omega = 1$ and total energy density is equal to the critical density,

$$\rho_{\mathrm{crit}} \equiv \frac{3H^2}{8\pi G}. \tag{A.48}$$

Returning to the Friedmann-Lemaître equations, in particular, equation A.26, recalling that $H = \dot{R}/R$, and using equations (A.34, A.35, A.36, A.40, A.43) we have an equation for the dynamical behavior of the Universe or structure of spacetime,

$$H^2 = \left(\frac{da}{dt}\right)^2 = H_0^2 \left(\Omega_{\mathrm{m},0} a^{-3} + \Omega_{\mathrm{r},0} a^{-4} + \Omega_{\Lambda,0} + \Omega_{k,0} a^{-2}\right), \tag{A.49}$$

Up until this point, we have only considered *cosmological constant* dark energy, but there are a variety of other more general possibilities for dark energy. In this work, we are especially interested in simple models for time evolution of the dark energy density, especially given by the form

$$\Omega_{\mathrm{DE}}(a) = \frac{\Omega_{\Lambda,0}}{a^{3(1+w(a))}}. \tag{A.50}$$

The subscript DE can also cover Ω_{DE} for a variety of dark energy models, while the subscript Λ implies cosmological constant ($w = -1$) dark energy. The most common time evolution model discussed is that from Linder (2003) where

$$w(a) = w_0 + w_a(1 - a) \tag{A.51}$$

which introduces additional interesting time evolution into the right hand side of equation A.49. Figure A.1 shows a calculated from $H(a)$ for several cosmologies.

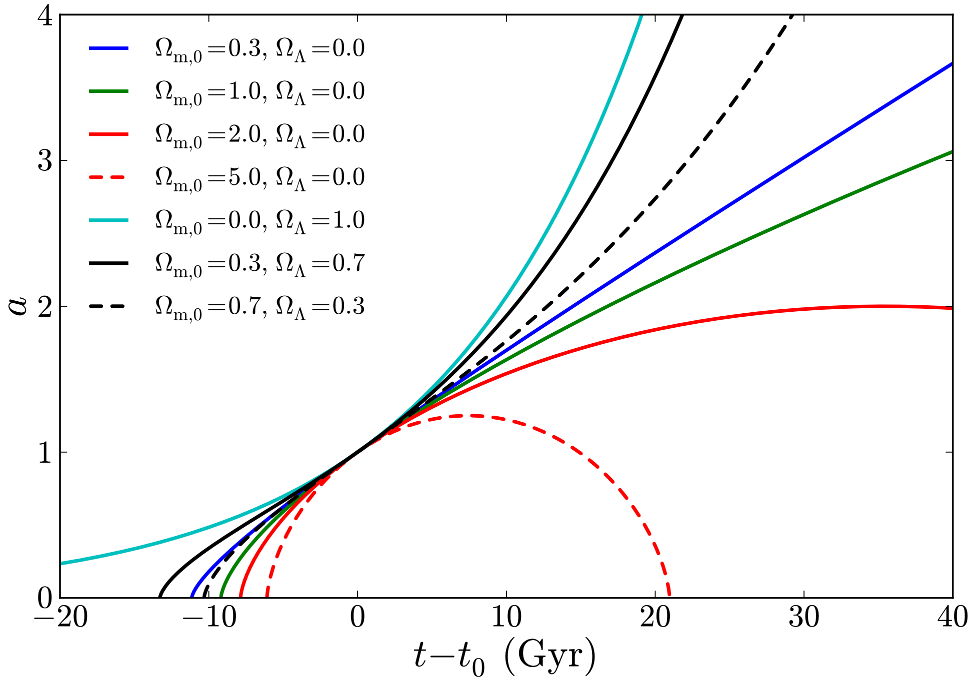


Figure A.1: This plot shows the scale factor, described by $H(a)$ for several cosmological models. The horizontal axis is given in time units in Gigayears, and each plot has the current age of the Universe t_0 subtracted. The *red* lines are closed universe models where the Universe collapses in a *big crunch*. All other models represent universes that end in a *big chill*, that is, they expand forever. The *green* line is a classic Einstein–de Sitter Universe with $\Omega_{\mathrm{m},0} = 1.0$ and no dark energy. The models with dark energy show accelerating expansion as the Universe get older. The *black* lines are ΛCDM models with low and high matter contents, and the cyan line represents a dark energy only Universe. For the dark energy only Universe, t_0 is taken to be the same as the ΛCDM model with $\Omega_{\mathrm{m},0} = 0.3$. The dark energy only universe expands exponentially.

A.2 Structure Formation

At some early time, there must have been regions in the Universe that must have differed in their density from the mean density of the Universe $\overline{\rho}$:

$$\delta_x(\vec{x}) = \frac{\rho(\vec{x}) - \langle\rho\rangle}{\langle\rho\rangle}. \tag{A.52}$$

Initial perturbations are expected to be Gaussian or close to Gaussian (Guth and Pi, 1982; Kolb and M. Turner, 1990), in which case they are isotropic and can be completely described by a power spectrum

$$P(k) = \langle|\delta_k|^2\rangle \tag{A.53}$$

where

$$\delta_k(k) = \int d^3x \delta(\vec{x}) e^{i\vec{k}\cdot\vec{x}}. \tag{A.54}$$

The power spectrum is generally assumed to obey a power-law $P(k) \propto k^n$, and is usually discussed in terms of its tilt, n and its normalization σ_8. The normalization is a constraint on the magnitude of the fluctuations in mass on an 8 Mpc scale. For a spherical top-hat function $W_x(\vec{r})$, a function that falls to zero outside radius $\vec{r}$, the mass perturbation within W is given by

$$\frac{\delta M}{M}(\vec{r}) = \int d^3x \delta(\vec{x}) W(|\vec{x} - \vec{r}|). \tag{A.55}$$

We can write down the variance of the mass perturbation $\sigma^2 = \langle|\delta M/M|^2\rangle$ in terms of the Fourier transform of W as

$$\sigma^2 = \int \frac{d^3k}{(2\pi)^3} P(k) |W_k(k)|^2 \tag{A.56}$$

where for σ_8, the integration is bounded above at 8 Mpc.

Time evolution

Birkhoff's theorem (Birkhoff and Langer, 1923; Peebles, 1993) tells us that for small perturbations, we can essentially treat each of them independently and in the Newtonian limit. As such, we can turn to standard Newtonian mechanics to describe the growth of perturbations. Here, I follow the method of Peebles (1993) though similar arguments, with varying degrees of depth can be found in Kolb and M. Turner (1990); Liddle and Lyth (2000); Voit (2005); Weinberg et al. (2012).

We can treat the density perturbation as an ideal pressureless fluid, with density ρ and velocity field $\vec{u}$. The collection of equations that describe the fluid are the continuity equation for Mass, the Euler equation of motion and the Poisson equation for the gravitational potential:

$$\left(\frac{\partial \rho}{\partial x}\right)_{\mathrm{r}} + \nabla_{\vec{r}} \cdot (\vec{u}\rho) = 0 \tag{A.57}$$

$$\left(\frac{\partial \vec{u}}{\partial t}\right)_{\mathrm{r}} + (\vec{u} \cdot \nabla_{\vec{r}})\vec{u} = -\nabla_{\vec{r}}\varphi \tag{A.58}$$

$$\nabla^2_{\vec{r}}\varphi = 4\pi G\rho. \tag{A.59}$$

These equations are set in an inertial coordinate system, $\vec{r}$, but they also apply in the dynamic Friedmann-Lemaître spacetime of section A.1.1, as long as we transform them into comoving with the expansion rate,

$$\vec{r} \equiv a(t)\vec{x}. \tag{A.60}$$

In this comoving coordinate system, the velocity field is

$$\vec{u} = \dot{\vec{r}} = \dot{a}\vec{x} + a\dot{\vec{x}} = \dot{a}\vec{x} + \vec{v} \tag{A.61}$$

where $\vec{v}$ is the peculiar velocity of the field, that is, the velocity ignoring the background expansion of the Universe. The differential operators transformed to the new coordinate system are as follows

$$\left(\frac{\partial}{\partial t}\right)_{\mathrm{r}} = \left(\frac{\partial}{\partial t}\right)_{\mathrm{x}} - H(t)(\vec{x} \cdot \nabla_{\vec{x}}) \tag{A.62}$$

$$\nabla_{\vec{r}} = \frac{1}{a}\nabla_{\vec{x}}. \tag{A.63}$$

The subscripts above refer to a coordinate system, but from here on, the operators will refer to the comoving coordinate system. After pushing through the changes to the velocity, and the change to the derivatives, and simplifying the expressions, we are left with the following forms for equations (A.57,A.58,A.59).

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot [(1+\delta)\vec{v}] = 0 \tag{A.64}$$

$$\nabla^2 \tilde{\varphi} = 4\pi G \langle \rho \rangle a^2 \delta \tag{A.65}$$

$$\frac{\partial \vec{v}}{\partial t} + H(t)\vec{v} + \frac{1}{a}(\vec{v} \cdot \nabla)\vec{v} = -\frac{1}{a}\nabla\tilde{\varphi} \tag{A.66}$$

In the linear regime of interest, terms of order δ^2 can be dropped, including terms like $\vec{v}\delta$ and $\vec{v}^2$. That leaves us with

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot \vec{v} = 0 \tag{A.67}$$

$$\nabla^2 \tilde{\varphi} = 4\pi G \langle \rho \rangle a^2 \delta \tag{A.68}$$

$$\frac{\partial \vec{v}}{\partial t} + H(t)\vec{v} + \frac{1}{a}\nabla\tilde{\varphi} = 0 \tag{A.69}$$

We can eliminate the peculiar velocity by subtracting the time derivative of the first equation from the divergence of the third. Combined with the Poisson equation for $\nabla^2\tilde{\varphi}$, we are left with an equation for the time evolution of density perturbations

$$\frac{\partial^2 \delta}{\partial t^2} + 2H(t)\frac{\partial \delta}{\partial t} = 4\pi G \langle \rho \rangle \delta \tag{A.70}$$

A.2.1 Linear growth

This differential equation for the growth has a growing and a decaying solution.

$$\delta(\vec{x},t) = A(\vec{x})D_1(t) + B(\vec{x})D_2(t). \tag{A.71}$$

The decaying solution is not interesting to us, because decaying perturbations in the early Universe will not form structures that we can observe today. We can rewrite equation A.70 in terms of the growth factor $D_1(t)$

$$\ddot{D}_1 + 2H(t)\dot{D} - 4\pi G\rho D = 0. \tag{A.72}$$

We are interested in matter perturbations, and we are interested mostly in a regime where dark energy and radiation are not contributing significantly to the density, ρ, so we can use our previous expression for $\rho_{\rm m}$ (A.34) to write this equation

$$\ddot{D}_1 + 2H(t)\dot{D} - \frac{3}{2}\Omega_{\rm m,0}H_0^2(1+z)^3 D = 0. \tag{A.73}$$

For a flat Universe with a cosmological constant ($w = -1$) the equation has the solution

$$D_1(z) = \frac{H(z)}{H_0}\int_z^\infty \frac{dz'(1+z')}{H^3(z')} \tag{A.74}$$

and can be normalized ($D_1 = 1$ at $z = 0$) by dividing by the same integral over all z. This expression is useful for evolving the power spectrum that describes the early density fluctuations to different epochs. While this expression is useful for a cosmological constant, a much more complicated form of the differential equation needs to be solved for more interesting dark energy scenarios. One common question about dark energy is if the equation of state parameter evolves with time, $w(a)$, in that case, the solution given by equation A.74 is invalid, and the full ODE needs

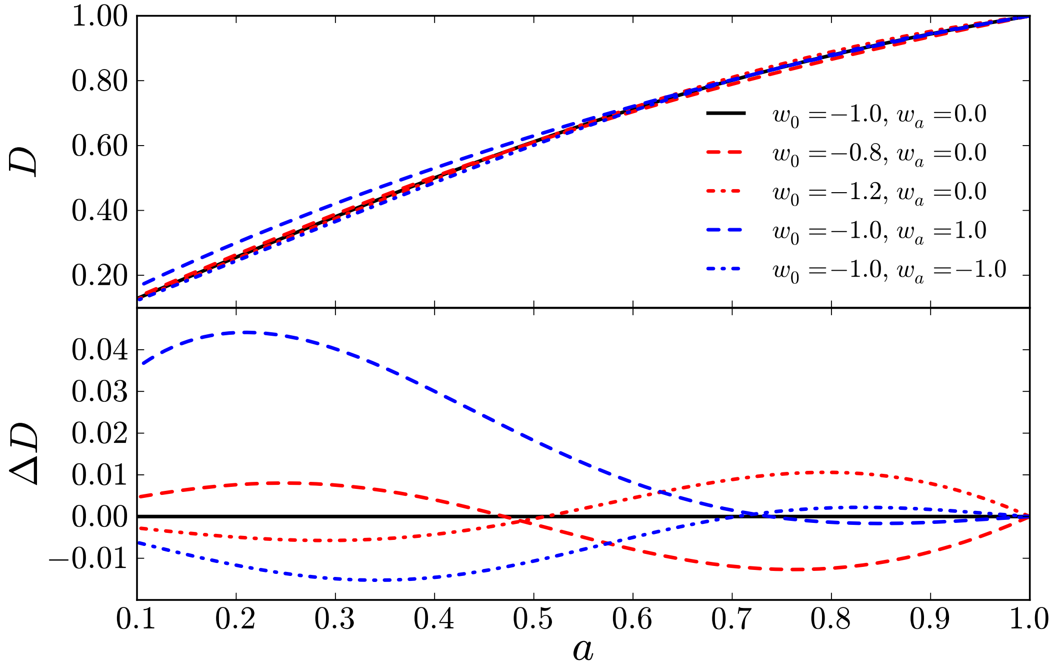


Figure A.2: The cosmological growth factor calculated by solving equation A.75 for $D(a)$. All of the calculations are for a ΛCDM model with $\Omega_{\mathrm{m},0} = 0.3, \Omega_\Lambda = 0.7$ while adjusting the values for w_0 and w_a described by equation A.51.

to be solved numerically. In this case, it is convenient to recast the equation in the following form (Komatsu et al., 2009; Linder and Jenkins, 2003)

$$\frac{d^2 g}{d\ln a^2} + \left[\frac{5}{2} + 12\left(\Omega_k(a) - 3w(a)\Omega_{\mathrm{DE}}(a)\right)\right]\frac{dg}{d\ln a} + \left[2\Omega_k(a) + \frac{3}{2}\left(1 - w(a)\right)\Omega_{\mathrm{DE}}(a)\right] g(a) = 0, \quad \text{(A.75)}$$

where

$$g(a) \equiv \frac{D(a)}{a}, \tag{A.76}$$

$$\Omega_k(a) \equiv \frac{\Omega_{k,0} H_0^2}{a^2 H^2(a)}, \tag{A.77}$$

$$\Omega_{\rm DE}(a) \equiv \frac{\Omega_{\Lambda,0} H_0^2}{a^{3[1+w_{\rm eff}(a)]} H^2(a)}, \tag{A.78}$$

$$w_{\rm eff}(a) \equiv \frac{1}{\ln a} \int_0^{\ln a} d \ln a' w(a'). \tag{A.79}$$

A.2.2 Spatial mass distribution

All of the discussions so far have considered linear large scale structure growth, but we can investigate a particular case of non-linear structure formation if we ask the question what happens if a density perturbation is sufficiently large to experience rapid gravitational collapse. First, we claimed in equation A.53 that the primordial density perturbations were Gaussian, which means the distribution of amplitudes ca be described by a Gaussian function

$$p(\delta) = \frac{1}{\sqrt{2\pi\sigma^2(M)}} \exp\left[-\frac{\delta^2}{2\sigma^2(M)}\right] \tag{A.80}$$

Press and Schechter (1974) address the question at hand. Their analysis begins by calculating the fraction $F(M)$ of objects of a given mass M which become bound at a particular epoch, because they have densities greater than some critical density $\delta_{\rm c}$,

$$F(M) = \frac{1}{\sqrt{2\pi\sigma^2(M)}} \int_{\delta_{\rm c}}^{\infty} d\delta \exp\left[-\frac{\delta^2}{2\sigma^2(M)}\right] = \frac{1}{2}\left[1 - \Phi(t_{\rm c})\right] \tag{A.81}$$

where $t_{\rm c} = \delta_{\rm c}/\sqrt{2\sigma^2}$ and $\Phi(x)$ is the integral

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{A.82}$$

Since the distribution of fluctuations is Gaussian, the mean is zero but the variance is

$$\left\langle \delta^2 \right\rangle = \sigma^2(M). \tag{A.83}$$

For a power-law power spectrum $P(k) \propto k^n$, we can estimate the mass variance by using the two point correlation function (see *e.g.*, Peebles (1993), Longair (2008)) smoother over a radius r,

$$\xi(r) \propto \int \frac{\sin kr}{kr} P(k) k^2 dk. \tag{A.84}$$

Since $\sin kr/kr$ is ≈ 1 for small kr and vanishes quickly for $kr \gg 1$, if we integrate this from 0 to $k_{\max} = 1/r$ we get $\xi(r) \propto r^{-(n+3)}$. Now, since M is proportional to $r^3 (M \approx \rho r^3$, we can write

$$\xi(M) \propto M^{-(n+3)/3}. \tag{A.85}$$

The two-point function ξ is related to the root-mean-square density fluctuations directly (again, see *e.g.*, Peebles (1993), Longair (2008)) so we have

$$\sigma^2(M) = AM^{-(n+3)/3}. \tag{A.86}$$

where A is a constant.

Now we can express $t_{\rm c}$ in terms of the mass fluctuations.

$$t_{\rm c} = \frac{\delta_{\rm c}}{\sqrt{2\sigma^2(M)}} = \frac{\delta_{\rm c}}{\sqrt{2A}} M^{3+n}/6 = \left(\frac{M}{M_*}\right)^{(3+n)/6} \tag{A.87}$$

where M_* is a reference mass, $M_* = (2A/\delta_{\rm c}^2)^{3/(3+n)}$.

Now, if we wish to estimate the mass distribution of collapsed objects, we need to consider the number of objects in a mass range $M + dM$. The fraction of objects falling in that mass range is given by $dF = (dF/dM)dM$. Also, the mass of the perturbation is $M = \bar{\rho} V$, where $\bar{\rho}$ is the mean background density. Once the perturbation begins non-linear growth, an object of mass M forms, and the density of such objects is $1/V$, so

$$N(M)dM = \frac{1}{V} = -\frac{\bar{\rho}}{M}\frac{dF}{dM}dM, \tag{A.88}$$

where the minus sign accounts for the fact that F is a decreasing function of M. Combining equations A.81 and A.88 and doing a stack of algebra not presented, one arrives at an expression for the space density of massive objects,

$$N(M) = \frac{\bar{\rho}}{2\sqrt{\pi}}\frac{\gamma}{M^2}\left(\frac{M}{M_*}\right)^{\gamma/2} \exp\left[-\left(\frac{M}{M_*}\right)^{\gamma}\right]. \tag{A.89}$$

In this expression, $\gamma = 1_{(}n/3)$. Information about time evolution is wrapped up into the reference mass M_* via δ_{c}. This expression undercounts the mass by a factor of 2, because once an object starts to collapse, it tends to accrete mass from its vicinity. This equation, with the corrected factor of 2 is often written in the form

$$\frac{dn(M,z)}{dM} = \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}(z)}{M^2}\left|\frac{d\ln\sigma(M,z)}{d\ln M}\right|\nu_{\mathrm{c}}(M,z)\exp\left[\frac{-\nu_{\mathrm{c}}^2(M,z)}{2}\right]. \tag{A.90}$$

In the above equation, ν is the critical overdensity in units of variance, $\delta_{\mathrm{c}}/\sigma$. Again, cosmological information is folded into the critical density.

Since the number density of large collapsed dark matter objects—halos—is observable because the halos host galaxy clusters, measurements of equation A.90 can be used to constrain cosmological models. In practice, this can be difficult, because estimating the mass and location of dark matter halos can be difficult. In the years since 1974, a number of new parameterizations for the mass function have been explored and calibrated using N-body simulations, including Jenkins et al. (2001); Warren et al. (2006) and Tinker et al. (2008).

APPENDIX B

N-body simulation

N-body simulations require at least two steps. First, one must generate the initial state particle set and then that state needs to be evolved by an appropriate N-body technique. In astrophysics, N-body simulation can be used to study how molecular clouds develop star systems (*e.g.,* Adams et al., 2006), how planetary systems evolve (*e.g.,* Ketchum, Adams, and Bloch, 2013), how galaxies form and evolve (*e.g.,* D'Onghia, Vogelsberger, and Hernquist, 2012; Shin and Ruszkowski, 2013), or, for our purposes, how the large-scale structure of the Universe evolves (Bertschinger, 1998). Each of these situations requires different initial conditions. For a planetary system, it is easy to imagine assigning about ten particles masses and velocities to evolve (planets and a star), though it may be less simple to imagine how to account for say, asteroids and moons. Cosmological simulations on one hand are more simple, because there are many particles and the starting time for cosmological simulations usually reaches far enough into the past that the positions are close to uniformly distributed and their velocities may be small. However, to guarantee that the simulation will produce a universe that matches the observed statistical properties of our Universe, some care needs to be taken. First, the large- and small-scale power of current-epoch structures are characterized in a matter power spectrum, and then an initial conditions generator uses that information to create an early Universe particle set that has statistically appropriate properties—plus randomization—to be evolved by an N-body code. For generating a matter power distribution, we use Code for Anisotropies in the Microwave Background (CAMB) , for our initial conditions, we use Second-order Langrangian Perturbation Theory Initial Conditions (2LPTic) and for our N-body evolution, we use LGadget. Each of these codes are described in this Appendix in sections B.1, B.2 and B.3

B.1 CAMB

We use the Code for Anisotropies in the Microwave Background written by Lewis, Challinor, and Lasenby (2000) to compute a matter power spectrum. As mentioned in Appendix A.2, the power spectrum for large scale structure is assumed to be smooth ($P(k) \propto k^n$), but that is not a sufficiently accurate assumption for modern high-precision sky surveys (Eisenstein and Hu, 1998). The matter power spectrum has a number of features due to interactions of all types of particles (baryons, photons, neutrinos) in the early Universe, which imprints some features beyond the smooth k^n form of the power spectrum. To fully account for those features, coupled linearized Boltzmann and Einstein equations with Thomson scattering need to be solved for perturbations in the early Universe. Fully solving all of the coupled differential equations in the Boltzmann equation can be computationally intensive, so CAMB uses a fast line-of-sight integration technique first presented by Seljak and Zaldarriaga (1996) to perform the same calculation quickly. The resulting power spectrum of fluctuations for each species can then be evolved forward in time and is often described by a transfer function. The transfer function relates the primordial power spectrum (still assumed to be $p(k) \propto k^n$) to the linear power spectrum at a later redshift,

$$P_{\text{lin}}(k,z) = AP_{\text{prim}}(k)T^2(k,z), \tag{B.1}$$

where A is a normalization constant that can be set by specifying σ_8 for the power spectrum. CAMB accepts cosmological parameters as inputs, as well as a normalization value. CAMB also outputs a number of properties of anisotropies of the Cosmic Microwave Background (CMB), but we do not use those features. CAMB is available for download at `http://camb.info/`.

B.2 2LPTic

To take the power spectrum of perturbations to an initial particle we use the Second-order Langrangian Perturbation Theory Initial Conditions code developed by Crocce, Pueblas, and Scoccimarro (2006). Using a linear power spectrum to seed the initial conditions implies that an N-body simulation should be started at the lowest scale factor (or highest redshift) possible, where the true perturbation spectrum is closer to the linear value. However, at smaller scale factors, any imprecision in the initial perturbations becomes more amplified in the late time particles properties, so it is imperative to have the most precise initial conditions possible. Traditionally, N-body simulations use the Zel'dovich Approximation (ZA) (Zel'dovich, 1970) to generate initial conditions (Dolag et al., 2004, 2009; Efstathiou et al., 1985; Evrard, 1990; Klypin and Shandarin, 1983; Metzler and Evrard, 1997), but the ZA has been shown to introduce transients—non-linear effects that survive to low redshift—due to second-order deviations in initial velocity fields because the ZA fails to conserve momentum (Scoccimarro, 1998). We have elected to use 2LPTic, because transients are greatly suppressed. 2LPTic perturbs particle positions from a grid according to second-order Lagrangian perturbation theory—the ZA is essentially first-order Lagrangian perturbation theory—as the name of the code would suggest. To generate a set of initial conditions, a *glass* distribution is first created. To generate a glass, a uniform random particle field is created, and then evolved with a repulsive gravitational force (*i.e.*, reverse the sign of G) until the particles reach stable positions. Then, the particles are perturbed according to

$$\vec{x} = \vec{q} - a\nabla_{\vec{q}}\phi^{(1)} - \frac{3}{7}a^2\nabla_{\vec{q}}\phi^{(2)} \tag{B.2}$$

$$\vec{v} = -aH\nabla_{\vec{q}}\phi^{(1)} - \frac{6}{7}a^2H\nabla_{\vec{q}}\phi^{(2)}, \tag{B.3}$$

where the potentials satisfy the Poisson equation,

$$\nabla_{\vec{q}}\phi^{(1)}(\vec{q}) = \delta(\vec{q}) \tag{B.4}$$

$$\nabla_{\vec{q}}\phi^{(2)}(\vec{q}) = \sum_{i>j}\left\{\phi_{ii}^{(1)}(\vec{q})\phi_{jj}^{(1)}(\vec{q}) - \left[\phi_{ij}^{(1)}(\vec{q})\right]^2\right\} \tag{B.5}$$

where $\vec{q}$ are unperturbed particle positions. Setting $\phi^{(2)} = 0$ recovers ZA initial conditions. For a full discussion of how to derive and implement these equations, see Scoccimarro (1998). 2LPTic is available for download at `http://cosmo.nyu.edu/roman/2LPT/`.

B.3 LGadget

For our N-body calculations, we use LGadget2, a modified version of the public code Gadget-2 Springel (2005) written by Volker Springel and used for the Millennium Simulation (Springel et al., 2005). The L stands for “Lean” because Gadget-2 incorporates smooth particle hydrodynmics along with a TreePM N-body code, and LGadget has stripped out all of the hydrodynamics for a pure N-body code as well as particle data for a number of different simulation options available in Gadget-2, reducing the memory footprint per particle from 80 bytes to 40 bytes.

B.3.1 Time Integration

To solve the equations describing a cosmological N-body system, LGadget uses a symplectic leapfrog integration method known as the kick-drift-kick (KDK) scheme. Since the Hamiltonian for an N-body system (equation 3.28) is separable $H = H_{\rm kin} + H_{\rm pot}$, time evolution operators for each part can be computed exactly (Quinn et al., 1997),

$$\mathcal{D}(\tau) \equiv \vec{x}_i(t+\tau) = \vec{x}_i(t) + \frac{\vec{p}_i(t)}{m_i}\int_t^{t+\tau}\frac{dt}{a^2} \quad \text{and} \quad \vec{p}_i(t+\tau) = \vec{p}_i(t) \tag{B.6}$$

$$\mathcal{K}(\tau) \equiv \vec{p}_i(t+\tau) = \vec{p}_i(t) + \vec{f}_i\int_t^{t+\tau}\frac{dt}{a} \quad \text{and} \quad \vec{x}_i(t+\tau) = \vec{x}_i(t) \tag{B.7}$$

where $\vec{f}_i = -\sum_j m_i m_j \frac{\partial\phi(\vec{x}_{ij})}{\partial x_i}$ is the force on particle i. Here, the gravitational potential ϕ takes on a special form.

$$\phi(\vec{x}_{ij}) = -G\sum_j m_j g(|\vec{x}_j - \vec{x}_i|), \tag{B.8}$$

where $g(u)$ is the spline potential of Monaghan and Lattanzio (1985),

$$g(u) = \begin{cases} \frac{16}{3}u^2 - \frac{48}{5}u^4 + \frac{32}{5}u^5 - \frac{14}{5}, & 0 \le u < \frac{1}{2}, \\ \frac{1}{15}u + \frac{32}{3}u^2 - 16u^3 + \frac{48}{5}u^4 - \frac{32}{15}u^5 - \frac{16}{5}, & \frac{1}{2} \le u < 1, \\ -\frac{1}{u}, & u \ge 1, \end{cases} \tag{B.9}$$

where $u = r/\varepsilon$ with ε being the gravitational softening length. The purpose of gravitational softening is meant to prevent hard collisions between particles in which there could be a large energy transfer, incorrectly leading to particles being ejected from dense regions. The gravitational softening length is usually set to 1/35 of the mean interparticle spacing for a simulation.

Using the Drift (equation B.6) and Kick (equation B.7) operators, a KDK time evolution operator can be constructed

$$\mathcal{U}(\Delta t) = \mathcal{K}\left(\frac{\Delta t}{2}\right)\mathcal{D}\left(\Delta t\right)\mathcal{K}\left(\frac{\Delta t}{2}\right). \tag{B.10}$$

This time-evolution operator has a number of nice properties detailed by Springel (2005), but one of note is how it can be used with adaptive timesteps. Wisely choosing timesteps can reduce the overall time to run a simulation by maintaining accuracy. Particles that are in low density regions can be integrated with larger timesteps than particles in highly dense regions. LGadget chooses the size of the next timestep based on the acceleration of the particle at this timestep, $\Delta t \propto 1/\sqrt{|\vec{a}|}$ and sub-cycles over particles in dense regions.

B.3.2 TreePM

For short-range force computations, LGadget uses a tree method. The purpose of the tree method is to reduce computation time by grouping distant particles into cells, allowing their gravitational potential to be computed using a multi-pole computed for the center of mass for the grid. LGadget uses the J. Barnes and Hut (1986) method. Here, a cubical root node that encompasses all of the particles in the simulation is recursively subdivided into eight child nodes, each of half the side length of the parent until there is only single particles remain at each node. Forces are then computed by traversing the tree. At each node, a decision is made as to whether the multipole potential will be sufficiently accurate at that point. If yes, then the multipole is used, and if not, then the child nodes are considered in turn. Force accuracy can be controlled by the opening criterion, which is what determines if a particular node's multi-pole will be used. In LGadget, the parameter is the opening angle from the particle for which the force is being computed to the edge of the cell in question. The result is that nearby cells, which have large opening angles will be opened to lower levels, while more distant cells which have a smaller opening angle will not. Because the force computation is not done for every pair of particles, the algorithm scales like $\mathcal{O}(N \log N)$.

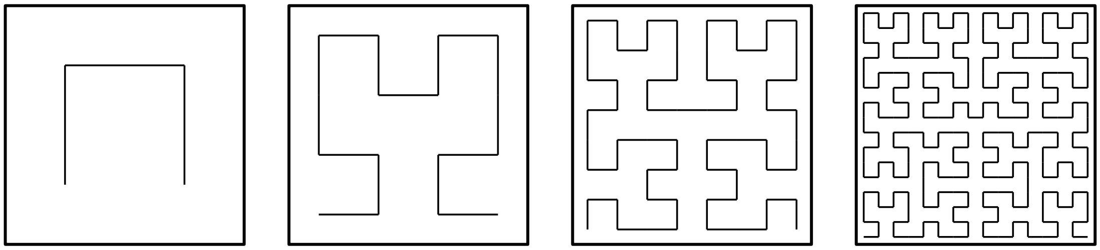
LGadget splits the gravitational potential into two parts, short range and long range. The short range portion of the gravitational potential is computed using the tree algorithm described above. For long range forces, LGadget uses the clouds-in-cells method of Hockney and Eastwood (1988), to construct a mass density on a Particle Mesh (PM). The fourier transform of the mass density, ρ_k can the be used to easily calculate the gravitational potential

$$\phi_{\vec{k}} = -4\pi G a^2 \frac{\rho_{\vec{k}}}{|\vec{k}|^2} \tag{B.11}$$

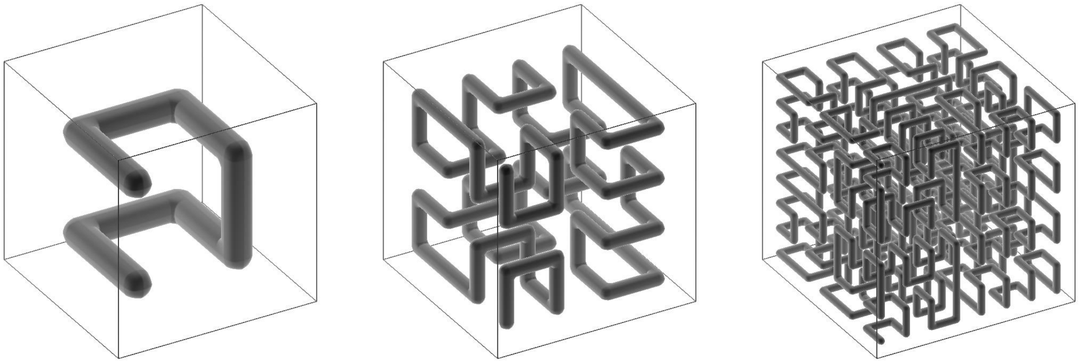
Then the Fourier transform of the potential is then transformed back to a grid and interpolated to give the potential for every simulation particle. This method is efficient for computing long range forces, and is immensely more computationally efficient than direct summation or Tree could be for long ranges.

B.3.3 Parallelization

Each particle requires 44 Bytes of memory, and the PM calculation and Barnes–Hut tree need additional memory. This means that for simulations with large particle number, the available memory to a single processing core can rapidly be consumed. For the Blind Cosmology Challenge (BCC) simulations, we used up to 2048^3 particles, which requires 352 GB of memory for particle data alone. About the same amount of memory is required for the PM calculation. Clearly, our computations require many processors, and so particles need to be distributed across processors in some way. This is done in LGadget by breaking the simulation volume into different domains along a space-filling fractal called a Peano–Hilbert curve. The Peano–Hilbert curve is nice for cosmological simulations because points that are close to each other along the length of the curve are also near each other in 3-dimensional space, which, reduces interprocessor communication time in a cosmological simulation when compared to other domain decomposition schemes (e.g. slab or grid). Also, the Peano–Hilbert curve has a nice property that it can be sliced in such a way as to be commiserate with the BH tree structure. The actual domain decomposition has a number of detailed steps, which are described in Springel (2005), but the general idea is shown in Figure B.2. The public version of Gadget-2 is available at `http://www.mpa-garching.mpg.de/gadget/` though LGadget is not a publicly available code.

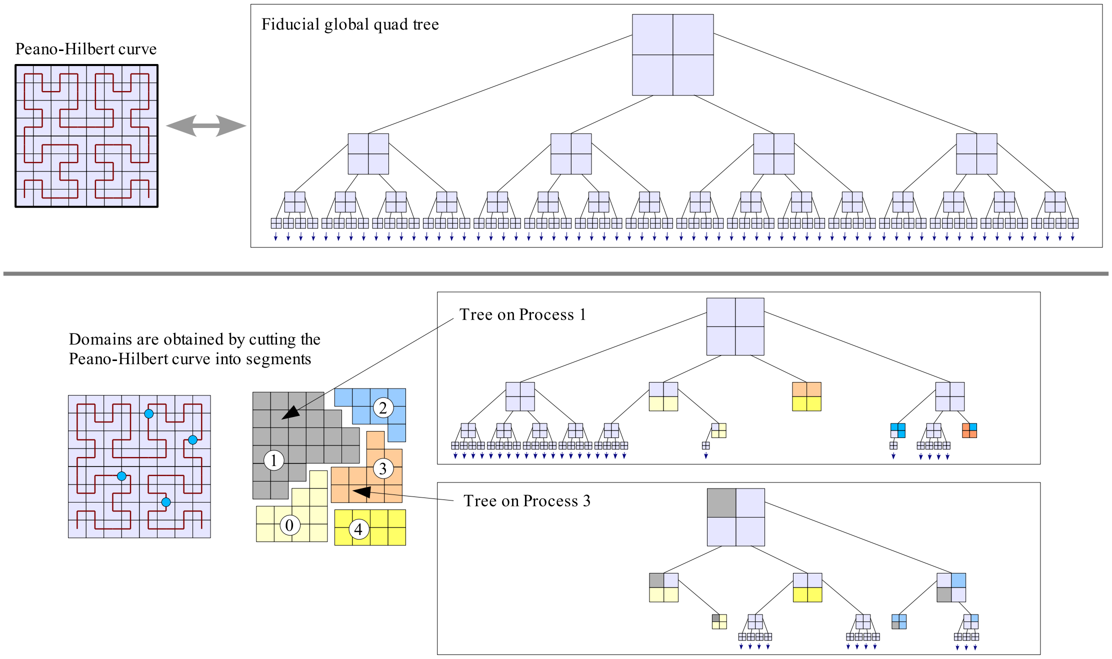


(a) Two-dimensional Peano–Hilbert curves



(b) Three-dimensional Peano–Hilbert curves

Figure B.1: A series of figures illustrating the Peano-Hilbert curve as well as the associated Barnes–Hut tree used by Gadget-2. The *top* row shows a Peano-Hilbert curve in two dimensions, while the *middle* row shows several examples in three dimensions.



(a) The Barnes–Hut tree across processors

Figure B.2: This figure shows the relationship between the Barnes-Hut tree and the Peano-Hilbert curve in two dimensions. The simulation volume is split into domains by cutting the Peano-Hilbert curve at locations indicated by the *blue circles*, with each computing core receiving particles associated with a single domain. Each of the domains corresponds to some branch of the global Barnes-Hut tree, and as such, all the particles for that branch reside on a single processor. Additionally each processor carries a top-level tree that represents higher levels of the global tree, with particles that reside on other processors marked with ghost particles. These figures come directly from Springel (2005).

B.3.4 Lightcones

LGadget contains an on-the-fly lightcone generator. An observer in the Universe, or in a simulation volume, observes photons emitted at time t_e that have traversed the comoving volume at time t, given by

$$r(t) = c\int_{t_\mathrm{e}}^{t} \frac{1}{a(t)} dt \tag{B.12}$$

where $a(t)$ is the scale factor for the spacetime metric. At each timestep in LGadget, particles are checked to see if they have crossed a spherical shell corresponding to $r(t)$ measured from an observers position. If so, the particle's position and velocity is used to quadratically interpolate the time and position at which passed through the lightcone surface and then written to disk. We place an observer at every corner of the simulation box, which means eight sky octants are dumped to disk. Those eight octants can then be stitched together exploiting the periodic box boundary conditions to produce a full sky lightcone with more volume than simply placing an observer at the origin. However, using the full-sky lightcone for statistical analysis can be misleading, because not all eight octants are independent. For Dark Energy Survey (DES), we only use two octants to produce a quartant for each simulation to minimize degeneracy effects.

APPENDIX C

Creating Observable Cluster Properties

C.1 CALCRNN

To estimate the local density of a dark matter particle δ_{DM}, we use a simple nearest neighbor calculation. The code CALCRNN uses a k-d tree to sort particles from a local slice of an LGadget Peano-Hilbert domain to quickly search for the nth nearest neighbor in the domain. A k-d tree splits a set of points into a binary search tree by partitioning the entire set of particles into two sets at each point, recursively, until all particles are assigned to the tree. Figure C.1 shows the spatial arrangement and resulting k-d tree for the set of points $\{(2,3),(5,4),(9,6),(4,7),(8,1),(7,2)\}$.

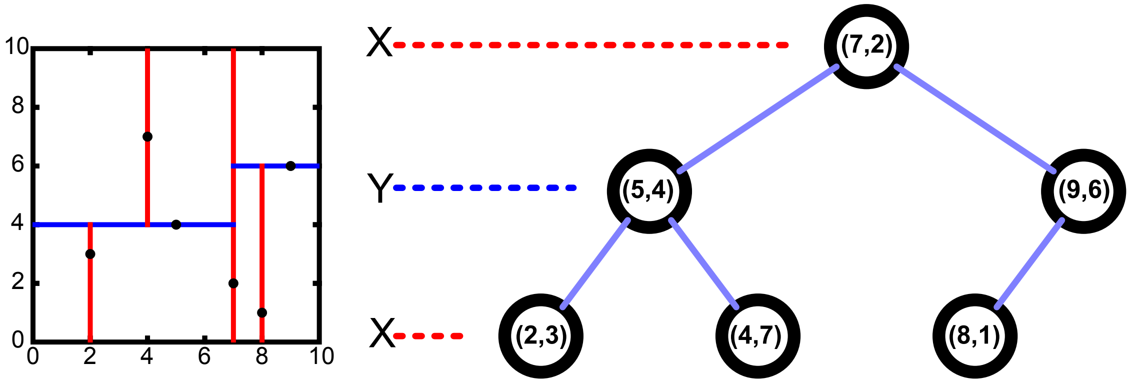


Figure C.1: The spatial cuts *left* and resulting 2 dimensional k-d tree *right* for the point set $\{(2,3),(5,4),(9,6),(4,7),(8,1),(7,2)\}$. Figure thanks to user KiwiSunset at wikimedia commons, CC: BY-SA.

C.2 ROCKSTAR

Observables related to the large scale structure of the Universe are related back to underlying cosmological models via the halo mass function, number density of dark matter halos as a function of mass and redshift. For modern sky surveys to reach their full potential in constraining dark energy, systematic errors in the halo mass function need to be understood at the same order as statistical errors in the observations (Cunha and Evrard, 2010). As such, a robust method for identifying dark matter halos in simulations that is consistent across simulation sizes, resolutions and time-steps is desirable. The Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement (ROCKSTAR) halo finder is designed to use full six-dimensional particle information (positions and velocities) as well as temporal information across simulation snapshots for robust identification of dark matter halos. The code is also written to run efficiently on massively parallel computers and on very large data sets (Behroozi, Wechsler, and Wu, 2013).

ROCKSTAR performs the following steps to identify dark matter halos and their properties. First, halos are approximated using a three-dimensional Friends-of-Friends (FoF) algorithm with a large linking length and distributes groups to individual processors for analysis. A FoF works by locating all particles that are within some distance, l (linking length), of a particle, then, all particles within the linking length of that particle are identified. This process is repeated until a group of friends is found, where all other simulation particles are separated by more than the linking length from all members of the group. Next, particle positions and velocities are normalized by the FoF group positions and velocities to give a phase-space metric. A hierarchy of FoF subgroups is built in phase space by adaptively reducing the six-dimensional linking length so that 70% of the parent group's particles are linked together in subgroups. Once all levels of substructure are identified, seed halos are placed at the lowest level of substructure and particles are assigned hierarchically to the closest seed halo in phase space. If multiple time-steps are available, ROCKSTAR then computes relationships between halos and subhalos. Once particles have been assigned to halos, halo properties are calculated (*e.g.*, position, velocity) excluding unbound particles. These steps are shown schematically in Figure C.2.

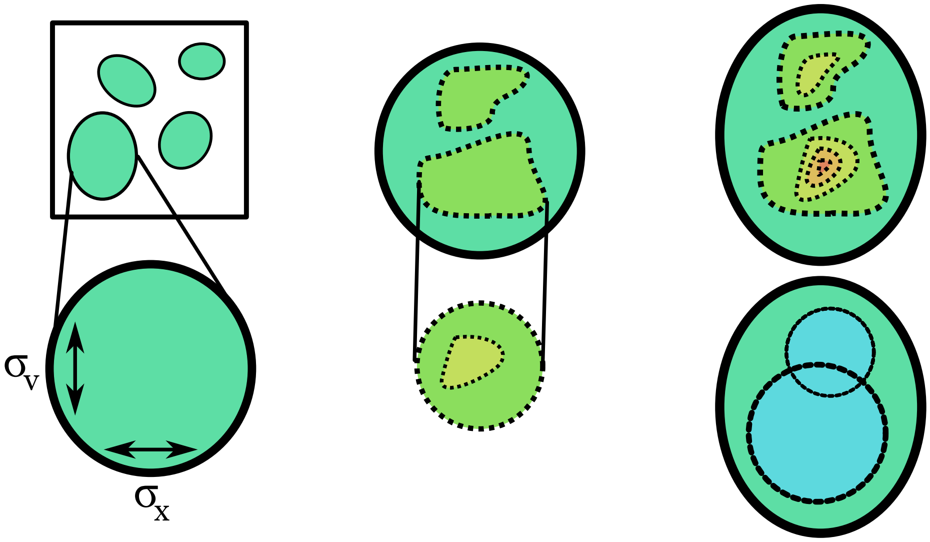


Figure C.2: The steps in ROCKSTAR. On the *left*, 3D FoF groups are identified and normalized. In the center, 6D FoF subgroups are identified and renormalized. On the *right*, all substructure is identified and the collected into dark matter halos. Figure from Behroozi, Wechsler, and Wu (2013).

C.3 CTREES

We also build merger trees and halo formation histories for some simulations. For this task we use the CTREES code by Behroozi et al. (2013) which is publically available at `http://code.google.com/p/consistent-trees/`. The merger trees ensure consistency in halo identification from snapshot to snapshot and reduce some sytematic effects in calculating subhalo properties. Merger trees can inform galaxy formation models (*e.g.*, Wechsler, 2001), and also inform the Sub-Halo Abundance Matching (SHAM) method used by Adding Density Determined GAlaxies to Lightcone Simulations (ADDGALS).

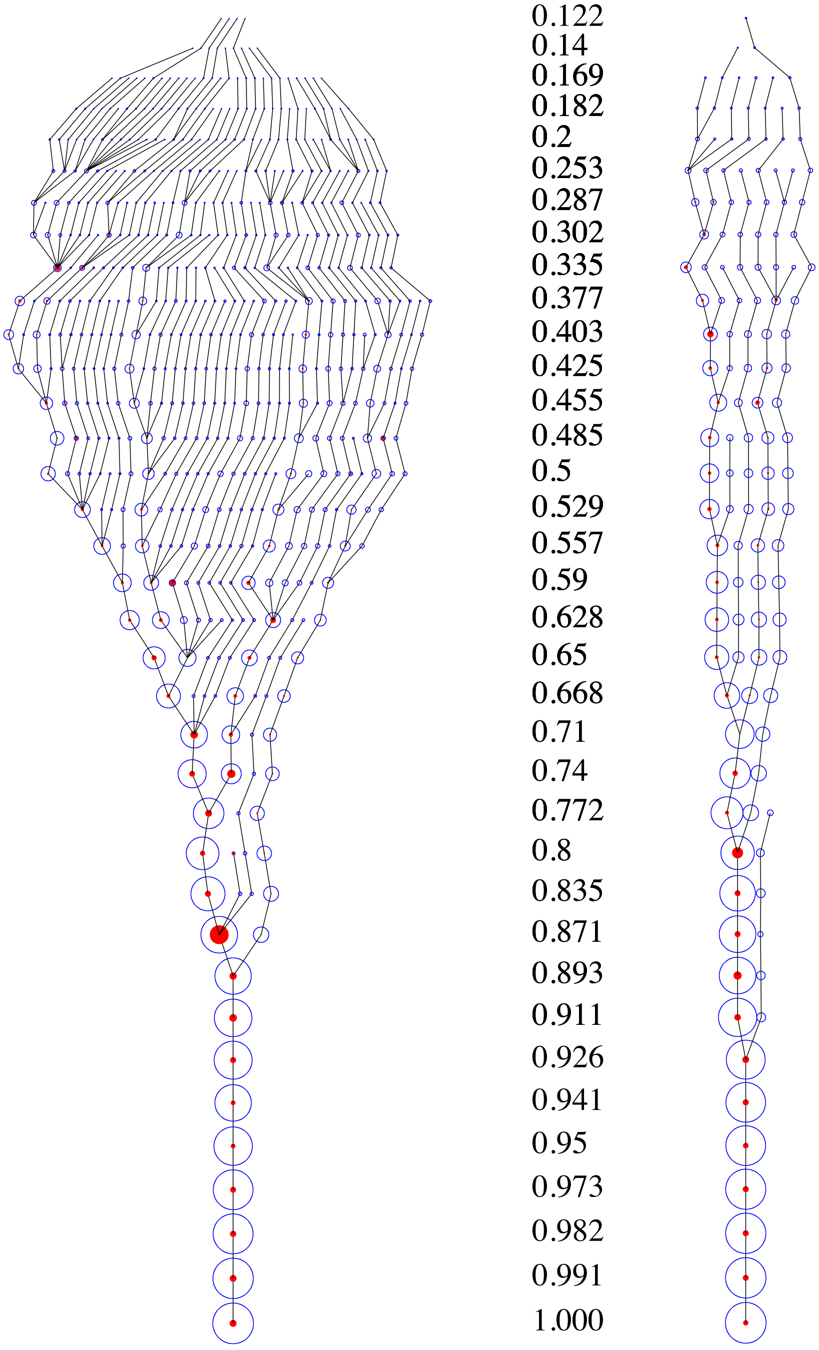


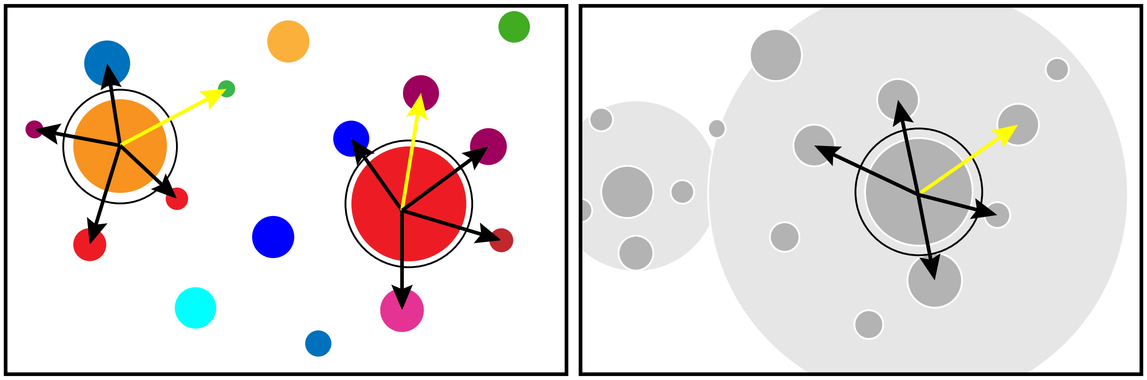
Figure C.3: Examples of merger trees for two halos. This diagram illustrates the merging history of a massive halo and a much smaller halo, with the vertical column of numbers indicating scale factor. The circles are directly related to the halo size, but rescaled to have the same size in the figure at $a = 1$. These trees show how two halos can have very different merger histories. History plays a role in galaxy formation and evolution, and is used in tuning ADDGALS. Figure from (Wechsler, 2001; Wechsler et al., 2002).

C.4 ADDGALS

To produce galaxy catalos from N-body simulations, Dark Energy Survey (DES) uses the ADDGALS algorithm. ADDGALS places synthetic galaxies on dark matter particles in a lightcone output N-body simulations. The galaxies are placed in such a way as to match the luminosities, colors, and clustering properties of galaxies from a variety of redshift surveys including Sloan Digital Sky Survey (SDSS). ADDGALS places galaxies in a simulation via a two step process: a galaxy population based on an input luminosity function are assigned to particles in the lightcone, then photometry is added to each galaxy using a training set of observed galaxies.

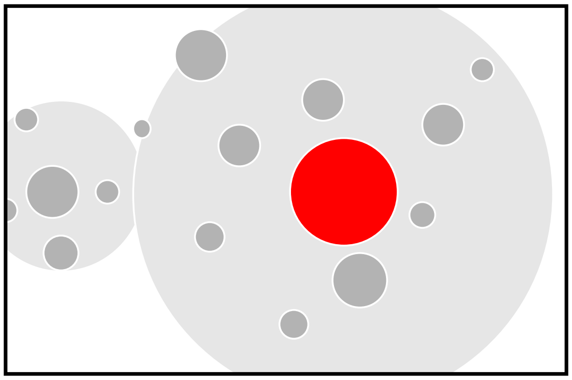
The first step starts by defining the probability that a galaxy with magnitude M_r and redshift z resides in a region with local density δ_DM, defined as a the redius of a sphere containing $1.8 \times 10^{13}\ h^{-1}\ M_\odot$ of dark matter—$P(\delta_\mathrm{DM}|M_\mathrm{r}, z)$. This relation can be tuned to reproduce the luminosity-dependent galaxy two-point function by using a high resolution tuning simulation and a technique known as subhalo abundance matching (Behroozi, Conroy, and Wechsler, 2010; Conroy, Wechsler, and Kravtsov, 2006; Wetzel and M. White, 2010). Once $P(\delta_\mathrm{DM}|M_\mathrm{r}, z)$ is measured using a tuning simulation, galaxies are added to the production simulation by integrating an r-band luminosity function to generate a list of galaxies with magnitudes and redshifts, then drawing a δ_DM for each galaxy from $P(\delta_\mathrm{DM}|M_\mathrm{r}, z)$ and attaching it to an appropriate dark matter particle. This process is applied to dark matter halo satellite galaxies and central galaxies separately to ensure that central galaxy properties match their host halo's properties in a detailed fashion.

Once galaxy positions have been assigned, a Spectral Energy Distribution (SED) is assigned to each galaxy by drawing a random training set galaxy with the appropriate r-band magnitude and local density. The SED is then K-corrected (Hogg et al., 2002) to the appropriate redshift and projected onto the DES survey band-pass filters. When assigning colors, the likelihood of assigning a red or blue galaxy is smoothly varied as a function of redshift to reproduce the observed red fraction of galaxies from various observational catalogs, including SDSS. Finally, photometric errors are applied to mimic DES photometric errors. A conceptual process for adding galaxies and color to a simulation is shown in Figure C.4.



(a) Data

(b) Mock



(c) Mock with color

Figure C.4: The steps in ADDGALS. The *top left* panel represents a collection of observed galaxies. The *top right* panel shows galaxies placed onto dark matter halos, assigned to particular dark matter particles with an appropriate δ_{DM}. To draw an appropriate training set galaxy to assign a spectral energy distribution to a mock galaxy, ADDGALS checks the observed training set for a similar galaxy by looking at the fifth nearest neighbor. In this example, there are two similar galaxies in the data for a particular mock galaxy, all circled in *black*. The five nearest neghbors have arrows pointing to them, with the fifth nearest neighbor indicated in *yellow*. The red galaxy on the right in the data has the most similar environment to the mock galaxy, and so the mock galaxy is assigned the same color as the data, illustrated in the *bottom* panel. This process is repeated to assign colors to all the ADDGALS galaxies. Figure is adapted from one presented by Michael Busha to the DES collaboration.

C.5 CALCLENS

For weak lensing calculations, we use the Curved-sky grAvitational Lensing for Cosmological Light conE simulatioNS (CALCLENS) algorithm. CALCLENS uses a multiple-plane ray tracing algorithm (Jain, Seljak, and S. White, 2000) adapted to a sphere (M. Becker, 2012).

The code proceeds in several steps. First, rays are initiated at the observer, then calculates the lensing potential for the mth lens plane $\psi^{(m)}$ in the multiple-plane and Limber approximations (Limber, 1953) via the two-dimensional Poisson equation. Then the derivative of the lensing potential and propogate the rays to the $m + 1$th lens plane. For galaxies in the $m + 1$th plane, the lensing Jacobian is interpolated onto them. These steps are repeated until the last lens plane is traversed. The code uses a spherical harmonic transform plus multigrid method to solve for the lensing potential. I never ran this application, so it is only mentioned here for completeness, details can be found in M. Becker (2012).

APPENDIX D

Parameter Bias

We often wish to estimate how well we can estimate a parameter from available data. The Cramér-Rao inequality (*e.g.,* Nielsen, 2013) states that a minimum variance, unbiased estimator can only do so well in estimating a target parameter, θ. In its one-dimensional form, the inequality is

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{F(\theta)} \tag{D.1}$$

where $F(\theta)$ is the Fisher information and $\hat{\theta}$ is an estimator for θ. The Fisher Information for a model can be used to calculate the minimum variance that the model can achieve in estimating some parameter. In multiple dimensions, the Fisher information is a matrix, the inverse of which is the covariance matrix for a model. The Fisher Matrix for a model is given by the expectation value of the negative second derivative of the log-likelihood function for a model with respect to parameters i and j,

$$\mathbf{F} = -\left\langle \frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j} \right\rangle . \tag{D.2}$$

In Chapter 5, we used the Fisher matrix to forecast the variance in a variety of cosmological parameters based on a variety of possible observational constraints on some parameters of the model. We additionally used the Fisher information to calculate how much parameter estimation can be biased by using an incorrect model for underlying data. We assumed a Gaussian likelihood, applying arguments made by Tegmark, A. Taylor, and Heavens (1997). Here, I derive expressions for the Fisher matrix given a log likelihood function and derive an expression for parameter bias in terms of the Fisher matrix.

D.1 Fisher Information Matrix

A Gaussian likelihood (pdf) is given by

$$L = \frac{1}{(2\pi)^{k/2} \det \mathbf{C}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T \mathbf{C}^{-1}(\vec{x}-\vec{\mu})\right] \tag{D.3}$$

for a random variable $\vec{x} = \langle x_1, x_2, \dots x_k \rangle$ where $\mathbf{C}$ is the covariance matrix and $\vec{\mu}$ is the vector of mean values.

The (negative) log-likelihood function for the gaussian is

$$\mathcal{L} = -\ln L(\vec{x}|\theta) = \frac{k}{2}\ln 2\pi + \frac{1}{2}\ln\det\mathbf{C} + \frac{1}{2}(\vec{x}-\vec{\mu})^T \mathbf{C}^{-1}(\vec{x}-\vec{\mu}) \tag{D.4}$$

To find the maximum likelihood estimator for θ, we take the derivative of the log likelihood with respect to θ. Also, multiply by 2 for convenience and drop the constant in front since when I get around to doing the derivatives, it of course, disappears.

$$2\mathcal{L} = \ln\det\mathbf{C} + (\vec{x}-\vec{\mu})^T \mathbf{C}^{-1}(\vec{x}-\vec{\mu}) \tag{D.5}$$

use $\ln\det\mathbf{C} = \mathrm{Tr}\ln\mathbf{C}$ to rewrite the first term. Also, we can rewrite the second term as a trace by noting that

$$\vec{v}^T \mathbf{M}\vec{v} = \mathrm{Tr}\{\mathbf{M}\vec{v}\vec{v}^T\} \tag{D.6}$$

where $\vec{v}\vec{v}^T$ is the kronecker product (outer product) of the two vectors, resulting in a matrix. So now we have

$$2\mathcal{L} = \mathrm{Tr}\{\ln \mathbf{C} + \mathbf{C}^{-1}\mathbf{D}\} \tag{D.7}$$

where $\mathbf{D} = (\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T$ is the data matrix. Now, we want to differentiate this expression with respect to the parameters θ_i and maximize to find the maximum likelihood. Looking at each term:

$$\frac{d}{d\theta_i} \ln \mathbf{C} = \mathbf{C}^{-1}\mathbf{C}_{,i} \tag{D.8}$$

$$\frac{d}{d\theta_i} \mathbf{C}^{-1}\mathbf{D} = -\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{D} + \mathbf{C}^{-1}\mathbf{D}_{,i} \tag{D.9}$$

so we have

$$2\mathcal{L}_{,i} = \mathrm{Tr}\{\mathbf{C}^{-1}\mathbf{C}_{,i} - \mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{D}_{,i}\} \tag{D.10}$$

where $_{,i}$ indicates differentiation with respect to parameter i. If we average over all $\vec{x}$ and then perform one more algebraic step, we have

$$2\langle\mathcal{L}_{,i}\rangle = \mathrm{Tr}\{\mathbf{C}^{-1}\mathbf{C}_{,i} - \mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\langle\mathbf{D}\rangle + \mathbf{C}^{-1}\langle\mathbf{D}_{,i}\rangle\} \tag{D.11}$$

$$= \mathrm{Tr}\{\mathbf{C}^{-1}\mathbf{C}_{,i}(1 - \mathbf{C}^{-1}\langle\mathbf{D}\rangle) + \mathbf{C}^{-1}\langle\mathbf{D}_{,i}\rangle\} \tag{D.12}$$

$$|\mathrm{Tr}\{\mathbf{C}^{-1}\mathbf{C}_{,i}(1 - \mathbf{C}^{-1}\langle\mathbf{D}\rangle) + \mathbf{C}^{-1}\langle\mathbf{D}_{,i}\rangle\}|_{\theta=\hat{\theta}} = 0 \tag{D.13}$$

If we differentiate equation D.10 again, and then average over $\vec{x}$ we will get an expression for the fisher information matrix

$$\begin{aligned} 2\mathcal{L}_{,ij} = \mathrm{Tr}\{&-\mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}_{,i} + \mathbf{C}^{-1}\mathbf{C}_{,ij} + \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{D} - \mathbf{C}^{-1}\mathbf{C}_{,ij}\mathbf{C}^{-1}\mathbf{D} \\ &+ \mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{D} - \mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{D}_{,j} - \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{D}_{,i} + \mathbf{C}^{-1}\mathbf{D}_{,ij}\}, \end{aligned} \tag{D.14}$$

where we can do some factoring to arrive at

$$\begin{aligned} 2\mathcal{L}_{,ij} = \mathrm{Tr}\{&-\mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}_{,i} + \mathbf{C}^{-1}\mathbf{C}_{,ij} + \mathbf{C}^{-1}(\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}_{,i} + \mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j})\mathbf{C}^{-1}\mathbf{D} \\ &- \mathbf{C}^{-1}(\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{D}_{,j} + \mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{D}_{,i}) - \mathbf{C}^{-1}\mathbf{C}_{,ij}\mathbf{C}^{-1}\mathbf{D} + \mathbf{C}^{-1}\mathbf{D}_{,ij}\}. \end{aligned} \tag{D.15}$$

If we average over $\vec{x}$, and note that $\langle\mathbf{D}\rangle = \mathbf{C}$ and $\langle\mathbf{D}_{,i}\rangle = 0$ we have

$$\begin{aligned} 2\mathcal{L}_{,ij} = \mathrm{Tr}\{&-\mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}_{,i} + \mathbf{C}^{-1}\mathbf{C}_{,ij} + \mathbf{C}^{-1}(\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}_{,i} + \mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j})\underbrace{\mathbf{C}^{-1}\mathbf{C}}_{\mathbb{1}} \\ &- \mathbf{C}^{-1}(\mathbf{C}_{,i}\mathbf{C}^{-1}0 + \mathbf{C}_{,j}\mathbf{C}^{-1}0) - \mathbf{C}^{-1}\mathbf{C}_{,ij}\underbrace{\mathbf{C}^{-1}\mathbf{C}}_{\mathbb{1}} + \mathbf{C}^{-1}\langle\mathbf{D}_{,ij}\rangle\}. \end{aligned} \tag{D.16}$$

Canceling terms we are left with

$$2\mathcal{L}_{,ij} = \mathrm{Tr}\{\mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}_{,i} + \mathbf{C}^{-1}\langle\mathbf{D}_{,ij}\rangle\}. \tag{D.17}$$

Next, we use the fact that $\mathbf{C}$ is symmetric, $\langle\mathbf{D}_{,ij}\rangle = \vec{\mu}_{,i}\vec{\mu}_{,j}^T + \vec{\mu}_{,j}\vec{\mu}_{,i}^T$ and $\mathrm{Tr}\{\mathbf{AB}\} = \mathrm{Tr}\{\mathbf{BA}\}$to write

$$2\mathcal{L}_{,ij} = 2\vec{\mu}_{,i}^T\mathbf{C}^{-1}\vec{\mu}_{,j} + \mathrm{Tr}\{\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j}\}, \tag{D.18}$$

and since the Fisher Information Matrix is $\mathbf{F}_{ij} = \mathcal{L}_{,ij}$, we know that

$$\mathbf{F}_{ij} = \vec{\mu}_{,i}^T\mathbf{C}^{-1}\vec{\mu}_{,j} + \frac{1}{2}\mathrm{Tr}\{\mathbf{C}^{-1}\mathbf{C}_{,i}\mathbf{C}^{-1}\mathbf{C}_{,j}\}. \tag{D.19}$$

D.2 Linearized Parameter Bias

Now, suppose we have a case where the properties of the distribution $\mathbf{C}(\theta), \vec{\mu}(\theta)$ are given by model A and these define the chosen likelihood function. Let us also suppose that the model does not exactly reflect the truth. Suppose that the true values for $\mathbf{C}$ and $\vec{\mu}$ come from a distribution with $\mathbf{C}^B$ and $\vec{\mu}^B$. That mean our observations (or model calculations) of the data matrix $\mathbf{D}$ will give different values than if our data were drawn or observed from model B.

If we average over $\vec{x}$ for our measurements, we find

$$\langle \mathbf{D} \rangle = \langle (\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T \rangle \tag{D.20}$$

$$= \langle \vec{x}\vec{x}^T \rangle - \langle \vec{x} \rangle \vec{\mu}^T - \mu \langle \vec{x}^T \rangle + \vec{\mu}\vec{\mu}^T \tag{D.21}$$

$$= \mathbf{C}^B + \vec{\mu}^B \vec{\mu}^{BT} - \langle \vec{x} \rangle \vec{\mu}^T - \mu \langle \vec{x}^T \rangle + \vec{\mu}\vec{\mu}^T \tag{D.22}$$

$$= \mathbf{C}^B + \vec{\mu}^B \vec{\mu}^{BT} - \vec{\mu}^B \vec{\mu}^T - \mu \vec{\mu}^{BT} + \vec{\mu}\vec{\mu}^T \tag{D.23}$$

$$= \mathbf{C}^B + (\vec{\mu}^B - \vec{\mu})(\vec{\mu}^B - \vec{\mu})^T \tag{D.24}$$

also, we have $\langle \mathbf{D}_{,i} \rangle$:

$$\langle \mathbf{D}_{,i} \rangle = \underbrace{(\langle \vec{x}\vec{x}^T \rangle)_{,i}}_{=0} - \langle \vec{x} \rangle \vec{\mu}_{,i}^T - \mu_{,i} \langle \vec{x}^T \rangle + (\vec{\mu}\vec{\mu}^T)_{,i} \tag{D.25}$$

$$= -\vec{\mu}^B \vec{\mu}_{,i}^T - \mu_{,i} \vec{\mu}^{BT} + \vec{\mu}_{,i} \vec{\mu}^T + \vec{\mu}\vec{\mu}_{,i}^T \tag{D.26}$$

$$= -2\vec{\mu}_{,i}(\vec{\mu}^B - \vec{\mu})^T \tag{D.27}$$

If we want to know how these observations B, coming from the *true* model, affect our maximum likelihood estimator, we can set $\hat{\theta} = \theta_t + \delta\theta$ and linearize equation D.13 (where θ_{t} is the true parameter value) with the $\langle \mathbf{D} \rangle$ and $\langle \mathbf{D}_{,i} \rangle$ given above containing the *true* parameter, and $\mathbf{C}$ and $\vec{\mu}$ containing $\hat{\theta}$.

Expanding relevant terms in equation D.13 one at a time and considering in some cases the results

$$\mathbf{C}^{-1}\mathbf{C}_{,i}(\hat{\theta}) \approx \mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t) + \sum_j \left[\mathbf{C}^{-1}\mathbf{C}_{,ij}(\theta_t) + \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}_{,i}\right]\delta\theta_j \tag{D.28}$$

$$1 - \mathbf{C}^{-1}\mathbf{C}^{B}(\hat{\theta}) \approx 1 - \mathbf{C}^{-1}\mathbf{C}^{B}(\theta_t) + \sum_j \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}^{B}(\theta_t)\delta\theta_j \tag{D.29}$$

$$\vec{\mu}(\hat{\theta}) \approx \vec{\mu}(\theta_t) + \sum_j \vec{\mu}_{,j}(\theta_t)\delta\theta_j \tag{D.30}$$

$$\mathbf{C}^{-1}(\hat{\theta}) \approx \mathbf{C}^{-1}(\theta_t) - \sum_j \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}(\theta_t)\delta\theta_j \tag{D.31}$$

Plugging $\langle\mathbf{D}\rangle$, $\langle\mathbf{D}_{,i}\rangle$ as well as these three expressions into equation D.13 we have this explosion:

$$\begin{gathered}
0 = \mathrm{Tr}\Bigg\{\Bigg(\mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t) + \sum_j \left[\mathbf{C}^{-1}\mathbf{C}_{,ij}(\theta_t) + \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}_{,i}\right]\delta\theta_j\Bigg) \\
\Bigg(1 - \mathbf{C}^{-1}\mathbf{C}^{B}(\theta_t) + \sum_j \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}^{B}(\theta_t)\delta\theta_j - \mathbf{C}^{-1}(\theta_t)\Big(\vec{\mu}^{B}(\theta_t) - \vec{\mu} - \sum_j \vec{\mu}_{,j}(\theta_t)\delta\theta_j\Big) \\
\Big(\vec{\mu}^{B}(\theta_t) - \vec{\mu} - \sum_j \vec{\mu}_{,j}(\theta_t)\delta\theta_j\Big)^{T} + \sum_j \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}(\theta_t)\delta\theta_j\Big(\vec{\mu}^{B}(\theta_t) - \vec{\mu} - \sum_j \vec{\mu}_{,j}(\theta_t)\delta\theta_j\Big) \\
\Big(\vec{\mu}^{B}(\theta_t) - \vec{\mu} - \sum_j \vec{\mu}_{,j}(\theta_t)\delta\theta_j\Big)^{T}\Bigg) - 2\mathbf{C}^{-1}\vec{\mu}_{,i}(\theta_t)\Big(\vec{\mu}^{B}(\theta_t) - \vec{\mu} - \sum_j \vec{\mu}_{,j}(\theta_t)\delta\theta_j\Big)^{T} \\
- 2\sum_j \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\vec{\mu}_{,i}(\theta_t)\delta\theta_j\Big(\vec{\mu}^{B}(\theta_t) - \vec{\mu} - \sum_j \vec{\mu}_{,j}(\theta_t)\delta\theta_j\Big)^{T} \\
- 2\mathbf{C}^{-1}(\theta_t)\Big(\sum_j \vec{\mu}_{,ij}(\theta_t)\delta\theta_j\Big)\Big(\vec{\mu}^{B}(\theta_t) - \vec{\mu} - \sum_j \vec{\mu}_{,j}(\theta_t)\delta\theta_j\Big)^{T} \\
- \Big(2\sum_j \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\vec{\mu}_{,i}(\theta_t)\delta\theta_j\Big)\Big(\sum_j \vec{\mu}_{,ij}(\theta_t)\delta\theta_j\Big)\Big(\vec{\mu}^{B}(\theta_t) - \vec{\mu} - \sum_j \vec{\mu}_{,j}(\theta_t)\delta\theta_j\Big)^{T}\Bigg\}
\end{gathered} \tag{D.32}$$

where I have noted (θ_t) to indicate evaluation at θ_t.

Now, if we drop every term that is $\mathcal{O}(\delta\theta^2)$ this expression reduces to

$$0 = \mathrm{Tr}\Bigg\{ \mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\Big(1 - \mathbf{C}^{-1}\mathbf{C}^B(\theta_t) + \sum_j \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}^B(\theta_t)\delta\theta_j\Big) \\ - 2\mathbf{C}^{-1}\vec{\mu}_{,i}(\theta_t)\Big(\vec{\mu}^B(\theta_t) - \vec{\mu} - \sum_j \vec{\mu}_{,j}(\theta_t)\delta\theta_j\Big)^T \Bigg\} \tag{D.33}$$

(note that $1 - \mathbf{C}^{-1}\mathbf{C}^B$ is $\mathcal{O}(\delta\theta)$). Reordering terms gives us a more suggestive form

$$0 = \mathrm{Tr}\Bigg\{ \mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\Big(1 - \mathbf{C}^{-1}\mathbf{C}^B(\theta_t)\Big) - 2\mathbf{C}^{-1}\vec{\mu}_{,i}(\theta_t)\Big(\vec{\mu}^B(\theta_t) - \vec{\mu}\Big)^T \\ + \mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\sum_j \mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}^B(\theta_t)\delta\theta_j - 2\mathbf{C}^{-1}\vec{\mu}_{,i}\sum_j \vec{\mu}^T_{,j}(\theta_t)\delta\theta_j \Bigg\}, \tag{D.34}$$

where we can reorder again, moving the $\sum$ to get

$$0 = \mathrm{Tr}\Bigg\{ \mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\Big(1 - \mathbf{C}^{-1}\mathbf{C}^B(\theta_t)\Big) - 2\mathbf{C}^{-1}\vec{\mu}_{,i}(\theta_t)\Big(\vec{\mu}^B(\theta_t) - \vec{\mu}\Big)^T \Bigg\} \\ + \sum_j \delta\theta_j \mathrm{Tr}\Bigg\{ \mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\mathbf{C}^{-1}\mathbf{C}_{,j}\mathbf{C}^{-1}\mathbf{C}^B(\theta_t) - 2\mathbf{C}^{-1}\vec{\mu}_{,i}\vec{\mu}^T_{,j}(\theta_t) \Bigg\}. \tag{D.35}$$

Next, using $\mathbf{C}^{-1}\mathbf{C}^B(\theta_t) \approx 1$ and taking the trace of the diagonal terms we have

$$0 = \mathrm{Tr}\Bigg\{ \mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\Big(1 - \mathbf{C}^{-1}\mathbf{C}^B(\theta_t)\Big) \Bigg\} - 2\mathbf{C}^{-1}\vec{\mu}_{,i}(\theta_t)\Big(\vec{\mu}^B(\theta_t) - \vec{\mu}\Big)^T \\ + \sum_j \delta\theta_j \left(\mathrm{Tr}\Bigg\{ \mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\mathbf{C}^{-1}\mathbf{C}_{,j} \Bigg\} - 2\mathbf{C}^{-1}\vec{\mu}_{,i}\vec{\mu}^T_{,j}(\theta_t) \right). \tag{D.36}$$

We can factor out a $\mathbf{C}^{-1}$ in the first trace, and divide through by 2 to arrive at

$$0 = \frac{1}{2}\mathrm{Tr}\left\{\mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\mathbf{C}^{-1}\Big(\mathbf{C} - \mathbf{C}^B(\theta_t)\Big)\right\} - \mathbf{C}^{-1}\vec{\mu}_{,i}(\theta_t)\Big(\vec{\mu}^B(\theta_t) - \vec{\mu}\Big)^T + \sum_j \delta\theta_j \left(\frac{1}{2}\mathrm{Tr}\left\{\mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\mathbf{C}^{-1}\mathbf{C}_{,j}\right\} - \mathbf{C}^{-1}\vec{\mu}_{,i}\vec{\mu}^T_{,j}(\theta_t)\right). \quad \text{(D.37)}$$

Recognizing that the last set of parenthesis contains an expression equal to equation D.19 and reordering the second term gives us

$$0 = \frac{1}{2}\mathrm{Tr}\left\{\mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\mathbf{C}^{-1}\Big(\mathbf{C} - \mathbf{C}^B(\theta_t)\Big)\right\} - \vec{\mu}_{,i}(\theta_t)\mathbf{C}^{-1}\Big(\vec{\mu}^B(\theta_t) - \vec{\mu}\Big)^T + \sum_j \delta\theta_j \mathbf{F}_{ij}, \quad \text{(D.38)}$$

where we can move terms to the other side of the equality

$$\sum_j \delta\theta_j \mathbf{F}_{ij} \; = \; \vec{\mu}_{,i}(\theta_t)\mathbf{C}^{-1}\Big(\vec{\mu}^B(\theta_t) - \vec{\mu}\Big)^T - \frac{1}{2}\mathrm{Tr}\left\{\mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\mathbf{C}^{-1}\Big(\mathbf{C} - \mathbf{C}^B(\theta_t)\Big)\right\}. \quad \text{(D.39)}$$

Finally, if we drop the Transpose that should be clear from simple vector multiplication rules, and reverse the order of the **C**s in the final term to change the sign, and multiply by the inverse Fisher matrix, leaving our parameter bias alone on the left hand side, we are finished.

$$\delta\theta_j \; = \; \sum_i \mathbf{F}^{-1}_{ij} \left[\vec{\mu}_{,i}(\theta_t)\mathbf{C}^{-1}\left(\vec{\mu}^B(\theta_t) - \vec{\mu}\right) + \frac{1}{2}\mathrm{Tr}\Big\{\mathbf{C}^{-1}\mathbf{C}_{,i}(\theta_t)\mathbf{C}^{-1}\left(\mathbf{C}^B(\theta_t) - \mathbf{C}\right)\Big\}\right] \quad \text{(D.40)}$$

D.3 Cluster Counts and Clustering

For the space density and clustering of halos, we follow conventions used in previous work (Cunha and Evrard, 2010). The mass function is

$$\frac{dn}{dM} = f(\sigma)\frac{\overline{\rho}_m}{M}\frac{d\ln\sigma^{-1}}{dM}, \tag{D.41}$$

and we adopt the Tinker parameterization of $f(\sigma)$ (Tinker et al., 2008),

$$f(\sigma) = A\left[\left(\frac{\sigma}{b}\right)^{-a} + 1\right]e^{-c/\sigma^2}. \tag{D.42}$$

Following Tinker et al., 2008 we allow three parameters of $f(\sigma)$ to vary with redshift:

$$A(z) = A_0(1+z)^{A_x}; \tag{D.43}$$

$$a(z) = a_0(1+z)^{a_x}; \tag{D.44}$$

$$b(z) = b_0(1+z)^{-\alpha}. \tag{D.45}$$

For fiducial parameters, we adopt the values of Tinker et al., 2008 at mean density contrast $\Delta = 200$: $A_0 = 0.186$, $A_x = -0.14$, $a_0 = 1.47$, $a_x = -0.06$, $b_0 = 2.57$, $\log_{10}(\alpha) = (\frac{0.75}{log(\Delta/75)})^{1.2}$, and $c = 1.19$. As Tinker et al. (2008) describe, A controls the overall amplitude of $f(\sigma)$, a controls the tilt and b sets the mass scale at which the power law in $f(\sigma)$ becomes significant.

The sample covariance of counts $N_{\alpha,i}$ is, given by (Hu and Kravtsov, 2003)

$$S_{ij}^{\alpha\beta} = \langle(N_{\alpha,i} - \overline{N}_{\alpha,i})(N_{\beta,j} - \overline{N}_{\beta,j})\rangle \tag{D.46}$$

$$= b_{\alpha,i}\overline{N}_{\alpha,i}b_{\beta,j}\overline{N}_{\beta,j}\int\frac{d^3k}{(2\pi)^3}W_i^*(\mathbf{k})W_j(\mathbf{k})\sqrt{P_i(k)P_j(k)}, \tag{D.47}$$

where $b_{\alpha,i}(z)$ is the average cluster linear bias parameter, defined as

$$b_{\alpha,i}(z) = \frac{1}{\overline{n}_{\alpha,i}(z)} \int_{M_{\rm obs}^{\alpha}}^{M_{\rm obs}^{\alpha+1}} \frac{dM_{\rm obs}}{M_{\rm obs}} \int \frac{dM}{M} \frac{d\overline{n}_{\alpha,i}(z)}{d{\rm ln}M} b(M;z)p(M_{\rm obs}|M). \tag{D.48}$$

$W_i^*(\mathbf{k})$ is the Fourier transform of the top-hat window function and $P_i(k)$ is the linear power spectrum at the centroid of redshift bin i.

We adopt the $b(M,z)$ fit of Sheth and Tormen (1999) for the halo bias

$$b(M,z) = 1 + \frac{a_c\delta_c^2/\sigma^2 - 1}{\delta_c} + \frac{2p_c}{\delta_c[1+(a\delta_c^2/\sigma^2)^{p_c}]}, \tag{D.49}$$

and choose the fiducial values for the parameters to be $a_c = 0.75$, $\delta_c = 1.69$, and $p_c = 0.3$.

Following Hu and Kravtsov (2003), we find that the window function $W_i^*(\mathbf{k})$ is given by

$$W_i(\mathbf{k}) = 2\exp\left[ik_{\parallel}\left(r_i\right)\right] \frac{\sin(k_{\parallel}\delta r_i/2)}{k_{\parallel}\delta r_i/2} \frac{J_1(k_{\perp}r_i\theta_s)}{k_{\perp}r_i\theta_s}. \tag{D.50}$$

Here $r_i = r(z_i)$ is the angular diameter distance to the $i^{\rm th}$ redshift bin, and $\delta r_i = r(z_{i+1}) - r(z_i)$. Similarly, $H_i = H(z_i) = H(z)$, which we assume to be constant inside each bin. The variables $k_{\parallel}$ and $k_{\perp}$ represent parallel and perpendicular components of the wavenumber $\mathbf{k}$ relative to the line of sight.

APPENDIX E

Clusters and Cluster Finding

Galaxy clusters are the observable counter-part to large dark matter halos, the largest gravitationally bound objects in the Universe. To extract cosmological information from galaxy cluster counts, an observed cluster population needs to be related to the underlying dark matter distribution. Cluster counting experiments measure the cluster counts in mass and redshift bins for an observable quantity X. The expected number of galaxy clusters, N, in an observable bin X_i and redshift bin z_j lying within a solid angle $\Delta\Omega_k$ can be written as a convolution

$$N_{ijk} = \frac{\Delta\Omega_k}{4\pi} \int_{z_j}^{z_{j+1}} \frac{dV}{dz}\, dz \int_{X_i}^{X_{i+1}} \int_0^{\infty} \frac{dn(M,z)}{dM} P(X|M,z)\, dM\, dX, \tag{E.1}$$

where M is the halo mass, $dn(M,z)/dM$ is the halo mass function, dV/dz is the cosmological volume element, and $P(X|M,z)$ is the probability of observing a particular X given an underlying halo mass and redshift.

An observational campaign designed to use cluster observations to address cosmology can measure N_{ijk}. Galaxy clusters can be detected in various wavebands such as optical, X-ray or microwave, and each band will be characterized by a different mass–observable relation—in addition to different astrophysics. So, the name of the game in cluster cosmology is measuring N_{ijk} and understanding $P(X|M,z)$. This dissertation is concerned almost exclusively with optical measurements.

Visible galaxies have long served as a means to identify clusters of galaxies (hence the name!), with the first systematic search for clusters being published in 1958 (Abell, 1958). Since 1958 many wide area surveys have provided large galaxy catalogs for cluster analysis, chronicles in Chapter 2, up to Sloan Digital Sky Survey (SDSS) and soon, Dark Energy Survey (DES). To extract quality cosmological constraints from a cluster survey requires not only identifying clusters, but identifying them in such a way that detected clusters are highly correlated with underlying dark matter halos and return a reliable mass estimate—characterized by small scatter in the mass–observable relation. The quality of a cluster finder can be described by its purity and completeness; purity being the probability that an identified cluster is a true cluster and completeness the probability that an identified cluster is associated with an underlying halo.

There are a variety of methods for detecting halos, discussed extensively in Hao (2009). The work in this dissertation is centered on one particular cluster finding algorithm, developed since Hao (2009), red-sequence Matched-filter Probabilistic Percolation (redMaPPer). I will give a short overview of redMaPPer, based on Section 3 of Rykoff et al. (2013) here, but for an extensive discussion, see Rykoff et al. (2013) and Rozo and Rykoff (2013).

E.1 redMaPPer

The redMaPPer cluster finder uses the λ richness estimator (Rozo, Rykoff, Koester, McKay, et al., 2009; Rykoff et al., 2012) for clusters which is approximately equal to the number of red galaxies cluster above $0.2L_*$, where L_* is the characteristic luminosity where the power-law part of the galaxy luminosity function cuts off (Longair, 2008; Schechter, 1976). The estimator is the sum over all galaxies of the probability of belonging to the cluster

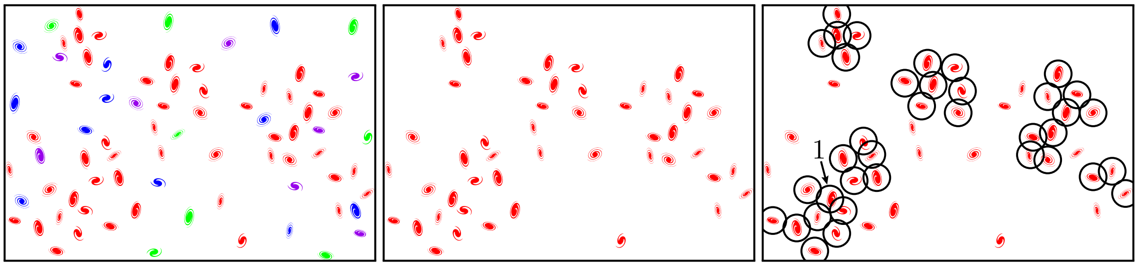
$$\lambda = \sum_{\text{galaxies}} p_{\text{membership}} \tag{E.2}$$

where calculating $p_{\text{membership}}$ is a complicated and iterative process described in Rykoff et al. (2013).

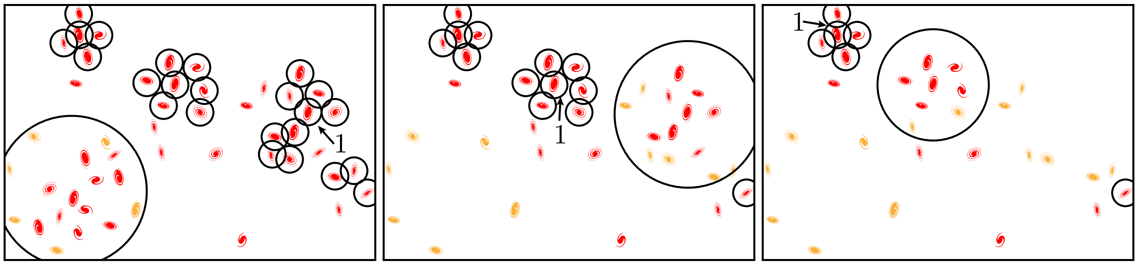
The algorithm has a training phase, where a set of training clusters, with spectroscopic central galaxy redshifts are used to calibrate a *red-sequence* model. Red-sequence galaxies are useful for cluster finding methods because all clusters contain a regular population of elliptical galaxies with a strong relationship between their color and magnitude. Since structure form hierarchically in the Universe, these cluster galaxies formed at high redshift and are now *red and dead.* With appropriate color filters, these galaxies stand out, and focusing on them is a useful method to identify clusters (Gladders and Yee, 2000). The redMaPPer algorithm calibrates a red-sequence model that can identify overdensities in red galaxies over a wide range of redshifts.

The cluster finder operates using the red-sequence model. The first step treats all galaxies in an input galaxy catalog as candidate cluster centers, and three cuts are applied. First, galaxies that deviate too far from the red-sequence or are dimmer than $0.2L_*$ are eliminated as possible cluster centers. Then, any remaining galaxy must have at least two others within $500\,h^{-1}$ kpc. These three cuts greatly reduce the number of possible central cluster galaxies, for example, reducing the SDSS Eighth Data Release catalog from $\sim$58 million galaxies to $\sim$9 million candidate centers (Rykoff et al., 2013).

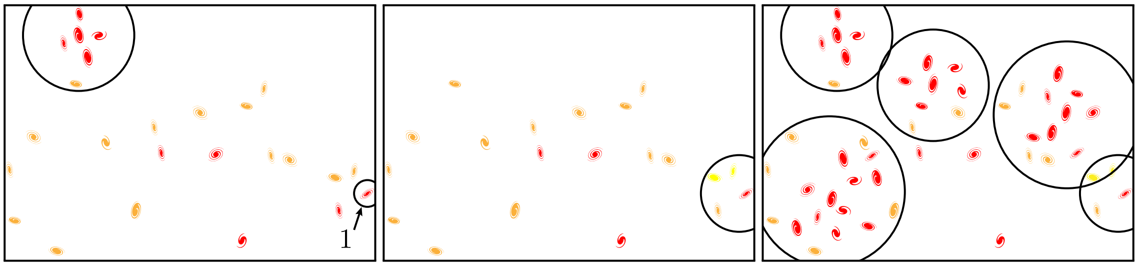
For each candidate center, all nearby galaxies are assigned a probability of belonging to a cluster centered on the candidate based on the red-sequence model. The resulting list of cluster candidates is then sorted according to likelihood, which is a sum of a term depending on probability sum of all nearby galaxies and a complicated term meant to account for the uncertainty in the cluster center. Since the galaxies are assigned a probability of membership, they can contribute fractionally to the total richness estimate for the cluster. Then, the list of galaxies is *percolated.* The



(a) Galaxy Field (b) Color/Luminosity Cut (c) Candidate Centers



(d) First Cluster (e) Second Cluster (f) Third Cluster



(g) Fourth Cluster (h) Fifth Cluster (i) All Clusters

Figure E.1: A conceptual redMaPPer percolation process. First, a set of input galaxies (E.1a) is cut based on how closely a galaxy's color follows the red-sequence model, and if it has sufficient luminosity to be a central galaxy (E.1b). Then, galaxies in overdense regions are identified as candidate cluster centers (E.1c), shown here in small *black circles*. The centers are ranked according to richness and centering likelihood, and we can imagine that the galaxy indicated by "1" here is the first ranked. Then, a cluster is built about that center (E.1d), shown in the large *black circle*, and the galaxies for that cluster are masked from further analysis. Galaxies in the cluster can not be centers of other clusters, but they can contribute to the richness of other clusters. In this case, I illustrate galaxies with a low probability of membership by leaving them in following panels, but colored *orange*, rather than *red*. In panels E.1e–E.1h, the galaxies are percolated, building a cluster catalog, in each case the highest ranked center is indicated by a "1." Note that in the fifth cluster (E.1h), two galaxies contribute to the richness that already contributed some of their membership probability to the second identified cluster (E.1e). Galaxies that contributed richness to two clusters are shown in *yellow*.

highest likelihood cluster is assigned the member galaxies associated with it. Any remaining lower ranked candidate centers that have been assigned to the first cluster are removed from the remaining list of candidate centers. Those galaxies can still contribute their membership probability to the richness for lower ranked clusters, but cannot be the center of a new cluster. Then, the process repeats based on the new *percolated* galaxy catalog.

E.2 RedMaPPer projection Model

Any measurement of λ will be the sum of richness from galaxies that are hosted by any dark matter halos in the vicinity of the target halo that are convolved with galaxies hosted by the target halo in the cluster finding process. So

$$\lambda = \sum_{\text{halos}} \lambda_i, \tag{E.3}$$

where $\lambda_{i+1} \leq \lambda_i$. Based on arguments presented in Chapter 6, the largest halo along the line of sight is the most significant source of projection, so we consider a two-halo model. As such, λ for the target will be the sum of the intrinsic richness and the richness of the next largest halo along the line of sight,

$$\lambda = \lambda_0 + \lambda_1. \tag{E.4}$$

We say a target halo suffers from projection effects if

$$\lambda_1 \geq \lambda_{\text{c}} \tag{E.5}$$

where λ_{c} is a constant. In other words, a halo suffers from projection effects if the contribution λ_1 to the total richness is larger than some critical value λ_{c}.

In section E.2.2, we derive an expression for the probability of a target halo to suffer from projection effects. As long as this probability is small, chance alignments of multiple halos will be correspondingly smaller, and therefore the distinction between λ_0 and $\tilde{\lambda}_0$ becomes irrelevant. Finally, we define $P_0(\lambda_0|M,z)$ as the probability distribution of λ_0. We assume P_0 is a Gaussian of mean $\bar{\lambda}_0(M,z)$ and standard deviation $\sigma_0(M,z) = \bar{\lambda}^{1/2}(M,z)$.

E.2.1 The λ_1 Distribution

We wish to determine the probability distribution for λ_1. The probability of a halo contributing richness λ_1 to our target halo is given by

$$P_1\left(\lambda|M,z\right) = \int dM'dz'P_{\mathrm{halo}}\left(M',z'\right)P_\rho\left(\lambda_1|M,M',z,z'\right) \tag{E.6}$$

where $P_{\mathrm{halo}}(M',z')dM'dz'$ is the probability that a halo of mass M' at redshift z' overlaps the target halo, and $P_\rho\left(\lambda_1|M,M',z,z'\right)$ is the probability that such an overlapping halo contributes λ_1 to richness λ. Assuming projection effects are rare, one has

$$P_{\mathrm{halo}}\left(M',z'\right)dM'dz' = dN \tag{E.7}$$

where dN is the number of halos of mass M' overlapping the target halo,

$$dN = \frac{dn}{dM'}\frac{dV}{dz'}\left(1+b(M)b(M')\xi(r)\right)dM'dz' \tag{E.8}$$

Here, dn/dM is the halo mass function and $b(M)$ the halo bias function. For the calculations in this work, we use the mass function from Tinker et al. (2008) and the bias function from Tinker et al. (2010). Additionally,

$$\frac{dV}{dz'}dz' = (1+z')^2\frac{D_A^2(z')}{D_A^2(z)}AcH^{-1}(z')dz' \tag{E.9}$$

is the cylindrical comoving volume at z' subtended by the area A used to estimate the richness of the target halo, r is the separation between the halos, and ξ is the linear matter correlation function. To simplify the calculations, we set $A = \pi$ Mpc^2—a 1 Mpc aperture radius—for all calculations. We ignore the transverse separation between halos when evaluating ξ, so that

$$r = cH^{-1}(z)|z - z'|. \tag{E.10}$$

It is useful to define the source function S as

$$S(M', z'|M, z) = \frac{dn}{dM'}\frac{dV}{dz'}\left(1 + b(M)b(M')\xi(r)\right) \tag{E.11}$$

so that

$$dN = S dM' dz' \tag{E.12}$$

Putting it all together, we arrive at

$$P_1\left(\lambda|M, z\right) = \int dN P_\rho\left(\lambda_1|M, M', z, z'\right) \tag{E.13}$$

$$= \int dM' dz' S(M', z'|M, z) P_\rho\left(\lambda_1|M, M', z, z'\right) \tag{E.14}$$

Equation E.14 is the fundamental result with which we can evaluate the probability distribution $P_1\left(\lambda|M, z\right)$. To do so, however, we need to know the distribution $P_\rho\left(\lambda_1|M, M', z, z'\right)$. Let then $\lambda_0(M', z')$ be the richness of halo (M', z') in the absence of projection effects. The probability distribution of $\lambda_0(M', z')$ is therefore simply $P_0\left(\lambda_0|M', z'\right)$. We assume that λ_1 scales with $\lambda_0(M')$, so that

$$\lambda_1 = \lambda_0(M')W(z, z') \tag{E.15}$$

where $W(z, z')$ is a weighting function. For the time being, we simply note that one must have $W(z,z) = 1$, and $W \to 0$ for $|z - z'| \to \infty$. Assuming the weighting function W is known, we have then

$$P_\rho\left(\lambda_1|M,M',z,z'\right) = P_0\left(\frac{\lambda_1}{W(z,z')}\middle|M',z'\right). \tag{E.16}$$

Inserting into equation E.14 we find

$$P_1\left(\lambda|M,z\right) = \int dM'dz'S(M',z'|M,z)P_0\left(\frac{\lambda_1}{W(z,z')}\middle|M',z'\right). \tag{E.17}$$

Before we move on, it is worth pausing here to consider the integration limits in the above equation. Specifically, in deriving equation E.17 we considered the line of sight contribution to the richness of a target halo. Observationally, if we were to find two overlapping halos along a line of sight, we would always consider the largest of the two halos as the target halo, and the smaller one as the projecting halo. Consequently, the the integral over halo mass for P_1 ought to extend only to masses as large as that of the target halo, so that putting the integration limits in equation E.17 explicitly, we find

$$P_1\left(\lambda|M,z\right) = \int_0^M dM' \int_0^\infty dz'S(M',z'|M,z)P_0\left(\frac{\lambda_1}{W(z,z')}\middle|M',z'\right). \tag{E.18}$$

E.2.2 Projection Effects and Completeness

Using equation E.18 requires some care. First, P_1 is *not* unity normalized:

$$1 \neq \int_0^\infty d\lambda_1 P_1\left(\lambda|M,z\right). \tag{E.19}$$

This is because P_1 is the probability that there exists a halo along the line of sight that contributes λ_1 to the richness of the target halo. That is, if we define

$$\tilde{c}(M,z) = \int_{\lambda_c}^{\infty} d\lambda_1 P_1\left(\lambda|M,z\right) \tag{E.20}$$

the $\tilde{c}$ is simply the probability that a target halo suffers from projection effects. The associated *completeness* is the probability that a halo does not suffer from projection effects,

$$c(M,z) = 1 - \tilde{c}(M,z). \tag{E.21}$$

The condition that projection effects are rare corresponds to

$$\tilde{c}(M,z) \ll 1. \tag{E.22}$$

If this condition is violated, our model is inapplicable. Indeed, this is precisely why we must introduce the parameter λ_c: the condition $\lambda_1 \geq \lambda_c$ can only be satisfied for halos relatively near to the target halo. If $\lambda_c \to 0$, any halo along the line of sight can contribute to $\tilde{c}$, which can quickly lead to the above inequality being violated.

For numerical purposes, it is worth limiting the integration volume by exploiting the properties of $\bar{\lambda}_0(M',z')$ and W. Specifically, W can only decrease as $|z-z'|$ increases, and $\bar{\lambda}_0(M',z')$ must increase with increasing M'. Consequently, we can write

$$\tilde{c}(M,z) = \int_{M_{\min}}^{M_{\max}} dM' \int_{z_{\min}}^{z_{\max}} dz' S(M',z'|M,z)\theta\left(\bar{\lambda}_0(M',z') - \frac{\lambda_c}{W(z,z')}\right) \tag{E.23}$$

where $M_{\max}$ is the target halo mass, M, and $M_{\min}(M,z)$ is given by

$$\bar{\lambda}_0(M_{\min},z) = \lambda_c, \tag{E.24}$$

and $z_{\min}(M,z)$ and $z_{\max}(M,z)$ are given by solving for z' in

$$W(z,z')\bar{\lambda}_0(M',z') = \lambda_c. \tag{E.25}$$

E.2.3 The Richness–Mass Relation

We now turn our attention towards the richness–mass relation of galaxy counts. Let then $\widetilde{P}_1$ denote the unity-normalized ($\int d\lambda_1 \widetilde{P}_1(\lambda_1) = 1$) probability distribution for λ_1. Since λ_1 is either identically zero or larger than λ_c, we have then

$$\widetilde{P}_1\left(\lambda_1|M,z\right) = c(M)\delta(\lambda_1) + P_1(\lambda_1|M,z)\theta(\lambda_1-\lambda_c) \tag{E.26}$$

where the θ term accounts for the fact that λ_1 must be greater than λ_c. Given that $\lambda = \lambda_0 + \lambda_1$ we find

$$P(\lambda|M,z) = \int d\lambda_0 d\lambda_1 P_0(\lambda_0)\widetilde{P}_1(\lambda_1)\delta(\lambda-\lambda_0-\lambda_1) \tag{E.27}$$

$$= \int d\lambda_0 P_0(\lambda_0)\left[c(M)\delta(\lambda_1) + P_1(\lambda_1|M,z)\theta(\lambda_1-\lambda_c)\right] \tag{E.28}$$

$$= c(M)P_0(\lambda) + \int d\lambda_0 P_0(\lambda_0)P_1(\lambda-\lambda_0)\theta(\lambda-\lambda_0-\lambda_c), \text{ use E.18} \tag{E.29}$$

$$= c(M)P_0(\lambda) + \int d\lambda_0 dM' dz' S(M',z'|M,z)P_0(\lambda_0)P_0'\left(\frac{\lambda-\lambda_0}{W(z,z')}\right)\theta(\lambda-\lambda_0-\lambda_c) \tag{E.30}$$

where $P_0(x)$ is shorthand for $P_0(x|M,z)$ and $P_0'(x)$ for $P_0(x|M',z')$.

To make this expression a little bit easier to examine, we can consider ignoring the variance contribution to λ due to the stochasticity of λ_1 by setting

$$P_0\left(\frac{\lambda-\lambda_0}{W(z,z')}\middle|M',z'\right) = \delta\left(\frac{\lambda-\lambda_0}{W(z,z')} - \bar{\lambda}_0(M',z')\right). \tag{E.31}$$

Doing so, we arrive at

$$P(\lambda|M,z) = c(M)P_0(\lambda|M,z) + \int dM'dz' S P_0(x)\theta(x)\theta(\bar{\lambda}_0(M',z')W(z,z') - \lambda_c) \tag{E.32}$$

where

$$x(M',z,z') = \lambda - \bar{\lambda}_0(M',z')W(z,z') \tag{E.33}$$

Note one may also write this expression as

$$P(\lambda|M,z) = c(M)P_0(\lambda) + \int dN P_0(x)\theta(x)\theta(\bar{\lambda}_0(M',z')W(z,z') - \lambda_c) \tag{E.34}$$

where dN is given by equation E.8. This illustrates the idea behind our result: there is a probability c that a halo doesnt suffer from projection effects in which case $\lambda = \lambda_0$ and λ is distributed according to $P_0(\lambda)$. On the other hand, there is a probability dN that a halo of mass M' at redshift z' projects onto the target halo, contributing $\bar{\lambda}_0'W$ to the richness, where $\bar{\lambda}_0' = \bar{\lambda}(M',z')$. The richness of the target halo must therefore be $x = \lambda - \bar{\lambda}_0'W$. The total probability for projection effects is therefore simply $dNP_0(x)$, summed over all possible projection halos, and subject to the restriction that $\bar{\lambda}_0' > \lambda_c$.

We can also use this result to derive simple expressions for the moments of λ. For instance, the mean richness is given by

$$\langle\lambda|M,z\rangle = \bar{\lambda}_0(M,z) + \int dN \bar{\lambda}_0(M',z')W(z,z')\theta(\bar{\lambda}_0(M',z')W(z,z') - \lambda_c). \tag{E.35}$$

This is exactly as we would expect: $\bar{\lambda}_0$ is the mean richness in the absence of projection effects. Projection effects simply add an additional contribution to the richness which is the mean number of projected halos dN times the mean richness of these objects, $\bar{\lambda}_0'W$.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aarseth, S. J. (1963). “Dynamical Evolution of Clusters of Galaxies, I”. In: *MNRAS* 126, p. 223. SAO–NASA ADS: `1963MNRAS.126..223A` (cit. on pp. 41, 42).

Aarseth, S. J., E. L. Turner, and J. R. Gott III (Mar. 1979). “N-body simulations of galaxy clustering. I - Initial conditions and galaxy collapse times”. In: *ApJ* 228, pp. 664–683. DOI: `10.1086/156892`. SAO–NASA ADS: `1979ApJ...228..664A` (cit. on p. 42).

Abbott, T., F. Abdalla, I. Achitouv, et al. (Dec. 2012). “First SN Discoveries from the Dark Energy Survey”. In: *The Astronomer’s Telegram* 4668, p. 1. URL: `http://www.astronomerstelegram.org/?read=4668` (cit. on p. 56).

Abdo, A. A. et al. (Apr. 2010). “The First Fermi Large Area Telescope Catalog of Gamma-ray Pulsars”. In: *ApJS* 187, pp. 460–494. DOI: `10.1088/0067-0049/187/2/460`. arXiv: `0910.1608 [astro-ph.HE]` (cit. on p. 27).

Abell, G. O. (May 1958). “The Distribution of Rich Clusters of Galaxies.” In: *ApJS* 3, p. 211. DOI: `10.1086/190036`. SAO–NASA ADS: `1958ApJS....3..211A` (cit. on pp. 15, 24, 89, 175).

Abell, G. O., H. G. Corwin Jr., and R. P. Olowin (May 1989). “A Catalog of Rich Clusters of Galaxies”. In: *ApJS* 70, pp. 1–138. DOI: `10.1086/191333`. SAO–NASA ADS: `1989ApJS...70....1A` (cit. on pp. 17, 24, 88).

Adami, C. et al. (Jan. 2010). “Galaxy Structure Searches by Photometric Redshifts in the CFHTLS”. In: *A&A* 509, A81, A81. DOI: `10.1051/0004-6361/200913067`. arXiv: `0910.3827 [astro-ph.CO]` (cit. on pp. 89, 107).

Adams, F. C., E. M. Proszkow, M. Fatuzzo, and P. C. Myers (Apr. 2006). “Early Evolution of Stellar Groups and Clusters: Environmental Effects on Forming Planetary Systems”. In: *ApJ* 641, pp. 504–525. DOI: `10.1086/500393`. eprint: `arXiv:astro-ph/0512330` (cit. on p. 147).

Ahn, C. P. et al. (Dec. 2012). “The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey”. In: *ApJS* 203, 21, p. 21. DOI: `10.1088/0067-0049/203/2/21`. arXiv: `1207.7137 [astro-ph.IM]` (cit. on p. 27).

Albrecht, A. et al. (Sept. 2006). “Report of the Dark Energy Task Force”. In: *ArXiv Astrophysics e-prints*. eprint: `arXiv:astro-ph/0609591` (cit. on p. 56).

Allen, S. W., A. E. Evrard, and A. B. Mantz (Sept. 2011). "Cosmological Parameters from Observations of Galaxy Clusters". In: *ARA&A* 49, pp. 409–470. DOI: `10.1146/annurev-astro-081710-102514`. arXiv: `1103.4829 [astro-ph.CO]` (cit. on pp. 29, 86).

Alpher, R. A., H. Bethe, and G. Gamow (Apr. 1948). "The Origin of Chemical Elements". In: *Physical Review* 73, pp. 803–804. DOI: `10.1103/PhysRev.73.803`. SAO–NASA ADS: `1948PhRv...73..803A` (cit. on p. 18).

Angulo, R. E., V. Springel, S. D. M. White, A. Jenkins, C. M. Baugh, and C. S. Frenk (Nov. 2012). "Scaling Relations for galaxy clusters in the Millennium-XXL simulation". In: *MNRAS* 426, pp. 2046–2062. DOI: `10.1111/j.1365-2966.2012.21830.x`. arXiv: `1203.3216 [astro-ph.CO]` (cit. on pp. 42, 45).

Annis, J., S. Bridle, et al. (Oct. 2005). "Constraining Dark Energy with the Dark Energy Survey: Theoretical Challenges". In: *ArXiv Astrophysics e-prints*. arXiv: `astro-ph/0510195` (cit. on pp. 28, 56).

Annis, J., F. J. Castander, et al. (Oct. 2005). "Dark Energy Studies: Challenges to Computational Cosmology". In: *ArXiv Astrophysics e-prints*. eprint: `arXiv:astro-ph/0510194` (cit. on p. 56).

Arnaud, M., G. W. Pratt, R. Piffaretti, H. Böhringer, J. H. Croston, and E. Pointecouteau (July 2010). "The universal galaxy cluster pressure profile from a representative sample of nearby systems (REXCESS) and the Y_{SZ}–M_{500} relation". In: *A&A* 517, A92, A92. DOI: `10.1051/0004-6361/200913416`. arXiv: `0910.1234 [astro-ph.CO]` (cit. on pp. 49, 82).

Bahcall, N. A. and X. Fan (Sept. 1998). "The Most Massive Distant Clusters: Determining Ω and σ_8". In: *ApJ* 504, p. 1. DOI: `10.1086/306088`. arXiv: `astro-ph/9803277` (cit. on p. 86).

Barkhouse, W. A. et al. (July 2006). "ChaMP Serendipitous Galaxy Cluster Survey". In: *ApJ* 645, pp. 955–976. DOI: `10.1086/504457`. eprint: `arXiv:astro-ph/0603521` (cit. on p. 79).

Barnes, D. G. et al. (Apr. 2001). "The Hi Parkes All Sky Survey: southern observations, calibration and robust imaging". In: *MNRAS* 322, pp. 486–498. DOI: `10.1046/j.1365-8711.2001.04102.x`. SAO–NASA ADS: `2001MNRAS.322..486B` (cit. on p. 27).

Barnes, J. and P. Hut (Dec. 1986). "A hierarchical $\mathcal{O}(N \log N)$ force-calculation algorithm". In: *Nature* 324, pp. 446–449. DOI: `10.1038/324446a0` (cit. on pp. 41, 43, 152).

Becker, A. B. (2012). "Is the Universe Normal? Constraining Scale-Dependent Primordial Non-Gaussianity". PhD thesis. University of Michigan, Ann Arbor. URL: `http://hdl.handle.net/2027.42/93893` (cit. on p. 45).

Becker, M. R. (Oct. 2012). "CALCLENS: Weak Lensing Simulations for Large-area Sky Surveys and Second-order Effects in Cosmic Shear Power Spectra". In: *ArXiv e-prints*. arXiv: `1210.3069 [astro-ph.CO]` (cit. on pp. 49, 163).

Behroozi, P. S., C. Conroy, and R. H. Wechsler (July 2010). "A Comprehensive Analysis of Uncertainties Affecting the Stellar Mass-Halo Mass Relation for $0 < z < 4$". In: *ApJ* 717, pp. 379–403. DOI: `10.1088/0004-637X/717/1/379`. arXiv: `1001.0015 [astro-ph.CO]` (cit. on p. 161).

Behroozi, P. S., R. H. Wechsler, and H.-Y. Wu (Jan. 2013). "The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores". In: *ApJ* 762, 109, p. 109. DOI: `10.1088/0004-637X/762/2/109`. arXiv: `1110.4372 [astro-ph.CO]` (cit. on pp. 47, 158, 159).

Behroozi, P. S., R. H. Wechsler, H.-Y. Wu, M. T. Busha, A. A. Klypin, and J. R. Primack (Jan. 2013). "Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology". In: *ApJ* 763, 18, p. 18. DOI: `10.1088/0004-637X/763/1/18`. arXiv: `1110.4370 [astro-ph.CO]` (cit. on pp. 47, 159).

Belhajjame, K., K. Wolstencroft, O. Corcho, T. Oinn, F. Tanoh, A. William, and C. Goble (May 2008). "Metadata Management in the Taverna Workflow System". In: *Cluster Computing and the Grid, 2008. CCGRID '08. 8th IEEE International Symposium on.* Lyon, France, pp. 651–656. ISBN: 978-0-7695-3156-4. DOI: `10.1109/CCGRID.2008.17` (cit. on p. 59).

Bell, A. E. (Aug. 1941). "The 'Horologium Oscillatorium' of Christian Huygens". In: *Nature* 148, pp. 245–248. DOI: `10.1038/148245a0` (cit. on p. 8).

Benoist, C. et al. (2013). "Wavelet Adapted Z Photometric (WAZP) cluster finder". In: *In Prep.* (Cit. on p. 79).

Berman, Fran (Nov. 2001). "Viewpoint: From TeraGrid to knowledge grid". In: *Commun. ACM* 44.11, pp. 27–28. ISSN: 0001-0782. DOI: `10.1145/384150.384156` (cit. on p. 53).

Bertschinger, E. (1998). "Simulations of Structure Formation in the Universe". In: *ARA&A* 36, pp. 599–654. DOI: `10.1146/annurev.astro.36.1.599`. SAO–NASA ADS: `1998ARA%26A..36..599B` (cit. on pp. 39–41, 147).

Biesiadzinski, T. et al. (2013). "Synthetic Sunyaev-Zel'dovich maps". In: *In Prep.* (Cit. on p. 81).

Biesiadzinski, T., J. J. McMahon, C. J. Miller, B. Nord, and L. Shaw (Sept. 2012). "Impact of Systematics on SZ-Optical Scaling Relations". In: *ApJ* 757, 1, p. 1. DOI: `10.1088/0004-637X/757/1/1`. arXiv: `1201.1282 [astro-ph.CO]` (cit. on p. 82).

Birkhoff, G. D. and R. E. Langer (1923). *Relativity and Modern Physics.* Cambridge MA, USA: Harvard University Press. Google Books: `SmAmAAAAMAAJ` (cit. on pp. 38, 140).

Bleem, L. E. et al. (July 2012). “A Measurement of the Correlation of Galaxy Surveys with CMB Lensing Convergence Maps from the South Pole Telescope”. In: *ApJ* 753, L9, p. L9. DOI: `10.1088/2041-8205/753/1/L9`. arXiv: `1203.4808 [astro-ph.CO]` (cit. on p. 82).

Bode, P., N. A. Bahcall, E. B. Ford, and J. P. Ostriker (Apr. 2001). “Evolution of the Cluster Mass Function: GPC^3 Dark Matter Simulations”. In: *ApJ* 551, pp. 15–22. DOI: `10.1086/320077`. arXiv: `astro-ph/0011376` (cit. on p. 42).

Borgani, S. and A. V. Kravtsov (June 2009). “Cosmological simulations of galaxy clusters”. In: *ArXiv e-prints*. arXiv: `0906.4370 [astro-ph.CO]` (cit. on p. 39).

Boué, G., F. Durret, C. Adami, G. A. Mamon, O. Ilbert, and V. Cayatte (Oct. 2008). “An optical view of the filament region of Abell 85”. In: *A&A* 489, pp. 11–22. DOI: `10.1051/0004-6361:20078972`. arXiv: `0806.4262` (cit. on p. 88).

Boulade, O. et al. (Mar. 2003). “MegaCam: the new Canada-France-Hawaii Telescope wide-field imaging camera”. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Ed. by M. Iye and A. F. M. Moorwood. Vol. 4841. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pp. 72–81. DOI: `10.1117/12.459890` (cit. on p. 89).

Bower, R. G., J. R. Lucey, and R. S. Ellis (Feb. 1992). “Precision photometry of early-type galaxies in the Coma and Virgo Clusters: a test of the universality of the colour–magnitude relation - II. Analysis”. In: *MNRAS* 254, p. 601. SAO–NASA ADS: `1992MNRAS.254..601B` (cit. on p. 90).

Busha, M. T. and R. H. Wechsler (2013). “ADDGALS”. In: *In Prep.* (Cit. on p. 48).

Cai, Y.-C., R. E. Angulo, C. M. Baugh, S. Cole, C. S. Frenk, and A. Jenkins (May 2009). “Mock galaxy redshift catalogues from simulations: implications for Pan-STARRS1”. In: *MNRAS* 395, pp. 1185–1203. DOI: `10.1111/j.1365-2966.2009.14402.x`. arXiv: `0810.2300` (cit. on p. 88).

Cappelluti, N. et al. (Mar. 2009). “XMM-Newton wide-field survey in COSMOS field (Cappelluti+, 2009)”. In: *VizieR Online Data Catalog* 349, p. 70635. SAO–NASA ADS: `2009yCat..34970635C` (cit. on p. 27).

Carlberg, R. G. and H. M. P. Couchman (May 1989). “Mergers and bias in a cold dark matter cosmology”. In: *ApJ* 340, pp. 47–68. DOI: `10.1086/167375`. SAO–NASA ADS: `1989ApJ...340...47C` (cit. on p. 42).

Carlstrom, J. E. et al. (May 2011). “The 10 Meter South Pole Telescope”. In: *PASP* 123, pp. 568–581. DOI: `10.1086/659879`. arXiv: `0907.4445 [astro-ph.IM]` (cit. on pp. 56, 108).

Carroll, S. M., W. H. Press, and E. L. Turner (1992). “The Cosmological Constant”. In: *ARA&A* 30, pp. 499–542. DOI: `10.1146/annurev.aa.30.090192.002435`. SAO–NASA ADS: `1992ARA%26A..30..499C` (cit. on p. 24).

Cellarius, A. (1661). *Harmonia Macrocosmica, sev Atlas Universalis et Novus, Totius Universi Creati Cosmographiam Generalem, et Novam Exhibens.* Amsterdam: Joannem Janssonium. URL: `http://www.rarebookroom.org/Control/gelmcs/index.html` (visited on Feb. 13, 2013) (cit. on pp. 7, 9).

Challener, S. (Dec. 1999). "Stonehenge: A Simple and Accurate Predictor of Lunar Eclipses". In: *American Astronomical Society Meeting Abstracts*. Vol. 31. Bulletin of the American Astronomical Society, p. 1511 (cit. on p. 4).

Challinor, A. and A. Lewis (May 2005). "Lensed CMB power spectra from all-sky correlation functions". In: *Phys. Rev. D* 71.10, 103010. DOI: `10.1103/PhysRevD.71.103010`. arXiv: `astro-ph/0502425` (cit. on p. 47).

Chen, A. (Aug. 2013). "The Projected Halo Model". PhD thesis. University of Michigan, Ann Arbor. (cit. on p. 118).

Claussius, R. J. E. (Aug. 1870). "On a Mechanical Theorem Applicable to Heat". In: *Phil. Mag.* 4th ser. 40, pp. 122–127. HathiTrust: `mdp.39015024087796;seq=622` (cit. on p. 14).

Clowe, D., M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky (Sept. 2006). "A Direct Empirical Proof of the Existence of Dark Matter". In: *ApJ* 648, pp. L109–L113. DOI: `10.1086/508162`. arXiv: `astro-ph/0608407` (cit. on pp. 22, 23).

Clowe, D., A. Gonzalez, and M. Markevitch (Apr. 2004). "Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter". In: *ApJ* 604, pp. 596–603. DOI: `10.1086/381970`. arXiv: `astro-ph/0312273` (cit. on p. 23).

Coc, A., E. Vangioni-Flam, P. Descouvemont, A. Adahchour, and C. Angulo (Jan. 2004). "Updated Big Bang Nucleosynthesis Compared with Wilkinson Microwave Anisotropy Probe Observations and the Abundance of Light Elements". In: *ApJ* 600, pp. 544–552. DOI: `10.1086/380121`. arXiv: `astro-ph/0309480` (cit. on p. 18).

Cohn, J. D., A. E. Evrard, M. White, D. Croton, and E. Ellingson (Dec. 2007). "Red-sequence cluster finding in the Millennium Simulation". In: *MNRAS* 382, pp. 1738–1750. DOI: `10.1111/j.1365-2966.2007.12479.x`. arXiv: `0706.0211` (cit. on pp. 87, 88, 90, 91, 94–96, 100, 101, 103, 109, 111, 125).

Cohn, J. D. and M. White (Feb. 2009). "Issues in joint SZ and optical cluster finding". In: *MNRAS* 393, pp. 393–405. DOI: `10.1111/j.1365-2966.2008.14215.x`. arXiv: `0809.0308` (cit. on p. 108).

Colberg, J. M. et al. (Nov. 2000). "Clustering of galaxy clusters in cold dark matter universes". In: *MNRAS* 319, pp. 209–214. DOI: `10.1046/j.1365-8711.2000.03832.x`. arXiv: `astro-ph/0005259` (cit. on p. 42).

Cole, S. et al. (Sept. 2005). "The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications". In: *MNRAS* 362, pp. 505–

534. DOI: `10.1111/j.1365-2966.2005.09318.x`. arXiv: `astro-ph/0501174` (cit. on p. 25).

Collister, A. A. and O. Lahav (Apr. 2004). “ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks”. In: *PASP* 116, pp. 345–351. DOI: `10.1086/383254`. arXiv: `astro-ph/0311058` (cit. on p. 49).

Conroy, C., R. H. Wechsler, and A. V. Kravtsov (Aug. 2006). “Modeling Luminosity-dependent Galaxy Clustering through Cosmic Time”. In: *ApJ* 647, pp. 201–214. DOI: `10.1086/503602`. arXiv: `astro-ph/0512234` (cit. on p. 161).

Copernicus, N. (1543). *De Revolutionibus Orbium Cœlestium.* Norimbergæ, Imperium Romanum Sacrum: J. Petreius. SAO–NASA ADS: `1543droc.book.....C` (cit. on p. 8).

Crocce, M., P. Fosalba, F. J. Castander, and E. Gaztañaga (Apr. 2010). “Simulating the Universe with MICE: the abundance of massive clusters”. In: *MNRAS* 403, pp. 1353–1367. DOI: `10.1111/j.1365-2966.2009.16194.x`. arXiv: `0907.0019 [astro-ph.CO]` (cit. on p. 87).

Crocce, M., S. Pueblas, and R. Scoccimarro (Nov. 2006). “Transients from initial conditions in cosmological simulations”. In: *MNRAS* 373, pp. 369–381. DOI: `10.1111/j.1365-2966.2006.11040.x`. eprint: `arXiv:astro-ph/0606505` (cit. on pp. 47, 58, 149).

Croton, D. J., V. Springel, S. D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. F. Navarro, and N. Yoshida (Jan. 2006). “The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies”. In: *MNRAS* 365, pp. 11–28. DOI: `10.1111/j.1365-2966.2005.09675.x`. arXiv: `astro-ph/0508046` (cit. on p. 91).

Cunha, C. E. (Mar. 2009). “Cross-calibration of cluster mass observables”. In: *Phys. Rev. D* 79.6, 063009. DOI: `10.1103/PhysRevD.79.063009`. arXiv: `0812.0583` (cit. on pp. 29, 81, 91, 95).

Cunha, C. E. and A. E. Evrard (Apr. 2010). “Sensitivity of galaxy cluster dark energy constraints to halo modeling uncertainties”. In: *Phys. Rev. D* 81.8, 083509. DOI: `10.1103/PhysRevD.81.083509`. arXiv: `0908.0526 [astro-ph.CO]` (cit. on pp. 92, 158, 172).

Cunha, C. E., D. Huterer, M. T. Busha, and R. H. Wechsler (June 2012). “Sample variance in photometric redshift calibration: cosmological biases and survey requirements”. In: *MNRAS* 423, pp. 909–924. DOI: `10.1111/j.1365-2966.2012.20927.x`. arXiv: `1109.5691 [astro-ph.CO]` (cit. on p. 57).

Cunha, C. E., D. Huterer, and J. A. Frieman (Sept. 2009). “Constraining dark energy with clusters: Complementarity with other probes”. In: *Phys. Rev. D* 80.6, 063532. DOI: `10.1103/PhysRevD.80.063532`. arXiv: `0904.1589 [astro-ph.CO]` (cit. on pp. 29, 86).

Cunha, C. E., M. Lima, H. Oyaizu, J. Frieman, and H. Lin (July 2009). “Estimating the redshift distribution of photometric galaxy samples - II. Applications and tests of a new method”. In: *MNRAS* 396, pp. 2379–2398. DOI: `10.1111/j.1365-2966.2009.14908.x`. arXiv: `0810.2991` (cit. on p. 49).

Curcin, V. and M. Ghanem (Dec. 2008). “Scientific workflow systems - can one size fit all?” In: *Biomedical Engineering Conference, 2008. CIBEC 2008. Cairo International.* Cairo, Egypt, pp. 1–9. ISBN: 978-1-4244-2694-2. DOI: `10.1109/CIBEC.2008.4786077` (cit. on p. 59).

Dalton, G. B., S. J. Maddox, W. J. Sutherland, and G. Efstathiou (Aug. 1997). “The APM Galaxy Survey - V. Catalogues of galaxy clusters”. In: *MNRAS* 289, pp. 263–284. SAO–NASA ADS: `1997MNRAS.289..263D` (cit. on p. 25).

Davis, M., G. Efstathiou, C. S. Frenk, and S. D. M. White (May 1985). “The evolution of large-scale structure in a universe dominated by cold dark matter”. In: *ApJ* 292, pp. 371–394. DOI: `10.1086/163168`. SAO–NASA ADS: `1985ApJ...292..371D` (cit. on pp. 24, 42).

de Lapparent, V., M. J. Kurtz, and M. J. Geller (May 1986). “The Shane-Wirtanen counts: Systematics and Two-Point Correlation Function”. In: *ApJ* 304, pp. 585–598. DOI: `10.1086/164194`. SAO–NASA ADS: `1986ApJ...304..585D` (cit. on p. 25).

de Sitter, W. (Nov. 1917). “On Einstein’s Theory of Gravitation and its Astronomical Consequences. Third Paper”. In: *MNRAS* 78, pp. 3–28. SAO–NASA ADS: `1917MNRAS..78....3D` (cit. on p. 18).

Dekel, A., G. R. Blumenthal, J. R. Primack, and S. Olivier (Mar. 1989). “Projection Contamination in Cluster Catalogs: Are the Abell Redshift Sample Correlations Significant?” In: *ApJ* 338, pp. L5–L8. DOI: `10.1086/185387`. SAO–NASA ADS: `1989ApJ...338L...5D` (cit. on p. 25).

DES Simulations Working Group (2013). “Dark Energy Survey Synthetic Catalogs”. In: *In Prep.* (Cit. on p. 122).

Descartes, R. (1677). *Le Monde, ou Traite de la lumiere (French) [The World, or Treatise on the Light].* Paris, FR: M. Bobin et N. Le Gras. URL: `http://www.princeton.edu/~hos/mike/texts/descartes/world/worldfr.htm` (visited on Feb. 13, 2013) (cit. on p. 10).

Dicke, R. H. and P. J. E. Peebles (1979). “The big bang cosmology - enigmas and nostrums.” In: *General Relativity: An Einstein Centenary Survey.* Ed. by S. W. Hawking and W. Israel, pp. 504–517. Google Books: `pxA4AAAAIAAJ` (cit. on p. 23).

Dodelson, S., E. I. Gates, and M. S. Turner (Oct. 1996). “Cold Dark Matter”. In: *Science* 274, pp. 69–75. DOI: `10.1126/science.274.5284.69`. arXiv: `astro-ph/9603081` (cit. on p. 24).

Dolag, K., M. Bartelmann, F. Perrotta, C. Baccigalupi, L. Moscardini, M. Meneghetti, and G. Tormen (Mar. 2004). “Numerical study of halo concentrations in dark-energy cosmologies”. In: *A&A* 416, pp. 853–864. DOI: `10.1051/0004-6361:20031757`. arXiv: `astro-ph/0309771` (cit. on p. 149).

Dolag, K., S. Borgani, G. Murante, and V. Springel (Oct. 2009). “Substructures in hydrodynamical cluster simulations”. In: *MNRAS* 399, pp. 497–514. DOI: `10.1111/j.1365-2966.2009.15034.x`. arXiv: `0808.3401` (cit. on p. 149).

Dolag, K., S. Borgani, S. Schindler, A. Diaferio, and A. M. Bykov (Feb. 2008). “Simulation Techniques for Cosmological Simulations”. In: *Space Sci. Rev.* 134, pp. 229–268. DOI: `10.1007/s11214-008-9316-5`. arXiv: `0801.1023` (cit. on p. 39).

Dong, F., E. Pierpaoli, J. E. Gunn, and R. H. Wechsler (Apr. 2008). “Optical Cluster Finding with an Adaptive Matched-Filter Technique: Algorithm and Comparison with Simulations”. In: *ApJ* 676, pp. 868–879. DOI: `10.1086/522490`. arXiv: `0709.0759` (cit. on p. 107).

D’Onghia, E., M. Vogelsberger, and L. Hernquist (Apr. 2012). “Self-Perpetuating Spiral Arms in Disk Galaxies”. In: *ArXiv e-prints*. arXiv: `1204.0513 [astro-ph.GA]` (cit. on p. 147).

Doroshkevich, A. G. and I. D. Novikov (Aug. 1964). “Mean Density of Radiation in the Metagalaxy and Certain Problems in Relativistic Cosmology”. In: *Soviet Physics Doklady* 9, p. 111 (cit. on p. 18).

Dreyer, J. L. E. (1888). “A New General Catalogue of Nebulae and Clusters of Stars, being the Catalogue of the late Sir John F. W. Herschel, Bart., revised, corrected, and enlarged.” In: *MmRAS* 49, pp. 1–237. HathiTrust: `mdp.39015077097916;seq=11` (cit. on p. 10).

— (1895). “Index Catalogue of Nebulae found in the years 1888 to 1894, with Notes and Corrections to the New General Catalogue”. In: *MmRAS* 51, pp. 185–228. HathiTrust: `mdp.39015006985512;seq=421` (cit. on p. 10).

— (1908). “Index Catalogue of Nebulae and Cluster Stars”. In: *MmRAS* 59, pp. 105–198. HathiTrust: `mdp.39015006985355;seq=121` (cit. on p. 10).

Eddington, A. S. (Apr. 1916). “The kinetic energy of a star-clusters”. In: *MNRAS* 76, pp. 525–528. SAO–NASA ADS: `1916MNRAS..76..525E` (cit. on p. 14).

Efstathiou, G., M. Davis, S. D. M. White, and C. S. Frenk (Feb. 1985). “Numerical techniques for large cosmological N-body simulations”. In: *ApJS* 57, pp. 241–260. DOI: `10.1086/191003`. SAO–NASA ADS: `1985ApJS...57..241E` (cit. on pp. 41, 43, 149).

Efstathiou, G. and J. W. Eastwood (Feb. 1981). “On the clustering of particles in an expanding universe”. In: *MNRAS* 194, pp. 503–525. SAO–NASA ADS: `1981MNRAS.194..503E` (cit. on pp. 41, 42).

Efstathiou, G., W. J. Sutherland, and S. J. Maddox (Dec. 1990). “The cosmological constant and cold dark matter”. In: *Nature* 348, pp. 705–707. DOI: `10.1038/348705a0` (cit. on p. 24).

Eisenstein, D. J. and W. Hu (Mar. 1998). “Baryonic Features in the Matter Transfer Function”. In: *ApJ* 496, p. 605. DOI: `10.1086/305424`. arXiv: `astro-ph/9709112` (cit. on p. 148).

Erickson, B. M. S., C. E. Cunha, and A. E. Evrard (Nov. 2011). “Influence of projection in cluster cosmology studies”. In: *Phys. Rev. D* 84.10, 103506, p. 103506. DOI: `10.1103/PhysRevD.84.103506`. arXiv: `1106.3067 [astro-ph.CO]` (cit. on pp. 3, 57).

Erickson, B. M. S., E. S. Rykoff, and E. Rozo (2013). “Projection effects in optical cluster surveys”. In: *In Prep.* (Cit. on p. 3).

Erickson, B. M. S., R. Singh, A. E. Evrard, M. R. Becker, M. T. Busha, A. V. Kravtsov, S. Marru, M. Pierce, and R. H. Wechsler (2012). “A high throughput workflow environment for cosmological simulations”. In: *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond*. XSEDE ’12. Chicago, Illinois: ACM, 34:1–34:8. ISBN: 978-1-4503-1602-6. DOI: `10.1145/2335755.2335830`. arXiv: `1210.3317 [astro-ph.CO]` (cit. on p. 3).

— (2013). “Enabling DES science with XSEDE resources”. In: *Proceedings of the 2nd Conference of the Extreme Science and Engineering Discovery Environment: Gateway to Discovery*. XSEDE ’13. San Diego, CA: ACM (cit. on pp. 3, 74).

Evrard, A. E. (Dec. 1986). “Dark Matter and the Dynamics of Galaxy Clustering”. PhD thesis. State University of New York at Stony Brook. URL: `http://search.proquest.com/docview/303547936` (visited on Apr. 4, 2013) (cit. on p. 41).

— (Dec. 1988). “Beyond N-body: 3D cosmological gas dynamics”. In: *MNRAS* 235, pp. 911–934. SAO–NASA ADS: `1988MNRAS.235..911E` (cit. on p. 54).

— (Nov. 1990). “Formation and Evolution of X-ray Clusters: A Hydrodynamic Simulation of the Intracluster Medium”. In: *ApJ* 363, pp. 349–366. DOI: `10.1086/169350`. SAO–NASA ADS: `1990ApJ...363..349E` (cit. on p. 149).

— (Dec. 1997). “The intracluster gas fraction in X-ray clusters - Constraints on the clustered mass density”. In: *MNRAS* 292, p. 289. SAO–NASA ADS: `1997MNRAS.292..289E` (cit. on p. 24).

Evrard, A. E., J. J. Mohr, D. G. Fabricant, and M. J. Geller (Dec. 1993). “A Morphology-Cosmology Connection for X-Ray Clusters”. In: *ApJ* 419, p. L9. DOI: `10.1086/187124`. arXiv: `astro-ph/9310002` (cit. on p. 24).

Evrard, A. E. and A. Yahil (Sept. 1985a). “Virialization in N-body Models of the Expanding Universe. I. Isolated Pairs”. In: *ApJ* 296, pp. 299–318. DOI: `10.1086/163449`. SAO–NASA ADS: `1985ApJ...296..299E` (cit. on p. 41).

Evrard, A. E. and A. Yahil (Sept. 1985b). “Virialization in N-Body Models of the Expanding Universe. II. All Pairs”. In: *ApJ* 296, p. 310. DOI: `10.1086/163450`. SAO–NASA ADS: `1985ApJ...296..310E` (cit. on p. 41).

Evrard, A. E. et al. (July 2002). “Galaxy Clusters in Hubble Volume Simulations: Cosmological Constraints from Sky Survey Populations”. In: *ApJ* 573, pp. 7–36. DOI: `10.1086/340551`. arXiv: `astro-ph/0110246` (cit. on pp. 58, 86, 87).

Flaugher, B. L. et al. (Sept. 2012). “Status of the Dark Energy Survey Camera (DECam) project”. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 8446. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. DOI: `10.1117/12.926216` (cit. on p. 28).

Friedmann, A. (1922). “Über die Krümmung des Raumes (German) [On the Curvature of Space]”. In: *Zeitschrift fur Physik* 10, pp. 377–386. DOI: `10.1007/BF01332580` (cit. on p. 18).

— (Dec. 1924). “Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes (German) [On the Possibility of a World with Constant Negative Curvature]”. In: *Zeitschrift fur Physik* 21, pp. 326–332. DOI: `10.1007/BF01328280` (cit. on p. 18).

Frieman, J. A., M. S. Turner, and D. Huterer (Sept. 2008). “Dark Energy and the Accelerating Universe”. In: *ARA&A* 46, pp. 385–432. DOI: `10.1146/annurev.astro.46.060407.145243`. arXiv: `0803.0982` (cit. on p. 30).

Gal, R. R., B. C. Lemaux, L. M. Lubin, D. Kocevski, and G. K. Squires (Sept. 2008). “The Complex Structure of the Cl 1604 Supercluster at $z \sim 0.9$”. In: *ApJ* 684, pp. 933–956. DOI: `10.1086/590416`. arXiv: `0803.3842` (cit. on p. 88).

Gelb, J. M. and E. Bertschinger (Dec. 1994). “Cold dark matter. 1: The formation of dark halos”. In: *ApJ* 436, pp. 467–490. DOI: `10.1086/174922`. arXiv: `astro-ph/9408028` (cit. on p. 42).

Geller, M. J. and J. P. Huchra (Nov. 1989). “Mapping the Universe”. In: *Science* 246, pp. 897–903. DOI: `10.1126/science.246.4932.897` (cit. on p. 25).

Gerdes, D. W., A. J. Sypniewski, T. A. McKay, J. Hao, M. R. Weis, R. H. Wechsler, and M. T. Busha (June 2010). “ArborZ: Photometric Redshifts Using Boosted Decision Trees”. In: *ApJ* 715, pp. 823–832. DOI: `10.1088/0004-637X/715/2/823`. arXiv: `0908.4085 [astro-ph.CO]` (cit. on p. 49).

Gerke, B. F. et al. (May 2005). “The DEEP2 Galaxy Redshift Survey: First Results on Galaxy Groups”. In: *ApJ* 625, pp. 6–22. DOI: `10.1086/429579`. arXiv: `astro-ph/0410721` (cit. on p. 108).

Gladders, M. D. and H. K. C. Yee (Oct. 2000). “A New Method For Galaxy Cluster Detection. I. The Algorithm”. In: *AJ* 120, pp. 2148–2162. DOI: `10.1086/301557`. arXiv: `astro-ph/0004092` (cit. on pp. 90, 176).

Gladders, M. D. and H. K. C. Yee (Mar. 2005). “The Red-Sequence Cluster Survey. I. The Survey and Cluster Catalogs for Patches RCS 0926+37 and RCS 1327+29”. In: *ApJS* 157, pp. 1–29. DOI: `10.1086/427327`. arXiv: `astro-ph/0411075` (cit. on p. 89).

Goecks, Jeremy, Anton Nekrutenko, and James Taylor (2010). “Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences”. In: *Genome Biology* 8.11, R86. DOI: `doi:10.1186/gb-2010-11-8-r86` (cit. on p. 59).

Governato, F., A. Babul, T. Quinn, P. Tozzi, C. M. Baugh, N. Katz, and G. Lake (Aug. 1999). “Properties of galaxy clusters: mass and correlation functions”. In: *MNRAS* 307, pp. 949–966. DOI: `10.1046/j.1365-8711.1999.02706.x`. arXiv: `astro-ph/9810189` (cit. on p. 42).

Groth, E. J. and P. J. E. Peebles (Oct. 1977). “Statistical Analysis of Catalogs of Extragalactic Objects. VII. Two- and Three-Point Correlation Functions for the High-Resolution Shane-Wirtanen Catalog of Galaxies”. In: *ApJ* 217, pp. 385–405. DOI: `10.1086/155588`. SAO–NASA ADS: `1977ApJ...217..385G` (cit. on p. 25).

— (Nov. 1986a). “The Shane-Wirtanen Counts: Observer and Time-Dependent Effects”. In: *ApJ* 310, pp. 507–517. DOI: `10.1086/164704`. SAO–NASA ADS: `1986ApJ...310..507G` (cit. on p. 25).

— (Nov. 1986b). “The Shane-Wirtanen Counts: Plate Correction Factors and Correlation Functions”. In: *ApJ* 310, pp. 499–506. DOI: `10.1086/164703`. SAO–NASA ADS: `1986ApJ...310..499G` (cit. on p. 25).

Grove, L. F., L. da Costa, and C. Benoist (Nov. 2008). “Multi-object spectroscopy of low-redshift EIS clusters. IV. Reliability of matched-filter results at $z \sim 0.3 - 0.4$”. In: *A&A* 490, pp. 945–963. DOI: `10.1051/0004-6361:20066830`. arXiv: `0809.4552` (cit. on pp. 88, 107).

Gunn, J. E. et al. (Dec. 1998). “The Sloan Digital Sky Survey Photometric Camera”. In: *AJ* 116, pp. 3040–3081. DOI: `10.1086/300645`. arXiv: `astro-ph/9809085` (cit. on p. 89).

Guth, A. H. (Jan. 1981). “Inflationary universe: A possible solution to the horizon and flatness problems”. In: *Phys. Rev. D* 23, pp. 347–356. DOI: `10.1103/PhysRevD.23.347` (cit. on p. 23).

Guth, A. H. and S.-Y. Pi (Oct. 1982). “Fluctuations in the new inflationary universe”. In: *Physical Review Letters* 49, pp. 1110–1113. DOI: `10.1103/PhysRevLett.49.1110` (cit. on p. 139).

Habib, S. et al. (Nov. 2012). “The Universe at Extreme Scale: Multi-Petaflop Sky Simulation on the BG/Q”. In: *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*. SC ’12. Salt Lake City, Utah: IEEE Computer Society Press, 4:1–4:11. ISBN: 978-1-4673-0804-

5. arXiv: `1211.4864 [cs.DC]`. URL: `http://dl.acm.org/citation.cfm?id=2388996.2389002` (visited on Feb. 13, 2013) (cit. on p. 54).

Hallman, E. J., B. W. O'Shea, J. O. Burns, M. L. Norman, R. Harkness, and R. Wagner (Dec. 2007). "The Santa Fe Light Cone Simulation Project. I. Confusion and the Warm-Hot Intergalactic Medium in Upcoming Sunyaev-Zel'dovich Effect Surveys". In: *ApJ* 671, pp. 27–39. DOI: `10.1086/522912`. arXiv: `0704.2607` (cit. on p. 87).

Hao, J. (Apr. 2009). "Optical Galacy Cluster Detection Across a Wide Redshift Range". PhD thesis. University of Michigan, Ann Arbor. (cit. on p. 175).

Hao, J. et al. (Dec. 2010). "A GMBCG Galaxy Cluster Catalog of 55,424 Rich Clusters from SDSS DR7". In: *ApJS* 191, pp. 254–274. DOI: `10.1088/0067-0049/191/2/254`. arXiv: `1010.5503 [astro-ph.CO]` (cit. on pp. 78, 89).

Harper, D. R. (Apr. 2013). "Search for physics beyond the Standard Model in events with a Z boson and large missing transverse momentum with the ATLAS detector". PhD thesis. University of Michigan, Ann Arbor. (cit. on p. 34).

Harrison, E. (1987). *Darkness at Night. A Riddle of the Universe.* Cambridge, MA, USA: Harvard University Press. ISBN: 0-674-19270-2. Google Books: `IRKYueWVftYC` (cit. on pp. 12, 13).

Hawkins, G. S. (Oct. 1963). "Stonehenge Decoded". In: *Nature* 200, pp. 306–308. DOI: `10.1038/200306a0` (cit. on p. 4).

— (1985). "Stonehenge Archives: Twenty Years After 'Stonehenge Decoded'". In: *Archaeoastronomy* 8, p. 6 (cit. on p. 4).

Heath, T. L. and Aristarchus of Samos (1913). *Aristarchus of Samos: The Ancient Copernicus; a History of Greek Astronomy to Aristarchus, Together with Aristarchus's Treatise on the Sizes and Distances of the Sun and Moon.* Oxford, UK: Clarendon Press. Google Books: `ZcLaAAAAMAAJ` (cit. on p. 6).

Heitmann, K et al. (Apr. 9, 2013). *Argonne Simulations.* DES LBNL Collaboration Meeting (cit. on p. 128).

Henry, J. P. and K. A. Arnaud (May 1991). "A Measurement of the Mass Fluctuation Spectrum from the Cluster X-Ray Temperature Function". In: *ApJ* 372, pp. 410–418. DOI: `10.1086/169987`. SAO–NASA ADS: `1991ApJ...372..410H` (cit. on p. 86).

Henry, J. P., A. E. Evrard, H. Hoekstra, A. Babul, and A. Mahdavi (Feb. 2009). "The X-Ray Cluster Normalization of the Matter Power Spectrum". In: *ApJ* 691, pp. 1307–1321. DOI: `10.1088/0004-637X/691/2/1307`. arXiv: `0809.3832` (cit. on p. 86).

Herschel, J. F. W. (1864). “Catalogue of Nebulae and Clusters of Stars.” In: *Royal Society of London Philosophical Transactions Series I* 154, pp. 1–137. JSTOR: `108864` (cit. on p. 10).

Herschel, W. (1785). “On the Construction of the Heavens.” In: *Royal Society of London Philosophical Transactions Series I* 75, pp. 213–266. JSTOR: `106755` (cit. on pp. 10, 12).

— (1786). “Catalogue of One Thousand New Nebulae and Clusters of Stars.” In: *Royal Society of London Philosophical Transactions Series I* 76, pp. 457–499. JSTOR: `106639` (cit. on p. 10).

Hetherington, N. S. (1993). *Cosmology: Historical, Literary, Philosophical, Religious, and Scientific Perspectives.* New York, NY, USA. ISBN: 0-8153-0934-1. Google Books: `8lGNQMUTr0gC` (cit. on p. 6).

Hey, Tony, Stewart Tansley, and Kristin M. Tolle, eds. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Microsoft Research. ISBN: 978-0982544204. URL: `http://research.microsoft.com/en-us/collaboration/fourthparadigm/` (cit. on p. 52).

Hillebrandt, W. and J. C. Niemeyer (2000). “Type IA Supernova Explosion Models”. In: *ARA&A* 38, pp. 191–230. DOI: `10.1146/annurev.astro.38.1.191`. arXiv: `astro-ph/0006305` (cit. on p. 30).

Hobson, M. P., G. Efstathiou, and A. Lasenby (Feb. 2006). *General Relativity: An Introduction for Physicists.* Cambridge, UK: Cambridge University Press. ISBN: 978-0-521-82951-9. Google Books: `xma1QuTJphYC` (cit. on p. 34).

Hockney, R. W. and J. W. Eastwood (1988). *Computer Simulation Using Particles.* New York, NY, USA: Taylor & Francis Group. ISBN: 978-0-05247-392-8. Google Books: `nTOFkmnCQuIC` (cit. on pp. 39, 153).

Hogg, D. W., I. K. Baldry, M. R. Blanton, and D. J. Eisenstein (Oct. 2002). “The K correction”. In: *ArXiv Astrophysics e-prints.* arXiv: `astro-ph/0210394` (cit. on p. 161).

Holder, G. P., Z. Haiman, and J. J. Mohr (Oct. 2001). “Constraints on Ω_m, Ω_Λ, and σ_8 from Galaxy Cluster Redshift Distributions”. In: *ApJ* 560, pp. L111–L114. DOI: `10.1086/324309`. arXiv: `astro-ph/0105396` (cit. on p. 97).

Holmberg, E. (Nov. 1941). “On the Clustering Tendencies among the Nebulae. II. A Study of Encounters Between Laboratory Models of Stellar Systems by a New Integration Procedure.” In: *ApJ* 94, p. 385. DOI: `10.1086/144344`. SAO–NASA ADS: `1941ApJ....94..385H` (cit. on p. 40).

Hoyle, F. (1948). “A New Model for the Expanding Universe”. In: *MNRAS* 108, p. 372. SAO–NASA ADS: `1948MNRAS.108..372H` (cit. on p. 17).

Hoyle, F. (Dec. 1966a). “Speculations on Stonehenge”. In: *Antiquity* 40, pp. 262–276. URL: `http://antiquity.ac.uk/ant/040/Ant0400262.htm` (visited on Mar. 16, 2013) (cit. on p. 4).

— (July 1966b). “Stonehenge-an Eclipse Predictor”. In: *Nature* 211, pp. 454–456. DOI: `10.1038/211454a0` (cit. on p. 4).

Hoyle, F., G. Burbidge, and J. V. Narlikar (June 1993). “A quasi-steady state cosmological model with creation of matter”. In: *ApJ* 410, pp. 437–457. DOI: `10.1086/172761`. SAO–NASA ADS: `1993ApJ...410..437H` (cit. on p. 17).

Hu, W. (Jan. 1995). “Wandering in the Background: A Cosmic Microwave Background Explorer”. PhD thesis. University of California, Berkeley. arXiv: `astro-ph/9508126` (cit. on p. 30).

Hu, W. and J. D. Cohn (Mar. 2006). “Likelihood Methods for Cluster Dark Energy Surveys”. In: *Phys. Rev. D* 73.6, 067301. DOI: `10.1103/PhysRevD.73.067301`. arXiv: `astro-ph/0602147` (cit. on p. 97).

Hu, W. and A. V. Kravtsov (Feb. 2003). “Sample Variance Considerations for Cluster Surveys”. In: *ApJ* 584, pp. 702–715. DOI: `10.1086/345846`. arXiv: `astro-ph/0203169` (cit. on pp. 172, 173).

Hubble, E. (Mar. 1929). “A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae”. In: *Proceedings of the National Academy of Science* 15, pp. 168–173. DOI: `10.1073/pnas.15.3.168`. SAO–NASA ADS: `1929PNAS...15..168H` (cit. on pp. 13, 14).

Huchra, J. P., M. Davis, D. Latham, and J. Tonry (June 1983). “A survey of galaxy redshifts. IV - The data”. In: *ApJS* 52, pp. 89–119. DOI: `10.1086/190860`. SAO–NASA ADS: `1983ApJS...52...89H` (cit. on p. 25).

Huterer, D. (Mar. 2002). “Weak lensing and dark energy”. In: *Phys. Rev. D* 65.6, 063001. DOI: `10.1103/PhysRevD.65.063001`. arXiv: `astro-ph/0106399` (cit. on p. 30).

Huterer, D. and M. Takada (May 2005). “Calibrating the nonlinear matter power spectrum: Requirements for future weak lensing surveys”. In: *Astroparticle Physics* 23, pp. 369–376. DOI: `10.1016/j.astropartphys.2005.02.006`. arXiv: `astro-ph/0412142` (cit. on p. 98).

Huygens, C. (1673). *Horologium Oscillatorium: sive, De Motu Pendulorum ad Horologia Aptato Demostrationes Geometricæ.* Paris, FR: F. Muguet. Google Books: `YgY8AAAAMAAJ` (cit. on p. 8).

Jain, B., U. Seljak, and S. D. M. White (Feb. 2000). “Ray-tracing Simulations of Weak Lensing by Large-Scale Structure”. In: *ApJ* 530, pp. 547–577. DOI: `10.1086/308384`. arXiv: `astro-ph/9901191` (cit. on p. 163).

Jenkins, A., C. S. Frenk, F. R. Pearce, P. A. Thomas, J. M. Colberg, S. D. M. White, H. M. P. Couchman, J. A. Peacock, G. Efstathiou, and A. H. Nelson (May 1998). "Evolution of Structure in Cold Dark Matter Universes". In: *ApJ* 499, p. 20. DOI: `10.1086/305615`. arXiv: `astro-ph/9709010` (cit. on p. 42).

Jenkins, A., C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida (Feb. 2001). "The mass function of dark matter haloes". In: *MNRAS* 321, pp. 372–384. DOI: `10.1046/j.1365-8711.2001.04029.x`. arXiv: `astro-ph/0005260` (cit. on pp. 87, 146).

Johnston, D. E., E. S. Sheldon, R. H. Wechsler, E. Rozo, B. P. Koester, J. A. Frieman, T. A. McKay, A. E. Evrard, M. R. Becker, and J. Annis (Sept. 2007). "Cross-correlation Weak Lensing of SDSS galaxy Clusters II: Cluster Density Profiles and the Mass–Richness Relation". In: *ArXiv e-prints*. arXiv: `0709.1159` (cit. on pp. 88, 92).

Kant, I. (1755). *Allgemeine Naturgeschichte und Theorie des Himmels (German) [Universal Natural History and Theory of the Heavens]*. Konigsberg, DE: J. F. Petersen. Google Books: `jT4OAAAAYAAJ`. URL: `http://records.viu.ca/~johnstoi/kant/kant2e.htm#twoseven` (visited on Feb. 13, 2013) (cit. on p. 10).

Ketchum, J. A., F. C. Adams, and A. M. Bloch (Jan. 2013). "Mean Motion Resonances in Exoplanet Systems: An Investigation into Nodding Behavior". In: *ApJ* 762, 71, p. 71. DOI: `10.1088/0004-637X/762/2/71`. arXiv: `1211.3078 [astro-ph.EP]` (cit. on p. 147).

Kim, J., C. Park, G. Rossi, S. M. Lee, and J. R. Gott III (Dec. 2011). "The New Horizon Run Cosmological N-Body Simulations". In: *Journal of Korean Astronomical Society* 44, pp. 217–234. DOI: `10.5303/JKAS.2011.44.6.217`. arXiv: `1112.1754 [astro-ph.CO]` (cit. on pp. 42, 45).

Klypin, A. A. and S. F. Shandarin (Sept. 1983). "Three-dimensional numerical model of the formation of large-scale structure in the Universe". In: *MNRAS* 204, pp. 891–907. SAO–NASA ADS: `1983MNRAS.204..891K` (cit. on p. 149).

Klypin, A. A., S. Trujillo-Gomez, and J. Primack (Oct. 2011). "Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation". In: *ApJ* 740, 102, p. 102. DOI: `10.1088/0004-637X/740/2/102`. arXiv: `1002.3660 [astro-ph.CO]` (cit. on p. 42).

Koester, B. P. et al. (May 2007). "A MaxBCG Catalog of 13,823 Galaxy Clusters from the Sloan Digital Sky Survey". In: *ApJ* 660, pp. 239–255. DOI: `10.1086/509599`. arXiv: `astro-ph/0701265` (cit. on pp. 86, 89, 92, 107).

Kolb, E. W. and M. S. Turner (1990). *The Early Universe*. Oxford, UK: Westview Press. ISBN: 0-201-62674-8. Google Books: `Qwijr-HsvMMC` (cit. on pp. 18, 20, 139, 140).

Komatsu, E. et al. (Feb. 2009). “Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation”. In: *ApJS* 180, pp. 330–376. DOI: `10.1088/0067-0049/180/2/330`. arXiv: `0803.0547` (cit. on pp. 92, 143).

Komatsu, E. et al. (Feb. 2011). “Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation”. In: *ApJS* 192, 18, p. 18. DOI: `10.1088/0067-0049/192/2/18`. arXiv: `1001.4538 [astro-ph.CO]` (cit. on p. 20).

Kuehn, K. et al. (Aug. 2012). “PreCam, a Precursor Observational Campaign for Calibration of the Dark Energy Survey”. In: *ArXiv e-prints*. arXiv: `1208.0865 [astro-ph.IM]` (cit. on p. 28).

Lanczos, C. (1922). “Bemerkung zur de Sitterschen Welt (German) [Remarks on the de Sitter universe]”. In: *Physikalische Zeitschrift* 23, pp. 539–543. ISSN: 0369-982X. HathiTrust: `mdp.39015086723239;seq=627` (cit. on p. 18).

Leauthaud, A., J. Tinker, P. S. Behroozi, M. T. Busha, and R. H. Wechsler (Sept. 2011). “A Theoretical Framework for Combining Techniques that Probe the Link Between Galaxies and Dark Matter”. In: *ApJ* 738, 45, p. 45. DOI: `10.1088/0004-637X/738/1/45`. arXiv: `1103.2077 [astro-ph.CO]` (cit. on p. 110).

Leauthaud, A. et al. (Jan. 2012). “New Constraints on the Evolution of the Stellar-to-dark Matter Connection: A Combined Analysis of Galaxy-Galaxy Lensing, Clustering, and Stellar Mass Functions from $z = 0.2$ to $z = 1$”. In: *ApJ* 744, 159, p. 159. DOI: `10.1088/0004-637X/744/2/159`. arXiv: `1104.0928 [astro-ph.CO]` (cit. on p. 110).

Lemaître, G. (1927). “Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques”. French. In: *Annales de la Societe Scietifique de Bruxelles*. A 47, pp. 49–59 (cit. on pp. 13, 18, 201).

— (Mar. 1931). “Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae (Translation of Lemaître, 1927)”. In: *MNRAS* 91, pp. 483–490. SAO–NASA ADS: `1931MNRAS..91..483L` (cit. on pp. 13, 18).

Levine, E. S., A. E. Schulz, and M. White (Oct. 2002). “Future Galaxy Cluster Surveys: The Effect of Theory Uncertainty on Constraining Cosmological Parameters”. In: *ApJ* 577, pp. 569–578. DOI: `10.1086/342119`. arXiv: `astro-ph/0204273` (cit. on p. 86).

Lewis, A., A. Challinor, and A. Lasenby (Aug. 2000). “Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models”. In: *ApJ* 538, pp. 473–476. DOI: `10.1086/309179`. arXiv: `astro-ph/9911177` (cit. on pp. 47, 58, 148).

Liddle, A. R. and D. H. Lyth (June 2000). *Cosmological Inflation and Large-Scale Structure*. Cambirdge, UK: Cambridge University Press. ISBN: 0-521-66022-X. Google Books: `XmWauPZSovMC` (cit. on p. 140).

Lima, M. and W. Hu (Aug. 2004). “Self-calibration of cluster dark energy studies: Counts in cells”. In: *Phys. Rev. D* 70.4, 043504. DOI: `10.1103/PhysRevD.70.043504`. arXiv: `astro-ph/0401559` (cit. on pp. 86, 91, 95, 97).

— (Aug. 2005). “Self-calibration of cluster dark energy studies: Observable-mass distribution”. In: *Phys. Rev. D* 72.4, 043006. DOI: `10.1103/PhysRevD.72.043006`. arXiv: `astro-ph/0503363` (cit. on p. 95).

— (Dec. 2007). “Photometric redshift requirements for self-calibration of cluster dark energy studies”. In: *Phys. Rev. D* 76.12, 123013. DOI: `10.1103/PhysRevD.76.123013`. arXiv: `0709.2871` (cit. on p. 95).

Limber, D. N. (Jan. 1953). “The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field.” In: *ApJ* 117, p. 134. DOI: `10.1086/145672`. SAO–NASA ADS: `1953ApJ...117..134L` (cit. on p. 163).

Linder, E. V. (Mar. 2003). “Exploring the Expansion History of the Universe”. In: *Phys. Rev. Lett.* 90.9, 091301, p. 091301. DOI: `10.1103/PhysRevLett.90.091301`. arXiv: `astro-ph/0208512` (cit. on pp. 37, 137).

Linder, E. V. and A. Jenkins (Dec. 2003). “Cosmic structure growth and dark energy”. In: *MNRAS* 346, pp. 573–583. DOI: `10.1046/j.1365-2966.2003.07112.x`. arXiv: `astro-ph/0305286` (cit. on p. 143).

Lockyer, N. (1906). *Stonehenge and other British Stone Monuments Astronomically Considered*. London, UK: Macmillan. Google Books: `RGITAAAAYAAJ` (cit. on p. 4).

Łokas, E. L., F. Prada, R. Wojtak, M. Moles, and S. Gottlöber (Feb. 2006). “The complex velocity distribution of galaxies in Abell 1689: implications for mass modelling”. In: *MNRAS* 366, pp. L26–L30. DOI: `10.1111/j.1745-3933.2005.00125.x`. arXiv: `astro-ph/0507508` (cit. on p. 88).

Longair, M. S. (2008). *Galaxy Formation*. Berlin, DE: Springer. ISBN: 978-3-540-84056-5. Google Books: `e-wJHSBOuZAC` (cit. on pp. 145, 175).

LSST Science Collaboration et al. (Dec. 2009). “LSST Science Book, Version 2.0”. In: *ArXiv e-prints*. arXiv: `0912.0201 [astro-ph.IM]` (cit. on p. 27).

Ludäscher, Bertram, Ilkay Altintas, Chad Berkley, Dan Higgins, Efrat Jaeger, Matthew Jones, Edward A. Lee, Jing Tao, and Yang Zhao (Aug. 2006). “Scientific workflow management and the Kepler system: Research Articles”. In: *Concurr. Comput. : Pract. Exper.* 18.10, pp. 1039–1065. ISSN: 1532-0626. DOI: `10.1002/cpe.v18:10` (cit. on p. 59).

Maddox, S. J., G. Efstathiou, and W. J. Sutherland (Oct. 1990). “The APM Galaxy Survey - II. Photometric corrections”. In: *MNRAS* 246, p. 433. SAO–NASA ADS: `1990MNRAS.246..433M` (cit. on p. 25).

Maddox, S. J., G. Efstathiou, W. J. Sutherland, and J. Loveday (Jan. 1990a). “Galaxy correlations on large scales”. In: *MNRAS* 242, 43P–47P. SAO–NASA ADS: `1990MNRAS.242P..43M` (cit. on p. 24).

— (Apr. 1990b). “The APM Galaxy Survey - I: APM measurements and star-galaxy separation”. In: *MNRAS* 243, pp. 692–712. SAO–NASA ADS: `1990MNRAS.243..692M` (cit. on p. 25).

Majumdar, S. and J. J. Mohr (Sept. 2004). “Self-Calibration in Cluster Studies of Dark Energy: Combining the Cluster Redshift Distribution, the Power Spectrum, and Mass Measurements”. In: *ApJ* 613, pp. 41–50. DOI: `10.1086/422829`. arXiv: `astro-ph/0305341` (cit. on pp. 86, 91).

Mantz, A. B., S. W. Allen, D. Rapetti, and H. Ebeling (Aug. 2010). “The observed growth of massive galaxy clusters - I. Statistical methods and cosmological constraints”. In: *MNRAS* 406, pp. 1759–1772. DOI: `10.1111/j.1365-2966.2010.16992.x`. arXiv: `0909.3098 [astro-ph.CO]` (cit. on p. 86).

Markevitch, M. (Jan. 2006). “Chandra Observation of the Most Interesting Cluster in the Universe”. In: *The X-ray Universe 2005*. Ed. by A. Wilson. Vol. 604. ESA Special Publication, p. 723. arXiv: `astro-ph/0511345` (cit. on pp. 22, 23).

Markevitch, M., A. H. Gonzalez, D. Clowe, A. Vikhlinin, W. Forman, C. Jones, S. Murray, and W. Tucker (May 2004). “Direct Constraints on the Dark Matter Self-Interaction Cross Section from the Merging Galaxy Cluster 1E 0657-56”. In: *ApJ* 606, pp. 819–824. DOI: `10.1086/383178`. arXiv: `astro-ph/0309303` (cit. on p. 23).

Marru, Suresh et al. (2011). “Apache airavata: a framework for distributed applications and computational workflows”. In: *Proceedings of the 2011 ACM workshop on Gateway computing environments*. GCE ’11. Seattle, Washington, USA: ACM, pp. 21–28. ISBN: 978-1-4503-1123-6. DOI: `10.1145/2110486.2110490` (cit. on p. 59).

Mather, J. C. et al. (Jan. 1994). “Measurement of the Cosmic Microwave Background Spectrum by the COBE FIRAS Instrument”. In: *ApJ* 420, pp. 439–444. DOI: `10.1086/173574`. SAO–NASA ADS: `1994ApJ...420..439M` (cit. on p. 20).

Maurogordato, S., J. L. Sauvageot, H. Bourdin, A. Cappi, C. Benoist, C. Ferrari, G. Mars, and K. Houairi (Jan. 2011). “Merging history of three bimodal clusters”. In: *A&A* 525, A79, A79. DOI: `10.1051/0004-6361/201014415`. arXiv: `1009.1967 [astro-ph.CO]` (cit. on p. 88).

Melott, A. L. (Feb. 1983). “Two-dimensional simulation of the gravitational superclustering of collisionless particles”. In: *MNRAS* 202, pp. 595–604. SAO–NASA ADS: `1983MNRAS.202..595M` (cit. on p. 42).

Messier, C. (1781). “Catalogue des Nébuleuses & des amas d’Étoiles (French) [Catalog of Nebulae and Star Clusters]”. In: *Connoissance des Temps for 1784*, pp. 227–267. URL: `http://messier.seds.org/xtra/Mcat/mcat1781.html` (visited on Feb. 13, 2013) (cit. on p. 10).

Metzler, C. A. and A. E. Evrard (Oct. 1997). “Simulations Of Galaxy Clusters With And Without Winds: I. The Structure Of Clusters”. In: *ArXiv Astrophysics e-prints*. arXiv: `astro-ph/9710324` (cit. on p. 149).

Milkeraitis, M., L. van Waerbeke, C. Heymans, H. Hildebrandt, J. P. Dietrich, and T. Erben (July 2010). “3D-Matched-Filter galaxy cluster finder - I. Selection functions and CFHTLS Deep clusters”. In: *MNRAS* 406, pp. 673–688. DOI: `10.1111/j.1365-2966.2010.16720.x`. arXiv: `0912.0739 [astro-ph.CO]` (cit. on pp. 90, 107).

Miller, C. J. et al. (Sept. 2005). “The C4 Clustering Algorithm: Clusters of Galaxies in the Sloan Digital Sky Survey”. In: *AJ* 130, pp. 968–1001. DOI: `10.1086/431357`. arXiv: `astro-ph/0503713` (cit. on pp. 78, 107).

Minkowski, R. L. and G. O. Abell (1963). “The National Geographic Society-Palomar Observatory Sky Survey”. In: *Basic Astronomical Data: Stars and Stellar Systems*. Ed. by K. A. Strand. Vol. 3. Chicago, IL: the University of Chicago Press, p. 481. Google Books: `TonvAAAAMAAJ` (cit. on p. 15).

Miyoshi, K. and T. Kihara (1975). “Development of the correlation of galaxies in an expanding universe”. In: *PASJ* 27, pp. 333–346. SAO–NASA ADS: `1975PASJ...27..333M` (cit. on p. 42).

Monaghan, J. J. and J. C. Lattanzio (Aug. 1985). “A refined particle method for astrophysical problems”. In: *A&A* 149, pp. 135–143. SAO–NASA ADS: `1985A%26A...149..135M` (cit. on p. 151).

Newham, C. A. (July 1966). “Stonehenge—a Neolithic ‘Observatory’”. In: *Nature* 211, pp. 456–458. DOI: `10.1038/211456a0` (cit. on p. 4).

Newton, I. (1687). *Philosophiae naturalis principia mathematica*. London, UK: Royal Society Press. Google Books: `exwAAAAAQAAJ`. URL: `http://cudl.lib.cam.ac.uk/view/PR-ADV-B-00039-00001/` (visited on Feb. 13, 2013) (cit. on p. 9).

Nielsen, F. (Jan. 2013). “Cramer-Rao Lower Bound and Information Geometry”. In: *ArXiv e-prints*. arXiv: `1301.3578 [cs.IT]` (cit. on p. 164).

Noh, Y. and J. D. Cohn (May 2011). “The geometry of the filamentary environment of galaxy clusters”. In: *MNRAS* 413, pp. 301–312. DOI: `10.1111/j.1365-2966.2010.18137.x`. arXiv: `1011.1000 [astro-ph.CO]` (cit. on p. 88).

Nord, B. et al. (2013). "Joint SZ-Optical cluster finding". In: *In Prep.* (Cit. on p. 82).

Oort, J. H. (Aug. 1932). "The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems". In: *Bull. Astron. Inst. Netherlands* 6, p. 249. SAO–NASA ADS: `1932BAN.....6..249O` (cit. on p. 14).

Ostriker, J. P. and P. J. Steinhardt (Oct. 1995). "The observational case for a low-density Universe with a non-zero cosmological constant". In: *Nature* 377, pp. 600–602. DOI: `10.1038/377600a0` (cit. on p. 24).

Pannekoek, A. (1961). *A History of Astronomy.* London, UK: G. Allen & Unwin Ltd. ISBN: 0-486-65994-1. Google Books: `O7D9AyU-nLYC` (cit. on pp. 5, 6).

Park, C. (Feb. 1990). "Large N-body simulations of a universe dominated by cold dark matter". In: *MNRAS* 242, 59P–61P. SAO–NASA ADS: `1990MNRAS.242P..59P` (cit. on p. 42).

— (Oct. 1997). "A Particle-Mesh Code for the Next Generation Cosmological N-Body Simulations". In: *Journal of Korean Astronomical Society* 30, pp. 191–200. SAO–NASA ADS: `1997JKAS...30..191P` (cit. on p. 42).

Park, C., M. S. Vogeley, M. J. Geller, and J. P. Huchra (Aug. 1994). "Power spectrum, correlation function, and tests for luminosity bias in the CfA redshift survey". In: *ApJ* 431, pp. 569–585. DOI: `10.1086/174508`. SAO–NASA ADS: `1994ApJ...431..569P` (cit. on p. 42).

Peebles, P. J. E. (Feb. 1970). "Structure of the Coma Cluster of Galaxies". In: *AJ* 75, p. 13. DOI: `10.1086/110933`. SAO–NASA ADS: `1970AJ.....75...13P` (cit. on pp. 41, 42).

— (1993). *Principles of Physical Cosmology.* Ewing, NJ, USA: Princeton University Press. ISBN: 0-691-07428-3. Google Books: `AmlEt6TJ6jAC` (cit. on pp. 38, 40, 140, 145).

Penzias, A. A. (July 1979). "The origin of the elements". In: *Reviews of Modern Physics* 51, pp. 425–432. DOI: `10.1103/RevModPhys.51.425` (cit. on p. 18).

Penzias, A. A. and R. W. Wilson (July 1965). "A Measurement of Excess Antenna Temperature at 4080 Mc/s." In: *ApJ* 142, pp. 419–421. DOI: `10.1086/148307`. SAO–NASA ADS: `1965ApJ...142..419P` (cit. on p. 18).

Perlmutter, S. and B. P. Schmidt (2003). "Measuring Cosmology with Supernovae". In: *Supernovae and Gamma-Ray Bursters.* Ed. by K. Weiler. Vol. 598. Lecture Notes in Physics, Berlin Springer Verlag, pp. 195–217. arXiv: `astro-ph/0303428` (cit. on p. 21).

Perlmutter, S. et al. (Dec. 1997). "Cosmology From Type IA Supernovae: Measurements, Calibration Techniques, and Implications". In: *American Astronomical So-*

ciety Meeting Abstracts. Vol. 29. Bulletin of the American Astronomical Society, p. 1351. arXiv: `astro-ph/9812473` (cit. on p. 20).

Perlmutter, S. et al. (June 1999). “Measurements of Omega and Lambda from 42 High-Redshift Supernovae”. In: *ApJ* 517, pp. 565–586. DOI: `10.1086/307221`. arXiv: `astro-ph/9812133` (cit. on pp. 20, 21).

Pierce, Marlon, Suresh Marru, Raminder Singh, Archit Kulshrestha, and Karthik Muthuraman (2010). “Open grid computing environments: advanced gateway support activities”. In: *Proceedings of the 2010 TeraGrid Conference*. TG ’10. Pittsburgh, Pennsylvania: ACM, 16:1–16:9. ISBN: 978-1-60558-818-6. DOI: `10.1145/1838574.1838590` (cit. on p. 59).

Planck Collaboration (Dec. 2011a). “Planck early results. I. The Planck mission”. In: *A&A* 536, A1, A1. DOI: `10.1051/0004-6361/201116464`. arXiv: `1101.2022 [astro-ph.IM]` (cit. on p. 27).

— (Dec. 2011b). “Planck early results. IX. XMM-Newton follow-up for validation of Planck cluster candidates”. In: *A&A* 536, A9, A9. DOI: `10.1051/0004-6361/201116460`. arXiv: `1101.2025 [astro-ph.CO]` (cit. on p. 108).

— (Dec. 2011c). “Planck early results. X. Statistical analysis of Sunyaev-Zeldovich scaling relations for X-ray galaxy clusters”. In: *A&A* 536, A10, A10. DOI: `10.1051/0004-6361/201116457`. arXiv: `1101.2043 [astro-ph.CO]` (cit. on p. 88).

— (Dec. 2011d). “Planck early results. XII. Cluster Sunyaev-Zeldovich optical scaling relations”. In: *A&A* 536, A12, A12. DOI: `10.1051/0004-6361/201116489`. arXiv: `1101.2027 [astro-ph.CO]` (cit. on p. 86).

— (Mar. 2013). “Planck 2013 results. I. Overview of products and scientific results”. In: *ArXiv e-prints*. arXiv: `1303.5062 [astro-ph.CO]` (cit. on p. 19).

Poole, G. B., M. A. Fardal, A. Babul, I. G. McCarthy, T. Quinn, and J. Wadsley (Dec. 2006). “The impact of mergers on relaxed X-ray clusters - I. Dynamical evolution and emergent transient structures”. In: *MNRAS* 373, pp. 881–905. DOI: `10.1111/j.1365-2966.2006.10916.x`. arXiv: `astro-ph/0608560` (cit. on p. 87).

Pratt, G. W., J. H. Croston, M. Arnaud, and H. Böhringer (May 2009). “Galaxy cluster X-ray luminosity scaling relations from a representative local sample (REXCESS)”. In: *A&A* 498, pp. 361–378. DOI: `10.1051/0004-6361/200810994`. arXiv: `0809.3784` (cit. on p. 87).

Press, W. H. and P. Schechter (Feb. 1974). “Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation”. In: *ApJ* 187, pp. 425–438. DOI: `10.1086/152650` (cit. on p. 144).

Quinn, T., N. Katz, J. Stadel, and G. Lake (Oct. 1997). “Time stepping N-body simulations”. In: *ArXiv Astrophysics e-prints*. arXiv: `astro-ph/9710043` (cit. on p. 151).

Rawlings, S. and R. Schilizzi (May 2011). “The Square Kilometre Array”. In: *ArXiv e-prints*. arXiv: `1105.5953 [astro-ph.IM]` (cit. on p. 28).

Reiprich, T. H. and H. Böhringer (Mar. 2002). “The Mass Function of an X-Ray Flux-limited Sample of Galaxy Clusters”. In: *ApJ* 567, pp. 716–740. DOI: `10.1086/338753`. arXiv: `astro-ph/0111285` (cit. on p. 88).

Reyes, R., R. Mandelbaum, U. Seljak, T. Baldauf, J. E. Gunn, L. Lombriser, and R. E. Smith (Mar. 2010). “Confirmation of general relativity on large scales from weak lensing and galaxy velocities”. In: *Nature* 464, pp. 256–258. DOI: `10.1038/nature08857`. arXiv: `1003.2185 [astro-ph.CO]` (cit. on p. 28).

Riess, A. G. et al. (Sept. 1998). “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant”. In: *AJ* 116, pp. 1009–1038. DOI: `10.1086/300499`. arXiv: `astro-ph/9805201` (cit. on pp. 20, 21).

Robertson, H. P. (Nov. 1935). “Kinematics and World-Structure”. In: *ApJ* 82, p. 284. DOI: `10.1086/143681`. SAO–NASA ADS: `1935ApJ....82..284R` (cit. on p. 18).

Rozo, E., J. G. Bartlett, A. E. Evrard, and E. S. Rykoff (Apr. 2012). “Closing the Loop: A Self-Consistent Model of Optical, X-ray, and SZ Scaling Relations for Clusters of Galaxies”. In: *ArXiv e-prints*. arXiv: `1204.6305 [astro-ph.CO]` (cit. on p. 86).

Rozo, E., A. E. Evrard, E. S. Rykoff, and J. G. Bartlett (Apr. 2012). “A Comparative Study of Local Galaxy Clusters: II: X-ray and SZ Scaling Relations”. In: *ArXiv e-prints*. arXiv: `1204.6292 [astro-ph.CO]` (cit. on p. 86).

Rozo, E. and E. S. Rykoff (Mar. 2013). “redMaPPer II: X-ray and SZ Performance Benchmarks for the SDSS Catalog”. In: *ArXiv e-prints*. arXiv: `1303.3373 [astro-ph.CO]` (cit. on pp. 78, 175).

Rozo, E., E. S. Rykoff, J. G. Bartlett, and A. E. Evrard (Apr. 2012). “A Comparative Study of Local Galaxy Clusters: I. Derived X-ray Observables”. In: *ArXiv e-prints*. arXiv: `1204.6301 [astro-ph.CO]` (cit. on p. 86).

Rozo, E., E. S. Rykoff, A. E. Evrard, et al. (July 2009). “Constraining the Scatter in the Mass-richness Relation of maxBCG Clusters with Weak Lensing and X-ray Data”. In: *ApJ* 699, pp. 768–781. DOI: `10.1088/0004-637X/699/1/768`. arXiv: `0809.2794` (cit. on pp. 92, 93).

Rozo, E., E. S. Rykoff, B. P. Koester, T. A. McKay, et al. (Sept. 2009). “Improvement of the Richness Estimates of maxBCG Clusters”. In: *ApJ* 703, pp. 601–613. DOI: `10.1088/0004-637X/703/1/601` (cit. on pp. 75, 175).

Rozo, E., E. S. Rykoff, B. P. Koester, B. Nord, H.-Y. Wu, A. E. Evrard, and R. H. Wechsler (Oct. 2011). “Extrinsic Sources of Scatter in the Richness-mass Relation of Galaxy Clusters”. In: *ApJ* 740, 53, p. 53. DOI: `10.1088/0004-637X/740/2/53`. arXiv: `1104.2090 [astro-ph.CO]` (cit. on pp. 88, 90, 91).

Rozo, E., A. Vikhlinin, and S. More (Nov. 2012). “The Y $_{SZ}$-Y$_X$ Scaling Relation as Determined from Planck and Chandra”. In: *ApJ* 760, 67, p. 67. DOI: `10.1088/0004-637X/760/1/67`. arXiv: `1202.2150 [astro-ph.CO]` (cit. on p. 86).

Rozo, E. et al. (Jan. 2010). “Cosmological Constraints from the Sloan Digital Sky Survey maxBCG Cluster Catalog”. In: *ApJ* 708, pp. 645–660. DOI: `10.1088/0004-637X/708/1/645`. arXiv: `0902.3702 [astro-ph.CO]` (cit. on p. 86).

Rykoff, E. S., B. P. Koester, E. Rozo, J. Annis, A. E. Evrard, S. M. Hansen, J. Hao, D. E. Johnston, T. A. McKay, and R. H. Wechsler (Feb. 2012). “Robust Optical Richness Estimation with Reduced Scatter”. In: *ApJ* 746, 178, p. 178. DOI: `10.1088/0004-637X/746/2/178`. arXiv: `1104.2089 [astro-ph.CO]` (cit. on pp. 75, 90, 93, 107, 124, 175).

Rykoff, E. S. et al. (June 2008). “The L$_X$–M relation of clusters of galaxies”. In: *MNRAS* 387, pp. L28–L32. DOI: `10.1111/j.1745-3933.2008.00476.x`. arXiv: `0802.1069` (cit. on p. 88).

Rykoff, E. S. et al. (Mar. 2013). “redMaPPer I: Algorithm and SDSS DR8 Catalog”. In: *ArXiv e-prints*. arXiv: `1303.3562 [astro-ph.CO]` (cit. on pp. 79, 111, 175, 176).

Saotome, R. (Apr. 2013). “Selected Studies in Classical and Quantum Gravity”. PhD thesis. University of Michigan, Ann Arbor. (cit. on p. 21).

Schechter, P. (Jan. 1976). “An analytic expression for the luminosity function for galaxies.” In: *ApJ* 203, pp. 297–306. DOI: `10.1086/154079` (cit. on p. 175).

Schmidt, B. (1938). “Ein lichtstarkes komafreies Spiegelsystem (German) [A bright coma-free mirror system]”. In: *Mitteilungen der Hamburger Sternwarte in Bergedorf* 7, pp. 15–17. SAO–NASA ADS: `1938MiHam...7...15S` (cit. on p. 15).

Schneider, D. P., M. Schmidt, and J. E. Gunn (Apr. 1994). “Spectroscopic CCD surveys for quasars at large redshift. 3: The Palomar Transit GRISM Survey catalog”. In: *AJ* 107, pp. 1245–1269. DOI: `10.1086/116937`. SAO–NASA ADS: `1994AJ....107.1245S` (cit. on p. 27).

Schneider, P. (2006). *Extragalactic Astronomy and Cosmology*. Berlin, DE: Springer-Verlag. ISBN: 978-3-540-82210-3. Google Books: `uP1Hz-6sHaMC` (cit. on p. 12).

Schneider, P., L. King, and T. Erben (Jan. 2000). “Cluster mass profiles from weak lensing: constraints from shear and magnification information”. In: *A&A* 353, pp. 41–56. eprint: `astro-ph/9907143` (cit. on p. 80).

Scoccimarro, R. (Oct. 1998). “Transients from initial conditions: a perturbative analysis”. In: *MNRAS* 299, pp. 1097–1118. DOI: `10.1046/j.1365-8711.1998.01845.x`. arXiv: `astro-ph/9711187` (cit. on pp. 149, 150).

Seljak, U. and M. Zaldarriaga (Oct. 1996). “A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies”. In: *ApJ* 469, p. 437. DOI: `10.1086/177793`. arXiv: `astro-ph/9603033` (cit. on pp. 92, 148).

Shane, C. D. and C. A. Wirtanen (1967). “The Distribution of Galaxies”. In: *Publ. of the Lick Obs.* 22, pp. 1–77. HathiTrust: `uc1.31822020237715;seq=13` (cit. on p. 25).

Shaw, L. D., G. P. Holder, and J. Dudley (June 2010). “Non-Gaussian Scatter in Cluster Scaling Relations”. In: *ApJ* 716, pp. 281–285. DOI: `10.1088/0004-637X/716/1/281`. arXiv: `0908.1978 [astro-ph.CO]` (cit. on p. 88).

Sheth, R. K. and G. Tormen (Sept. 1999). “Large-scale bias and the peak background split”. In: *MNRAS* 308, pp. 119–126. DOI: `10.1046/j.1365-8711.1999.02692.x`. arXiv: `astro-ph/9901122` (cit. on p. 173).

Shin, M.-S. and M. Ruszkowski (Jan. 2013). “Ram pressure stripping in elliptical galaxies - I. The impact of the interstellar medium turbulence”. In: *MNRAS* 428, pp. 804–814. DOI: `10.1093/mnras/sts071`. arXiv: `1206.6947 [astro-ph.CO]` (cit. on p. 147).

Skrutskie, M. F. et al. (Feb. 2006). “The Two Micron All Sky Survey (2MASS)”. In: *AJ* 131, pp. 1163–1183. DOI: `10.1086/498708` (cit. on p. 27).

Smoot, G. F. et al. (Sept. 1992). “Structure in the COBE differential microwave radiometer first-year maps”. In: *ApJ* 396, pp. L1–L5. DOI: `10.1086/186504`. SAO–NASA ADS: `1992ApJ...396L...1S` (cit. on p. 20).

Soares-Santos, M., R. R. de Carvalho, J. Annis, R. R. Gal, F. La Barbera, P. A. A. Lopes, R. H. Wechsler, M. T. Busha, and B. F. Gerke (Jan. 2011). “The Voronoi Tessellation Cluster Finder in 2+1 Dimensions”. In: *ApJ* 727, 45, p. 45. DOI: `10.1088/0004-637X/727/1/45`. arXiv: `1011.3458 [astro-ph.CO]` (cit. on pp. 79, 107).

Soares-Santos, M. and the DES Collaboration (July 2012). “Dark energy physics expectations at DES”. In: *Journal of Physics Conference Series* 375.3, p. 032006. DOI: `10.1088/1742-6596/375/1/032006`. arXiv: `1202.1848 [astro-ph.CO]` (cit. on p. 31).

Song, J. et al. (2013). “The Cluster Finder Comparison Project”. In: *In Prep.* (Cit. on p. 78).

Song, J., J. J. Mohr, W. A. Barkhouse, M. S. Warren, K. Dolag, and C. Rude (Mar. 2012). “A Parameterized Galaxy Catalog Simulator for Testing Cluster Finding, Mass Estimation, and Photometric Redshift Estimation in Optical and Near-infrared Surveys”. In: *ApJ* 747, 58, p. 58. DOI: `10.1088/0004-637X/747/1/58`. arXiv: `1104.2332 [astro-ph.CO]` (cit. on p. 107).

Spergel, D. N. et al. (Sept. 2003). “First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters”. In: *ApJS*

148, pp. 175–194. DOI: `10.1086/377226`. arXiv: `astro-ph/0302209` (cit. on p. 20).

Springel, V. (Dec. 2005). “The cosmological simulation code GADGET-2”. In: *MNRAS* 364, pp. 1105–1134. DOI: `10.1111/j.1365-2966.2005.09655.x`. arXiv: `astro-ph/0505010` (cit. on pp. 44, 47, 150, 152, 154, 155).

Springel, V., C. S. Frenk, and S. D. M. White (Apr. 2006). “The large-scale structure of the Universe”. In: *Nature* 440, pp. 1137–1144. DOI: `10.1038/nature04805`. arXiv: `astro-ph/0604561` (cit. on p. 26).

Springel, V. et al. (June 2005). “Simulations of the formation, evolution and clustering of galaxies and quasars”. In: *Nature* 435, pp. 629–636. DOI: `10.1038/nature03597`. arXiv: `astro-ph/0504097` (cit. on pp. 26, 42–44, 58, 91, 150).

Standish Jr., E. M. (1968). “Numerical Studies of the Gravitational Problem of N Bodies.” PhD thesis. Yale University. URL: `http://search.proquest.com/docview/302358887` (visited on Mar. 19, 2013) (cit. on p. 42).

Stanek, R. (2009). “Population Statistics of Galaxy Halos in Cosmological Simulations”. PhD thesis. University of Michigan, Ann Arbor. URL: `http://hdl.handle.net/2027.42/63677` (cit. on p. 39).

Stanek, R., A. E. Evrard, H. Böhringer, P. Schuecker, and B. Nord (Sept. 2006). “The X-Ray Luminosity-Mass Relation for Local Clusters of Galaxies”. In: *ApJ* 648, pp. 956–968. DOI: `10.1086/506248`. arXiv: `astro-ph/0602324` (cit. on p. 88).

Stanek, R., E. Rasia, A. E. Evrard, F. Pearce, and L. Gazzola (June 2010). “Massive Halos in Millennium Gas Simulations: Multivariate Scaling Relations”. In: *ApJ* 715, pp. 1508–1523. DOI: `10.1088/0004-637X/715/2/1508`. arXiv: `0910.1599 [astro-ph.CO]` (cit. on p. 87).

Sullivan, M. et al. (Aug. 2011). “SNLS3: Constraints on Dark Energy Combining the Supernova Legacy Survey Three-year Data with Other Probes”. In: *ApJ* 737, 102, p. 102. DOI: `10.1088/0004-637X/737/2/102`. arXiv: `1104.1444 [astro-ph.CO]` (cit. on p. 28).

Sunyaev, R. A. and Y. B. Zel’dovich (Nov. 1972). “The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies”. In: *Comments on Astrophysics and Space Physics* 4, p. 173. SAO–NASA ADS: `1972CoASP...4..173S` (cit. on p. 56).

Suto, Y. and T. Suginohara (Mar. 1991). “Redshift-space correlation functions in the cold dark matter scenario”. In: *ApJ* 370, pp. L15–L18. DOI: `10.1086/185966`. SAO–NASA ADS: `1991ApJ...370L..15S` (cit. on p. 42).

Swetz, D. S. et al. (June 2011). “Overview of the Atacama Cosmology Telescope: Receiver, Instrumentation, and Telescope Systems”. In: *ApJS* 194, 41, p. 41. DOI: `10.1088/0067-0049/194/2/41`. arXiv: `1007.0290 [astro-ph.IM]` (cit. on p. 108).

Szabo, T., E. Pierpaoli, F. Dong, A. Pipino, and J. Gunn (July 2011). "An Optical Catalog of Galaxy Clusters Obtained from an Adaptive Matched Filter Finder Applied to Sloan Digital Sky Survey Data Release 6". In: *ApJ* 736, 21, p. 21. DOI: `10.1088/0004-637X/736/1/21`. arXiv: `1011.0249 [astro-ph.CO]` (cit. on p. 89).

Tegmark, M., A. N. Taylor, and A. F. Heavens (May 1997). "Karhunen-Loeve Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets?" In: *ApJ* 480, p. 22. DOI: `10.1086/303939`. arXiv: `astro-ph/9603021` (cit. on p. 165).

Teyssier, R., S. Pires, S. Prunet, D. Aubert, C. Pichon, A. Amara, K. Benabed, S. Colombi, A. Refregier, and J.-L. Starck (Apr. 2009). "Full-sky weak-lensing simulation with 70 billion particles". In: *A&A* 497, pp. 335–341. DOI: `10.1051/0004-6361/200810657`. arXiv: `0807.3651` (cit. on p. 42).

The Dark Energy Survey Collaboration (Oct. 2005). "The Dark Energy Survey". In: *ArXiv Astrophysics e-prints*. arXiv: `astro-ph/0510346` (cit. on pp. 28, 56).

Tinker, J., A. V. Kravtsov, A. Klypin, K. Abazajian, M. S. Warren, G. Yepes, S. Gottlöber, and D. E. Holz (Dec. 2008). "Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality". In: *ApJ* 688, pp. 709–728. DOI: `10.1086/591439`. arXiv: `0803.2706` (cit. on pp. 87, 95, 146, 172, 179).

Tinker, J., B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yepes, and S. Gottlöber (Dec. 2010). "The Large-scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests". In: *ApJ* 724, pp. 878–886. DOI: `10.1088/0004-637X/724/2/878`. arXiv: `1001.3162 [astro-ph.CO]` (cit. on p. 179).

Towns, J. (Oct. 2011). "Evolving from TeraGrid to XSEDE". In: *APS Southeastern Section Meeting Abstracts*, p. C2 (cit. on p. 53).

Vikhlinin, A. et al. (Feb. 2009). "Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints". In: *ApJ* 692, pp. 1060–1074. DOI: `10.1088/0004-637X/692/2/1060`. arXiv: `0812.2720` (cit. on pp. 32, 86).

Voges, W. et al. (Sept. 1999). "The ROSAT all-sky survey bright source catalogue". In: *A&A* 349, pp. 389–405. arXiv: `astro-ph/9909315` (cit. on p. 124).

Voit, G. M. (Apr. 2005). "Tracing cosmic evolution with clusters of galaxies". In: *Reviews of Modern Physics* 77, pp. 207–258. DOI: `10.1103/RevModPhys.77.207`. arXiv: `astro-ph/0410173` (cit. on pp. 29, 86, 140).

von Hoerner, S. (1960). "Die numerische Integration des n-Körper-Problemes für Sternhaufen. I (German) [Numerical integration of n-body problems for star clusters]". In: *ZAp* 50, pp. 184–214. SAO–NASA ADS: `1960ZA.....50..184V` (cit. on pp. 40, 42).

Walker, A. G. (1937). "On Milne's Theory of World-Structure". In: *Proceedings of the London Mathematical Society* s2-42.1, pp. 90–127. DOI: `10.1112/plms/s2-42.1.90` (cit. on p. 18).

Wambsganss, J., P. Bode, and J. P. Ostriker (May 2004). "Giant Arc Statistics in Concord with a Concordance Lambda Cold Dark Matter Universe". In: *ApJ* 606, pp. L93–L96. DOI: `10.1086/421459`. arXiv: `astro-ph/0306088` (cit. on p. 42).

Warren, M. S., K. Abazajian, D. E. Holz, and L. Teodoro (Aug. 2006). "Precision Determination of the Mass Function of Dark Matter Halos". In: *ApJ* 646, pp. 881–885. DOI: `10.1086/504962`. arXiv: `astro-ph/0506395` (cit. on pp. 87, 146).

Warren, M. S., P. J. Quinn, J. K. Salmon, and W. H. Zurek (Nov. 1992). "Dark halos formed via dissipationless collapse. I - Shapes and alignment of angular momentum". In: *ApJ* 399, pp. 405–425. DOI: `10.1086/171937`. SAO–NASA ADS: `1992ApJ...399..405W` (cit. on p. 42).

Wechsler, R. H. (June 2001). "Dark Halo Merging and Galaxy Formation". PhD thesis. University of California, Santa Cruz. URL: `http://search.proquest.com/docview/275853759` (visited on Apr. 4, 2013) (cit. on pp. 159, 160).

Wechsler, R. H., J. S. Bullock, J. R. Primack, A. V. Kravtsov, and A. Dekel (Mar. 2002). "Concentrations of Dark Halos from Their Assembly Histories". In: *ApJ* 568, pp. 52–70. DOI: `10.1086/338765`. arXiv: `astro-ph/0108151` (cit. on p. 160).

Weinberg, D. H., M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Rozo (Jan. 2012). "Observational Probes of Cosmic Acceleration". In: *ArXiv e-prints*. arXiv: `1201.2434 [astro-ph.CO]` (cit. on p. 140).

Weller, J., R. A. Battye, and R. Kneissl (June 2002). "Constraining Dark Energy with Sunyaev-Zel'dovich Cluster Surveys". In: *Physical Review Letters* 88.23, 231301. DOI: `10.1103/PhysRevLett.88.231301`. arXiv: `arXiv:astro-ph/0110353` (cit. on pp. 86, 91).

Wen, Z. L., J. L. Han, and F. S. Liu (Aug. 2009). "Galaxy Clusters Identified from the SDSS DR6 and Their Properties". In: *ApJS* 183, pp. 197–213. DOI: `10.1088/0067-0049/183/2/197`. arXiv: `0906.0803 [astro-ph.CO]` (cit. on pp. 89, 107).

Wetzel, A. R. and M. White (Apr. 2010). "What determines satellite galaxy disruption?" In: *MNRAS* 403, pp. 1072–1088. DOI: `10.1111/j.1365-2966.2009.16191.x`. arXiv: `0907.0702 [astro-ph.CO]` (cit. on p. 161).

White, M., J. D. Cohn, and R. Smit (Nov. 2010). "Cluster galaxy dynamics and the effects of large-scale environment". In: *MNRAS* 408, pp. 1818–1834. DOI: `10.1111/j.1365-2966.2010.17248.x`. arXiv: `1005.3022 [astro-ph.CO]` (cit. on p. 88).

White, M., L. Hernquist, and V. Springel (Nov. 2002). "Simulating the Sunyaev-Zeldovich Effect(s): Including Radiative Cooling and Energy Injection by Galactic Winds". In: *ApJ* 579, pp. 16–22. DOI: `10.1086/342756`. arXiv: `astro-ph/0205437` (cit. on p. 87).

White, M. et al. (Feb. 2011). “The Clustering of Massive Galaxies at z ∼0.5 from the First Semester of BOSS Data”. In: *ApJ* 728, 126, p. 126. DOI: `10.1088/0004-637X/728/2/126`. arXiv: `1010.4915 [astro-ph.CO]` (cit. on p. 88).

White, S. D. M. (Dec. 1976). “The dynamics of rich clusters of galaxies”. In: *MNRAS* 177, pp. 717–733. SAO–NASA ADS: `1976MNRAS.177..717W` (cit. on pp. 41, 42).

White, S. D. M., C. S. Frenk, M. Davis, and G. Efstathiou (Feb. 1987). “Clusters, filaments, and voids in a universe dominated by cold dark matter”. In: *ApJ* 313, pp. 505–516. DOI: `10.1086/164990`. SAO–NASA ADS: `1987ApJ...313..505W` (cit. on p. 42).

White, S. D. M., J. F. Navarro, A. E. Evrard, and C. S. Frenk (Dec. 1993). “The baryon content of galaxy clusters: a challenge to cosmological orthodoxy”. In: *Nature* 366, pp. 429–433. DOI: `10.1038/366429a0` (cit. on p. 24).

Wootten, A. (Feb. 2003). “Atacama Large Millimeter Array (ALMA)”. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Ed. by J. M. Oschmann and L. M. Stepp. Vol. 4837. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pp. 110–118. DOI: `10.1117/12.456705`. arXiv: `astro-ph/0209630` (cit. on p. 28).

Wright, T. (1750). *An Original Theory or New Hypothesis of the Universe*. London, UK: H. Chapelle. Google Books: `PI8LAQAAIAAJ` (cit. on p. 10).

Wu, H.-Y., E. Rozo, and R. H. Wechsler (Dec. 2008). “The Effects of Halo Assembly Bias on Self-Calibration in Galaxy Cluster Surveys”. In: *ApJ* 688, pp. 729–741. DOI: `10.1086/591929`. arXiv: `0803.1491` (cit. on p. 98).

Wu, H.-Y., A. R. Zentner, and R. H. Wechsler (Apr. 2010). “The Impact of Theoretical Uncertainties in the Halo Mass Function and Halo Bias on Precision Cosmology”. In: *ApJ* 713, pp. 856–864. DOI: `10.1088/0004-637X/713/2/856`. arXiv: `0910.3668 [astro-ph.CO]` (cit. on p. 81).

Xu, G. (May 1995). “A New Parallel N-Body Gravity Solver: TPM”. In: *ApJS* 98, p. 355. DOI: `10.1086/192166`. arXiv: `astro-ph/9409021` (cit. on p. 43).

Yang, H.-Y. K., S. Bhattacharya, and P. M. Ricker (Dec. 2010). “The Impact of Cluster Structure and Dynamical State on Scatter in the Sunyaev-Zel’dovich Flux-mass Relation”. In: *ApJ* 725, pp. 1124–1136. DOI: `10.1088/0004-637X/725/1/1124`. arXiv: `1010.0249 [astro-ph.CO]` (cit. on p. 87).

Yang, H.-Y. K., P. M. Ricker, and P. M. Sutter (July 2009). “The Influence of Concentration and Dynamical State on Scatter in the Galaxy Cluster Mass-Temperature Relation”. In: *ApJ* 699, pp. 315–329. DOI: `10.1088/0004-637X/699/1/315`. arXiv: `0808.4099` (cit. on p. 87).

Zehavi, I. et al. (July 2011). “Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity”. In: *ApJ* 736, 59, p. 59. DOI:

`10.1088/0004-637X/736/1/59`. arXiv: `1005.2413 [astro-ph.CO]` (cit. on p. 116).

Zel'dovich, Y. B. (Mar. 1970). "Gravitational instability: An approximate theory for large density perturbations." In: *A&A* 5, pp. 84–89. SAO–NASA ADS: `1970A%26A.....5...84Z` (cit. on p. 149).

Zheng, Z., I. Zehavi, D. J. Eisenstein, D. H. Weinberg, and Y. P. Jing (Dec. 2009). "Halo Occupation Distribution Modeling of Clustering of Luminous Red Galaxies". In: *ApJ* 707, pp. 554–572. DOI: `10.1088/0004-637X/707/1/554`. arXiv: `0809.1868` (cit. on p. 88).

Zwicky, F. (1933). "Die Rotverschiebung von extragalaktischen Nebeln (German) [The redshift of extragalactic nebulae]". In: *Helvetica Physica Acta* 6, pp. 110–127. SAO–NASA ADS: `1933AcHPh...6..110Z` (cit. on pp. 14, 29).

— (Oct. 1937). "On the Masses of Nebulae and of Clusters of Nebulae". In: *ApJ* 86, p. 217. DOI: `10.1086/143864` (cit. on pp. 14, 29).

Zwicky, F., E. Herzog, P. Wild, M. Karpowicz, and C. T. Kowal (1961–1968). *Catalogue of galaxies and of clusters of galaxies, 6 Vols.* Pasadena: California Institute of Technology (cit. on pp. 17, 25).