Probing the New Cosmology

DISSESTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Andrew Ronald Zentner, B.S.E.E.

* * * * *

The Ohio State University

2003

Dissertation Committee:
Professor Terrence P. Walker, Adviser
Professor Robert J. Scherrer
Professor Samir D. Mathur
Professor Michael A. Lisa
Professor Edward A. Overman

Approved by

______________________________
Adviser
Department of Physics
ABSTRACT

Improvements in observational techniques have transformed cosmology into a field inundated with ever-expanding, high-quality data sets and driven cosmology toward a standard model where the classic cosmological parameters are accurately measured. I briefly discuss some of the methods used to determine cosmological parameters, particularly primordial nucleosynthesis, the magnitude-redshift relation of supernovae, and cosmic microwave background anisotropy. I demonstrate how cosmological data can be used to complement particle physics and constrain extensions to the Standard Model. Specifically, I present bounds on light particle species and the properties of unstable, weakly-interacting, massive particles.

Despite the myriad successes of the emerging standard cosmological model, unanswered questions linger. Numerical simulations of structure formation predict galactic central densities that are considerably higher than observed. They also reveal hundreds of satellites orbiting Milky Way-like galaxies while the Milky Way has only eleven known satellites within 300 kpc. I explore the possibility that these conundrums may have a common remedy in the form of the power spectrum of initial density fluctuations that seed structure growth. To address the substructure issue, I develop a semi-analytic method that suffers from no inherent resolution limits and can therefore be used to complement numerical simulations.
I find that tilted initial power spectra and spectra with running tilts provide for an intriguing possibility. In these models, the amplitude of initial fluctuations can be normalized against cosmic microwave background measurements on large scales. Yet, the reduction in small-scale power brings galactic central densities down to acceptable levels and allows the Milky Way satellite population to be accounted for without invoking differential feedback mechanisms. Furthermore, substructure mass fractions are not significantly altered in these models so probes of substructure via gravitational lensing do not disfavor them. The primordial fluctuations are thought to be generated during an early epoch of inflation and one implication is that galaxy properties may convey information about inflation. I also address alternative proposals, such as warm dark matter and broken scale-invariant inflation, in light of lensing probes of substructure and find these models to be disfavored. I close with a few words on refining the model and alternative applications.
To my brother, Christopher, and my sisters, Shanna and Camille.
ACKNOWLEDGMENTS

I am grateful to my advisor, Terry Walker, for guidance, ideas, and helpful discussions and to his son, Cameron, whose dedication to hallway football provided a perfect break from physics. I am equally indebted to James “Daddy” Bullock, who has been nothing short of a second advisor to me. I am grateful to Gary Steigman for his invaluable help and guidance as I learned to be a researcher. I would also like to thank Stuart Raby, Bob Scherrrer, and David Weinberg for many helpful conversations.

I thank the occupants of 4012 Smith Lab for putting up with the shenanigans that I am wont to instigate, and more importantly, I thank Louis Strigari and Savvas Koushiappas for participating in those shenanigans. I also thank Louis, Savvas, Michael Schirber, and James Kneller for regular attendance at Friday Happy Hours, for tons of patience and help, and for allowing me to bounce ideas off of them.

Most of all, I am grateful to Carrie. Without her love, support, patience, entertainment, and computer lessons this work would certainly not have been possible.

I thank my collaborators Terry Walker, Gary Steigman, James Bullock, and Savvas Koushiappas for allowing me to present our work in this dissertation. I thank James Kneller for providing Figure 3.3 and Andrey Kravtsov for Figure 1.4. I am grateful to James Bullock and Kathryn Johnston for allowing me to include preliminary results in Figures 6.28 and 6.29. I thank Rob Swaters for providing me with his unpublished rotation curve data [291]. I acknowledge the use of CMBFAST [262].
VITA

February 24, 1976 ................. Born - New York, NY, USA

1998 .................................. B.S. Electrical Engineering,
The Cooper Union for the Advancement of Science and Art
New York, NY, USA

1998-2001 ............................. William A. Fowler Fellow,
Department of Physics,
The Ohio State University
Columbus, OH, USA

1999-2000 ............................. Graduate Teaching Associate,
Department of Physics,
The Ohio State University
Columbus, OH, USA

2001-2002 ............................. Graduate Research Associate,
Department of Physics,
The Ohio State University
Columbus, OH, USA

2002-present ......................... Distinguished University Fellow,
The Ohio State University
Columbus, OH, USA

PUBLICATIONS

Research Publications


Major Field: Physics

Studies in Theoretical Astrophysics and Cosmology: Professor Terrence P. Walker
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction and Summary</td>
<td>1</td>
</tr>
<tr>
<td>2. Toward A Standard Cosmology</td>
<td>17</td>
</tr>
<tr>
<td>2.1 Big Bang Nucleosynthesis and The Baryon Density</td>
<td>18</td>
</tr>
<tr>
<td>2.2 The Magnitude-Redshift Relation of Supernovae</td>
<td>22</td>
</tr>
<tr>
<td>2.3 The Growth of Large-Scale Structure and The Matter Density of the</td>
<td>27</td>
</tr>
<tr>
<td>Universe</td>
<td></td>
</tr>
<tr>
<td>2.3.1 The Power Spectrum</td>
<td>27</td>
</tr>
<tr>
<td>2.3.2 The Behavior of Density Perturbations</td>
<td>30</td>
</tr>
<tr>
<td>2.3.3 Measuring the Mass Density With the Power Spectrum</td>
<td>34</td>
</tr>
<tr>
<td>2.4 The Cosmic Microwave Background</td>
<td>38</td>
</tr>
<tr>
<td>2.4.1 The Angular Power Spectrum</td>
<td>39</td>
</tr>
<tr>
<td>2.4.2 Anisotropy on Large Scales: The Sachs-Wolfe Plateau</td>
<td>40</td>
</tr>
<tr>
<td>2.4.3 Small-Scale Fluctuations: Acoustic Peaks</td>
<td>45</td>
</tr>
<tr>
<td>2.4.4 CMB Power Spectra and Cosmological Parameters: A Pictorial</td>
<td>49</td>
</tr>
<tr>
<td>2.5 The Parameters of a Standard Cosmology</td>
<td>54</td>
</tr>
</tbody>
</table>
3. Cosmological Constraints on Relativistic and Unstable Particle Species . 64
   3.1 Introduction ....................................................... 64
   3.2 Constraints on Relativistic Energy ............................... 65
       3.2.1 Big Bang Nucleosynthesis Constraint ....................... 65
       3.2.2 The CMB Constraint ...................................... 72
       3.2.3 Constraints from Type Ia Supernovae ......................... 75
   3.4 Large Scale Structure Constraints ............................... 80
   3.3 Constraining Relic Decays ..................................... 83
       3.3.1 CMB Constraints on Relic Properties ....................... 85
       3.3.2 Large-scale Structure Constraints on Relic Decays ........... 87
       3.3.3 SN1a Constraint on Relic Decays .......................... 91
   3.4 Discussion and Conclusions .................................. 91

4. Problems with Cold Dark Matter on Galactic and Sub-galactic Scales . 93
   4.1 Introduction ...................................................... 93
   4.2 The Structure of Dark Matter Halos ............................. 94
   4.3 The Cuspy Halo Issue .......................................... 98
   4.4 The Concentration Problem and the Central Density Problem ... 103
   4.5 The Dwarf Satellite Problem .................................. 107

5. Inflation, Cold Dark Matter, and Galactic Central Densities ............ 110
   5.1 Introduction ...................................................... 110
   5.2 Inflationary Power Spectra .................................... 117
       5.2.1 Inverted Power Law Potentials ............................. 120
       5.2.2 Running-mass Inflation .................................. 122
       5.2.3 Broken Scale-Invariance in the Primordial Power Spectrum . 125
   5.2.4 The Power Spectrum of WMAP ................................ 127
   5.3 Massive Neutrinos and Hot Dark Matter .......................... 128
   5.4 The Central Densities of Dark Matter Halos ..................... 130
   5.5 Results .......................................................... 134
   5.6 Central Densities and the Power Spectrum: Conclusions and Discussi
       on ............................................................. 143

6. Dark Matter Halo Substructure and the Primordial Power Spectrum .... 150
   6.1 Halo Substructure: Introduction and Background ................ 150
   6.2 Modeling Halo Substructure ................................... 162
       6.2.1 Overview .................................................. 162
Appendices:

A. Robertson-Walker Metric and Kinematics .......................... 248

B. The Friedmann-Robertson-Walker Model and the Standard Cosmological Parameters .......................... 254

Bibliography ............................................................. 257
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>WMAP Best-fit parameters</td>
</tr>
<tr>
<td>6.1</td>
<td>Properties of various primordial power spectra</td>
</tr>
<tr>
<td>6.2</td>
<td>Properties of the satellite galaxies of the Milky Way</td>
</tr>
<tr>
<td>6.3</td>
<td>The maximum circular velocities of the halos of the Milky Way satellites assuming isotropic velocities</td>
</tr>
<tr>
<td>6.4</td>
<td>The maximum circular velocities of the halos of the Milky Way satellites with anisotropic stellar velocities</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Two-dimensional analogues of the open, closed and flat geometries of the three-dimensional spatial hypersurfaces of FRW models.</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Evidence for dark matter in galactic dynamics.</td>
<td>5</td>
</tr>
<tr>
<td>1.3 The central density problem illustrated by the observed rotation curve of F583-1.</td>
<td>11</td>
</tr>
<tr>
<td>1.4 A pair of Local Group-like halos being orbited by hundreds of smaller satellite halos in an N-body simulation.</td>
<td>13</td>
</tr>
<tr>
<td>2.1 Big Bang Nucleosynthesis Abundance Predictions vs. Observations</td>
<td>21</td>
</tr>
<tr>
<td>2.2 The high-redshift portion of the magnitude-redshift relation for the data of the Supernova Cosmology Project.</td>
<td>25</td>
</tr>
<tr>
<td>2.3 Confidence limits on cosmological parameters based on observations of high-redshift supernovae</td>
<td>26</td>
</tr>
<tr>
<td>2.4 Joint confidence contours on cosmological parameters based on high-redshift supernovae and the universal baryon fraction.</td>
<td>28</td>
</tr>
<tr>
<td>2.5 The power spectrum of galaxy clustering as a probe of the universal matter density.</td>
<td>36</td>
</tr>
<tr>
<td>2.6 The sensitivity of CMB power spectra to ( \Omega_{M,0} ).</td>
<td>51</td>
</tr>
<tr>
<td>2.7 The sensitivity of CMB power spectra to ( \Omega_{b,0} h^2 ).</td>
<td>52</td>
</tr>
<tr>
<td>2.8 The problem of degeneracy in the CMB power spectrum.</td>
<td>53</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>2.9</td>
<td>The sensitivity of CMB acoustic peak positions to the spatial curvature of the Universe.</td>
</tr>
<tr>
<td>2.10</td>
<td>The CMB power spectrum measured by WMAP along with the best-fit models with power law primordial power spectra.</td>
</tr>
<tr>
<td>2.11</td>
<td>The WMAP observed temperature-polarization power spectrum and the WMAP best-fit models.</td>
</tr>
<tr>
<td>2.12</td>
<td>The CMB power spectrum measured by WMAP along with the best-fit models with non-zero running of the spectral index.</td>
</tr>
<tr>
<td>3.1</td>
<td>The big bang nucleosynthesis bound on the abundance of relativistic species in the Universe.</td>
</tr>
<tr>
<td>3.2</td>
<td>A summary of constraints on the relativistic energy density of the Universe at various epochs.</td>
</tr>
<tr>
<td>3.3</td>
<td>Neutrino degeneracy provides a loophole for any BBN constraint on the relativistic energy density of the Universe.</td>
</tr>
<tr>
<td>3.4</td>
<td>The CMB power spectrum and the relativistic energy density of the Universe at decoupling.</td>
</tr>
<tr>
<td>3.5</td>
<td>Constraints on relativistic energy from the magnitude-redshift relation of high-redshift, type Ia supernovae.</td>
</tr>
<tr>
<td>3.6</td>
<td>The high-redshift portion of the magnitude-redshift relation and SNIa constraints on relativistic energy.</td>
</tr>
<tr>
<td>3.7</td>
<td>Constraints on massive relic properties in the $M_{keV} - \tau_{yr}$ plane.</td>
</tr>
<tr>
<td>4.1</td>
<td>The observed rotation curve of ESO 2060140: cusp vs. core.</td>
</tr>
<tr>
<td>4.2</td>
<td>The best-fit inner profile slopes for sixty dwarf and low surface-brightness galaxies.</td>
</tr>
<tr>
<td>4.3</td>
<td>The $\chi^2$ per degree of freedom of rotation curve fits to NFW and pseudo-isothermal profiles provides another illustration of the cuspy halo problem.</td>
</tr>
</tbody>
</table>
4.4 The best-fit values of \( c_{\text{vir}} \) for sixty observed rotation curves. ........ 104
4.5 The central densities of dark matter halos: theory vs. observation. ... 106
4.6 The observed velocity function of satellites of the Milky Way compared with the predicted number of CDM halos with the MW virial radius. 108
5.1 The rms overdensity as a function of mass scale \( \sigma(M) \), for several inflation-derived primordial power spectra. ......................... 128
5.2 The effect of neutrino hot dark matter on the power spectrum of CDM density perturbations ............................................................... 131
5.3 The \( c_{\text{vir}}-M_{\text{vir}} \) relation as a function of the primordial power spectrum in ΛCDM cosmologies. .......................... 138
5.4 The \( c_{\text{vir}}-M_{\text{vir}} \) relation in cold + hot dark matter cosmologies. ........... 139
5.5 The distribution of \( c_{\text{vir}} \) in the \( \sigma_8 = 0.75 \) model. .................. 140
5.6 The distribution of \( c_{\text{vir}} \) in the \( \frac{dn}{d\ln k} = -0.03 \) running index (RI) model. .................................................. 141
5.7 The distribution of \( c_{\text{vir}} \) in the broken scale-invariant (BSI) inflation model. .................................................. 142
5.8 \( V_{\text{max}} \) vs. \( \Delta_{V/2} \) predictions compared with observations for several primordial power spectra. ............................................. 143
5.9 \( V_{\text{max}} \) vs. \( \Delta_{V/2} \) predictions in hot + cold dark matter models compared to observational data ...................... 144
6.1 Reionization feedback as a mechanism for solving the dwarf satellite problem. ........................................ 159
6.2 The rotation curves of two halos that lead to the same observed line-of-sight central velocity dispersion for Carina. ..................... 161
6.3 Two results for halo concentrations. ........................................ 174
6.4 The probability distribution of the orbital circularities of merging and surviving substructure. ........................................ 178
6.5 The effect of mass and concentration on the orbits and mass loss rates of satellite halos. ........................................ 185
6.6 The velocity function predicted by semi-analytic models vs. the velocity function observed in N-body simulations. ................. 186
6.7 The radial distribution of substructure in a Galaxy-sized halo. ....... 188
6.8 Inflation motivated power spectra........................................ 193
6.9 Warm dark matter power spectra......................................... 194
6.10 The distribution of accreted halos with mass.......................... 196
6.11 The mass accretion rates of halos that grow to be approximately Milky Way-sized by $z = 0$. ................................. 198
6.12 The cumulative mass function $N(> M_{\text{mt}})$ of surviving subhalos as a function of the primordial power spectrum. ............ 200
6.13 The cumulative velocity function $N(> V_{\text{max}})$ of satellite halos of a fixed mass host as a function of the power spectrum. ........... 201
6.14 The cumulative velocity function of satellite halos of a host of fixed maximum circular velocity as a function of the power spectrum. ... 203
6.15 Differential mass fractions in substructure and the self-similar nature of substructure distributions as a function of host halo mass. .... 204
6.16 The mass fraction of dark matter halos bound up in substructure. . 206
6.17 The mass fraction in substructure projected along a halocentric cylinder of radius 10 kpc. ................................................. 207
6.18 Spherically averaged mass fraction profiles compared to mass fraction profiles in two-dimensional projection. ......................... 208
6.19 Projected substructure mass fraction profiles as a function of mass cut in standard $\Lambda$CDM. .................................................. 209

6.20 Projected substructure mass fraction profiles as a function of mass cut in the broken scale-invariant model. ............................... 210

6.21 The effect of warm dark matter-like power spectra on the total mass fraction in substructure. .............................. 216

6.22 Projected mass fractions and warm dark matter. ................ 218

6.23 The effect of WDM phase space constraints on substructure mass fractions. ............................................................. 219

6.24 Comparing measured velocity dispersion to the maximum circular velocity of a dark matter halo. ................................. 224

6.25 Observed and predicted satellite halo velocity functions assuming isotropic velocity ellipsoids. .............................. 227

6.26 Observed and predicted satellite halo velocity functions assuming an anisotropy parameter of $\beta = 0.15$. ......................... 230

6.27 The number of Milky Way subhalos directly detectable via neutralino annihilation. .................................................. 234

6.28 Testing the analytic code using idealized N-body simulations. ................................................................. 238

6.29 Tidal stripping at pericenter: improvements to the analytic model. ........................ 240
CHAPTER 1

INTRODUCTION AND SUMMARY

Cosmology is the study of the structure, origin, and evolution of the Universe that we inhabit. Over the last twenty years, cosmology has developed significantly into a mature science. The quality of cosmological data has improved dramatically with the consequence that cosmology is now a data-driven field and much less a forum for open speculation. The so-called “era of precision cosmology” [259, 307] has seen cosmology thrust towards a very successful and highly predictive standard or “concordance” model with remarkable rapidity [228, 301]. However, with new data come unforeseen quandaries and new questions to be answered.

Cosmology begins with the cosmological principle, which states that the Universe is homogeneous and isotropic when averaged over large enough scales. The gross evolution of the Universe is determined by its mean energy density and pressure. Einstein’s general relativity is the theory of gravity that provides the link between the contents of the Universe and the dynamics of space and time. The Universe today is observed to be expanding and the basic premise of cosmology is that Einstein’s equations can be used to trace the expansion history of the Universe backward and/or forward in time. As we shall see, there is good reason to believe that the history of
our expanding Universe can be traced back to a time ~ 14 billion years ago, when the Universe was merely one ten-billionth its present size, rather reliably.

Perhaps the most fundamental questions of cosmology pertain to the energy density and spatial geometry of the Universe. These quantities are inextricably intertwined: if the energy density is equal to the so-called critical density $\rho_{\text{crit}}$, the geometry of the Universe is flat and ordinary Euclidean geometry is valid. Contrarily, if the density of the Universe is greater than the critical density, the Universe is “closed” and has finite volume, while if it is less than the critical density, the Universe is “open” or hyperbolic. In Figure 1.1, I show two dimensional analogues of these three options for the three-dimensional geometry of the Universe. The critical density is determined in terms of the expansion rate of the Universe today, or the Hubble constant $H_0$. The Hubble constant has been measured to be $H_0 = 72 \pm 8$ km s$^{-1}$ Mpc$^{-1}$ [109] but it is standard practice to express uncertainty in the Hubble constant by writing $H_0 = 100h$ km s$^{-1}$ Mpc$^{-1}$ (so $h \approx 0.72$). The critical density is then $\rho_{\text{crit}} \approx 1.9h^2 \times 10^{-29}$ g cm$^{-3}$, roughly six protons per cubic meter or about three bottles of beer per hundred earth volumes.

The critical density is therefore a natural density scale in cosmology and it is conventional to denote the density of any species in units of the critical density. For instance, the density species “$i$”, $\rho_i$, is typically written as $\Omega_i$ and the cosmological densities of matter $\Omega_{M,0}$, radiation $\Omega_{R,0}$, and “vacuum energy” or cosmological constant $\Omega_{\Lambda,0}$ are typically expressed in this way. The subscript “0” indicates that the density is the value measured today. A flat Universe then requires that the sum of all energy densities equal the critical density or $\Omega_{\text{tot}} = \sum_i \Omega_i = 1$. 

2
Figure 1.1: Two-dimensional analogues of the three-dimensional spatial hypersurfaces of FRW models. The flat geometry is represented at the left, the open geometry (resembling a saddle) in the center, and the closed geometry at the right.

The geometry of the Universe can be measured fairly robustly. The Universe is permeated by a nearly uniform bath of thermal radiation at a temperature of $T = 2.726$ K [201]. This thermal bath is a directly observable relic of the hot, early Universe and its isotropy to a few parts in $10^5$ is a striking observational confirmation of the assumption of isotropy. At $T \approx 2.7$ K, the radiation itself contributes a negligible amount to the critical density, $\Omega_{R,0} \sim 10^{-5} - 10^{-4}$. This cosmic microwave background (CMB) radiation is not perfectly isotropic. Anisotropies in the observed CMB yield information about the small, primeval inhomogeneities that seeded the formation of all the structures we observe today. As I discuss in Chapter 2, the angular pattern of anisotropy in the CMB is a robust probe of the geometry of the Universe, and measurements of this pattern indicate that the Universe is flat, $\Omega_{\text{tot}} \simeq 1$, to within a few percent [274]. The total energy density of the Universe is very nearly equal to the critical density.

Do protons, neutrons, and all the baryonic matter of everyday life account for the critical density? This question can be addressed using the theory of big bang
nucleosynthesis (BBN). The Universe 14 billion years ago was a hot, dense sea of photons and neutrinos with a sprinkling of protons and neutrons. As the Universe expanded and cooled, many of the protons and neutrons bound into light nuclei. Almost all neutrons were incorporated into $^4$He so that about 25% of all baryons were in $^4$He after BBN, making it the most common nucleus aside from the single proton of hydrogen. Trace amounts of D, $^3$He, $^7$Li and other elements were also formed. In the standard model of BBN, the light nuclide yields are a function of only one parameter, the baryon density. Observations of $^4$He and D abundances indicate that the baryon density is only 4% of the critical density, $\Omega_b,0 \approx 0.04$ [225]. Independent estimates of the baryon density from CMB analyses confirm this result [274], while inventories of luminous baryons count only $\sim 1/10$ of the BBN estimate [111]. Only about 4% of the energy density of the Universe is in the form of commonplace, baryonic matter.

Evidence for non-baryonic dark matter has been known for some time. Zwicky gave the first evidence when he pointed out that the relative velocities of galaxies in the Coma cluster are too large to be accounted for by the gravitational pull of the luminous matter only [341]. Almost forty years later, Rubin and Ford found that the rotational speed of gas in nearby Andromeda is too large to be accounted for by only the luminous matter in the galaxy [253]. Now, evidence for dark matter in the rotation speeds of hydrogen gas is ubiquitous. In particular, rotation speeds seen at radial distances that contain most of the light of the galaxy do not follow the Newtonian expectation of $v^2 \propto 1/r$. Rather, rotation speeds seem to obey $v \propto$ constant implying that the mass enclosed by a given radius increases as $M \propto r$ well beyond the radius that encloses most of the light. An example of such a rotation curve is shown in Figure 1.2. As is evident in Figure 1.2, the light from the baryons alone cannot account for
Figure 1.2: Measurements of the rotation speeds of gas in the galaxy F571-8 by de Blok, McGaugh, and Rubin [88] (points with error bars). Also shown are the best fit models to the full rotation curve (solid) and the estimated contribution from baryons assuming a reasonable and constant ratio of mass per unit luminosity. Baryons alone cannot account for the observed rotation speeds unless a large, radially-dependent ratio of baryonic mass-to-luminosity is assumed.

the dynamics of the gas at large radii. The galaxy seems to be surrounded by a much more massive cloud or “halo” of “dark matter.”

The existence of dark matter is now almost unimpeachable and dark matter constitutes one of the key elements of the modern cosmological model. After the initial work on rotation curves, it was quickly realized that dark matter could help to aid the growth of structure and galaxies [235, 330] and stabilize galactic disks against an instability that would tend to drive the matter of the disk into a bar shape [227]. As I discuss in Chapter 2, there are many methods that can be employed to obtain an estimate of the matter density of the Universe. The abundance of galaxy clusters as
a function of redshift [16], the baryon fraction of clusters (coupled with knowledge of \( \Omega_{b,0} \) from BBN) [329], the coherent distortion of galaxy images due to lensing [314, 146], and the statistics of galaxy clustering all point toward a total matter density in the Universe of \( \Omega_{M,0} \sim 0.3 \). The baryons contribute to \( \Omega_{M,0} \) but this leaves 26% of the energy density of the Universe in the form of some unknown, massive species.

The energy budget of the Universe is then balanced by \( \sim 4\% \) normal, baryonic matter, about 26% of some unknown form of weakly-interacting “dark” matter, and \( \sim 70\% \) something even more exotic. What are the properties of this exotic “dark energy”? One strange property can be determined by examining the light from “standard candles.” A standard candle is an object of known intrinsic luminosity which is useful because given the flux observed at earth and the intrinsic luminosity of the object, one can determine its distance. Over very large distances, this relationship between observed flux and distance is not so trivial. Photons “redshift” as the Universe expands (see Appendix A) so that the observed wavelength of a photon is a factor \( (1 + z) \) longer than when it was emitted. The factor \( (1 + z) \) is the amount by which length scales in the Universe have expanded since the emission of the photon and the quantity \( z \) is called the “cosmological redshift.” The redshift of an object can be determined by observing the shifts in well-known spectral lines relative to their wavelengths measured in the laboratory. The mapping between redshift and the observed flux of an object of known luminosity is a function of the geometry and dynamics of the Universe and thus, observations of standard candles at varied redshifts can be used to constrain the properties of the universal energy density.
A particular type of supernovae (type Ia) have been found to be dependable standard candles and two collaborations have embarked on ventures to measure the expansion history of the Universe by observing these objects [239, 258]. These groups determine that the expansion of the Universe is accelerating. The measurement of an accelerated expansion was received with astonishment by the cosmological community, the implication is that the dark energy obeys an equation of state that relates its energy density $\rho$, to its pressure $P$, of $\rho = wP$, where $w < -1/3$. The dark energy has negative pressure, or in other words, tension. If $w = -1$, the dark energy has the same effect dynamically as the cosmological constant, $\Lambda$, that Einstein invoked in order to enforce a static Universe. This is the simplest hypothesis and is consistent with all available cosmological data, so cosmologists are wont to say that $\Omega_{\Lambda,0} = 0.7$; however, alternatives have been proposed such as the idea that dark energy may be the potential energy of some unknown scalar field [249].

Measurements of the cosmological parameters using many independent techniques that converge on compatible values constitute a tremendous achievement and a great triumph for the framework of the standard cosmology. In the standard model, the contemporary Universe is dominated by a cosmological constant ($\Lambda$) and cold dark matter (CDM). Consequently the standard model has been dubbed the $\Lambda$CDM cosmological model. In Chapter 2, I give a brief review of some of the standard probes of cosmology. This Chapter serves to motivate the standard model that will be assumed throughout much of the remainder of this Thesis and to introduce notation and basic concepts. Introductory material on the Friedmann-Robertson-Walker (FRW) cosmological model can be found in Appendices A and B.
These measurements pose two profound challenges to particle physicists and cosmologists, namely, to determine the nature and properties of the dark matter and the dark energy. Nevertheless, these cosmological achievements also complement particle physics. Having inventoried the contents of the Universe, it is possible to determine what room is left for as yet undetected particle species. These constraints can be applied to extensions of the standard model of particle physics and provide a non-trivial set of boundaries that particle theorists must work within in order to build consistent theories. In Chapter 3, I discuss a set of cosmological bounds on any relativistic species that may be added to the Standard Model of particle physics, such as extra light neutrinos, and anything that may mimic relativistic energy, such as the dark radiation of some extra-dimensional brane cosmologies [248, 79, 32, 121]. In this Chapter, I also discuss constraints on the properties of unstable, massive particles that decay into relativistic species after being produced in the early Universe. I present the bounds in a way that does not specify the interactions of the particle and is therefore applicable to any Standard Model extension.

With robust estimates of the fundamental cosmological parameters in place it is possible to compute the expected pattern of galaxy clustering, the anisotropy in the CMB, and to carry out detailed numerical simulations of structure formation in the context of the standard cosmology. The missing ingredient is the spectrum of fluctuations that seed structure formation. The density perturbation field is typically denoted by \( \delta(\vec{x}) = (\rho(\vec{x}) - \bar{\rho})/\bar{\rho} \), where \( \bar{\rho} \) is the mean background density. The statistical properties of the density field can be expressed in terms of the power spectrum \( P(k) \equiv |\delta_k|^2 \), where \( \delta_k \) is the Fourier transform of the density field. Toeing the standard line, the initial density fluctuations are generated during a period of extremely
rapid cosmological expansion in the very early Universe called “inflation.” After the end of inflation, the length scales that correspond to observable cosmic structure are much larger than the horizon scale, which determines the length scale on which causal physical processes may act. As the Universe expands, cosmologically observable scales become smaller than the horizon and perturbations on these scales begin to grow. A special form for the power spectrum is the “scale-invariant” form, with $P(k) \propto k$. The name scale-invariant derives from the fact that the rms overdensity on all length scales as they cross the horizon are equal, in other words, all scales enter the horizon with the same typical overdensity. Inflation generally produces an initial spectrum of density perturbations that is very nearly scale invariant, $P(k) \propto k^n$ with $n \approx 1$ [122, 123, 134, 20]. The power law index $n$ is known as the “tilt” of the primordial power spectrum. The normalization of the spectrum can be determined by observations of anisotropy in the cosmic microwave background as I discuss in the next Chapter.

Finally, with all of these ingredients, it is possible to trace the growth of cosmic structure. On large scales, where the density perturbations are small and linear perturbation theory can be used, the agreement between theory and observation is rather impressive as I've already mentioned. On smaller scales, the density perturbations are large and numerical simulations and semi-analytic approximations are the preferred tools. On these small scales, particularly galactic and sub-galactic scales, the agreement between theory and observation is not apparent. The enigma is twofold. One aspect of the problem concerns the central mass distribution of halos and the second mystery concerns the abundance of satellite galaxies around Milky Way-like
halos. I briefly introduce these problems with small-scale structure in the ΛCDM model here. I discuss these issues in much more detail in Chapter 4, 5, and 6.

The issue of central mass distributions is illustrated in Figure 1.3. The data points represent the observed rotation curve of a low surface brightness galaxy by de Blok, McGaugh, and Rubin [88]. Cosmological N-body simulations reveal a correlation between the maximum circular velocity achieved in a halo and the shape of the rotation curve [220, 48, 103]. A typical rotation curve that achieves the same maximum circular velocity in N-body simulations has the shape of the solid line in Figure 1.3. Notice that the N-body halo has its mass more strongly concentrated toward its center than the observed galaxy. This is not a peculiar case, similar behavior is prevalent among observed rotation curves (see [5, 86, 41, 205, 337]). This discrepancy is commonly referred to as the “central density” or “concentration” problem.

I tackle the central density problem in Chapter 5. I first emphasize that the paradigm of inflation does not necessarily demand that the primeval spectrum of density fluctuations that seed the growth of structure is exactly scale-invariant. Rather, inflation implies an approximately scale-invariant spectrum with primordial index $n \approx 1$, yet almost all simulations of cosmological structure growth are based on the scale-invariant assumption. I work within the context of several inflationary models that naturally predict $n < 1$ in which case the power spectrum is tilted to favor large scales (small $k$) over small scales. This is important because if the power spectrum is tilted to favor large scales, the initial fluctuations on galactic scales become correspondingly smaller. Therefore, relative to the standard scale-invariant model fluctuations on these scales require more time to grow before objects like dark matter halos can form. The delay in halo formation means that halos typically collapse and
Figure 1.3: This Figure illustrates the central density problem with observed galaxies. The data points represent the observed rotation curve of F583-1 by de Blok, McGaugh, and Rubin [88]. The solid line represents the shape of a rotation curve of a typical halo that has the same maximum circular velocity as observed in numerical simulations [220, 48, 103]. The mismatch between theory and observation is called the concentration or central density problem.

...virialize at a later time, when the Universe is less dense due to the dilution of the cosmological expansion, and halos are correspondingly less dense. In this way, the central density problem can be alleviated. In particular, I show that models with a tilt of $n \approx 0.9$ have the attractive feature that the primordial power spectrum can be normalized properly on the large scales probed by CMB anisotropy while having small enough power on galactic scales to bring the predicted central densities of dark matter halos into line with observations. I also demonstrate how models of inflation with more drastic features in their primordial spectra, such as a sudden break or a running spectral index (i.e., $dn/d\ln k \neq 0$), can have a similar effect.
The Milky Way galaxy has eleven known satellite galaxies within 300 kpc: Carina, Draco, Fornax, Leo I, Leo II, Sagittarius, Sculptor, Sextans, the Large Magellanic Cloud (LMC), the Small Magellanic Cloud (SMC), and Ursa Minor. The Milky Way’s largest companion in the Local Group, Andromeda, has thirteen similar satellites orbiting it (for a recent census of Local Group occupants, consult [200]). Figure 1.4 shows a Local Group-like clustering of galaxies observed in an N-body simulation. The two dominant galaxies, similar to the Milky Way and Andromeda, are surrounded by hundreds of smaller satellite halos. A large amount of halo substructure is a fundamental prediction of CDM. Halo masses are not directly observable so number counts are typically made as a function of the maximum circular velocities achieved within the satellite halos. In this way, maximum circular velocity serves as a measure of halo size. The “dwarf satellite problem” can then be formulated in the following way: the number of observed satellites of the Milky Way with maximum circular velocities in the range 10 – 20 km s\(^{-1}\) is roughly an order of magnitude smaller than what is expected from the results of cosmological N-body simulations.

The issue of galactic substructure is particularly relevant at present because much work is being done to measure substructure populations in distant galaxies by the gravitational lensing effect of these objects [81, 215]. A popular solution to the dwarf satellite puzzle is that a feedback mechanism, like the heating of baryons by a photo-ionizing background [302], prevents star formation in small halos so that the halos may be present, but invisible [50]. Gravitational lensing is sensitive to the mass fraction of a halo that is contained in substructure, whether or not the subhalos host luminous galaxies, so these measurements may become fundamental tests of the
Figure 1.4: A pair of halos, similar to the Milky Way and Andromeda which are the largest members of the Local Group, as observed in an N-body simulation. The cosmology is the standard ΛCDM cosmology, the box is 1h⁻¹ Mpc on a side, and color is proportional to density. Notice that the two dominant halos are being orbited by hundreds of smaller satellite halos. Simulation by A. V. Kravtsov and A. A. Klypin. Rendering by A. V. Kravtsov.

ΛCDM cosmological model. If lensing measurements uncover significant substructure mass fractions, the dwarf satellite problem becomes a matter of understanding feedback mechanisms that inhibit star formation in shallow potential wells, while if substructure is not found, the CDM paradigm will be faced with a grave challenge. Preliminary results indicate that the substructure is, indeed, there [81].

I deal with the problem of halo substructure in Chapter 6. There are four primary motivations for this work. First and foremost, this is an extension of the work on
the central density problem. By lowering small-scale power in the primordial power spectrum, it is possible to alleviate the central density problem while maintaining the overall success of the $\Lambda$CDM paradigm. However, it is important to understand the consequences of reduced small-scale power for halo substructure as lensing measurements emerge as a test of structure formation scenarios. If by reducing small-scale power substructure is also erased and the model becomes inconsistent with lensing measurements, the attractiveness of this solution is lost. Second, the population of substructure in galactic halos may serve as a diagnostic with which to test alternatives to the CDM paradigm such as “warm dark matter,” in which substructure populations have significantly different characteristics. Third, it is important to understand the nature of the dwarf satellite problem, particularly the role of feedback mechanisms in galaxy formation, in models with reduced small-scale power. Fourth, the recent analysis of the measurements of CMB anisotropy by the Wilkinson Microwave Anisotropy Probe (WMAP) team [274] indicate that the primeval power spectrum may deviate from scale-invariance with a running that reduces small-scale power $dn/d\ln k < 0$. This lends credence to the aforementioned solution to the central density problem, while conversely, exploring the properties of galactic densities and substructure and comparing them with observations is one of the few promising possibilities for probing the initial power spectrum at very large wavenumber (and thus further constraining the potential of the field that drives inflation). N-body simulations do not have the resolution to make studies of this kind feasible, so I begin Chapter 6 by developing a semi-analytic model that can be used to estimate the characteristics of substructure populations as a function of various cosmological parameters. I then apply this
model to study substructure as a function of the primordial power spectrum of density fluctuations.

I find that a number of competing effects conspire in a way that makes the substructure mass fraction of a galactic halo relatively insensitive to the tilt, $n$, and running, $dn/d\ln k$, of the primordial power spectrum. I find that current observational limits on substructure mass fractions seem to disfavor warm dark matter-like alternatives to CDM and models with sharp features in the power spectrum that are designed to remedy the central density problem, but it is presently difficult to derive rigorous constraints. Upcoming lensing studies that utilize new observational techniques [215, 209] will greatly improve these limits. I also find that the mapping between the mass of a halo and the maximum circular velocity of the halo is altered in tilted models in such a way that radically changes the dwarf satellite problem. In the tilted models the observed satellites can be accounted for without the need for differential feedback to squelch star formation in $\sim 90\%$ of halos at circular velocities $\sim 10$ km s$^{-1}$. Taken together, these results are provocative. Tilting the power spectrum can serve as a common solution to both the central density problem and the dwarf satellite problem while it does not reduce substructure mass fractions so that tilted models are not in danger of under-predicting future lensing constraints on substructure. That a single modification may alleviate all of the small-scale woes of the standard cosmology, leading to a remarkably consistent and unified model, rather than a loosely-knit patchwork of ad hoc remedies, is an interesting and attractive result. Moreover, it is generally believed that the properties of the primeval power spectrum are fixed during the early inflationary epoch which begets the standard hot big bang cosmology, and it is intriguing to speculate that the properties of galaxies
may reveal information about the Universe in its infancy and physics at the highest energies.
CHAPTER 2

TOWARD A STANDARD COSMOLOGY

In recent years, dramatic progress has been made in honing in on the values of the fundamental parameters of cosmology. Many of these parameters simply express how much "stuff" there is in the Universe like $\Omega_{b,0}$, $\Omega_{M,0}$, $\Omega_{\Lambda,0}$, $\Omega_{\gamma,0}$, $\Omega_{\nu,0}$. Another parameter, measures the current rate of expansion of the Universe. This is the Hubble parameter $H_0$, and it is convenient to use the dimensionless parameter $h \equiv H_0/100$ km s$^{-1}$ Mpc$^{-1}$ to express the current expansion rate. These suffice to determine the behavior of the large-scale, homogeneous Universe. On smaller scales, the Universe is clearly inhomogeneous as galaxies, galaxy groups, and galaxy clusters abound. These structures must have grown from some primordial seed perturbations as the Universe expanded and there is another set of parameters that describes the nature of the seed perturbations. We come to these in Section 2.3.

It seems that cosmology is converging toward a highly-predictive and successful standard model. It is imperative to understand the state of this standard model in order to attack the new questions that confront this new cosmology. In this Chapter, I review the pillars of standard cosmology, focusing on the determination of cosmological parameters from numerous observations and the establishment of a standard
parameter set. The material collected in this Chapter can be found in several standard texts [235, 324, 172, 236, 233, 230, 185, 64] and the probes that I discuss by no means exhaust the techniques that have been applied. Some introductory material is collected in Appendices. The spacetime framework for cosmology is the Robertson-Walker (RW) metric described in Appendix A. The dynamics of the spacetime are given by Einstein’s equations and complete the Friedmann-Robertson-Walker (FRW) cosmological model as described in Appendix B.

2.1 Big Bang Nucleosynthesis and The Baryon Density

The Universe is observed to be expanding and is filled with a thermal bath of photons at temperature\(^1\) \(T_0 = 2.348 \times 10^{-4}\) eV [201]. Extrapolating backward in time, this implies an early Universe that was very hot, dense, and dominated by relativistic species. In this environment, \(~25\%\) of the protons and neutrons in the Universe were fused into light nuclides, primarily \(^4\text{He}\), in a process known as Big Bang Nucleosynthesis (BBN).

At temperatures \(T \gg 1\) MeV, weak interactions like \(\nu_e + p \leftrightarrow e^+ + n\), \(\nu_e + n \leftrightarrow p + e^-\), and \(n \leftrightarrow p + e^- + \bar{\nu}_e\) maintained the ratio of the number density of neutrons, \(n_n\), to the number density of protons, \(n_p\), at its equilibrium value

\[
\frac{n}{p} \equiv \frac{n_n}{n_p} \simeq \exp\left(-\frac{Q}{T}\right),
\]  

where \(Q \simeq 1.3\) MeV, is the mass difference between the neutron and proton. As the Universe expands and cools, the weak interaction rates decline. At a temperature \(T_F \simeq 0.7\) MeV, the weak interaction rates become less than the expansion rate \(H\),

\(^1\)Throughout I use units in which \(\hbar = k_B = c = 1\) unless otherwise noted.
and these interactions effectively stop occurring. Thus the \((n/p)\) ratio becomes frozen in at its value at \(T_F\), \((n/p)_F \simeq 0.17\). After the freeze out of the weak interactions, neutrons undergo free decay with lifetime \(\tau_n \simeq 887\) s until stable nuclei are formed.

All roads to heavier nuclides go through deuterium. One route to synthesize \(^4\)He is \(n(p, \gamma)D(D, \gamma)^4\)He. Although the binding energy of deuterium \(B_D \sim 2.2\) MeV, is smaller than \(T_F\), the number density of baryons \(n_B\), is much smaller than the number density of photons, \(n_\gamma \sim (10^9 - 10^{10})n_B\), so there are plenty of photons on the high-energy tail of the Planck distribution capable of dissociating deuterium. Only when the temperature approaches \(T_N \simeq 70\) keV does it become thermodynamically favorable to produce deuterium, at which point the road to heavier nuclei is opened. In the intervening time, free neutron decay has reduces the neutron-to-proton ratio to \((n/p)_N \simeq 0.13\). \(^4\)He is much more strongly bound than all other nuclei, so as a rough approximation, one may estimate the amount of \(^4\)He produced in the early universe by assuming that all neutrons left at the time of nucleosynthesis are incorporated into \(^4\)He. Using \((n/p)_N \simeq 0.13\), the mass fraction of baryons incorporated into \(^4\)He is \(Y_p \simeq 0.23\). As the cosmic expansion dilutes and cools the Universe, nuclear reaction rates slow and nucleosynthesis effectively ceases at a temperature of \(T_E \simeq 20\) keV.

In addition to \(^4\)He, trace amounts of \(^3\)He, \(^6\)Li are produced. Any nuclear abundance \(Y_i\), obeys a rate equation of the form \(dY_i/dt = J - \Gamma Y_i\), where \(J\) and \(\Gamma\) are source and sink terms respectively. The source term \(J\) simply represents a sum over all production rates of element \(i\) and depends upon the abundance of other species. \(\Gamma\), represents a sum over all destruction rates and naturally, the destruction term is proportional to the abundance of \(i\) itself. Precise predictions require detailed numerical calculations that take into account the interplay of many nuclear reactions. The
first detailed numerical study was performed by Wagoner, Fowler, and Hoyle [319], and most contemporary calculations are based on this. The case of D is qualitatively simple. Once it becomes thermodynamically favorable to produce D, the D is quickly incorporated into more strongly bound nuclides like \(^4\)He. This destruction process is more efficient the more baryons there are around, so the amount of deuterium produced during BBN is a rapidly declining function of the baryon density. Assuming photons and the standard three neutrino species to be the only particles with masses \(\ll\) MeV, the only input parameter that determines primordial element abundances is the baryon density \(\Omega_b,0h^2\), yet BBN predicts observable abundances of \(^4\)He, \(^3\)He, D, and \(^7\)Li so the theory is over-constrained and serves as the earliest and one of the most robust tests of the standard cosmology.

In Figure 2.1, I show theoretical predictions for the abundances of the light nuclides \(^4\)He, D, \(^3\)He, and \(^7\)Li. Comparison between BBN theory and observations of nuclide abundances is not trivial. For one thing, one must account for any net production or destruction of nuclides in the \(\sim\) 13 Gyr since BBN. The net production or destruction of both \(^7\)Li and \(^3\)He are poorly-understood and estimates of primordial \(^7\)Li and \(^3\)He are subject to significant systematic uncertainties (see Ref. [55] for more details). For this reason, \(^4\)He and D are the probes usually used to confront theory with observation. The chemical evolution of D is fairly simple; astrophysical processes only destroy deuterium. Additionally, the predicted abundance of deuterium is a rapidly decreasing function of the baryon density and so the observed deuterium abundance is an excellent probe of the baryon abundance. The deuterium abundance is primarily measured from an absorption feature in absorption spectra along the line-of-sight to high redshift quasars. Likewise, the post-BBN evolution of the \(^4\)He abundance is
Figure 2.1: BBN Abundance Predictions vs. Observations. The solid lines represent the theoretical predictions for light element yields during nucleosynthesis as a function of baryon density. \( Y_p \) is the mass fraction of baryons in \(^4\)He and is shown in the upper panel. The lower panel shows, from top to bottom, the number density of D, \(^3\)He, and \(^7\)Li relative to the number density of \(^1\)H. The dashed horizontal lines in the upper panel represent the allowed observationally range of \( Y_p \). The upper set of dashed horizontal lines in the lower panel represent the observational range of D/H inferred from measurements along lines of sight to five high redshift quasar absorption-line systems. The lower set of dashed horizontal lines represents the inferred primordial \(^7\)Li abundance from observations of low-metallicity stars. Primordial abundances inferred from observations are taken from a compilation in Ref. [55].

simple; hydrogen burning in stars results in a net production of \(^4\)He. The primordial abundance of \(^4\)He can be determined from observations of the \(^4\)He abundance in low-metallicity HII regions and extrapolating to zero-metallicity.\(^2\) In Figure 2.1, I show

\(^2\)Metallicity is a measure of the relative amount of heavy elements in a system. Often the iron or oxygen abundances are used as measures of metallicity. Metals are an indication that the material has been processed extensively in stars while low metallicities indicate minimal processing.
estimates of primordial abundances taken from the compilation by Burles, Nollett, and Turner [55] and the deuterium measurements of Refs. [56, 184, 226, 240] in quasar absorption systems. I have been particularly generous in demarcating the boundary of the observed $^4$He abundance. The region shown accommodates two conflicting estimates of the primordial $^4$He abundance as determined by Olive, Skillman, and Steigman [224] and by Izotov and Thuan [145]. I do not show $^3$He abundances because its poorly-understood post-BBN evolution makes it a suspect BBN diagnostic. Notice that observations of both D, $^4$He and $^7$Li appear to be consistent with a single value of the baryon density. This is a striking success of the theory. A detailed statistical analysis yields an inferred baryon density of [55]

$$\Omega_{b,0} h^2 = 0.020 \pm 0.002$$  \hspace{1cm} (2.2)

at the 90% confidence level. BBN along with inferred primordial abundances of the light nuclides based on observational data yield a measurement of the baryon density.

### 2.2 The Magnitude-Redshift Relation of Supernovae

The flux received at earth from a distant source is a complicated function of the cosmological density parameters because the geometry of the FRW spacetime is not the static, Euclidean spacetime on which we base our intuition. The total flux $F$, received at the earth from an object of total luminosity $L$, is given by

$$F = \frac{L}{4\pi d_L^2},$$  \hspace{1cm} (2.3)

where $d_L$ is the luminosity distance defined in equation (A.19). It is conventional to express the brightness of an object in terms of its magnitude. **Apparent magnitude** $m$,
is a logarithmic measure of the flux received at earth, $m = -2.5 \log(F/F_0)$, where $F_0$
sets the zero point of the magnitude system. The absolute magnitude $M$, of an object
is a measure of the luminosity of an object and is simply the apparent magnitude
that the object would be assigned if it were at a luminosity distance of 10 pc. Thus
the apparent and absolute magnitudes are related via

$$m = M + 5 \log(d_L/10 \text{ pc}).$$

(2.4)

A common convention sets the zero point of the magnitude system such that the ab-
solute magnitude of the sun is $M_\odot = 4.74$. In this discussion we have been referring
to total fluxes and luminosities, that is, integrated over all wavelengths. This defines
what is called the bolometric magnitude. In practice, observations are made in par-
ticular frequency bands and corrections must be made to account for the response
of the filter over different frequencies and the fact that the emitted and observed
frequencies are not the same due to the cosmological redshift.

In terms of bolometric magnitudes, equation (2.3) can be recast as

$$m = M + 5 \log(d_L/ \text{ Mpc}) + 25.$$  

(2.5)

Combining equations (A.19) and (B.6), the luminosity distance is

$$d_L = \frac{1 + z}{H_0 \sqrt{\Omega_{k,0}}} \Sigma(\chi(z)), \text{ with}$$

(2.6)

$$\chi(z) = \sqrt{\Omega_{k,0}} \int_0^z \frac{dz'}{\sqrt{(1 + z')^2[1 + \Omega_{M,0}z' + (2 + z')z'\Omega_{R,0}] - z'(2 + z')\Omega_{\Lambda,0}}}.$$

$\Sigma(x)$ is defined in (A.4-A.6). Equation (2.5) is the so-called “magnitude-redshift”
relation. Notice that the different energy densities enter into $d_L$ with different powers
of redshift. This implies that given a population of “standard candles” of known luminosity at high redshift ($z > 0.5$), it is possible to determine the values of the different cosmological parameters (for a recent review, see [117]).

Type Ia Supernovae (SNIa) are known to be fairly standard candles with known absolute magnitude in a particular frequency band of $M_B \approx -19$ and dispersion $\sigma_{M_B} \approx 0.3$ [43]. In practice, SNIa can be made even more effective standard candles by exploiting a correlation between the timescale for the supernova to dim and the intrinsic brightness of the object [251]. At low redshift ($z \lesssim 0.05$), the magnitude-redshift relation is insensitive to the values of the cosmological energy densities and provides a measure of the Hubble parameter, $h = 0.72 \pm 0.08$ [109].

The Supernova Cosmology Project (SCP) [239] and the High-$z$ Supernovae Search Team (HZT) [258] have exploited SNIa standard candles in order to constrain the cosmological energy density parameters. In Figure 2.2, I show the data compiled by the SCP along with several theoretical lines representing the predicted magnitude-redshift relations for different choices of cosmological parameters. Clearly, a universe dominated by radiation is disfavored by the data, but this is no surprise given that the measured CMB temperature implies that $\Omega_{R,0} \sim 10^{-5}$. Note that universes dominated by either matter ($\Omega_{M,0} = 1$) or by a cosmological constant ($\Omega_{\Lambda,0} = 1$) are also disfavored. Given that $\Omega_{R,0}$ is expected to be small at redshifts $z \lesssim 10^3$, it is useful to neglect this parameter and use the magnitude-redshift relation to constrain the matter and vacuum energy densities of the Universe. The result of a likelihood analysis performed on the data set collected by the SCP is shown in Figure 2.3. Some results are clear: a matter dominated universe with $\Omega_{M,0} = 1$ is ruled out at very
Figure 2.2: The high-redshift portion of the magnitude-redshift relation for the SCP data (with 1σ error bars) [239] alongside the theoretical magnitude-redshift relation in several flat cosmologies: \( \Omega_{M,0} = 0.3, \Omega_{\Lambda,0} = 0.7, \) and \( \Omega_{R,0} = 0 \) (heavy, solid); \( \Omega_{M,0} = 0, \Omega_{\Lambda,0} = 0.8, \) and \( \Omega_{R} = 0.2 \) (dashed); \( \Omega_{\Lambda} = 1 \) and \( \Omega_{R,0} = \Omega_{M,0} = 0 \) (dotted); \( \Omega_{M,0} = 1 \) and \( \Omega_{R,0} = \Omega_{\Lambda,0} = 0 \) (dash-dot); and \( \Omega_{R,0} = 1, \Omega_{M,0} = \Omega_{\Lambda,0} = 0 \) (light, solid).

...high confidence and any Universe with \( \Omega_{M,0} > 0.2 \) must contain a significant fraction of vacuum energy.

The contours in Figure 2.3 select a degenerate region in parameter space. It is possible to go further given that BBN yields a reliable estimate of the baryon density. Rich galaxy clusters are the objects that have collapsed and virialized most recently (relative to, \textit{e.g.}, galaxies and/or groups). As such, it is thought that the characteristics of rich clusters may be fairly representative of the Universe at large. In particular, the ratio of the baryonic mass of a cluster to its total mass may provide...
Figure 2.3: Confidence contours in the $\Omega_{M,0} - \Omega_{\Lambda,0}$ plane inferred from the SNIa magnitude-redshift relation. The solid contour marks the 64% confidence level and the dotted contour marks the 95% confidence level. The shaded portion in the upper left corresponds to parameter values that do not trace back to high enough redshift to accommodate big bang nucleosynthesis. These universes oscillate rather than expand monotonically so there is no “big bang.” Also shown is the locus of points that corresponds to a flat Universe $(k = 0)$.

A good estimate of the fraction of all matter in the Universe contained in baryons, $f_B$. The hot X-rays from intra-cluster gas can be used to determine both the baryonic mass of the cluster and the total gravitating mass of the cluster, but the relations between X-ray temperature, cluster mass, and total baryonic mass must be calibrated using high-resolution numerical simulations. The simulations of Frenk et al. [329] along with observations of X-rays from clusters indicate that $f_B h^2 = 0.065 \pm 0.016$
The BBN estimate of the baryon density can then be used to obtain an estimate of the matter density, \( \Omega_{M,0}h^2 = \Omega_{b,0}h^2/f_Bh^2 \).

Including this information on the universal baryon fraction and the baryon density from BBN in the likelihood analysis of the SNIa supernovae breaks the degeneracy in the \( \Omega_{M,0} - \Omega_{\Lambda,0} \) plane and yields the confidence contours shown in Figure 2.4 [281, 278]. This yields a much more restrictive bound on the mass density,

\[
\Omega_{M,0} = 0.26^{+0.08}_{-0.06} \tag{2.7}
\]

Interestingly, this analysis implies the necessity of a large, non-zero value for the cosmological constant or vacuum energy density, \( \Omega_{\Lambda,0} = 0.74^{+0.16}_{-0.17} \).

### 2.3 The Growth of Large-Scale Structure and The Matter Density of the Universe

#### 2.3.1 The Power Spectrum

By dint of gravitational instability, small, primordial density fluctuations give rise to all of the structure in the Universe. The fluctuations in density are usually described in terms of the density contrast \( \delta(\vec{x}) \equiv (\rho(\vec{x}) - \bar{\rho})/\bar{\rho} \), where \( \bar{\rho} \) is the mean density. Throughout, \( \vec{x} \) represents a comoving coordinate. In flat Universes, the proper distance from the origin to the comoving coordinate \( \vec{y} \) is related to the comoving coordinate by \( d_{\text{proper}} = a|\vec{y}| \). The quantity \( a \) is the scale factor normalized to unity today. When typical density contrasts are small, they can be treated as a perturbation on top of the smooth background and linear perturbation theory can be applied to describe the behavior of the fluctuations. It is convenient to deal with the Fourier transform of the density contrast,
Figure 2.4: Confidence contours in the $\Omega_{M,0} - \Omega_{\Lambda,0}$ plane inferred from the SNIa magnitude-redshift relation, BBN bounds on the universal baryon fraction, and an estimate of the universal baryon fraction from rich clusters. The solid contour marks the 64% confidence level and the dotted contour marks the 95% confidence level.

\[ \delta_k = V^{-1} \int d^3 x e^{i \vec{k} \cdot \vec{x}} \delta(\vec{x}), \]  

(2.8)

with inverse transform given by

\[ \delta(\vec{x}) = \frac{V}{(2\pi)^3} \int d^3 k e^{-i \vec{k} \cdot \vec{x}} \delta_k. \]  

(2.9)

The quantity $V$ is a fiducial large volume over which the boundary conditions are supposed to be periodic. In the limit that the $V$ is very large compared to any observable quantities, the Fourier integral is taken as a good approximation to the Fourier sum. The rms density contrast in spheres of radius $R$ is then
\[ \sigma^2(R) = \int \Delta^2(k)|VW(k; R)|^2 \, d\ln k, \]  

(2.10)

where \( W(k; R) \) is the Fourier transform of a spherical top-hat of radius \( R \), \( \Delta^2(k) \equiv \sqrt{2\pi}k^3|\delta_k|^2 \). The factor of \( V \) results from the Fourier transform definition of Equation (2.8), but \( W(k; R) \propto V^{-1} \) so that \( VW(k; R) \) is independent of the fiducial volume. I have assumed isotropy such that the \textit{power spectrum} \( P(k) \equiv |\delta_k|^2 \) depends upon only the magnitude of the wavenumber. Note that it is also a common practice to absorb the factor of \( V \) into the definition of the power spectrum, \( i.e., P(k) = V|\delta_k|^2 \) so that the power spectrum has units of volume, because this is the quantity that is related most directly to observation. Assuming that the density field is a Gaussian random field, any statistical quantity related to the density field can be expressed in terms of \( P(k) \). It is commonly assumed that prior to any processing by causal physical processes, the seed fluctuations were Gaussian and had a power spectrum of the form \( P(k) \propto k^n \) with \( n \approx 1 \). These assumptions are justified by observations and are predicted by the leading paradigm for the production of seed perturbations, inflation \([122, 123, 134, 20] \). The quantity \( \Delta(k) \propto k^{(n+3)/2} \) is important because it represents the rms overdensity per logarithmic interval in wavenumber. In other words, the rms overdensity on a scale \( k \) is \( \sim \Delta(k) \).

The density field is modified by causal physical processes on scales smaller than the horizon, \( R \lesssim 1/aH \) or \( k \gtrsim aH \). As the Universe expands, larger and larger scales constantly come within the horizon and are processed by causal physics. A power spectrum with \( n = 1 \) has the special property that \( \Delta^2(k) \) is constant when evaluated on the scale of the horizon, namely \( k = aH \). Physically, this means that each scale enters the horizon with approximately the same rms overdensity and then begins to be
processed. A power spectrum with $n = 1$ is therefore referred to as a scale-invariant power spectrum.

2.3.2 The Behavior of Density Perturbations

The formalism for studying the growth of density fluctuations due to gravity is well-understood (see, for example Refs. [235, 195, 230]). It is instructive to consider the simplified case of a two-component, flat universe with matter component ($M$), and radiation component ($R$) that represents all relativistic species. The energy density in matter scales as $\rho_M \propto a^{-3}$ while the energy density in radiation scales like $\rho_R \propto a^{-4}$ so a universe with these two components will under go an early radiation-dominated phase when $\rho_R \gg \rho_M$ and the dynamics of the spacetime are dominated by the relativistic component, and a late matter-dominated phase when the dynamics are dominated by the matter component. The matter and radiation densities are equal when the scale factor is $a_{\text{EQ}} = \Omega_{R,0}/\Omega_{M,0}$ and it is convenient to use $x = a/a_{\text{EQ}}$ as the independent variable perturbation analyses.

The analysis of perturbations is most simply accomplished in Fourier space because different $k$ modes decouple from each other at linear order. The behavior of a specific mode is given by

$$\frac{d^2 \delta_R}{d(\ln x)^2} - \left( \frac{x}{2(1 + x)} - 1 \right) \frac{d \delta_R}{d \ln x} + \left( \frac{2x^2 k^2}{3H_{\text{EQ}}^2 a_{\text{EQ}}^2} + \frac{4x}{3(x + 4/3)^2} - \frac{8}{3(x + 4/3)} \right) \delta_R$$

$$= \frac{2x \delta_M}{1 + x} - \frac{x}{1 + 3x/4} \frac{d \delta_M}{d \ln x}$$

(2.11)

and
\[
\frac{d^2 \delta_M}{d(\ln x)^2} + \frac{x}{2(1 + x)} \frac{d \delta_M}{d \ln x} - \frac{3x}{2(1 + x)} \delta_M = \frac{1}{x + 4/3} \left[ \frac{d \delta_R}{d \ln x} - \left( \frac{x}{x + 4/3} - 2 \right) \delta_R \right]
\]

(2.12)

where I have suppressed the argument \( k \) in \( \delta_M(k) \) and \( \delta_R(k) \). The quantities \( \delta_R \) and \( \delta_M \) are the overdensities in the radiation and matter components respectively and I have expressed the expansion rate at a given time \( H(a) \), in terms of the expansion rate at matter-radiation equality, \( H^2 = H_{\text{EQ}}^2(1 + x)/2x^4 \). It is necessary to specify a coordinate gauge in order to define perturbations in general relativity. Equations (2.11) and (2.12) describe perturbations in the comoving gauge. The detail of gauge choice is an unnecessary complication for the present purpose. Density perturbations in all gauges reduce to the standard Newtonian conception on scales much smaller than the horizon. Further, because I assume a nearly scale-invariant primordial spectrum, where the rms fluctuations on all scales are equal at horizon-crossing, it is only the behavior of perturbations after horizon-crossing that is important in this discussion. Restricting consideration to the behavior of perturbations on scales much smaller than the horizon requires \( k \ll aH \).

Equations (2.11) and (2.12) appear formidable, yet it is not difficult to understand the basic physics that describe the evolution of perturbations. Consider first the behavior of perturbations to the radiation density in equation (2.11). Unlike the matter perturbations, perturbations in radiation experience a scale dependent source term \( \propto k^2 \), the first term in the parentheses. This term contains the effect of pressure in the
relativistic fluid. The pressure of a relativistic fluid is significant compared to its energy density, \( P = \rho/3 \), and allows the fluid to resist compression due to gravity. Focusing on scales much smaller than the horizon, corresponds to taking \( k^3 x^2 / H_{\text{EQ}}^2 a_{\text{EQ}}^2 \gg 1 \) so that the term proportional to \( \delta_R \) reduces to \( \approx (2x^2 k^2 / 3 H_{\text{EQ}}^2 a_{\text{EQ}}^2) \delta_R \). This term is always positive in this limit and suggests that the overdensity in the radiation component has an oscillatory behavior with a frequency \( \sim \sqrt{2k/3H_{\text{EQ}}a_{\text{EQ}}} \). A full solution of these equations reveals that this is indeed the case. On scales smaller than the horizon, perturbations in relativistic species oscillate and cannot grow.

Next, consider perturbations in the matter component. In light of the above result and in order to simplify the analysis, ignore perturbations in the relativistic component as a source for perturbations in the matter component and take \( \delta_R \approx 0 \). The equation governing the growth of the matter component then reduces to

\[
\frac{d^2 \delta_M}{d(\ln x)^2} + \frac{x}{2(1 + x)} \frac{d \delta_M}{d \ln x} - \frac{3x}{2(1 + x)} \delta_M \approx 0.
\]

(2.13)

Notice that the growth of perturbations in the matter component is independent of wavenumber for scales well within the horizon. The terms in this equation are easy to interpret as well. The term proportional to \( \delta_M \) represents the source of perturbations: existing perturbations source the growth of perturbations. The term proportional to \( d\delta_M/d\ln x \) represents the suppression of perturbation growth due to the effect of the cosmological expansion, which tends to dilute. It is instructive to consider two limits. First consider the behavior of perturbations during radiation domination, \( x \ll 1 \). In this limit, \( \delta_M \) obeys

\[
\frac{d^2 \delta_M}{dx^2} + \frac{1}{x} \frac{d \delta_M}{dx} \approx 0,
\]

(2.14)
which has solutions

$$\delta_M \propto \ln x \quad \text{and} \quad \delta_M \propto \text{constant.} \quad (2.15)$$

Perturbations in matter only grow logarithmically during radiation domination because the rate of cosmic expansion is set by the radiation density and is large enough that cosmological dilution overcomes perturbation growth. Second, consider the growth of structure during the matter domination, $x >> 1$. The matter perturbation now obeys

$$\frac{d^2 \delta_M}{dx^2} + \frac{3}{2x} \frac{d \delta_M}{dx} - \frac{3}{2x^2} \delta_M \approx 0. \quad (2.16)$$

This equation has two power law solutions,

$$\delta_M \propto x \propto a \quad \text{and} \quad \delta_M \propto x^{-3/2}. \quad (2.17)$$

During matter domination, $\delta_M$ grows in proportion to the scale factor. In this case, it is the matter density that sets the rate of cosmic expansion and the rate of perturbation growth so the cosmic expansion rate cannot outstrip the rate of perturbation growth and quash the formation of structure.

In the modern picture of cosmology, the contemporary Universe is dominated by a cosmological constant, but the energy density in radiation is negligibly small. In a toy universe composed only of matter and a cosmological constant, perturbations grow as

$$\frac{d^2 \delta_M}{da^2} + \frac{3}{2a} \left[ 1 + \frac{\rho_\Lambda}{\rho_M + \rho_\Lambda} \right] \frac{d \delta_M}{da} = \frac{3}{2a} \left[ \frac{\rho_M}{\rho_M + \rho_\Lambda} \right] \delta_M. \quad (2.18)$$
In the limit of cosmological constant domination ($\rho_\Lambda \gg \rho_M$),

$$\frac{d^2 \delta_M}{da^2} + \frac{3}{a} \frac{d\delta_M}{da} = 0$$

(2.19)

which has solutions $\delta_M \propto \text{constant}$ and $\delta_M \propto a^{-3}$, so structure growth ceases once the cosmological constant comes to be the dominant energy component in the Universe. At redshifts lower than $z \sim 1100$ (the epoch of decoupling, see below), the growth of perturbations on scales smaller than the horizon is essentially independent of scale.

The growing mode of cosmological perturbations can then be described by a function of the scale factor only, $\delta_M(a) \propto D(a)\delta_M(a_{\text{init}})$, where $a_{\text{init}}$ is some initial epoch. The function $D(a)$ is the growth function.

This brief foray into perturbation theory has revealed two key results regarding structure formation. First, on scales smaller than the horizon only perturbations in the matter component can grow. Second, even matter perturbations cannot grow under arbitrary circumstances. During radiation domination the timescale for the cosmic expansion is much shorter than the timescale for the growth of matter perturbations and so the growth of structure is quelled by the rapid expansion of the Universe. Likewise, matter perturbations stop growing during $\Lambda$ domination.

### 2.3.3 Measuring the Mass Density With the Power Spectrum

The suppression of structure growth during radiation domination has important consequences for the shape of the power spectrum today. The overdensities on all scales that enter the horizon during radiation domination are temporarily frozen at their horizon-crossing values, $\Delta^2(k) \approx \Delta^2(k)|_{k=aH} \approx \text{constant}$. Once matter comes to dominate the energy budget of the Universe, overdensities on all scales grow uniformly
as $\delta_k \propto a$ and the shape of the power spectrum on large scales is unaffected. Thus the comoving scale of the horizon at the epoch of matter-radiation equality $k_{\text{EQ}}$, is imprinted on the power spectrum. On scales smaller than the comoving horizon scale at matter-radiation equality (i.e., $k > k_{\text{EQ}}$), the power spectrum is nearly flat, $\Delta^2(k > k_{\text{EQ}}) \sim$ constant. On scales larger than the horizon scale at equality (i.e., $k < k_{\text{EQ}}$), the shape of the power spectrum is unaltered and still reflects its pristine primordial form, $\Delta^2(k < k_{\text{EQ}}) \propto k^4$.

Assuming that the only relativistic species in the Universe are photons and the three standard neutrinos, the comoving scale of the horizon at matter-radiation equality can be computed straightforwardly from the Friedmann equation and is given by\(^3\)

\begin{equation}
    k_{\text{EQ}} \simeq 0.39(\Omega_{M,0} h) \, h \, \text{Mpc}^{-1}.
\end{equation}

Roughly speaking, the power spectrum today has the form of a broken power law with $\Delta^2(k) \propto k^4$ at $k \ll k_{\text{EQ}}$ and $\Delta^2(k) \approx$ constant for $k \gg k_{\text{EQ}}$. The value of $k_{\text{EQ}}$, and thus the shape of the power spectrum, depends on the combination $\Gamma = \Omega_{M,0} h$ which is known as the \textit{shape parameter}. The shape parameter can be determined from measurements of the contemporary power spectrum. The clustering properties of galaxies are well-determined on scales from $0.01 h \, \text{Mpc}^{-1} \lesssim k \lesssim 0.5 h \, \text{Mpc}^{-1}$, where overdensities are in the linear regime, so the power spectrum of galaxy clustering is a fruitful way to constrain the shape parameter. The matter power spectrum may be inferred from the measured clustering properties of galaxies by assuming that the overdensities in galaxy number are proportional to matter overdensities, $\delta_{\text{galaxy}} \propto \delta_M$.

\(^3\)Distances are expressed in units of $h^{-1}\text{Mpc}$ in order to explicitly contain the uncertainty in the Hubble parameter.

35
Figure 2.5: The power spectrum of galaxy clustering as a probe of the matter density of the Universe. The points and error bars show estimates of the galaxy power spectrum as extracted from the 2dF 100k galaxy redshift survey by Tegmark, Hamilton, and Xu [300]. The lines show theoretical matter power spectra for three representative values of the shape parameter. The data seem to favor $\Gamma \sim 0.2$.

In Figure 2.5, I demonstrate the efficacy of determining the matter density via the galaxy power spectrum. In this Figure, I show the power spectrum of galaxy clustering from the analysis of the 2dF 100k galaxy redshift survey by Tegmark, Hamilton, and Xu [300]. I also exhibit theoretical power spectra for three representative choices of the shape parameter: $\Gamma = 0.07$; $\Gamma = 0.21$; and $\Gamma = 0.42$. The measured power spectrum of galaxies clearly places a strong constraint on the shape parameter. Observe that the overall normalizations of the measured and theoretical spectra can be allowed to float relative to one another in order to accommodate any offset, or bias, between
the clustering of matter and the clustering of galaxies and yet the $\Gamma = 0.07$ and $\Gamma = 0.42$ models would still be disfavored. Merely the shape of the power spectrum is enough to constrain $\Omega_{M,0} h$. Using the galaxy power spectrum measured by the automated plate-measuring machine survey of galaxies, Padilla and Baugh [229] find $\Gamma = 0.19^{+0.13}_{-0.04}$. Combined with the Hubble Space Telescope Key Project measurement of the Hubble parameter, $h = 0.72 \pm 0.08$, this implies a matter density of

$$\Omega_{M,0} = 0.26^{+0.18}_{-0.06}. \tag{2.21}$$

This estimate of the matter density of the Universe agrees remarkably well with the independent estimate of $\Omega_{M,0} = 0.26^{+0.08}_{-0.06}$ coming from the combination of BBN and SNIa data in Section 2.2.

Other measurements of structure growth are also sensitive to the matter density of the Universe. The abundance of clusters, for instance, is sensitive to the matter density of the Universe. Naturally, there is a degeneracy between $\Omega_{M,0}$ and the amplitude of initial density fluctuations if one counts the abundance of clusters at only one redshift; however, measuring the cluster abundance at various redshifts breaks this degeneracy. Two efforts to do just this find $\Omega_{M,0} = 0.26 \pm 0.07$ [207], and the somewhat lower value $0.17 \pm 0.05$ [14]. The density field of the Universe also serves to gravitationally lens the photons emitted by distant objects. This results in an apparent shear or correlated orientation of the long axes of galaxies toward certain directions on the sky and is sensitive to the amount of matter in the Universe. Measurements of this weak lensing effect in the Red-sequence Cluster Survey yield an estimate of $\Omega_{M,0} = 0.3 \pm 0.04$ [71].

37
2.4 The Cosmic Microwave Background

As I mentioned in Chapter 1, the universe is permeated with a nearly thermal bath of photons at a temperature $T \approx 2.7 \text{ K}$. Early in the history of the Universe, the photons and baryons were tightly coupled and in nearly thermal equilibrium at a common temperature due to rapid Thomson scattering of photons and free electrons and the electromagnetic coupling between electrons and ionized hydrogen and helium. As the Universe expanded and cooled, the electron density fell for two reasons. First, the expansion of the spacetime diluted the electron density. Second, as the temperature of the CMB dropped below $\sim 1 \text{ eV}$, it became thermodynamically favorable to incorporate electrons and nuclei into neutral atoms rather than maintain an ionized plasma. As the electron density declined, so did the rate of Thomson scattering, $\Gamma_T \propto n_e \sigma_T$, where $n_e$ is the free electron density and $\sigma_T$ is the Thomson scattering cross section. Finally, when the temperature fell below $\sim 0.3 \text{ eV}$, at a redshift of $z_d \sim 1100$, the rate of Thomson scattering fell below the expansion rate of the Universe and the photons and baryons effectively decoupled from each other. After the “epoch of decoupling,” photons essentially propagated freely without scattering any further off of the matter in the Universe. Consequently, CMB photons today contain a veritable treasure trove of information about the early Universe and measurements of these photons can be used to place strong constraints on cosmology. In this section I discuss the important features of CMB anisotropy by examining several highly idealized examples (see [139]).
2.4.1 The Angular Power Spectrum

As discussed in Section 2.3 the early Universe is filled with density fluctuations that seed the formation of structure. These fluctuations are reflected in the effective temperature of the CMB at different locations. As such, when we observe the CMB we not only see the thermal radiation with mean temperature $T \approx 2.7 \text{ K}$, but we also see fluctuations in intensity as a function of the direction on the sky. These fluctuations are of order one part in $\sim 10^5$ and may be analyzed using linear perturbation theory. Furthermore, to first order in perturbations, the Planckian shape of the distribution function of photons is preserved and the fluctuations can be described by small changes in the observed temperature

$$
\Delta T(\hat{n}) = \frac{T(\hat{n}) - \bar{T}}{\bar{T}},
$$

(2.22)

where $\hat{n}$ is a unit vector indicating a direction on the sky, $T(\hat{n})$ is the measured temperature in the direction $\hat{n}$, $\bar{T}$ is the mean temperature, and $\Delta T$ is the temperature anisotropy.

The generation of density perturbations, and thus temperature anisotropies, is a statistical process so theory and observations must be compared in a statistical manner. The first step is to expand the measured temperature anisotropy in spherical harmonics, which form a basis set for directions on the sky,

$$
\Delta T(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\theta, \phi).
$$

(2.23)

Using the addition theorem of spherical harmonics the correlation function of temperature anisotropy between two points on the sky is
\[ C(\alpha) \equiv \langle \Delta T(\hat{n}_1) \Delta T(\hat{n}_2) \rangle = \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell + 1)|a_{\ell m}|^2 P_\ell(\cos \alpha), \]  

(2.24)

where the average is over directions on the sky, \( \alpha \) is the angle separating the two directions and \( P_\ell(x) \) are the Legendre polynomials. In comparing theory with observation, it is necessary to imagine an ensemble of observers. Each observer would measure different \( a_{\ell m} \) because the realization of a particular density perturbation field is a statistical process. The expectation value of \( |a_{\ell m}|^2 \) is typically denoted by \( C_\ell \) such that

\[ C(\theta) = \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell + 1)C_\ell P_\ell(\cos \theta). \]  

(2.25)

The values of the \( C_\ell \) coefficients reflect the variance in CMB temperature fluctuations over typical angular separations \( (\theta/1^\circ) \sim \ell^{-1} \). The \( C_\ell \) are used to compare theory with observation and are commonly referred to as the CMB anisotropy power spectrum.

The monopole term in equation (2.25) gives the background temperature and the dipole term is unobservable because the motion of the earth through the cosmic rest frame generates a dipole that swamps any temperature fluctuations intrinsic to the CMB. As such, these terms are not useful for studying anisotropy in the CMB. The remaining multipoles reflect physical processes that are important on each scale. In what follows, I give a brief discussion of the CMB anisotropy on several scales.

### 2.4.2 Anisotropy on Large Scales: The Sachs-Wolfe Plateau

In order to analyze perturbations on scales comparable to the size of the horizon, a general relativistic treatment is necessary. It is simplest to neglect tensor mode perturbations to the metric, which decouple from scalar density perturbations and
are expected to be small in the standard paradigm. Vector perturbations decay as the Universe expands and so one can neglect these as well. It is simplest then to consider only scalar perturbations and to work in the Newtonian gauge. For $\ell > 1$, the $C_\ell$ are gauge-invariant.

In Newtonian gauge, the perturbed metric is

$$ds^2 = a^2(\tau)[-(1 + 2\Psi)d\tau^2 + (1 - 2\Phi)h_{ij}dx^i dx^j],$$

(2.26)

where the timelike variable $\tau$ is known as *conformal time* and is related to the coordinate time $t$ of equation (A.1) by $d\tau = dt/a(t)$. It is another great simplification to consider only Universes that are spatially flat in the absence of perturbations so that $h_{ij} = \delta_{ij}$. In this case calculations are simple because the metric is diagonal. The perturbation $\Psi$ represents a time-lag between the proper time experienced by a comoving observer and coordinate time, while $\Phi$ represents a perturbation to the curvature of spatial hypersurfaces. A great advantage of the Newtonian gauge is that $\Psi$ reduces to the gravitational potential in the Newtonian limit, making calculations more intuitive and easy to interpret. Furthermore, in the case when the Universe is dominated by matter, the perturbed Einstein equations simplify tremendously yielding $\Psi \approx \Phi$ and $\partial^2 \Psi / \partial a^2 + 3/2a(\partial \Psi / \partial a) = 0$. This gives $\Psi \propto a^{-5/2}$ and $\Psi \propto$ constant. In what follows, I ignore decaying modes and take $\Psi = \Phi = \text{constant}$ during matter domination.

To solve for the temperature perturbation in all generality requires a solution of the coupled Boltzmann and Einstein equations for the evolution of the photon distribution function. In general the distribution function is a function of $\tau$, $x^i$, and $p_i$, where $p_i$ is the momentum conjugate to the $x^i$ coordinates. It is convenient
to express the momentum as $p_i = a(1 - \Phi)E n_i$ where $E$ is the proper momentum measured by a comoving observer. Writing the Boltzmann equation in terms of the temperature anisotropy, rather than the distribution function, yields (see e.g., [138] for details):

$$\frac{d(\Delta T + \Psi)}{d\tau} = \frac{\partial \Phi}{\partial \tau} - \frac{\partial \Psi}{\partial \tau} + \left( \frac{\partial \Delta T}{\partial \tau} \right)_C. \quad (2.27)$$

The last term represents the effect of any scattering processes (in this case, Thomson scattering off of electrons). The first term is due to the gravitational redshift of photons in climbing out of a potential well of depth $\Psi$, while the second term is a temperature perturbation due to the effect of spatial curvature. As I mentioned above, Einstein’s equations close the loop with expressions for $\partial \Psi/\partial \tau$ and $\partial \Phi/\partial \tau$.

Again, the analysis is typically performed in Fourier space. The comoving horizon size at decoupling is $\lambda_D \sim 180/\sqrt{\Omega_{M,0} h^2}$ Mpc, and if we restrict our analysis to modes that are well outside the horizon ($k \ll \lambda_D^{-1}$) causal physical processes cannot act to alter the temperature perturbations and Thomson scattering can be neglected. Because scattering processes are unimportant in determining the temperature fluctuations on these scales, these super-horizon-sized perturbations reflect the properties of the primordial power spectrum of fluctuations. Neglecting Thomson scattering and approximating decoupling as an instantaneous process, equation (2.27) can be integrated to yield an observed anisotropy, $\Delta T_o$, of

$$\Delta T_o \simeq \Delta T_{LS} + \Psi_{LS} - \Psi_o + \int_{\tau_{LS}}^{\tau_o} d\tau \left[ \frac{\partial (\Phi + \Psi)}{\partial \tau} \right]. \quad (2.28)$$

$\Delta T_{LS}$ corresponds to the intrinsic temperature anisotropy at the moment of last-scattering, $\Psi_{LS}$ is the potential at last-scattering, and $\Psi_o$ represents the potential.
at the observer which adds an overall, unobservable constant and can therefore be neglected. The terms are easy to understand. The $\Delta T_{LS}$ simply gives the anisotropy in temperature that existed at the last-scattering surface, while the $\Psi_{LS}$ term gives the redshift of the photon energy as it climbs out of the potential well. The integral is taken along the null geodesics back to the last-scattering surface and accounts for any time-variation in gravitational potentials. Taken together, the first two terms are known as the Sachs-Wolfe (SW) effect [254]. The last term is correspondingly referred to as the integrated Sachs-Wolfe (ISW) effect. The simplest models of inflation predict the initial fluctuations to be adiabatic in the sense that they maintain a constant entropy per matter particle. The entropy density is dominated by the radiation component, so this implies that on scales larger than the horizon, $\delta_M = 3\delta_R/4$. In the limit of complete matter domination, $\Phi = \Psi = \text{constant}$, and $\Psi \approx -\delta_M/2$, so that the temperature anisotropy on the last scattering surface is $\Delta T_{LS} \approx -2\Psi_{LS}/3$ and

$$\Delta T_o \approx \frac{\Psi_{LS}}{3}. \quad (2.29)$$

Notice that overdense regions actually appear to be cool to an observer today due to the effect of the redshift.

It is easy to compute the CMB power spectrum due to the Sachs-Wolfe effect. Expanding the anisotropy in plane waves gives and using the spherical Bessel function expansion for plane waves gives

$$\Delta T_o = \frac{1}{3} \int d^3k \sum_{\ell=0}^{\infty} (-i)^\ell (2\ell + 1) \Psi_{LS}(\vec{k}) j_\ell(k \eta_{LS}) P_\ell(\hat{k} \cdot \hat{n}). \quad (2.30)$$
In (2.30), $\eta_{LS}$ represents the comoving distance to the last scattering surface. Rewriting the Legendre polynomial by using the addition theorem for spherical harmonics gives the $a_{\ell m}$:

$$a_{\ell m} = (-i)^{\ell} \frac{4\pi}{3} \int d^3 k Y_{\ell m}^* (\mathbf{k}) j_\ell (k \eta_{LS}) \Psi_{LS}(\mathbf{k}).$$  \hspace{1cm} (2.31)

Finally, using this we can obtain the $C_\ell$ by taking $|a_{\ell m}|^2$. In order to calculate the $C_\ell$, it is necessary make an assumption about the expectation value of the quantity $k^3 |\Psi_{LS}(k)|^2$. The gravitational potential is given by the relativistic analog of Poisson’s equation, so $\delta_M(k) \propto k^3 \Psi(k)$. A primordial power spectrum of the form $P(k) \propto k^n$ implies that $k^3 |\Psi(k)|^2 = A k^{n-1}$, where $A$ is a normalization to be fixed by observations. In this case,

$$C_\ell \simeq \frac{2A}{9\pi} \eta_{LS}^{1-n} \int dx x^{n-2} j_\ell^2 (x)$$

$$= \frac{4A}{9H_0^{1-n} \Gamma^2 (2-n/2)} \frac{\Gamma (3-n) \Gamma (\ell + (n - 1)/2)}{\Gamma (\ell + (5-n)/2)}. \hspace{1cm} (2.32)$$

In the case of a scale-invariant primordial spectrum with $n = 1$,

$$\ell (\ell + 1) C_\ell = \frac{16A}{9\pi}. \hspace{1cm} (2.33)$$

If $n = 1$, the Sachs-Wolfe effect leads to a temperature anisotropy spectrum that is flat in the variable $\ell (\ell + 1) C_\ell$. This is referred to as the Sachs-Wolfe plateau.

The contribution of the anisotropy in each $k$-mode to a particular $C_\ell$ is a complicated integral over all $k$, but roughly speaking, a given mode $k$ contributes the most on a scale $\ell \sim k \eta_{LS}$. The Sachs-Wolfe effect is the dominant effect on scales larger than the horizon at decoupling, $k \ll 2\pi/\lambda_D$. The comoving distance to the
last-scattering surface is $\eta_{LS} \sim 4000$ Mpc, and so the Sachs-Wolfe plateau dominates the power spectrum for $\ell \lesssim 100$.

The relevance of the Sachs-Wolfe plateau is manifest. Measurements of the low multipoles of the CMB anisotropy, $\ell < 100$, allow for the determination of the amplitude of the primordial power spectrum of density fluctuations and the spectral index of the power spectrum $n$. As I discuss below, CMB anisotropy measurements are consistent with $n \approx 1$.

More generally, the ISW effect cannot be neglected. If the Universe is not matter-dominated, the potentials are not constant. This can be seen using Newtonian intuition (see [235]). A density perturbation on comoving scale $\lambda$ induces a potential fluctuation

$$\Phi \sim \frac{4\pi(a\lambda)^2\delta}{3}\bar{\rho}.$$  \hspace{1cm} (2.34)

As I discussed in Section 2.3.2, $\delta \propto a$ during matter domination, while $\bar{\rho} \propto a^{-3}$ so that $\Phi \sim$ constant. During radiation and vacuum energy domination, perturbations cease growing while $\bar{\rho} \propto a^{-4}$ and $\bar{\rho} \propto$ constant respectively, and the potential is not constant. The ISW effect then adds an additional component to the observed temperature fluctuation. This is important at low redshift, when the cosmological constant becomes dynamically important, and at the epoch of decoupling when the radiation content of the Universe is non-negligible.

2.4.3 Small-Scale Fluctuations: Acoustic Peaks

On scales smaller than the scale of the horizon at decoupling, scattering processes become important and determining the precise anisotropy spectrum becomes
an involved problem that must be solved by integrating the differential equations of perturbation growth numerically. Again, the essential features of CMB anisotropy can be gleaned by taking a simplified approach (see the work of Hu and Sugiyama [140] and Hu [138] for a detailed exposition). In particular, I make the approximation that decoupling happens instantaneously so that the problem can be broken into two steps: (1) calculate the intrinsic anisotropy due to conditions on the last-scattering surface; and (2) calculate any additional anisotropy due to the propagation of photons since decoupling. The intrinsic anisotropy seen due to the state of the photons on the last-scattering surface can be written

\[ \Delta T_{\text{int}} = \Delta T_{\text{LS}} + V_{B,\text{LS}}^i n_i, \quad (2.35) \]

where \( \Delta T_{\text{LS}} \) is the intrinsic temperature anisotropy. The second term gives the effect of the Doppler shift of photons due to the velocity of the baryon fluid \( V_B \) at last-scattering along the line-of-sight, \( \hat{n} \).

To calculate the intrinsic temperature fluctuations on the surface of last-scattering, I take the limit of tight-coupling between photons and baryons. In this limit the Thomson scattering rate is high and the perturbation equations are taken only to first order in the mean free path of photons to Thomson scattering. The photons and baryons then behave as a single photon-baryon fluid with a single sound speed, \( c_s = \frac{dP}{d\rho} \approx \sqrt{1/3(1 + R)} \), where \( R \equiv 3\rho_B/4\rho_\gamma \), \( P \) is the total pressure of the fluid, \( \rho \) is the total energy density, \( \rho_B \) is the baryon energy density, and \( \rho_\gamma \) is the photon energy density. Roughly, \( R \approx 0.3(1100/1 + z)(\Omega_{b,0} h^2/0.02) \). The photon temperature fluctuations are described by
\[ \Delta^2 T + \frac{\dot{R}}{1+R} \Delta T + k^2 c_s^2 \Delta T = \dot{\Phi} + \frac{\dot{R}}{1+R} \Phi - \frac{k^2}{3} \Psi, \]  

(2.36)

where \( \dot{x} \equiv \partial x/\partial \tau \). Taking the limit of matter domination gives \( \Psi \approx \Phi \) and \( \dot{\Phi} \approx 0 \).

For scales much smaller than the horizon, the timescale for the variation of \( R \) is much smaller than the the oscillation timescale \( \sim 1/kc_s \). The final approximation is to ignore any variation in \( R \) (a better approximation is to assume only that \( R \) varies slowly compared with the expansion rate and apply the WKB approximation [140], but this analysis suffices for the present purpose). The solution to Equation (2.36) with these simplifications and adiabatic initial conditions \((\Delta T \approx 0)\) is simply

\[ \Delta T_{LS} + \Psi_{LS} \approx \frac{\Psi_{LS}}{3} (1 + 3R) \cos(kc_s\tau_{LS}) - R\Psi_{LS}. \]  

(2.37)

The redshift of the photons leaving the potential wells adds \( \Psi_{LS} \) to the observed anisotropy, so I have written this factor explicitly on the left hand side of (2.37). It is the combination \( \Delta T_{LS} + \Psi_{LS} \) that contributes to the observed anisotropy as discussed in section 2.4.2.

The continuity equation gives the velocity as \( \Delta T = -kV_B/3 \). The rms velocity along the line of sight is then

\[ \frac{V_{B,LS}}{\sqrt{3}} \approx -\Psi_{LS} \frac{1 + 3R}{\sqrt{1+R}} \sin(kc_s\tau_{LS}), \]  

(2.38)

and contributes to the observed anisotropy through the doppler effect.

We could go through the procedure of the last section in order to produce a multipole spectrum, but for our purposes this is unnecessary. We know that each wavenumber contributes to multipoles \( \ell \sim k\eta_{LS} \) and that the \( C_\ell \) correspond to integrals over \( |\Delta T_{LS}(k)|^2 \) and are positive. The power spectrum has two oscillatory
contributions that yield acoustic peaks. From equation (2.37), the intrinsic temperature perturbations contribute to anisotropy peaks at \( kc_s \tau_{LS} = n\pi \). The peaks at odd integer multiples of \( \pi \), \( kc_s \tau_{LS} = (2m + 1)\pi \), correspond to compression of the fluid, while peaks with \( kc_s \tau_{LS} = 2m\pi \) correspond to rarefaction. The zero-point of these oscillations is shifted to \(-R\Psi_{LS}\) so that the compressional peaks are larger than the rarefaction peaks. The relative difference in height between the peaks of rarefaction and compression therefore yield a measure of the baryon density that complements the BBN measurement. The peaks due to the Doppler shift are \( \pi/2 \) out of phase with the compression-rarefaction peaks and are smaller by a factor \( \sim \sqrt{1 + R} \). Consequently, the anisotropy does not go to zero at \( kc_s \tau_{LS} = (m + 1/2)\pi \), but the Doppler peaks are sub-dominant to the compressional peaks.

The peak positions occur for wavenumbers \( k = n\pi/c_s \tau_{LS} \). Thus the first peak occurs at a multipole \( \ell \simeq \pi \eta_{LS}/c_s \tau_{LS} \simeq \pi/c_s a_{LS}^{1/2} \simeq 220 \) in this idealized cosmology. In a more general cosmological model with \( \Omega_{M,0} \neq 1 \), the same ideas hold. The sound horizon at last-scattering, \( r_s \sim c_s \tau_{LS} \) serves as a fundamental distance scale on the last scattering surface because it is a measure of the distance that pressure waves could have traveled prior to decoupling, and thus it gives the length scale of compressions and rarefactions in the photon-baryon fluid. The sound horizon at last-scattering is roughly,

\[
r_s \sim 0.06 \frac{c_s}{\sqrt{\Omega_{M,0} H_0}}.
\]

(2.39)

The angle subtended by this distance on the sky is given by \( \theta \simeq r_s/d_A^{LS} \), so the peak position of the multipole scales as \( \ell_{\text{peak}} \propto d_A^{LS}/r_s \), where the angular diameter distance to the last scattering surface is
\[ d^{\text{LS}}_\Lambda = \frac{1}{H_0 \sqrt{|\Omega_{k,0}|}} \Sigma(\chi(z_{\text{LS}})), \quad (2.40) \]

and \( z_{\text{LS}} \approx 1100 \) is the redshift of last scattering. Thus, peak positions remain roughly fixed only for models with similar values of \( \sqrt{\Omega_{M,0} h^2 d^{\text{LS}}_\Lambda} \). As a consequence, the multipole position of the first acoustic peak can serve as a measure of cosmological parameters and the large-scale curvature of spacetime. The form of Equation (2.40) is far from transparent. Hu et al. [141] have found that the peak position scales approximately as

\[ \ell_{\text{peak}} \propto h^{-0.34} \Omega_{M,0}^{-0.15} (1 - \Omega_{k,0})^{-1.4} \quad (2.41) \]

using numerical solutions. Notice the strong scaling of \( \ell_{\text{peak}} \) with the spatial curvature of the FRW spacetime. This occurs because a fixed length scale, or “standard ruler,” subtends a larger angle in a spherically curved space than in a flat space, and a smaller angle in an hyperbolically curved space. Measurements of the peak positions in the CMB power spectrum provide a robust estimate of curvature.

### 2.4.4 CMB Power Spectra and Cosmological Parameters: A Pictorial

Having given a brief introduction to the physical nature of features in the CMB power spectrum, I now illustrate these features through numerical solutions to the equations of the coupled Einstein-Boltzmann hierarchy. I have performed all of the computations in this section using the publicly available CMBFAST code by Seljak and Zaldarriaga [262], which is based on the earlier COSMICS program by Ma and Bertschinger [194, 195].
First, the CMB power spectrum is sensitive to the matter density of the Universe. The multipole peak positions are only affected by a small amount \([\text{cf.}, \text{Eq. (2.41)}]\), but the amplitudes are affected significantly. In Section 2.4.2, we saw that the ISW effect is ineffective during matter domination. Although, just after decoupling, the Universe is not completely matter-dominated as the cosmological backgrounds of photons and light neutrinos may contribute \(\sim 30\%\) to the energy of the Universe. Thus the second term of Equation (2.28) is effective near the epoch of decoupling. This shows up in the CMB as enhanced anisotropy on the scale corresponding to the horizon at photon-baryon decoupling. On smaller scales, the potential has already had time to decay before decoupling, while larger scales only enter the horizon at later epochs, when the Universe has become increasingly matter-dominated. As such, the height of the first peak increases as \(\Omega_{M,0}/\Omega_{R,0}\) decreases. Subsequent peaks are effected in a similar way thanks to a process called “radiation driving,” where the decay of the gravitational potentials drives the acoustic oscillations, which is also sensitive to the ratio \(\Omega_{M,0}/\Omega_{R,0}\) \([140]\). This is illustrated with four representative models in Figure 2.6. All other parameters have been fixed at \(n = 1, \Omega_{R,0} = 8.06 \times 10^{-5}, \Omega_{\Lambda,0} = 1 - \Omega_{M,0}\) to maintain flatness \(\Omega_{M,0} + \Omega_{\Lambda,0} = 1\), and \(h = 0.72\).

In addition to the \(\Omega_{M,0}\) dependence of peak height, other features are also evident in the spectrum of Figure 2.6. First, the for higher multipoles \(\ell \gtrsim 600\), anisotropy falls off rapidly. This effect is due to the damping of anisotropies during the epoch of photon-baryon decoupling. Decoupling is not the instantaneous process of our idealized models. As the rate of Thomson scattering declines (over a redshift interval of order \(\Delta z \sim 100\)), photons diffuse farther and farther out of the overdensities and underdensities that they occupied previously. The result is a damping of anisotropy on
Figure 2.6: The power spectrum of CMB anisotropy as a probe of the matter density of the Universe. Four power spectra are shown with matter densities of $\Omega_{\text{M},0} = 0.2$ (dotted), $\Omega_{\text{M},0} = 0.3$ (solid), $\Omega_{\text{M},0} = 0.6$ (dashed), and $\Omega_{\text{M},0} = 1$ (dash-dot). In each case, the cosmology is flat and the remaining parameters are fixed as discussed in the text. Notice the rise in the first peak height as the matter density decreases. The CMB power spectrum is plotted as $\sqrt{\ell(\ell+1)C_\ell/2\pi}$.

small scales due to the finite “depth” of the last-scattering surface [266]. Second, the Sachs-Wolfe plateau at low-$\ell$ is evident; however, it is not flat in each model. This is another manifestation of the ISW effect known as the late ISW effect. Anisotropies are enhanced because the Universe transitions out of matter domination and into vacuum energy domination at low redshift, $z \lesssim 0.5$. This causes a boost in the observed anisotropy on scales of order the horizon today due to the second term in Equation (2.28).
Figure 2.7: The power spectrum of CMB anisotropy has the ability to probe the Universal baryon density. Three power spectra are shown with baryon densities $\Omega_{b,0} h^2 = 0.036$ (dotted), $\Omega_{b,0} h^2 = 0.02$ (solid), and $\Omega_{b,0} h^2 = 0.01$ (dashed). The remaining parameters are fixed as discussed in the text. Notice the difference in the relative heights of even and odd peaks.

The CMB power spectrum is also sensitive to the baryon density through the ratios of the amplitudes of the odd acoustic peaks to the even acoustic peaks. I illustrate this in Figure 2.7. I have chosen a fiducial, flat cosmological model with $\Omega_{M,0} = 1 - \Omega_{\Lambda,0} = 0.3$, $h = 0.72$, the standard radiation density, and only varied the physical baryon density $\Omega_{b,0} h^2$ in order to demonstrate this.

It might appear that these effects are not easily separable given a necessarily imperfect data set, and in fact, this is true. There are many well-known degeneracies that must be dealt with in CMB analysis (see [39, 190]). In Figure 2.8, I show two
Figure 2.8: The power spectrum of CMB anisotropy is degenerate in many directions in parameter space. I show two power spectra. The first has $\Omega_{M,0} = 0.3$, $\Omega_{\Lambda,0} = 0.7$, and $h = 0.65$ (solid). The second has $\Omega_{M,0} = 0.7$, $\Omega_{\Lambda,0} = 0.3$, and $h = 0.47$ (dashed). This degenerate set was pointed out by Lineweaver [190].

significantly different models that yield remarkably similar power spectra despite the fact that the matter density in each case is drastically different. Clearly, CMB data alone cannot measure $\Omega_{M,0}$ or $\Omega_{\Lambda,0}$. This is the case for many parameter combinations. In Section 2.4.2 I touted the ability of the CMB to measure the spectral tilt $n$, through the Sachs-Wolfe effect; however, the tilt is also degenerate with several parameters. The Universe reionized at some redshift (because we can see light that would not have passed through neutral hydrogen) so at some point photons scattered off electrons once again. This damps all anisotropy on scales of order the horizon size at reionization and can mimic the effect of a spectral tilt. Furthermore, since the
baryon density effects the relative heights of peaks, it too is degenerate with \( n \) unless data is available out to very large \( \ell \).

One may ask if there is a parameter that the CMB alone provides an unmatched estimate of and there is. Notice that both of the models in Figure 2.8 are spatially flat, \( \Omega_{M,0} + \Omega_{\Lambda,0} = 1 \). The CMB power spectrum, particularly the position of the first acoustic peak, has the ability to place a strong limit on the degree of spatial curvature. We saw this in the form of Equation (2.41) and I demonstrate this explicitly in Figure 2.9. In this Figure, I assume that the galaxy power spectrum and the SNIa data are correctly telling us that \( \Omega_{M,0} = 0.3 \), and I vary \( \Omega_{\Lambda,0} \) to induce spatial curvature. Note that a measurement of the first acoustic peak around \( \ell \approx 220 \), would provide a strong indication of spatial flatness and placed against indications from the galaxy power spectrum that \( \Omega_{M,0}h \approx 0.2 \), and \( h \approx 0.7 \) [109], this would be a strong indication that the Universe is filled with a vacuum energy or cosmological constant.

### 2.5 The Parameters of a Standard Cosmology

The Wilkinson Microwave Anisotropy Probe (WMAP) has recently measured the spectrum of temperature anisotropy in the cosmic microwave background with unprecedented accuracy [24, 25]. The power spectrum measured by WMAP is depicted in Figure 2.10. The points represent the binned and decorrelated multipole moments from the WMAP Explanatory Supplement [331]. Moreover, the WMAP team also measured the cross-correlation between temperature anisotropy and CMB polarization that allow for the construction of a temperature-polarization (TE) power spectrum which breaks some of the degeneracies in the temperature power spectrum. Net polarization of the CMB is another effect that occurs due to the finite time for
Figure 2.9: The positions of the acoustic peaks are very sensitive to the spatial curvature of the Universe. I demonstrate this with three power spectra: one flat model with $\Omega_{M,0} = 0.3$, $\Omega_{\Lambda,0} = 0.7$ (solid); one positively curved model with $\Omega_{M,0} = 0.3$, $\Omega_{\Lambda,0} = 0.9$ (dotted); and one negatively curved model with $\Omega_{M,0} = 0.3$, $\Omega_{\Lambda,0} = 0$ (dashed). first has $\Omega_{M,0} = 0.3$, $\Omega_{\Lambda,0} = 0.7$, and $h = 0.65$ (solid).

the transition between the tightly-coupled limit of the photon-baryon fluid and decoupling. It effectively arises from velocity gradients in the photon-baryon fluid at the epoch of decoupling and as such, the polarization power spectrum has periodic peaks that occur out of phase with the compressional peaks of the $C_\ell$, at wavenumbers where velocities peak [cf., Equations (2.37) and (2.38)]. The measurement of the TE cross-correlation power spectrum yields an additional observable with which
to determine the physical conditions at last-scattering. The observed temperature-polarization angular power spectrum from the WMAP mission is shown in Figure 2.11.

The first and second acoustic peaks are clearly resolved in the WMAP data. The WMAP team measure the position of the first peak with astonishing accuracy, \( \ell_1 = 220 \pm 1 \) [231]. Again, taking \( h = 0.72 \pm 0.08 \) together with this measurement is enough to provide a strong indication that the Universe is very nearly flat [cf., Equations (2.41) and (2.40)]. This breaks a major degeneracy in the determinations of cosmological parameters using the SNIa data set. In Figure 2.3, I show both the confidence contours from the SNIa data analysis and the line in the \( \Omega_{\Lambda,0} - \Omega_{M,0} \) plane corresponding to a flat Universe (\( \Omega_{\Lambda,0} + \Omega_{M,0} = 1 \)). Immediately two consistency checks on the standard model framework can be performed. First, marginalizing the SNIa likelihoods along the line \( k = 0 \) yields an estimate of \( \Omega_{M,0} \) from the SNIa data and only the input of flatness from the CMB. The result is \( \Omega_{M,0} = 1 - \Omega_{\Lambda,0} = 0.28^{+0.08}_{-0.07} \) [239, 281]. Note the excellent agreement with both the determination of \( \Omega_{M,0} \) from SNIa and the baryon fraction of clusters in Equation (2.7) and the shape of the galaxy power spectrum in Equation (2.21). Using this estimate and the baryon fraction, we can then infer the baryon density of the Universe at low redshift. This yields [281, 278]

\[
\Omega_{b,0}h^2 = \Omega_{M,0}f_Bh^2 = 0.019^{+0.008}_{-0.006}, \tag{2.42}
\]

a value that is in remarkable agreement with determinations of the baryon density from primordial nucleosynthesis.

The WMAP team performed a detailed statistical analysis on their CMB data set in order to obtain precise measurements of the cosmological parameters [274, 316].
Figure 2.10: The points show an estimate of the binned and decorrelated CMB power spectrum as measured by the WMAP team. The errorbars reflect the estimated errors at the 1σ level. The two lines represent model fits. The solid line is the best fit to the WMAP data only, with a pure power law primordial power spectrum. The dashed line represents the best fit to a combined data set of WMAP data plus high multipole data from the Arcminute Cosmology Bolometer Array Receiver (ACBAR) [180] and Cosmic Background Imager (CBI) [234] CMB experiments, galaxy power spectrum data from the 2dF Galaxy Redshift Survey [68], and estimates of the power spectrum from the Ly-α forest [77]. As in the previous section, the theoretical predictions were computed using the CMBFAST Einstein-Boltzmann code [262].

The best-fit power spectrum, assuming a power law initial power spectrum is shown in Figure 2.10. In order to break degeneracies within the CMB, they also performed an analysis in which they added data from several other data sets. First, they added data from the ground-based CMB experiments Arcminute Cosmology Bolometer Array Receiver (ACBAR) [180] and Cosmic Background Imager (CBI). They added data on
Figure 2.11: The points show an estimate of the binned and decorrelated CMB temperature-polarization cross-correlation power spectrum as measured by the WMAP team. The errorbars reflect the estimated errors at the 1σ level. The two lines represent model fits. The solid line is the best fit to a data set consisting of WMAP, ACBAR, and CBI CMB data, 2dFGRS galaxy power spectrum data and Ly-α forest power spectrum data, assuming a pure power law primordial power spectrum. The dashed line represents the best fit to a cosmological model with a running spectral index as discussed in the text. All theoretical calculations have been performed using a modified version of the CMBFAST code of Seljak and Zaldarriaga [262].

the galaxy power spectrum from the 2 Degree Field Galaxy Redshift Survey (2dFGRS) [68]. They also added data from the Ly-α forest measurement of absorption of photons at the Ly-α transition along the line-of-sight to distant quasars. These data provide a mapping of the neutral hydrogen density as a function of redshift along the line-of-sight and can be used to estimate the power spectrum on scales $k \sim 1$ Mpc. The WMAP team defined their six parameter set as: (1) the physical matter density
\( \Omega_{M,0} h^2 \); (2) the physical baryon density \( \Omega_{b,0} h^2 \); (3) the Hubble parameter \( h \); (4) the amplitude of the matter power spectrum on the scale \( k_0 = 0.05 \) Mpc\(^{-1}\), \( A \); (5) the power law index of the power spectrum, \( n \) (\( P(k) = A(k/k_0)^n \)); and (6) the optical depth to the surface of last-scattering, \( \tau \). The optical depth to the last-scattering surface provides a measure of the epoch of reionization as a particular redshift maps onto a particular optical depth and anisotropies are damped on all scales larger than the horizon size at reionization by an amount \( \sim e^{-2\tau} \). The WMAP team explicitly assumed a flat Universe \( \Omega_{\Lambda,0} = 1 - \Omega_{M,0} \) in their primary analysis. In other fits, the WMAP team found the Universe to be very nearly flat, \( \Omega_{k,0} = -0.02 \pm 0.02 \), and that allowing for non-flat models did not significantly improve the fits.

The best fits to the WMAP only and combined WMAP + ACBAR + CBI + 2dFGRS + Ly-\( \alpha \) data sets are collected in Table 2.1. The results show a remarkable consistency that is emerging between various determinations of cosmological parameters. The WMAP determination of the Hubble parameter is in superb agreement with the determination from the Hubble Space Telescope Key Project value \( h = 0.72 \pm 0.08 \) [109]. The independent determination of the baryon density by the WMAP team agrees at 1\( \sigma \) with determinations from BBN (2.2) and SNIa plus the cluster baryon fraction (2.42). The prediction of the inflation paradigm that \( n \approx 1 \) on observable scales is also confirmed by this analysis and the WMAP team’s estimate of \( \Omega_{M,0} \) also agrees with all of the other estimates that I have discussed so far!

The fact that large, independent sets of cosmological data all seem to indicate a small acceptable region in parameter space justifies the statement made in the Introduction that cosmology is rapidly converging toward a standard model where the classic cosmological parameters are rather well-measured. The Universe seems to
<table>
<thead>
<tr>
<th>Parameter</th>
<th>WMAP</th>
<th>ALL DATA</th>
<th>WMAP+Hi-ℓ</th>
<th>ALL DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Omega_b,oh^2$</td>
<td>0.024 ± 0.001</td>
<td>0.0226 ± 0.0008</td>
<td>0.022 ± 0.001</td>
<td>0.0224 ± 0.0009</td>
</tr>
<tr>
<td>$\Omega_M,oh^2$</td>
<td>0.14 ± 0.02</td>
<td>0.133 ± 0.006</td>
<td>0.14 ± 0.01</td>
<td>0.135^{+0.008}_{-0.009}</td>
</tr>
<tr>
<td>$h$</td>
<td>0.72 ± 0.05</td>
<td>0.72 ± 0.03</td>
<td>0.71 ± 0.06</td>
<td>0.71^{+0.04}_{-0.03}</td>
</tr>
<tr>
<td>$n(k_0)$</td>
<td>0.99 ± 0.04</td>
<td>0.96 ± 0.02</td>
<td>0.91 ± 0.06</td>
<td>0.93 ± 0.03</td>
</tr>
<tr>
<td>$\frac{dn}{d\ln k}$</td>
<td>N/A</td>
<td>N/A</td>
<td>-0.055 ± 0.038</td>
<td>-0.031 ± 0.016</td>
</tr>
<tr>
<td>$A$</td>
<td>0.9 ± 0.1</td>
<td>0.75 ± 0.08</td>
<td>0.9 ± 0.1</td>
<td>0.83^{+0.09}_{-0.08}</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.166^{+0.070}_{-0.071}</td>
<td>0.117^{+0.057}_{-0.053}</td>
<td>0.20 ± 0.07</td>
<td>0.17 ± 0.06</td>
</tr>
</tbody>
</table>

Table 2.1: WMAP best-fit parameters to flat ΛCDM models. Columns (2) and (3) reflect best-fit parameters assuming a power law primordial power spectrum. Columns (4) and (5) contain the best-fit parameters assuming a “running” of the spectral index, $dn/d\ln k \neq 0$, but no “running-of-running,” $d^2n/d(\ln k)^2 = 0$. Column (2) corresponds to a fit to the WMAP data only. Columns (3) and (5) correspond to fits of the combined data set from WMAP, ACBAR, CBI, 2dFGRS, and the Ly-α forest. Column (4) corresponds to a fit to the WMAP data plus the high-\ℓ CMB data from ACBAR and CBI. The quoted values are from Spergel et al. [274].
be expanding at a rate of $H_0 \approx 70$ km s$^{-1}$/ Mpc today and the Universe is very nearly spatially flat, as predicted by the simplest inflation models. In this model, “ordinary” baryonic matter makes up only $\sim 4\%$ of the total energy density of the Universe. The energy budget of the Universe seems to be dominated by about $26\%$ cold dark matter (CDM) and about $70\%$ cosmological constant, $\Lambda$, or vacuum energy (I will use the terms cosmological constant and vacuum energy interchangeably because they are equivalent in their effects on the dynamics of the FRW spacetime). This model has therefore been dubbed the standard $\Lambda$CDM model. Given the Hubble expansion rate and the cosmological parameters, the age of the Universe (or at least the amount of proper time experienced by comoving observers since the earliest time we have reliable information about the Universe at BBN), is about $t_0 \approx 13.5$ Gyr. Furthermore, in this model, the spectrum of primordial fluctuations is nearly scale-invariant with $n \approx 1$, as predicted by inflation.

The analysis performed by the WMAP team did yield surprises, however. The epoch of reionization imprints a polarization signal on the CMB because photons can once again scatter off of electrons at this epoch. The WMAP experiment detected the signature of very early reionization ($i.e.$, a large optical depth) at $z \sim 17$. This can have non-trivial consequences on structure formation models because it means that something (stars, quasars, ...) must have been producing enough photons at high redshift to ionize the neutral hydrogen of the Universe [270]. Second, notice that the best-fit value of the tilt parameter decreases when data on small-scale fluctuations are added to the data set (Table 2.1). The WMAP team found that their fits were improved if they allowed the tilt to run, adding a second parameter $dn/d\ln k$ (they explicitly assumed that $d^2n/d(\ln k)^2 = 0$ so that the running is constant) [274, 316].
The best-fit models with running are also shown in Table 2.1 and in Figure 2.12. The theoretical spectra in Figure 2.12 were calculated using a version of the CMBFAST code [262] that was modified in order to accommodate spectral index running.

Inflation models do predict some non-zero running; however, this is typically too small to be measured \(|dn/d\ln k| \ll 0.01\) so confirmation of a strong running in the spectrum would severely constrain models of inflation [237]. Interestingly, if the spectral index does run from \(n > 1\) on large scales to \(n < 1\) on small scales, this implies that the inflaton potential should have a point of inflection on a nearby scale in the context of slow-roll inflation (see Chapter 5). What is more, the properties of observed galaxies are determined by fluctuations on scales \(k_{\text{galaxy}} \gg 1 \text{ Mpc}\) that are not accessible to CMB or other measurements. The predicted properties of galaxies can be significantly affected by this addition to the scale-invariant standard paradigm and, conversely, the measured properties of galaxies and sub-galactic structure may yield information about the power spectrum, inflation, and physics at the highest energies and earliest epochs of the Universe.
Figure 2.12: The points show an estimate of the binned and decorrelated CMB power spectrum as measured by the WMAP team. The errorbars reflect the estimated errors at the 1σ level. The lines represent model fits. The solid line is the best fit the the WMAP, ACBAR, and CBI CMB data only, with a primordial power spectrum with spectral index running (d\(n/d\ln k\) = −0.055. The dashed line represents the best fit to a combined data set of WMAP data plus high multipole data from the ACBAR and CBI experiments, galaxy power spectrum data from 2dFGRS [68], and estimates of the power spectrum from the Ly-\(\alpha\) forest [77]. The dotted line represents the best fit pure power law model (\(dn/d\ln k = 0\)) for all of these data sets (also shown in Figure 2.10). Again, the theoretical curves were calculated using a modified version of the CMBFAST code [262].
CHAPTER 3

COSMOLOGICAL CONSTRAINTS ON RELATIVISTIC
AND UNSTABLE PARTICLE SPECIES

3.1 Introduction

One of the advantages of a well-constrained standard model of cosmology is that it allows one to place stringent and non-trivial constraints on any physics that may lead to modifications in the standard scenario. In this way, the framework of the ΛCDM cosmology constitutes a bridgehead from which one can cross into studies of physics at the highest energies. One can use cosmology to produce robust constraints on modifications to the Standard Model of particle physics or to learn about the properties of any scalar fields that may participate in symmetry breaking at the GUT or higher scales [96, 98]. It may even be possible to probe Planck-scale physics with accurate measurements of the primordial power spectrum [97, 10].

Cosmological constraints on the relativistic energy density (RED) of the Universe, particularly the bound derived from the successful prediction of the primordial abundances of D, 4He, and 7Li, provide perhaps the strongest connections between cosmology and particle physics. In this Chapter I examine constraints on the RED (To be specific, I consider constraints on the energy density of any species that obeys

64
an equation of state \( P = \rho/3 \) at various cosmological epochs and discuss how these bounds can be used to limit the mass and lifetime of an unstable big bang relic particle. With the vast amounts of cosmologically relevant observational data now available, these bounds are more restrictive than ever imagined.

In this Chapter, I review the classic big bang nucleosynthesis bound and the bound from the cosmic microwave background anisotropy power spectrum. I present a new bound derived from the magnitude-redshift relation of Type Ia supernovae, and I update the bound from the clustering of large scale structure (LSS). Although I will attempt to derive bounds that are independent of additional cosmological parameters, where necessary I will borrow constraints from other techniques in order to further limit the relativistic energy density of the Universe. For example, when needed I adopt the hard bound \( 0.1 \leq \Omega_{M,0} \leq 0.5 \), consistent with recent measurements of the matter density from weak lensing [315], the rich cluster baryon fraction [202, 281, 118], mass-to-light ratio estimates [15], and the power spectrum of the Ly\( \alpha \) forest [77]. I also adopt a limit on the Hubble parameter \( (h \equiv H_0/100 \text{ km s}^{-1}\text{Mpc}^{-1}) \) of \( 0.56 \leq h \leq 0.88 \) in accordance with the 2\( \sigma \) range of the HST Key Project [109]. These bounds primarily serve to break degeneracies that we already encountered in the discussions of Chapter 2.

3.2 Constraints on Relativistic Energy

3.2.1 Big Bang Nucleosynthesis Constraint

It has long been known that an increase in the RED of the Universe during BBN leads to an overproduction of \(^4\text{He} \) \([279]\) (BBN reviews include \([225]\)). Prior to freeze-out of the weak interactions, the neutron-to-proton ratio \((n/p)\) tracks its
equilibrium value as discussed in Section 2.1. As the Universe expands and cools, the weak interactions freeze-out and \((n/p)\) freezes in. Freeze-out occurs when the weak interaction rates become comparable to the expansion rate \(H\), given by \(H^2 = 8\pi G\rho/3\), where \(\rho\) is the mean cosmological energy density. Once deuterium can be synthesized nearly all available neutrons are incorporated into \(^4\)He. Increasing the energy density of the Universe during BBN increases the expansion rate, causing the weak interactions to freeze-out at a higher temperature and therefore, at a higher value of \((n/p)_F\), in the notation of Section 2.1. An increased expansion rate also allows less time for free neutron decay. Both effects leave more neutrons available for \(^4\)He synthesis when deuterium comes to be thermodynamically favored (\(T \sim 70\) keV).

In addition, BBN production of \(^4\)He slowly increases with the baryon density or equivalently, the baryon-to-photon ratio, \(\eta \equiv n_B/n_\gamma\). As I've shown, BBN production of deuterium is a rapidly decreasing function of \(\eta\) and the baryon density can be deduced from the D/H ratio observed in QSO absorption line systems (see Figure 2.1). Furthermore, because \(^4\)He has been produced in stars since BBN, the observed \(^4\)He abundance places an upper limit on its primordial abundance. Armed with this upper bound to primordial \(^4\)He and an independent measure of \(\eta\), we can place a bound on the RED during the BBN epoch (\(z \sim 10^9\)).

I illustrate the current status of the BBN bound to the RED as follows (see the recent analyses of [55, 191] for detailed calculations using maximum-likelihood techniques). I adopt a very liberal range for the primordial D/H ratio that accommodates recent observations of deuterium in several QSO absorption line systems [56, 184, 226, 240]: \(1.75 \leq 10^6(D/H) \leq 5.0\). I then bound the RED by taking a
Figure 3.1: The BBN bound on the RED. The region bounded by the dashed curves delimits the parameter space for which the theoretically predicted and observed values of D/H are consistent. Likewise, the data on $Y_P$ are consistent with predictions in the region below the solid curve. The intersection, marked by the circle, shows the maximum value of $N_{\nu,\text{eff}}$ that can be made consistent with these limits on $Y_P$ and D/H.

A conservative upper limit on the primordial $^4\text{He}$ mass fraction, $Y_P \leq 0.25$. This bound is consistent with both the low [224] and high [145] zero-metallicity extrapolations of low-metallicity HII region data on $^4\text{He}$ (but see [120]).

It is conventional to express limits on relativistic energy in terms of the energy carried by an effective number of light neutrino species $N_{\nu,\text{eff}}$, and I follow this convention. In order to calculate the energy density in neutrino species, it is necessary to get a handle on the neutrino temperature, $T_\nu$. Then, the energy density in neutrinos is simply
\[ \rho_\nu \simeq \frac{7\pi}{120} T_\nu^4. \]  

(3.1)

The neutrino temperature relative to the photon (\textit{i.e.}, CMB) temperature is easy to estimate using the fact that the cosmological expansion is very nearly adiabatic. The entropy density of the Universe is dominated by relativistic species and can be written as

\[ s = \frac{2\pi^2}{45} g_* s T^3, \]  

(3.2)

where \( T \) is the CMB temperature (at the redshift of interest) and \( g_* s \) is a parameter that counts the number of relativistic species,

\[ g_* s = \frac{7}{8} \sum_{i = \text{fermions}} g_i \left( \frac{T_i}{T} \right)^3 + \sum_{j = \text{bosons}} g_j \left( \frac{T_j}{T} \right)^3. \]  

(3.3)

In Equation (3.3), \( g_i \) represents the number of internal degrees of freedom of species \( i \) and \( T_i \) is the temperature that describes the phase space distribution of \( i \). Adiabatic expansion then implies that the combination \( g_* s T^3 \) = constant. At temperatures \( T \gg 1 \text{ MeV} \) and \( T \gg m_e \), where \( m_e = 0.511 \text{ MeV} \) is the electron mass, the relativistic species that contribute to the entropy density are electrons, neutrinos, and photons and they are all in thermal equilibrium. As discussed in the previous Chapter, neutrinos decouple from the thermal bath when the weak interactions freeze out at a temperature \( T \sim 1 \text{ MeV} \). After neutrino decoupling, the temperature falls below the electron mass, at which time it becomes thermodynamically favorable to have photons rather than electron-positron pairs, so the \( e^+ - e^- \) pairs annihilate, giving their entropy to the photons only (recall, the weak interaction rates are too slow compared to the expansion rate and the electromagnetic interactions to couple the

68
neutrinos to the annihilating pairs). Prior to annihilation, the photons and electrons are coupled, and they contain an entropy given by (3.2) with an effective $g_{\ast S}^{\text{init}} = 11/2$. After $e^+ - e^-$ annihilation, all of this entropy is dumped into $g_{\ast S}^{\text{final}} = 2$ states corresponding to the photons. Conservation of entropy therefore implies that the photon temperature is raised relative to the neutrino temperature by a factor $(11/4)^{1/3}$ due to this process. After $e^+ - e^-$ annihilation (temperatures $T < m_\nu$), the neutrino temperature is thus

$$T_\nu = (4/11)^{1/3}T. \quad (3.4)$$

This is a specific example of a more general consequence of entropy conservation. Any particle that decouples when there are $g_{\ast,\text{D}}$ effective degrees of freedom has its relic abundance (by number) suppressed by a factor of $2/g_{\ast,\text{D}}$ relative to photons.

Using the neutrino temperature in (3.4), the RED at some redshift $z$, after $e^+ e^-$ annihilation, is related to $N_{\nu,\text{eff}}$ by

$$\theta(z) \equiv \frac{\rho_R(z)}{\rho_{\gamma\nu}(z)} = [1 + 0.135(N_{\nu,\text{eff}} - 3)], \quad (3.5)$$

where I scale the RED by $\rho_{\gamma\nu}(z)$ which is the standard model RED carried by photons and three massless neutrino species (with a current CMB temperature of 2.726 K [201], $\Omega_{\gamma\nu,0} \equiv \rho_{\gamma\nu}(z = 0)/\rho_{\text{crit}} = 4.18h^{-2} \times 10^{-5}$).

In Figure 3.1, I show regions of the $\eta - N_{\nu,\text{eff}}$ plane allowed by the observed $^4$He and D abundances. The consistency with observational data requires that $N_{\nu,\text{eff}} \lesssim 3.4$ or, in terms of the RED during nucleosynthesis $\theta_{\text{BBN}} \equiv \theta(z = 10^9)$,

$$\theta_{\text{BBN}} \lesssim 1.05. \quad (3.6)$$

In Figure 3.2, I show the amount of extra relativistic energy permitted during BBN.
Figure 3.2: Constraints on the RED. The horizontal lines denote the maximum value of $\theta(z)$ allowed by BBN, CMB, LSS and SNIa. Notice that the BBN constraint in the lower right-hand corner is hardly noticeable (i.e., very strict). By extending the limits to higher redshift, I have made the assumption that after BBN relativistic energy is only injected and never removed. The curved line shows $\theta(z)$ in a scenario where a massive big bang relic with $M_{\text{keV}}Y = 1.2 \times 10^{-3}$ decays with lifetime $\tau_{\nu} = 2 \times 10^9$. There are additional CMB and LSS constraints on relativistic energy injected after $z \sim 1000$; however, they do not rule out the decay shown above (see sections 3.3 and 3.4).

There are two important caveats regarding the BBN RED bound. First, although (3.6) is a rather stringent bound on the RED, corresponding to an increase of $\sim 5\%$ over the standard $\rho_{\gamma \nu}$, it only applies during the epoch of nucleosynthesis. Relativistic energy injected after BBN is not subject to this bound (in Section 3.3, I discuss a toy model in which relativistic energy is injected into the Universe after the light nuclide abundances have been fixed). Second, the BBN RED bound is flavor
dependent in the sense that extra relativistic energy in the form of degenerate electron neutrinos changes the rates of the weak interactions that inter-convert protons and neutrons thereby changing theoretical BBN yields. Introducing an appropriate amount of electron neutrino degeneracy can always compensate for excess relativistic energy present during nucleosynthesis [223, 153, 166], nullifying the BBN RED constraint. This effect can be understood by applying Le Chatelier’s principle to the thermodynamics that determine the neutron-to-proton ratio. The \( (n/p) \) ratio is kept at its equilibrium value prior to weak interaction freeze out by interactions like \( \bar{\nu}_e + p \leftrightarrow e^+ + n, \nu_e + n \leftrightarrow p + e^- \), and \( n \leftrightarrow p + e^- + \bar{\nu}_e \). Introducing an excess of \( \nu_e \) over \( \bar{\nu}_e \) pushes the equilibrium state toward a state of fewer neutrons. In particular, the neutron-to-proton ratio becomes

\[
\left( \frac{n}{p} \right) \approx \exp \left( -\frac{Q}{T} - \frac{\mu_{\nu_e}}{T}\right),
\]

where \( \mu_{\nu_e} \) is the chemical potential of the electron neutrino. Thus, the introduction of an appropriate electron neutrino chemical potential counteracts the addition (or subtraction) of relativistic energy. This degeneracy has been discussed by several author [223, 153, 166]. I illustrate this in Figure 3.3 which was kindly provided by J. P. Kneller. Such models seem contrived, however, because it is necessary to introduce two new parameters, \( N_{\nu,e,f} \) and \( \xi_e = \mu_{\nu_e}/T \), to a model that works well without them. Furthermore, if lepton number is similar to the baryon number of the Universe, \( n_B/n_\gamma \sim 10^{-9} \), as one might suppose by naturalness arguments, the non-zero electron chemical potential has a negligibly small value. As such, the BBN bound on the RED represents a non-trivial bound that models must work within.
Figure 3.3: The four panels in this figure show how neutrino degeneracy provides a loophole for any bound on the RED from BBN. Each panel shows the observationally acceptable region of the $\eta - N_{\nu,\text{eff}}$ plane for different values of the electron neutrino degeneracy parameter, $\zeta_e = \mu_{\nu_e}/T_\nu$. In each panel, the solid lines delimit the boundary for observational acceptable $^4\text{He}$, D, and $^7\text{Li}$ [166]. This figure is courtesy J. P. Kneller.

unless the models also provide a mechanism for the production of a very large lepton number.

3.2.2 The CMB Constraint

The CMB anisotropy power spectrum is also sensitive to the RED of the Universe at recombination, primarily through the early ISW effect (there is a small shift of peaks to angular scales as well due to the increased energy density and later epoch
of matter-radiation equality, but this is a much smaller effect). The erosion of gravitational potentials due to incomplete matter domination at recombination leads to a boost in power, particularly near the first acoustic peak (for a review see [299]). I illustrate this in Figure 3.4, where I show several model power spectra with different values of $N_{\nu,\text{eff}}$. Adding radiation effectively boosts the amplitude of fluctuations at the first acoustic peak. This is essentially the opposite of the effect of increasing $\Omega_{M,0}$ seen in Figure 2.6. This fact indicates the degeneracy within those parameters: very similar CMB power spectra are obtained if one varies $\Omega_{\gamma,0}$, but keeps the ratio $\Omega_{\gamma,0}/\Omega_{M,0}$ approximately constant (so that the epoch of matter-radiation equality happens at the same redshift). As a result, a reasonable prior must be taken on $\Omega_{M,0}$ in order to obtain stringent bounds on $\Omega_{\gamma,0}$ or, equivalently, $N_{\nu,\text{eff}}$ [166].

CMB constraints on relativistic energy complement those from BBN in two ways. First, the CMB constraints are flavor independent; the power spectrum measures relativistic energy regardless of its form whereas the BBN bounds treat electron neutrinos and all other forms of energy distinctly. Second, the CMB constrains the RED at later epochs so that the CMB can be used to study the injection of relativistic energy after the epoch of nucleosynthesis, for example, by the decays of extremely massive particles.

Wang, Tegmark & Zaldarriaga [320] have compiled a combined CMB data set including the recent results from the BOOMERANG [84], MAXIMA [18] and DASI [124] experiments. Hannestad [128] has performed a likelihood analysis on this combined data set and found, with weak priors on the Hubble and tilt parameters and the baryon density, $N_{\nu,\text{eff}} \leq 19$ with 95% confidence. In terms of the RED present at recombination, $\theta_{\text{CMB}} \equiv \theta(z \approx 1100)$, this bound is
Figure 3.4: The power spectrum of CMB anisotropy as a probe of $N_{\nu,\text{eff}}$. The points represent the power spectrum measured by the WMAP CMB anisotropy experiment. The solid line represents the best fit to the WMAP data with $N_{\nu,\text{eff}} = 3$ as in the standard model. The dashed line represents the predicted spectrum for the same model with $N_{\nu,\text{eff}} = 15$, while the dotted line shows the $N_{\nu,\text{eff}} = 0$ case. The purpose of this Figure is to illustrate the effect only. One should not draw conclusions directly from this Figure because I have kept all other parameters. In fitting CMB data, degeneracies (particularly between $N_{\nu,\text{eff}}$, $\Omega_{M,0}$, and $h$) can compensate for each other. As a result the bound on $N_{\nu,\text{eff}}$ is far weaker than it would appear from this Figure.

\[ \theta_{\text{CMB}} \lesssim 3.2 \] (3.8)

as depicted in Figure 3.2. A Preliminary analysis of the WMAP data indicates that the limit may be as low as $N_{\nu,\text{eff}} \leq 13$ or $\theta_{\text{CMB}} \lesssim 2.4$ [129].

A cautionary note is in order. Although the priors chosen by Hannestad are quite conservative, it must be borne in mind that the bound in (3.8) does depend upon these
priors. For example, by adopting tighter, yet reasonable, prior constraints on $h$ and $\eta$, Hannestad shrinks the above limit to $N_{\nu,\text{eff}} \leq 14$ at 95% confidence [128, 129] (see also [131]). Kneller et al., [166] have explored the prior dependence of these bounds in detail, showing that the bound on $N_{\nu,\text{eff}}$ scales with the upper bound chosen for $\Omega_M$ (i.e., interesting CMB bounds on the RED require a prior constraint on $\Omega_M$).

### 3.2.3 Constraints from Type Ia Supernovae

As I discussed in Section 2.2, given a distant population of so-called “standard candles”, the magnitude-redshift relation is a powerful way to determine cosmological parameters directly [256]. In a standard FRW cosmology, the apparent bolometric magnitude, $m(z)$, of a standard candle is related to its absolute bolometric magnitude, $M$, and redshift by

$$m(z) = M + 5 \log(D_L(z, \Omega_{M,0}, \Omega_{\Lambda,0}, \Omega_{R,0})) - 5 \log(H_0) + 25, \quad (3.9)$$

where $H_0$ is the Hubble parameter in kms$^{-1}$Mpc$^{-1}$. $D_L$ is the “Hubble-constant-free” luminosity distance in Mpc, given by

$$D_L = c(1+z)\sqrt{\frac{1}{|\Omega_k|} \Sigma \left( \sqrt{|\Omega_k|} \int_0^z \frac{d\bar{z}}{(1+z)^2(1+\Omega_{M,0}z) + (1+z)^2(\Omega_{R,0} - \bar{z}(2+\bar{z})\Omega_{\Lambda,0}} \right)}$$

where $c$ is in kms$^{-1}$, $\Omega_k \equiv 1 - \Omega_{M,0} - \Omega_{\Lambda,0} - \Omega_{R,0}$, and

$$\Sigma(x) = \begin{cases} \sin(x) & \text{if } \Omega_k < 0 \\ x & \text{if } \Omega_k = 0 \\ \sinh(x) & \text{if } \Omega_k > 0. \end{cases}$$

This relation is the same as the one given in Equation (2.6) except in this case the explicit dependence on the Hubble parameter has been removed from the term that depends on the cosmological parameters. In this case, $H_0$ is treated as nuisance
parameter. In what follows, I treat the quantity $M - 5 \log(H_0)$ as an effective absolute luminosity and marginalize the likelihoods over this nuisance parameter. The matter, radiation and cosmological constant energy densities enter $D_L$ with different powers of $z$, making it possible to utilize observations of standard candles over a range of redshifts to determine the cosmological density parameters (for a review, see [117]).

Two groups, the Supernova Cosmology Project [239] and the High-z Supernova Search Team [258], have been engaged in a systematic study of the magnitude-redshift relation of high-redshift, type Ia supernovae in an effort to constrain $\Omega_{M,0}$ and $\Omega_{\Lambda,0}$. I have re-analyzed the data of Perlmutter et al. [239] allowing for a non-negligible contribution from $\Omega_{R,0}$. I have assumed flatness to be a robust result of CMB measurements [231, 274] and have performed a maximum-likelihood analysis. I assigned likelihoods as $\mathcal{L} \propto e^{-\chi^2/2}$ subject to the priors $\Omega_{M,0} \geq 0$ and $\Omega_{R,0} \geq 0$ in order to derive constraints on $\Omega_{M,0}$ and $\Omega_{R,0}$ with $\Omega_{\Lambda,0} \equiv 1 - \Omega_{M,0} - \Omega_{R,0}$\footnote{Relaxing flatness yields a less restrictive bound on $\Omega_{R,0}$ but much of the favored region of parameter space would be ruled out by age considerations, CMB measurements, or the aforementioned bounds on $\Omega_{M,0}$.}. The $\chi^2$ function is computed from the data points $m_i^{\text{data}}$, and theoretical calculations $m_i^{\text{theory}}$ via

$$\chi^2 = \sum_{i=\text{data points}} \frac{(m_i^{\text{data}} - m_i^{\text{theory}})^2}{\sigma_i^2}. \quad (3.10)$$

The 1σ measurement error is denoted by $\sigma_i$.

In Figure 3.5, I show the constraints on $\Omega_{M,0}$ and $\Omega_{R,0}$ from SNIa. The projection $\Omega_{R,0} = 0$ is consistent with earlier analyses that found $\Omega_{M,0} \sim 0.3$ and $\Omega_{\Lambda,0} \sim 0.7$ assuming the RED to be negligible. Allowing for relativistic energy density, I find a degenerate set of $\Omega_{M,0}$ and $\Omega_{R,0}$ that are consistent with the SNIa data: the high-matter-content (i.e., 30% matter, 70% cosmological constant) flat Universe and the
Figure 3.5: 68% (filled) and 95% (open) confidence contours in the \( \Omega_{M,0}\)-\( \Omega_{R,0} \) plane. For illustration, I show several labeled isochrones (dotted) and the elliptical 95% confidence contour obtained by adding an additional prior constraint of \( \Omega_{M,0} = 0.3 \pm 0.1 \) (solid, heavy). Also shown are contours of constant effective \( m_B \) for SNIa at redshifts of \( z = 0.5 \) (dashed at \( m_B = 23, 23.2 \)) and \( z = 1.5 \) (dash-dot at \( m_B = 26, 26.2 \)).

A high-radiation-content (i.e., 20% radiation, 80% cosmological constant) flat Universe are equally good fits. Observe that regardless of whether relativistic energy is allowed or not, the SNIa data require a large cosmological constant \( \Omega_{\Lambda,0} \ll 0.9 \). The need for a large cosmological constant or vacuum energy density to make the Universe flat seems unavoidable!

The best-fit line and 95% confidence contour are approximately fit by the relation [340]

\[
\Omega_{R,0} + 0.62\Omega_{M,0} \simeq 0.17 \pm 0.08. 
\] (3.11)
It is not surprising that the data pick out this degenerate valley in the $\Omega_{M,0}$-$\Omega_{R,0}$ plane because most of the SNIa data are from $z \sim 0.5$ and the degenerate valley represents the parameters with approximately constant luminosity distance at this redshift. The degeneracy can be further understood with the help of Figures 3.5 & 3.6. Dashed and dot-dashed lines in Figure 3.5 depict contours of constant apparent magnitude at redshifts $z = 0.5$ and $z = 1.5$ respectively. As most of the SNIa data lie near $z \sim 0.5$, the confidence region is nearly parallel to lines of constant $m_B$ at $z = 0.5$. At higher redshifts, lines of constant apparent magnitude have a more shallow slope and are closer together, thus observations of SNIa at $z > 0.5$ can break the matter-radiation degeneracy.

Figure 3.6 (which is a copy of Figure 2.2 reprinted here for convenient reference) further illustrates the large lever arm of high-$z$ SNIa for cosmological parameter estimation. Notice that the magnitude-redshift relation in a high-matter-content Universe ($\Omega_{M,0} = 0.3$, $\Omega_{R,0} = 0$) is quite similar to that in a high-radiation-content Universe ($\Omega_{R,0} = 0.2$, $\Omega_{M,0} = 0$) at $z \lesssim 0.5$. At higher redshift, one begins to probe the epoch prior to matter-$\Lambda$ and/or radiation-$\Lambda$ equality; the cosmological constant becomes increasingly unimportant compared to radiation and/or matter and the two curves begin to diverge. The SCP data only extend to $z = 0.83$, therefore they cannot be used to scrutinize this earlier phase and they cannot distinguish between a high-matter-content Universe and a high-radiation-content Universe. The proposed Supernova Acceleration Probe (SNAP) [290] may have the ability to break this degeneracy by observing many more SNIa at significantly higher redshift.

Marginalizing over $\Omega_{M,0}$ by integrating the likelihood, I obtain a bound on the RED at low redshift $\theta_{\text{SNIa}} \equiv \theta(z \lesssim 0.5)$,
Figure 3.6: The high-redshift portion of the Hubble diagram for the SCP data (with 1σ error bars) alongside the magnitude-redshift relation in several flat cosmologies: $\Omega_{M,0} = 0.3$ and $\Omega_{R,0} = 0$ (heavy, solid), $\Omega_{M,0} = 0$ and $\Omega_{R,0} = 0.2$ (dashed), $\Omega_{\Lambda,0} = 1$ (dotted), $\Omega_{M,0} = 1$ (dash-dot), $\Omega_{R,0} = 1$ (light, solid).

$$\theta_{\text{SNIa}} \lesssim 3.4h^2 \times 10^3 (68\%) \text{ or } 4.8h^2 \times 10^3 (95\%).$$

(3.12)

In terms of the RED today, these bounds are $\Omega_{R,0} \leq 0.14$ (68\%) and $\Omega_{R,0} \leq 0.20$ (95\%).

The SNIa constraint on the RED applies during recent epochs, namely, $z \lesssim 0.5$, as can be seen in Figure 3.2. Of course, in light of other estimates of cosmological parameters, the allowed region in Fig. 3.5 is not equally probable. Estimates of the contemporary matter density of the Universe favor the region $\Omega_{M,0} = 0.3 \pm 0.1$ and would slightly reduce the SNIa upper bound on the RED. This is illustrated in Figure
3.5 where I also plot the 95% confidence contour obtained with the additional prior constraint $\Omega_{M,0} = 0.3 \pm 0.1$. This prior constraint on the matter density reduces the 95% bound on $\Omega_{R,0}$ by about 25% to $\Omega_{R,0} \leq 0.15 [340]$.

Note also that the SNIa bound on the contemporary RED is more stringent than bounds that follow from the requirement that the expansion age of the Universe be at least as large as the ages of the oldest objects in the Universe. In Figure 3.5, the entire region delineated by SNIa data corresponds to an acceptable age. In particular, a high-radiation-content Universe with $\Omega_{M,0} = 0.04$, $\Omega_{R,0} = 0.20$, $\Omega_{\Lambda,0} = 0.76$ and a Hubble parameter at the extreme lower limit, $h = 0.56$, is 13.8 Gyr old. Even with $h = 0.72$, $t_0 \approx 10.7$ Gyr, a value that is not grossly inconsistent with the ages of the oldest stars and globular clusters [59, 58, 304].

### 3.2.4 Large Scale Structure Constraints

The effect of additional relativistic energy on the growth of large scale structure (LSS) has been studied by numerous authors [305, 280, 326, 206, 30, 306]. Typically, the introduction of “hot dark matter” was considered a means to suppress power on small scales and thus reconcile an $\Omega_{M,0} = 1$, Cold Dark Matter (CDM) cosmology with the observed power spectrum derived from galaxy surveys [188]. Conversely, too much relativistic energy adversely affects the growth of structure and therefore LSS can be used to constrain the RED. In light of mounting evidence, the CDM paradigm has given way to the so-called $\Lambda$CDM paradigm with $\Omega_{M,0} \sim 0.3$ and $\Omega_\Lambda \sim 0.7$. With this in mind, I revise previous work in order to constrain the relativistic energy content of the Universe using LSS.
As discussed in Section 2.3, the most striking feature of a CDM or ΛCDM type power spectrum is a break in the power law at, roughly speaking, the comoving horizon scale at matter-radiation equality

\[ \lambda_{\text{EQ}} \simeq 16(\Omega_{\text{M,0}}h)^{-1} \, h^{-1} \, \text{Mpc}. \]  

(3.13)

This feature arises because the growth of sub-horizon sized perturbations is quelled by the cosmological expansion during radiation domination. Hence, perturbations on scales smaller than \( \lambda_{\text{EQ}} \) are suppressed by a factor \( \approx (\lambda/\lambda_{\text{EQ}})^2 \) relative to scales that were super-horizon sized at matter-radiation equality.

If the RED of the Universe is contained entirely in photons and three light neutrino species, and if the primordial power spectrum is nearly scale-invariant \( (P(k) \propto k^n, \ n \approx 1) \), the ΛCDM matter power spectrum can be expressed in terms of only one quantity, the shape parameter \( \Gamma \simeq \Omega_{\text{M,0}}h \) [19] (this neglects the effect of baryons, see [289]). Several authors have used LSS observations on linear scales to infer acceptable values of the shape parameter, for example Ref. [100]. Our goal is to present robust, conservative bounds on the RED, so I adopt one of the more permissive of these determinations, the 95\% range

\[ 0.06 \leq \Gamma \leq 0.46, \]  

(3.14)

quoted by Efstathiou & Moody [99].

In the presence of excess relativistic energy, the horizon scale at matter-radiation equality is no longer given by (3.13). Rather,

\[ \lambda_{\text{EQ}} \simeq 16(\Omega_{\text{M,0}}h)^{-1} \sqrt{\Omega_{\text{R,0}}/\Omega_{\gamma,0}} \, h^{-1} \, \text{Mpc} \]  

(3.15)
and the effective shape parameter is therefore given by [326]

\[
\Gamma \simeq \Omega_{M,0} h \theta^{-1/2}.
\] (3.16)

Taking the lower bound \( \Gamma > 0.06 \) we immediately come upon a generic constraint on the mean RED [340]:

\[
\theta \leq \left( \frac{\Omega_{M,0} h}{0.06} \right)^2.
\] (3.17)

With my conservative assumptions that \( \Omega_{M,0} \leq 0.5 \) and \( h \leq 0.88 \), the corresponding restriction on the RED during (and prior to) the epoch of matter-radiation equality, \( \theta_{\text{LSS}} \equiv \theta(z = z_{\text{EQ}}) \), is

\[
\theta_{\text{LSS}} \lesssim 54
\] (3.18)

where \( 1 + z_{\text{EQ}} \equiv \Omega_{M,0}/\Omega_{R,0} \) [340]. The relative weakness of this bound is due to my conservative lower bound on \( \Gamma \). Taking the 95\% band of Eisenstein & Zaldarriaga [100], \( 0.15 \leq \Gamma \leq 0.58 \), results in \( \theta_{\text{LSS}} \lesssim 9 \).

Note that this constraint (3.18) allows the epoch of matter-radiation equality to be at redshifts as low as \( z_{\text{EQ}} \approx 120 \). One can obtain a more stringent bound on the RED by following an argument invoked by Turner, Steigman & Krauss [305], Steigman & Turner [280] and Turner & White [306]. Assuming a nearly scale-invariant primordial power spectrum, data from the Cosmic Background Explorer (COBE) Differential Microwave Radiometer (DMR) experiment [23] indicate that the rms density contrast at horizon crossing is on the order of \( \delta_H \sim \text{few} \times 10^{-5} \) [52, 53]. Meanwhile, measurements of the galaxy correlation function reveal nonlinear clustering on scales smaller than a critical scale, \( \lambda_{\text{NL}} \sim 5h^{-1} \text{Mpc} \) [334, 150] (see also Section 2.3). I
adopt the conservative constraint that perturbations on scales smaller than $\lambda_{NL}$ must have grown by at least a factor of $\gamma_{\text{min}} \equiv 10^3$ in order for the rms perturbation on these scales to be nonlinear. In linear perturbation theory, density fluctuations grow as $\delta \propto (1 + z)^{-1}$ during matter domination and only logarithmically during radiation domination (see Section 2.3.1). This implies that the matter-dominated epoch must span at least three orders of magnitude in redshift. With $\Omega_{M,0} \leq 0.5$ this, in turn, imposes the limitation

$$\theta_{\text{LSS}} h^2 \lesssim 12 \quad \text{or} \quad \theta_{\text{LSS}} \lesssim 38,$$

(3.19)

where I have taken $h \geq 0.56$. This constraint is shown in Figure 3.2 where I summarize the constraints on relativistic energy imposed by BBN, the CMB, LSS and the SNIa magnitude-redshift relation. I now proceed to show how these constraints can be used to bound the properties of massive big bang relics in a way that is model-independent and can be applied to an arbitrary Standard Model extension.

### 3.3 Constraining Relic Decays

In the preceding, Section I reported limits on the RED at various epochs. In the absence of electron neutrino degeneracy, the BBN constraint on the RED is, by far, the most stringent; it leaves little room for any non-standard relativistic energy during the BBN epoch (with certain caveats). One way to circumvent the BBN constraint is to inject relativistic energy after nucleosynthesis has ended, for instance, through the decay of a particle that was non-relativistic during BBN ($M \gg 1$ MeV) into relativistic products. As such, the study of relic particle decays comes part and parcel with constraints on relativistic energy. I examine the simple case of a massive,
unstable, big bang relic which may be pertinent to physics “beyond the Standard Model” and I discuss constraints on the relic’s mass and lifetime that follow from the bounds on the RED in Section 3.2. It is important to note that, aside from the CMB constraint, the assumption that the relic is very massive is not critical. In general, the particle must be non-relativistic at decay in order for its decay product’s energy density to be comparable to or greater than \( \rho_\gamma \).

Consider the decay of a massive relic particle \( X \), with lifetime \( \tau \), into relativistic products. Had \( X \) not decayed, the energy density in these particles today would be

\[
\Omega_X h^2 \simeq 274 M_{\text{keV}} Y.
\]  

In (3.20), \( M_{\text{keV}} \) is the mass of the particle in keV and \( Y \) is the ratio of the number density of the particle to the entropy density, \( Y \equiv n_X/s \) (\( Y \simeq 0.039 \) for a light neutrino, \( Y \simeq 2 \times 10^{-10} \) for a 5 GeV, Dirac neutrino). It is conventional to scale the relic density of particles by the entropy density because it scales out the \( n_X \propto (1+z)^3 \) dependence in number density, making it easier to calculate the relic abundance. It is also convenient because the quantity \( Y \) then remains constant after \( X \) decouples from the thermal bath. Given a specific particle physics model in which \( X \) is produced in the early Universe, \( Y \) is fixed (in the absence of subsequent entropy production) and can be calculated using the Boltzmann equation in a homogeneous spacetime:

\[
\frac{dn_X}{dt} = -3 \frac{d \ln a}{dt} - <\sigma_{\text{ann}}|v|> [n_X^2 - n_{\text{EQ}}].
\]  

The first term on the left simply represents the dilution due to the cosmological expansion. The quantity \( <\sigma_{\text{ann}}|v|> \) represents the annihilation cross section averaged over the phase space of the particles and \( n_{\text{EQ}} \) is the equilibrium abundance of the
species. Therefore the term proportional to \( n_X^2 \) gives the annihilation rate and the
term proportional to \( n_{\text{EQ}}^2 \) gives the rate of production. Casting this equation in terms
of \( Y \) and the independent variable \( x \equiv M_X/T \) results in (e.g., [172])

\[
\frac{1}{Y_{\text{EQ}}} \frac{dY}{d \ln x} = -n_{\text{EQ}} \left< |v| \sigma_{\text{ann}} \right> \frac{Y^2}{H(x)} \left[ \left( \frac{Y^2}{Y_{\text{EQ}}^2} \right) - 1 \right]. \tag{3.22}
\]

This form of the Boltzmann equation makes the freeze out condition apparent: when
the interaction rate \( \Gamma = n_{\text{EQ}} \left< |v| \sigma_{\text{ann}} \right> \) is less than the Hubble expansion rate
\( H(x) \), the abundance of the particle changes very slowly compared with the age of
the Universe.

In order to make the constraints on heavy particle decays as generic as possible,
I have chosen to keep \( Y \) as an explicit, free parameter. Further, I assume that the
daughter particles are weakly-interacting (constraints on radiative decays are quite
severe [247]). In all analytic calculations, I assume that decays occur simultaneously
at \( t = \tau \). Energy conservation during decay demands that the present energy density
in decay products be \( \Omega_D = \Omega_X/(1 + z_D) \), where \( z_D \) is the redshift at decay (i.e., \( z \) at
\( t = \tau \)). Assuming the Universe to be X-dominated prior to decay [280],

\[
\Omega_D h^2 \simeq 5.1 \times 10^{-4} M_{\text{keV}}^{4/3} Y^4/3 \tau_{\text{yr}}^{2/3}, \tag{3.23}
\]

where \( \tau_{\text{yr}} \) is the lifetime of X in years.

### 3.3.1 CMB Constraints on Relic Properties

One constraint on relic properties follows from the requirement that the total
energy density in relativistic particles \( \Omega_{\text{R,0}} = \Omega_{\gamma,0} + \Omega_D \), fall under the CMB bound in
(3.8) during the epoch of recombination. If the particle decays prior to recombination,
the CMB constraint (3.8), along with the analytic approximation of (3.23), implies that\textsuperscript{5}

$$M_{\text{keV}}^2 Y^2 \tau_{\text{yr}} \lesssim 7.5 \times 10^{-2}. \quad (3.24)$$

For pertinent lifetimes, $10^{-3} \lesssim \tau_{\text{yr}} \lesssim 10^5$, the excluded region of parameter space is displayed in Figure 3.7.

Post-recombination X decays modify the CMB power spectrum through the ISW effect and through shifts in the multipole positions of the acoustic peaks due to a change in the angular diameter distance to the surface of last-scattering. It is possible to use this modification of the observed CMB anisotropy power spectrum to constrain post-recombination X decays [155, 127], but these constraints would not be generic because the gravitational dynamics of the relic as well as its decay scheme can contribute to the ISW effect. It then becomes necessary to specify a specific particle physics model and mass. Such constraints would have to be developed on a case-by-case basis considering the relic mass and relic abundance separately; however, I mention a specific case that evades the CMB bound and contributes a significant RED at the present epoch in section 3.4. In the following subsection, I show that the growth of LSS can place severe, yet generic, constraints on post-recombination decays.

\textsuperscript{5}The CMB constraint on relativistic energy is strict enough that the assumption of X-domination prior to decay is untenable, making it necessary to integrate the equations governing heavy WIMP decay in a flat, Friedmann cosmology. I have performed the necessary integration and find that the above analytic bound is typically accurate to within $\leq 20\%$. 

86
Figure 3.7: The regions of parameter space for a massive relic decay that are excluded by CMB (heavily shaded), LSS (moderately shaded) and SNIa (lightly shaded) arguments. Each excluded region extends to arbitrarily high $M_{keV}$.  

3.3.2 Large-scale Structure Constraints on Relic Decays

Large scale structure considerations also lead to bounds on decaying relics. Utilizing (3.23), $\theta$ can be expressed as $\theta \simeq 1 + 12.2 M_{keV}^{4/3} Y^{4/3} \tau_{\text{yr}}^{2/3}$. Consequently, the effective shape parameter (3.16) may be written

$$\Gamma \simeq \Omega_{M,0} h [1 + 12.2 M_{keV}^{4/3} Y^{4/3} \tau_{\text{yr}}^{2/3}]^{-1/2} \quad (3.25)$$

from which we conclude that

$$M_{keV}^2 Y^2 \tau_{\text{yr}} \simeq 188 [\Omega_{M,0} h]^2 - \Gamma^2]^{3/2}. \quad (3.26)$$

Again, taking $\Omega_{M,0} \leq 0.5$, $h \leq 0.88$ and $\Gamma \geq 0.06$, we deduce that
\[ M_{\text{keV}}^2 Y^2 \tau_{\text{yr}} \lesssim 16. \quad (3.27) \]

Observe that (3.27) restricts the combination \( M_{\text{keV}}^2 Y^2 \tau_{\text{yr}} \). As has been underscored by McNally & Peacock and Bharadwaj & Sethi [206, 30], the power spectrum will also exhibit a feature on small scales corresponding to the transition between radiation domination and an early matter-dominated (domination by X particles) phase. As the comoving horizon scale during the first epoch of matter-radiation equality is given by

\[ \lambda_X \simeq \frac{1}{17.1 M_{\text{keV}}} \text{Mpc}, \quad (3.28) \]

this small-scale feature can, in principle, constrain the combination \( M_{\text{keV}} Y \) alone. In practice, however, much of the interesting region of relic parameter space would correspond to nonlinear scales and such a constraint would require a better theoretical handle on nonlinear clustering and bias.

We can strengthen the bound in (3.27) by requiring that structure grow sufficiently. As Steigman and Turner [280] have noted, there are two ways in which this can occur. One way was mentioned in Section 3.2.4, namely, that the relic decay early enough so that the most recent epoch of matter domination began at a redshift \((1 + z_{\text{EQ}}) \geq 10^3\). The redshift of equality is \((1 + z_{\text{EQ}}) \simeq \Omega_{M,0}/\Omega_D\) and using (3.20) and (3.23) we find that this scenario requires

\[ M_{\text{keV}}^2 Y^2 \tau_{\text{yr}} \lesssim 0.66 \quad (3.29) \]

and is relevant for lifetimes in the range \(10^{-3} \lesssim \tau_{\text{yr}} \lesssim 10^5\).
Alternatively, in the presence of a massive unstable relic, there can be two phases of matter domination: there may be an early X-dominated phase and a second matter-dominated phase after relic decay. It may be possible for perturbations on scales $\lambda < \lambda_{\text{NL}}$ to take advantage of both of these periods of growth and thereby grow by a factor greater than $\gamma_{\text{min}}$. X domination begins at redshift $(1 + z_X) = \Omega_X/\Omega_{\gamma,0}$, followed by decay at redshift $(1 + z_D) = \Omega_X/\Omega_D$. Thus the total growth factor for scales smaller than the horizon scale at X domination, $\lambda_X$, is

$$\gamma \simeq \frac{(1 + z_X)}{(1 + z_D)}(1 + z_{\text{EQ}}) \simeq 2.4 \times 10^4 \Omega_{M,0} h^2.$$  \hspace{1cm} (3.30)

With the aforementioned limits on $\Omega_{M,0}$ and $h$, these scales grow by a maximum of $\gamma \simeq 9.3 \times 10^3$. Perturbations that enter the horizon after X domination begins, grow by a smaller factor:

$$\gamma(\lambda > \lambda_X) \simeq 9.3 \times 10^3 \left(\frac{\lambda_X}{\lambda}\right)^2.$$  \hspace{1cm} (3.31)

Compelling $\gamma(\lambda_{\text{NL}})$ to be greater than $\gamma_{\text{min}}$ forces

$$M_{\text{keV}} Y \lesssim 3.6h \times 10^{-2} \lesssim 3.2 \times 10^{-2}.$$  \hspace{1cm} (3.32)

Combining (3.27), (3.29) and (3.32), I summarize the LSS constraints on early relic decays as [340]

$$M_{\text{keV}}^2 Y^2 \tau_{\text{yr}} \lesssim 0.66 \quad \text{or the pair of constraints}$$

$$M_{\text{keV}}^2 Y^2 \tau_{\text{yr}} \lesssim 16 \quad \text{and} \quad M_{\text{keV}} Y \lesssim 3.2 \times 10^{-2}.$$  \hspace{1cm} (3.33)

The excluded region is shown in Figure 3.7.
The above constraints (3.33) are pertinent for lifetimes in the range $10^{-3} \lesssim \tau_{yr} \lesssim 10^6$ because I assumed that the turnover in the power spectrum is indicative of $\lambda_{EQ}$, the horizon scale at the epoch of matter-radiation equality after X decay. Alternatively, the X particles may have very large lifetimes, $\tau_{yr} \gg 10^6$, in which case the turnover in the power spectrum would be indicative of $\lambda_X$, the horizon scale at the first epoch of matter-radiation equality, prior to X decay. In this case, the effective shape parameter is simply

$$\Gamma \simeq (\Omega_X + \Omega_{M,0}) h.$$  \hfill (3.34)

If I adopt the limiting case that $\Omega_{M,0} \geq 0.1$ in order to be consistent with various measures of the contemporary matter density, I find that the restriction $\Gamma \leq 0.46$ from (3.14) asserts that

$$M_{keV} Y \lesssim 1.2 \times 10^{-3}.$$  \hfill (3.35)

I illustrate this bound in Figure 3.7 by the vertical boundary for $\tau_{yr} \gtrsim 10^7$. Lastly, because $(1 + z_{EQ}) \sim 10^4$ in this scenario and $\Omega_X$ is not more than a factor of five larger than the lower bound on $\Omega_{M,0}$, requiring $\gamma_{\text{min}} \gtrsim 10^3$ provides only a weak restriction on X lifetimes. It may be more useful to take advantage of the large-scale feature that would be present in the power spectrum due to the injection of relativistic energy in order to limit relic properties. I do not explore such bounds as this would require specifying $M_{keV}$ and $Y$ separately. In other words, this requires knowledge of the interactions of X and the mass of X independently.
3.3.3 SNIa Constraint on Relic Decays

With the SNIa bound on relativistic energy from Section 3.2.3, it is now easy to obtain a SNIa bound on the relic decay properties. Using (3.23) and the 95% upper limit in (3.12), produces

\[ M_{keV}^2 Y^2 \tau_{yr} \lesssim 5.3 \times 10^3, \]  

where I have once again assumed \( h \leq 0.88 \). Notice that this follows from a bound on the contemporary RED and, therefore, is most pertinent to late decays (\( i.e., \tau_{yr} \gtrsim 10^8 \)). Again, this constraint is shown in Figure 3.7 where these bounds on the properties of decaying relics are summarized.

3.4 Discussion and Conclusions

Constraints on the cosmological RED can provide a fundamental probe of particle physics beyond the standard model. In this Chapter I have discussed constraints on the RED during four distinct epochs arising from BBN, the CMB, LSS and SNIa (see Figure 3.2). Further, I have shown how these bounds constrain the mass and lifetime of a hypothetical big bang relic (see Figure 3.7). Somewhat surprisingly, the RED at the current epoch is relatively unconstrained: I have shown that the magnitude-redshift relation for SNIa is consistent with a flat universe comprised of up to 20% relativistic energy. Conventional wisdom suggests that the RED today must be small to allow for sufficient growth of large scale structure and not appreciably alter the CMB anisotropy power spectrum. The LSS bound does, in fact, significantly limit the RED near the epoch of matter-radiation equality. Any RED consistent with subsequent growth of LSS, redshifted to the current epoch, would be quite small.
Conversely, a RED that is large, yet acceptable, with respect to the SNIa bound would clearly inhibit the growth of LSS if it were redshifted to the past. However, a long-lived particle that decays sufficiently late as to avoid the LSS constraint could nevertheless contribute substantially to the RED today. In this case, the relevant constraint would be the aforementioned CMB bound (see Section 3.3.1). Avoiding the CMB constraint requires the X particles to be very long-lived if they are to contribute appreciably to the RED today. Consider a big bang relic with $M_{\text{keV}} Y = 1.2 \times 10^{-3}$ (a 30 eV neutrino is an example of a particle with the necessary abundance, but a 30 eV neutrino would be ruled out by BBN constraints) and a very long lifetime, $\tau_{\text{yr}} = 2 \times 10^9$. The decay products of this relic contribute $\Omega_D h^2 \sim 0.1$ and its properties are marginally consistent with the growth of LSS. In addition, the decay products are produced sufficiently late so as to contribute an unconstrained ISW perturbation at low multipole moments, peaking around $\ell \sim 10$, and to change the angular diameter distance to the last-scattering surface by only $\sim 7\%$ when other cosmological parameters are held fixed (see Figures 3 and 4 of Kaplinghat et al., [155]). These effects cannot be ruled out by current CMB data. In Figure 3.2 I show the evolution of the RED including the decay products of this hypothetical, long-lived big bang relic.
CHAPTER 4

PROBLEMS WITH COLD DARK MATTER ON GALACTIC AND SUB-GALACTIC SCALES

4.1 Introduction

In the standard model of structure formation ($\Lambda$CDM) the Universe is dominated by cold, collisionless, dark matter (the CDM), made flat by a cosmological constant ($\Omega_{\Lambda,0}$), and endowed with initial density perturbations via quantum fluctuations during the early epoch of cosmological inflation. The “cold” in “cold dark matter” indicates that the dark matter particles are supposed to have a negligibly small velocity dispersion prior to gravitational collapse and virialization. The need for the cosmological constant is an unexpected surprise; however the $\Lambda$CDM+inflation paradigm is strongly motivated, and with the parameter choices $\Omega_{M,0} = 1 - \Omega_{\Lambda,0} \approx 0.3$, $h \approx 0.7$, and $\Omega_{b,0}h^2 \approx 0.02$, it can account for an utterly impressive range of astronomical observations on large scales (e.g., [274, 238]). This is the standard paradigm that I motivated extensively in Chapter 2. However, on galactic and sub-galactic scales, this model faces some potentially ruinous difficulties when confronted with observational
data. In this Chapter, I give a brief introduction to the basic results of N-body simulations of structure growth in the ΛCDM model and I discuss some of the challenges being faced by the ΛCDM model on galactic and sub-galactic scales.

4.2 The Structure of Dark Matter Halos

The absolute size of a virialized dark matter halo can be quantified in terms of its virial mass $M_{\text{vir}}$, or equivalently its virial radius $R_{\text{vir}}$, or virial velocity $V_{\text{vir}}^2 \equiv GM_{\text{vir}}/R_{\text{vir}}$. The virial radius of a halo is defined as the radius within which the mean mass density is equal to the virial overdensity $\Delta_{\text{vir}}$, multiplied by the mean matter density of the universe, $\rho_M$. By this definition, $M_{\text{vir}}$ and $R_{\text{vir}}$ must be related by

$$M_{\text{vir}} = \frac{4\pi}{3} \rho_M(z) \Delta_{\text{vir}}(z) R_{\text{vir}}^3. \quad (4.1)$$

The virial overdensity $\Delta_{\text{vir}}$ is an estimate of the relative overdensity of objects that have achieved virial equilibrium after gravitational collapse. Very roughly, material outside of $R_{\text{vir}}$ is still infalling while material inside of $R_{\text{vir}}$ is in virial equilibrium. The virial overdensity can be determined within the context of the spherical top-hat collapse approximation (e.g., [235]). $\Delta_{\text{vir}}$ is generally a function of the matter and vacuum energy densities as well as redshift. For flat cosmologies with a cosmological constant, $\Delta_{\text{vir}}(z)$ can be approximated by [45]

$$\Delta_{\text{vir}} \simeq \frac{18\pi^2 + 82x - 39x^2}{x + 1}, \quad (4.2)$$

with $x + 1 \equiv \Omega_M(z) = \Omega_{M,0}(1 + z)^3/\Omega_{M,0}(1 + z)^3 + \Omega_{\Lambda,0}$. In the standard cosmology considered here, $\Omega_{M,0} = 0.3$, $\Omega_{\Lambda,0} = 0.7$, and $\Delta_{\text{vir}}(z = 0) \simeq 337$. At high redshift
\( \Delta_{\text{vir}} \rightarrow 178 \) which is the value in the so-called “standard cold dark matter” cosmology with \( \Omega_{\text{M,0}} = 1 \).

Several analytic density profiles have been proposed as good approximations to the gross, spherically-averaged density structure of dark matter halos observed in high-resolution N-body simulations. Moore et al. [214] found that the density profiles in the inner regions of dark matter halos vary as \( \rho(r) \propto r^{-3/2} \); however, in a recent numerical convergence study, Power et al. [243] found that the density profiles vary as \( \rho(r) \propto r^{-\alpha} \) with \( 0.8 < \alpha < 1.2 \). This result and other theoretical arguments [298, 91, 92] have discredited the result of Moore et al. and bolstered the case for the universal profile proposed by Navarro, Frenk, and White [218, 219, 220] (hereafter NFW profile):

\[
\rho(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}.
\]  

(4.3)

The NFW profile seems to provide a good description of dark matter halos observed in high-resolution cosmological N-body simulations. Of course, the divergent \( \rho \propto r^{-1} \) scaling is not expected to hold down to arbitrarily small radii where, perhaps the finite velocities of the dark matter particles or dark matter particle-anti-particle annihilations will become important. However, the break in this power law is expected only on scales that are far below the resolution of numerical simulations and the mass contained within the “break” region would be negligibly small so that the NFW relation is a very useful profile for practical calculations. Moreover, in galaxies that host luminous galaxies, baryonic cooling and collapse and angular momentum transfer between infalling baryons and dark matter a far larger effect on the structure of galaxies and halos than the above limitations.
The parameters of the NFW profile are related by \( \rho_s \approx \rho(r \approx 0.47r_s) \). The amount of mass contained within a radius \( r \), is given by

\[
M(< r) = M_{\text{vir}} \frac{g(x)}{g(c_{\text{vir}})}
\]  

where \( x \equiv r/r_s \), and the function \( g(y) \equiv \ln(1 + y) - y/(1 + y) \). The so-called concentration parameter \( c_{\text{vir}} \), provides a useful criterion for assessing the relative central concentration of a halo. The concentration parameter is defined as

\[
c_{\text{vir}} \equiv R_{\text{vir}}/r_s.
\]  

Restating the NFW density profile of Equation (4.3) in terms of a circular velocity profile yields

\[
V_c^2(r) = V_{\text{vir}}^2 \frac{c_{\text{vir}} g(x)}{x g(c_{\text{vir}})}.
\]  

The circular velocity profile provides a mapping of the halo potential as \( V_c^2(r) = -r d\Phi(r)/dr \). The potential of an NFW halo is given by

\[
\Phi(r) = -V_{\text{vir}}^2 \frac{c_{\text{vir}}}{g(c_{\text{vir}})} \left[ \frac{g(x)}{x} + \frac{1}{1 + x} \right].
\]  

The maximum circular velocity of an NFW halo is given by

\[
V_{\text{max}}^2 \approx 0.216 V_{\text{vir}}^2 \frac{c_{\text{vir}}}{g(c_{\text{vir}})}
\]  

and occurs at a radius \( r_{\text{max}} \approx 2.16r_s \). The halo achieves half of its maximum circular velocity at \( r_{V/2} \approx 0.13r_s \).
It should be expected that the transition scale of the NFW profile, parameterized by \( r_s \) or \( c_{\text{vir}} \) for instance, is not a free parameter, but is correlated with halo properties and with the parameters of the background cosmology. In fact, such correlations have been observed in simulations. According to the studies by Wechsler et al. [321] and Bullock et al. [48, 46], dark matter halo concentrations are determined almost exclusively by their mass assembly histories. The reasoning is as follows. The amplitude of perturbations is a function scale, and a length scale \( \lambda \) can be related to the mass of a collapsed, non-linear object at any redshift roughly as \( M(z) \sim 4\pi \rho_M(z) \lambda^3 / 3 \). \(^6\)

With this identification Equation 2.10 then yields the rms overdensity as a function of mass scale, \( \sigma(M) \). In this way, objects of a given mass can be associated with a typical collapse redshift when \( \sigma(M) \sim 1.69 \) and a typical density that is related to the background density at collapse. Bullock et al. [48] and Wechsler et al. [321] find that the median relation between halo virial mass and the NFW concentration parameter \( c_{\text{vir}} \) in the standard \( \Lambda \)CDM cosmology is

\[
c_{\text{vir}}(M_{\text{vir}}) \simeq 9 \left( \frac{M_{\text{vir}}}{M_*} \right)^{-0.13}.
\]

\( M_* \) is the typical collapsing mass defined such that \( \sigma(M_*) = 1.69 \) and in the standard \( \Lambda \)CDM model, \( M_* \approx 1.6 \times 10^{13} \, M_\odot \). Of course, the overdensity on a given scale is distributed statistically so there is some scatter in the relation. Wechsler et al. and Bullock et al. find that the distribution of \( c_{\text{vir}} \) at a given \( M_{\text{vir}} \) can be well-described by a log-normal distribution with standard deviation \( \sigma(\log(c_{\text{vir}})) \approx 0.18 \), while Jing [149] finds a log-normal distribution with standard deviation \( \sigma(\log(c_{\text{vir}})) \approx 0.14 \).

The \( c_{\text{vir}} - M_{\text{vir}} \) relationship will be discussed in detail in the following chapters.

\(^6\)The background density at a redshift \( z \) is given by \( \rho_M(z) = \rho_{\text{crit}} \Omega_M \rho_0 (1 + z)^3 \).
For the time being, it suffices to note that halos observed in numerical studies of gravitational clustering have typical density profiles and that the parameters of these profiles depend upon cosmological parameters in a known way so that they are not free parameters with which to fit density profiles.

### 4.3 The Cuspy Halo Issue

One of the most well-publicized challenges to the standard paradigm is known as the “cuspy halo problem.” The problem can be stated as follows. If the dark matter truly is a cold species, then it is expected that the halos that provide the potential wells in which galaxies and clusters of galaxies sit should be described by NFW profiles of Equation (4.3). In particular, one expects that $\rho(r) \propto r^{-1}$ on very small length scales, $r \ll r_s$. However, many studies claim that rotation curves of observed galaxies indicate that density profiles are much flatter towards the halo center [88, 86, 41]. Comparisons between theory and observation are usually restricted to dwarf and low surface brightness (LSB) galaxies because these galaxies have far less luminous material than can account for their measured rotation curves at almost all radii and are consequently believed to be dynamically dominated by the mass of the cold dark matter [87]. In these galaxies, the luminous baryons are thought to be faithful tracers of the potential of the host dark matter potential. The implication is that the dark matter density profiles, may not be well described by the theoretically-motivated NFW profile and that the nature of the dark matter may not be that of the cold, collisionless species that is typically assumed.

98
Figure 4.1: The points show the rotation curve of ESO 2060140 as measured by de Blok, McGaugh, and Rubin [88, 85]. The solid curve represents the best-fit NFW profile to the observed rotation curve. The dashed curve depicts the best fit \( \alpha \)-profile fit to the observed rotation curve (with \( \alpha = 0.01 \)). The values of the chi\(^2\) per degree of freedom for each fit are also shown. Notice that the introduction of the extra parameter greatly increases the value of the chi\(^2\) per degree of freedom.

An example rotation curve is shown in Figure 4.1. Consider the exercise of fitting rotation curves with the NFW profile and also with the “\( \alpha \)-profile” with an additional free parameter:

\[
\rho(r) \propto \frac{1}{[r/r_s]^\alpha[1 + (r/r_s)]^{(3-\alpha)}}.
\]  

(4.10)

The \( \alpha \)-profile agrees with the NFW profile at large radii where \( \rho(r) \propto r^{-3} \); however, it has the added flexibility of a variable inner profile power law index, \( \rho(r) \propto r^{-\alpha} \), when \( r \ll r_s \). The results of this exercise for the galaxy ESO 2060140 are shown
in Figure 4.1. Notice that the best-fit value of $\alpha = 0.01$. The observed rotation curve indicates a nearly flat inner density profile, or “core,” with $\rho(r) \propto$ constant when $r \ll r_s$. Notice also that the $\chi^2$ per degree of freedom of the $\alpha$-profile fit is greatly reduced relative to the NFW fit suggesting that the $\alpha$-profile truly does give a superior description of the rotation curve.

In fact, this tendency for observed rotation curves to prefer low values of $\alpha$ generic. In Figure 4.2, I show the results of fitting a set of sixty dwarf and LSB galaxy rotation curves observed by de Blok, McGaugh, and Rubin [88], de Blok and Bosma [86], and Swaters [291]. I present the results in the form of a histogram of best-fit values of alpha. I have restricted the values of $\alpha$ to the range $\alpha > 0$ so that density cannot increase with radius. Notice that the histogram of best-fit inner power law indices is strongly peaked at $\alpha < 0.1$ rather than at $\alpha \approx 1$ as would be expected if observed galaxies were well-described by the NFW profile. In fact, observed rotation curves are almost universally better-fit by a rotation curve with a core, such as the “pseudo-isothermal” profile [88]:

$$\rho(r) \propto \frac{1}{1 + (r/r_s)^2}. \quad (4.11)$$

This is illustrated quite plainly in Figure 4.3. For the same set of galaxies, I plot the values of the $\chi^2$ per degree of freedom for the NFW fits against the $\chi^2$ per degree of freedom for the pseudo-isothermal profile fits. Notice that most galaxies have $\chi^2_{\text{NFW}} > \chi^2_{\text{ISO}}$, and that $\chi^2_{\text{NFW}} > 1$ for almost $1/3$ of all of the data points.

In the end, the cuspy halo problem is rather controversial and it is unclear if it actually does pose a robust challenge to the cold dark matter paradigm. First, most rotation curves are acceptably fit by NFW profiles. This, coupled with the fact that
Figure 4.2: The best-fit values of the inner power law index for dark matter halo density profiles. This is a compilation of best fit values of the parameter $\alpha$ in equation (4.10 for sixty galactic rotation curves reported in Refs. [88, 86, 291].

the “pseudo-isothermal” profile is not motivated by any theory, indicates that there is not yet good reason to reject the theoretically-motivated NFW profile hypothesis. Second, the inner slopes of the density profiles are extremely difficult to determine observationally. As was emphasized by Swaters et al. [292] and van den Bosch and Swaters [311], all observational errors such as slit-offset, errors in the location of the dynamical center of the observed galaxy, and the limits of spatial resolution tend to make the apparent observed rotation curves look more like those of cored profiles. They also point out that observing galaxies at low inclination causes gas at all radii to contribute so that it is more difficult to assign a rotation velocity to a particular radius and that non-circular motions in galaxies also tend to make profiles appear
Figure 4.3: Observed rotating curves are generally better-fit by cored pseudo-isothermal profiles than NFW profiles. Here I show a scatter plot of the $\chi^2$ per degree of freedom of fits to NFW profiles (horizontal axis) versus fits to pseudo-isothermal profiles (vertical axis). The data are from Refs. [88, 86, 291]. The dashed line represents $\chi^2_{\text{NFW}} = \chi^2_{\text{P-ISO}}$. Notice that most data points lie below the line.

more core-like. Accounting for all of these biases and uncertainties greatly reduces the statistical argument against the NFW profile [292]. In summation, there may be a discrepancy with the centers of dark matter-dominated galaxies, but the hue and cry over the cuspy halo problem as a challenge to the CDM paradigm may be somewhat unwarranted.
4.4 The Concentration Problem and the Central Density Problem

Consider another exercise in rotation curve fitting. Assume that the NFW profile is the proper profile shape. The virial radius and mass are related through Equation (4.1) and consistency with Equation (4.3) demands that the NFW parameters are related by

\[ \rho_s = \frac{\rho_M \Delta_{\text{vir}}}{3} \frac{c_{\text{vir}}^3}{g(c_{\text{vir}})}, \]

(4.12)

so the profile of an NFW halo observed in an N-body simulation or in the sky is completely specified by \( M_{\text{vir}} \) and \( c_{\text{vir}} \). Fitting the sixty observed rotation curves in the previous section, one can compose a histogram of best-fit values of \( c_{\text{vir}} \) and compare them with the theoretical expectations.

I have performed this exercise and the resulting histogram is shown in Figure 4.4. I also show the theoretically expected value of \( c_{\text{vir}} \) as the solid curve in Figure 4.4. Notice that observed rotation curves are significantly less concentrated than they are predicted to be in the \( \Lambda \)CDM model of cosmology. This is one formulation of the concentration problem. The concentration problem is considered a more serious and more robust challenge to the standard cosmology. The concentration problem exists even if one assumes that the shape of profiles agree with theory. Moreover, the concentration problem is not subject to the large amount of uncertainty associated with trying to measure an inner profile power law index. The concentration parameter can be estimated from many other measurements because it serves as a measure of the amount of mass contained within a fixed radius as in Equation (4.4). Consequently, there is evidence for a concentration problem from observations of the rotation rates.
Figure 4.4: The best-fit values of $c_{\text{vir}}$ for sixty observed rotation curves. The broken line represents a histogram of the best-fit values of the concentration parameter $c_{\text{vir}}$ for the sixty galactic rotation curves reported in Refs. [88, 86, 291]. The solid curve denotes the expected distribution of $c_{\text{vir}}$ values based on the profiles of halos observed in N-body numerical simulations.

of bars in spirals [90], the Einstein radii and magnifications in observed gravitational lenses [157], the rotation curves LSB galaxies, and even those of large spiral galaxies where baryons play an important role [41].

This discrepancy can be placed on even firmer ground by characterizing observed rotation curves without reference to a particular density profile. A more directly observable measure of halo central density is the central density parameter $\Delta v/2$, proposed by Alam, Bullock, and Weinberg [5]. The central density parameter is defined as
\[ \Delta V/2 \equiv \frac{\bar{\rho}(r_{V/2})}{\rho_{\text{crit}}} = \frac{1}{2} \left( \frac{V_{\text{max}}}{H_0 r_{V/2}} \right)^2, \]  

(4.13)
or the mean density within the radius \( r_{V/2} \), where the rotation curve falls to half of its maximum value. In practical units, the central density parameter is given by

\[ \Delta V/2 \simeq 5 \times 10^5 \left( \frac{V_{\text{max}}}{100 \text{ km s}^{-1}} \right)^2 \left( \frac{r_{V/2}}{h^{-1} \text{ Mpc}} \right)^2. \]  

(4.14)

There are distinct advantages to using the central density parameter to characterize the central densities of dark matter halos. For one, \( \Delta V/2 \) is much more robustly determined, both observationally and in numerical data from N-body simulations, than is the inner slope of a density or velocity profile because the central density parameter represents an integrated density within a well-defined radius rather than a differential measure. Furthermore, the central density parameter can be estimated directly from observed rotation curves using only Newton’s law as a theoretical input.

Yet, the definition of \( \Delta V/2 \) is such that this parameter still probes densities on scales close enough to galactic centers to betray the conflict between theory and observation. Halos in N-body simulations are well-described by NFW profiles with a known relation between mass and \( c_{\text{vir}} \). The central density parameter for an NFW halo in the standard \( \Lambda \)CDM cosmology is given by

\[ \Delta_{V/2}^{\text{NFW}} \simeq 341 \frac{c_{\text{vir}}^3}{g(c_{\text{vir}})}. \]  

(4.15)

I have used the measured rotation curves of the same set of sixty galaxies to estimate values of the central density parameter for each galaxy. I elaborate on the method used to estimate the central density parameter from observations in Chapter 5. To compare theory with observations, I have made a scatter plot of these values
Figure 4.5: The points represent the data of de Blok, McGaugh, and Rubin [88] (triangles and stars), de Blok and Boëma [86] (squares), and Swaters [291] (pentagons). The central solid line represents the median value of the predicted relation between $V_{\text{max}}$ and $\Delta v/2$ in a standard $\Lambda$CDM cosmology. The dotted lines represent the scatter in the relation according to Jing [149], while the dashed lines represent the theoretical scatter according to the study of Bullock et al. [48].

as a function of the maximum circular velocity of the measured rotation curve $V_{\text{max}}$, which serves as a measure of the absolute size of the halos. I have also computed the theoretical relationship between $V_{\text{max}}$ and $\Delta v/2$ using Equations (4.15) and (4.8).

The comparison is shown in Figure 4.5 [337]. A familiar pattern is evident, the central regions of observed galaxies seem underdense relative to their counterparts in N-body simulations. We are left with the following conclusion. The standard cosmology does not correctly predict the density structure of galaxy-sized halos and
additional physics is needed to reconcile the standard model of ΛCDM plus scale-invariant primordial power spectrum with the observed central densities and concentrations of dark matter-dominated galaxies. This is the “central density problem” and this problem does seem to be a robust challenge to the standard model.

4.5 The Dwarf Satellite Problem

There is also an apparent discrepancy related to the observed and predicted number counts of satellite galaxies in the vicinity of the Milky Way. The Milky Way has eleven known satellite galaxies within 300 kpc of the Galactic center. The known dwarf satellite galaxies are Carina, Draco, Fornax, the Large Magellanic Cloud (LMC), Leo I, Leo II, Sagittarius, Sculptor, Sextans, the Small Magellanic Cloud (SMC), and Ursa Minor. Compared with this relatively small census of Milky Way guests, N-body simulations reveal ~ 100 – 200 satellite halos of comparable size orbiting within 300 kpc of MW-sized halos (refer back to Figure 1.4). The “dwarf satellite problem” is the name given to the apparent gross mismatch between the number of halos orbiting around galaxy-sized objects in numerical data and the number of observed satellite galaxies.

Number counts of dwarf satellites are typically expressed in terms of the cumulative velocity function, \( N(\geq V_{\text{max}}) \). The cumulative velocity function is the number of halos with maximum circular velocities greater than \( V_{\text{max}} \) as a function of \( V_{\text{max}} \). The maximum circular velocity is conventionally used as a measure of the size of the satellite halo because it is generally believed to be more closely related to the one-dimensional velocity dispersion, which is the quantity that is directly observable. In Figure 4.6, I show the observed velocity function of MW dwarf satellite galaxies
Figure 4.6: The solid line represents the median cumulative velocity function for satellites within the virial radius of a cold dark matter halo of mass $M = 1.4 \times 10^{12}$ $M_\odot$. The errorbars represent the statistical scatter from halo to halo. The squares represent the observed cumulative velocity function of the Milky Way dwarf satellites.

along with the predictions of a semi-analytic model that is in excellent agreement with the results of cosmological N-body simulations [339]. I describe the model in detail in Chapter 6.

Whether or not the dwarf satellite problem presents a serious challenge to the standard cosmological framework is not so clear-cut, either. First, it has long been known that there are feedback mechanisms that prevent the baryons in low-mass systems from cooling, condensing, forming stars and lighting up [330]. Only luminous galaxies are directly observable, so the dwarf satellite discrepancy may simply be explained by the fact that only about one-in-ten halos hosts a luminous galaxy. One
natural source of feedback would be the same background of radiation that leads to the reionization of the Universe at some fairly low redshift, \( z_{\text{rec}} \sim 6 - 15 \) or so. This radiation serves to heat the baryons in small halos to a temperature high enough such that they are not strongly bound by the potential wells of their host halos. Therefore, the baryons in small halos that collapse after reionization never collapse, interact, cool, and condense towards the center of halos and form galaxies. Bullock, Kravtsov, and Weinberg showed that a feedback mechanism of this kind could naturally reproduce the observed velocity function of luminous Local Group satellites for a reasonable range in the redshift of reionization [49].

Nevertheless, there are many uncertainties associated with the dwarf satellite problem. The mapping between observed velocity dispersions and halo \( V_{\text{max}} \) is not as simple as is often supposed, moreover the shape of the velocity ellipsoid of stars in dwarf satellite galaxies is not well-understood and this has important consequences for the observed velocity function. In fact, it is not even known if a “differential” feedback mechanism that causes \( \sim 1/10 \) of all low-mass halos to host luminous galaxies is the type of mechanism that is needed to resolve this issue. It may be that a “sharp” feedback mechanism that causes all halos below a particular mass scale to be completely dark is what is needed in order to reconcile theory with observations. I elaborate on all of these issues in Chapter 6. At present, it suffices to say that the population of dwarf satellites that surround the Galaxy is not understood in the context of CDM theory.
CHAPTER 5

INFLATION, COLD DARK MATTER, AND GALACTIC CENTRAL DENSITIES

5.1 Introduction

In Chapter 4, I discussed two problems with the distribution of matter in the central regions of galaxies. The first problem concerns the fact that the integrated mass densities within well-defined central radii of observed galaxies seem to be a factor of $\sim 4 - 6$ larger than the central densities predicted by standard $\Lambda$CDM [33, 157, 90, 5, 157, 72, 41, 199, 337]. The second, often referred to as the cuspy halo problem, highlights the fact that CDM halo density profiles are predicted to diverge at small $r$ ($\rho(r) \propto r^{-\alpha}, \alpha \sim 0.8 - 1.2$), while galaxy rotation curves are often better fit with constant density cores [107, 212, 54, 34, 214, 88, 86]. While the second of these issues has received the most attention, as I discussed above, it is related to the first problem and it is markedly more controversial. For example, de Blok and collaborators [86] have argued that their sample of low-surface brightness galaxies favor fits to density profiles with a constant density central core over those with cusps; however, van den Bosch and Swaters argue that a majority of galactic rotation curves are acceptably fit by divergent profiles as long as they are much less
centrally concentrated than typical halos in standard ΛCDM [311]. Furthermore, all observational errors (e.g., slit offset) tend to favor constant apparent central density over cusps. At present, it is not clear that the cuspy halo problem presents a serious challenge to ΛCDM, although it appears almost certain that the data do prefer halos that are less centrally concentrated than typical halos in the standard LCDM model.

These problems with central densities have triggered a growing concern that cosmologists are missing some fundamental ingredient necessary for a basic understanding of galaxy formation. (This is in spite of the fact that some of the problematic claims are disputed [311, 164, 148, 252].) Proposed solutions to these problems span a wide range of physical possibilities. In large, spiral galaxies part of the solution may well be baryonic. For instance, supernovae may blow baryons out of the centers of galaxies and these baryons may drag dark matter particles with them [196], while Weinberg and Katz [323] have proposed that baryonic bars in spirals can transfer significant amounts of angular momentum to the dark matter particles leading to a transfer of mass from the inner regions of the halo to the outer regions of the halo (see Refs. [164, 217, 92, 48, 321, 49] for other proposed solutions that make use of baryonic physics). While such “astrophysical” solutions are reasonably well-motivated and likely due play some role in determining the properties of observed galaxies, the fact that these problems seem to exist for relatively small, dim, dark matter-dominated dwarf galaxies all the way up in scale to very large, bright, elliptical galaxies suggests that a single baryonic solution is not sufficient to address all of our concerns.

Other proposed solutions include those that rely on altering the nature of the dark matter. The dark matter may have self-interactions lead to annihilations in high-density regions and cause subhalos within large hosts to disintegrate [273, 154, 106].
Another solution is to endow the dark matter particles with a large primordial velocity dispersion that allows it to resist gravitational clustering [137, 187]. The altered dark matter solutions could possibly be made to match the range of observations, but only by invoking unmotivated candidates. While supersymmetric extensions to the standard model provide a natural candidate for a cold dark matter particle, particularly the lightest supersymmetric particle which is usually taken to be the neutralino or gravitino, there is no well-motivated warm dark matter candidate. Furthermore, any successful model requires the parameters to be fine-tuned if a single solution is to alleviate both the central density and dwarf satellite problems [5, 339].

In this Chapter, I explore a different form of solution to the central density problem. Rather then modify dark matter physics, I investigate the possibility that the key may lie in the primordial power spectrum. The work presented here builds on the success of several precursors. This study is principally inspired by Alam, Bullock, and Weinberg [5] who suggested that the central density problem would be reduced significantly in $\Lambda$CDM model if the initial inflationary power spectrum were tilted to favor large scales. Recall from Chapter 2 that the term “tilted” is defined in terms of the assumed form of the primordial power spectrum of density fluctuations. It is assumed that the power spectrum can be written as a power law over the observable range in wavenumber, $P(k) \propto k^n$, which corresponds to a mass variance per logarithmic interval in wavenumber of $\Delta(k) = k^3P(k)/2\pi^2$. “Tilted” power spectra refer to those with $n \neq 1$. In the “standard” $\Lambda$CDM model, it is generally assumed that $n$ is exactly 1, corresponding to a scale-invariant, Harrison-Zel’dovich [132, 335] primordial power spectrum. In the following Chapters I use the term “standard $\Lambda$CDM” to refer to the choice of cosmological parameters advocated in Chapter 2 plus the additional
assumption that $n = 1$. I showed in Chapter 2 that the name “scale-invariant” for a spectrum with $n = 1$ derives from the fact that in this case, the rms overdensity on all scales are equal at the time that each scale enters the horizon. Qualitatively, one may say that density fluctuations enter the horizon with the same initial conditions before being processed by causal physical processes.

The effect of a tilt is easy to understand. Most of the observations that yield the dramatic confirmations of the ΛCDM model that I discussed in Chapter 2 are based on measurements of clustering on scales $k \sim 10^{-3} - 10^{-1}$ Mpc$^{-1}$, but the mass assembly histories of dark matter-dominated galaxies are determined by the typical amplitude of fluctuations on scales $k > 10$ Mpc$^{-1}$. Introducing a tilt (or even a running tilt in which $dn/d\ln k < 0$) makes it possible to fix the amplitude of fluctuations on the large scales probed by the CMB and/or galaxy clustering while reducing the power on the small scales that determine the properties of individual dark matter halos.

The choice of a scale-invariant initial power spectrum is often justified by the tendency for inflation models to predict *nearly* scale-invariant spectra. I emphasize below that while inflation does predict a *nearly* scale-invariant initial spectrum of density fluctuations, $P(k) \propto k^n$ with $n \approx 1$ (the normalization can be fixed by measurements of the Sachs-Wolfe plateau in the CMB power spectrum), it does not predict an *exactly* scale-invariant power spectrum. In fact, inflation is a paradigm within which many of the features of the Universe we live in, such as flatness and the lack of monopoles and other topological objects, can be understood but there is no specific theory of inflation. Rather there are many equally well-motivated models that are poorly-constrained.
In this Chapter, I begin by computing the primordial power spectra predicted by several models of inflation that exhibit measurable deviations from scale-invariance. I then show the ramifications for the central densities of dark matter-dominated galaxies. This work can be interpreted in two ways: I employ specific models of inflation as a conceptual device to show how the different ingredients of the standard paradigm may fall together into one consistent picture; however, one may adopt a more empirical view and consider the spectra studied here merely as a set of spectra that span an observationally viable range. Though this is similar in spirit to the agenda of Kamionkowski and Liddle [152], who suggested that the small-scale crises confronting CDM might reflect a sharp feature in the inflationary power spectrum, my mind-set here is to look at models that are not particularly fine-tuned. I simply choose fairly representative, simple, single field inflationary models and we examine a range of predicted power spectra. In the context of slow roll inflation, models that predict significant tilt generally yield effective spectral indices that are scale-dependent or exhibit significant running of the spectral index. Consequently, when comparing observational data that span a wide range of scales ($\Delta \ln k > 12$ in this case) it makes sense to account for the variation of $n(k)$ with scale in addition to the tilt of the spectrum. I account for the running of the spectral index by calculating it in specific inflationary models and show that the running can have an important effect on structure on galactic scales.

Another way to reduce power on small-scales is to introduce a small admixture of a very light species, with a significant primordial velocity dispersion, to the mix of CDM and baryons. This possibility is motivated by the fact that neutrinos likely have non-zero rest mass and therefore may play a role in gravitational clustering. For
completeness, I also estimate the effect of a “hot dark matter” component in the form of massive neutrinos on the central densities of dark matter halos.

Although I examine models with varying amounts of small-scale power, I am not free to alter the spectrum by an arbitrary amount. The amplitude of power on small (\( \sim 8 \, h^{-1} \) Mpc) scales is often quantified in terms of \( \sigma_8 \), which is the rms overdensity smoothed with a top-hat filter of radius \( 8 \, h^{-1} \) Mpc [cf., Equation (2.10)]. Observationally, this quantity can be determined in a number of ways including techniques that rely on the abundance of rich x-ray clusters, the cosmic shear from weak gravitational lensing, and galactic peculiar velocity flows. However, these estimates do not converge on a definitive value (even when the same method is used by different authors) and many recent estimates seem to advocate surprisingly low values of \( \sigma_8 \) [208, 317, 16, 146, 13, 314, 44, 6, 125, 40, 241, 136]. Roughly speaking, recent estimates yield values that span the range \( 0.55 \lesssim \sigma_8 \lesssim 1.2 \) for \( 0.2 \lesssim \Omega_m,0 \lesssim 0.5 \). In the following, I only consider models with \( \sigma_8 > 0.55 \) because models that imply a smaller value of \( \sigma_8 \) do not have a good chance of being able to match recent observations. I acknowledge that even this limit is pushing the observational bounds (although it is consistent with the findings of Melchiorri and Silk [208] and Viana, Nichol, and Liddle [317]) but I feel that it is best to explore all possibilities for the sake of alleviating the tension between theory and observation on sub-galactic scales. For reference, a scale-invariant spectrum (i.e., with spectral index \( n = 1 \)) that is normalized to the Cosmic Background Explorer (COBE) measurements [23] of the large-scale cosmic microwave background anisotropy via the fitting forms of Bunn, Liddle, and White [52, 53], has \( \sigma_8 \sim 0.95 \), assuming that the gravitational wave contribution to CMB anisotropy is negligible.

115
The inflation-based or hot dark matter-based solutions to the central density problem are “non-astrophysical” in that they do not invoke the poorly-understood details of baryonic cooling and collapse and so they may alleviate problems on all scales, from dark matter-dominated LSB galaxies to large, bright ellipticals. Furthermore, this solution works entirely within the desirable tenets of the CDM paradigm and the ΛCDM model that seems to work so well on large scales. Within the context of this model, there is no need to develop poorly motivated or highly fine-tuned dark matter candidates such as warm particles [137] or particles endowed with large primordial velocities due to the fact that they are the products of the decay of even more massive, unstable relics [187]. The comparably well-motivated neutralino can still serve as the bulk of the dark matter in this case. This approach is conservative in that it concentrates on the integrated density within some well-defined radius, which is certainly more robustly determined in simulated halos than the central slope of the density profile (see Section 4.4). As I shall discuss below, the same is likely true for observed galaxies.

The remainder of this Chapter is organized as follows. In Sec. 5.2 I introduce several models of inflation and calculate the power spectra predicted by each model. I give a short description of the effects of massive neutrinos on the evolved, linear power spectrum in Sec. 5.3. I discuss the properties of dark matter halos and describe our semi-analytic model for estimating halo central densities in Sec. 5.4. In Sec. 5.5 I present our results and compare them with the observed central densities of dwarf and low surface brightness (LSB) galaxies. Lastly, I summarize present the conclusions of this study and indicate directions for future work in Sec. 5.6. Throughout this work
I assume fixed cosmological parameters with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_B h^2 = 0.020$, and $h = 0.72$, in agreement with the parameter choices discussed in Chapter 2.

5.2 Inflationary Power Spectra

It is widely believed that the primordial density perturbations that seeded the growth of all the multifarious structure in the Universe were produced during an early epoch of cosmological inflation: quantum fluctuations in the inflaton field were frozen in as the extraordinary rapid cosmological expansion stretched these fluctuations to length scales larger than the horizon [122, 123, 134, 20]. The power spectrum of primordial perturbations can be calculated analytically via the slow roll approximation (for a review, see Refs. [185, 193, 186] and references therein). The standard calculation, to lowest order in slow roll, yields expressions for the amplitude of the spectrum of density perturbations at horizon crossing,

$$\delta^2_H(k) \simeq \frac{1}{75\pi^2 m^6_{pl}} \frac{V^3(\phi)}{V''(\phi)}, \quad (5.1)$$

the effective spectral index of the primordial spectrum,

$$n(k) \simeq 1 + 2\eta - 6\epsilon, \quad (5.2)$$

and the running of the spectral index,

$$\frac{dn}{d\ln k} \simeq -16\epsilon\eta + 24\epsilon^2 + 2\xi^2 \quad (5.3)$$

in terms of the inflaton potential $V(\phi)$ ($V'(\phi) \equiv dV/d\phi$) and the slow roll parameters $\epsilon \equiv m_{pl}^2/2(V'/V)^2$, $\eta \equiv m_{pl}^2(V''/V)$, and $\xi^2 \equiv m_{pl}^4 V'''/V^2$. Viable models of inflation require slow roll in order to the solve the flatness and monopole problems.
The slow roll approximation amounts to the assumption that the energy in the infla-
ton field is dominated by potential, rather than kinetic energy, and that the rate at
which the inflaton field rolls down its potential is slow compared to the expansion rate
of the Universe. These conditions require \( \epsilon \ll 1 \), \( \eta \ll 1 \), and \( \xi^2 \ll 1 \). The reduced
Planck mass is defined in terms of Newton’s constant as \( m_{\text{pl}} \equiv 1/\sqrt{8\pi G_N} \simeq 2.4 \times 10^{18} \)
GeV and and, as usual, I work in units in which \( c = k_B = \hbar = 1 \). As usual, these
expressions are to be evaluated at horizon crossing (i.e., when \( k = aH \)).

In the limit of exact de Sitter space during inflation, the predicted primordial
power spectrum would approach exact scale-invariance; however, any model in which
a scalar field is slowly rolling towards a minimum of its potential will predict some
small deviation from \( n = 1 \). A significant deviation from \( n = 1 \), say \( n = 0 \) or \( n = 3 \),
would be damning for the inflationary paradigm and the fact that observations suggest
that \( n \approx 1 \) is, perhaps, the strongest observationally-grounded statement in favor of
inflation. Of course, in the context of many inflationary models, the deviation from
scale-invariance is quite small, and scale-invariance is an excellent approximation.
One frequently-cited example of this sort is power law inflation, for which there is an
exact solution [1]. The reason why approximate scale-invariance is expected in this
model has to do with estimates of the gravitational wave contribution to the CMB
quadrupole. In addition to scalar, density fluctuations, inflation also produces tensor
or gravity wave fluctuations. In the power law inflation case, the ratio of the tensor to
scalar contribution to the CMB quadrupole, \( r \equiv C_{2}^{\text{tensor}}/C_{2}^{\text{scalar}} \), increases with the tilt
as \( r \simeq 6.9(1 - n) \). A similar result also applies to chaotic inflation models (e.g., Ref.
[189]) because \( \phi \sim m_{\text{pl}} \) in these models. Recent CMB measurements indicate that
the tensor contribution is small \( r \lesssim 0.2 \) [320, 130, 274, 237] so power-law inflation requires \( n \gtrsim 0.97 \).

However, this case does not exemplify general inflationary predictions. Tensor perturbations can be negligible even if the tilt is not. This is because the tensor wave amplitude depends only on the energy scale of inflation, \( \delta h_{ij} \sim 1 / \pi \sqrt{3 / 2} (V^{1/2} / \mpk^2) \). The gravity wave contribution is negligible if the inflaton field remains far below the Planck scale, as would be expected in well-motivated models such as the running-mass case discussed below. In fact, most popular models fulfill this criterion because the epoch of inflation is associated with the phase transition corresponding to the breaking of a grand unification symmetry, where scalar fields may be important and the epoch of inflation can dilute away any resultant topological objects, and the energy scale of inflation is thus \( \sim M_{\text{GUT}} \sim 10^{15} \) GeV. Models meeting this requirement can naturally produce modest tilts and spectral index running without violating any bounds on the tensor perturbations. Moreover, there are reasonably well-motivated cases that can yield dramatic departures from scale-invariance. Most models can be characterized simply with a tilt because the running is weak and the range of cosmologically observable scales is fairly small. Kosowsky and Turner [174] were the first to point out the potential observability of a spectral index running.

In the balance of this section, I briefly outline the predictions of several models of inflation that lead to deviations from the standard \( n = 1 \), scale-invariant primordial spectrum and present the \( z = 0 \) linear power spectra in each case. Included in this set of models is a more extreme example that exhibits so-called “broken scale-invariance” and for which the slow roll approximation cannot be used. In all other cases, I calculate the primordial power spectrum to second order in slow roll using
the method of Stewart and Gong [286] which is sufficiently accurate for our purposes [119]. In this way, I explicitly account for both the tilt of the power spectrum and the running of the spectral index.

To derive the low-redshift power spectra, we use the fitting form for the transfer function given by Eisenstein and Hu [101] and the exact relation for the linear growth factor in flat cosmologies with a cosmological constant given by Bildhauer, Buchert, and Kasai [31]. In all cases we normalize the power spectrum to COBE using the fitting functions of Bunn, Liddle and White [52, 53]. We consider several models in which the effective spectral index varies significantly with scale. In these cases, we follow the prescription of Ref. [52, 53] and evaluate the normalization at the scale \( k_* = 7H_0 \simeq 0.0023 \text{ h Mpc}^{-1} \), which is approximately the pivot scale of the COBE data, using the effective spectral index at that scale, \( n(k_*) \).

### 5.2.1 Inverted Power Law Potentials

We begin with the illustrative example of the inverted power law (IPL) potential (or “small field polynomial” in the language of Ref. [128]) which has the basic characteristics of “new inflation” [282]. The general form is

\[
V(\phi) = V_0(1 - c\phi^p)
\]

(5.4)

with \( p > 2 \). This potential implies that the effective spectral index of the primordial power spectrum on the scale \( k_* \) is given by

\[
n(k_*) \simeq 1 - 2 \left( \frac{p-1}{p-2} \right) \frac{1}{N_*},
\]

(5.5)
where \( N_* \) is the number of e-folds of inflation that occur between the epoch when the scale \( k_* \) leaves the horizon and the end of inflation. This is another parameter that finds its way into all inflationary calculations. The reason is that we do not know the exact physics of the end of inflation and reheating and so, we do not know exactly how to map a comoving horizon size onto the value of the inflaton field at a particular epoch. The parameter \( N_* \) embodies this uncertainty and can be allowed to vary at the factor of \( \sim 2 \) level. We can obtain a rough estimate of \( N_* \) in terms of the energy density at reheating, \( \rho_{RH} \), the value of the inflaton potential when \( k_* \) leaves the horizon, \( V_* \), and the value of the inflaton potential at the end of inflation, \( V_F \), by assuming instantaneous transitions between vacuum domination and matter domination at the end of inflation and matter domination and radiation domination at reheating. This exercise gives

\[
N_* \approx 57 - \ln \left( \frac{10^{15} \text{ GeV}}{V_*^{1/4}} \right) + \ln \left( \frac{V_*^{1/4} \rho_{RH}^{3/4}}{V_F} \right). \tag{5.6}
\]

If the details of the end of inflation and the process of reheating were known, \( N_* \) would be known precisely; however, these details are not known. In order to obtain definite predictions, we take \( N_* = 50 \) which is a fairly standard working hypothesis. Using this in Eq. (5.5), we see that with \( p = 4 \) (we refer to this model as IPL4) this model predicts a mild deviation from scale-invariance, namely \( n(k_*) \approx 0.94 \). Accordingly, the spectral index is mildly scale-dependent, \( |dn(k)/dk| \approx 0.002 \).

Figure 5.1 depicts a typical power spectrum at \( z = 0 \) with the choice \( p = 4 \). Rather than \( P(k) \) or \( \Delta^2(k) \), I plot the rms overdensity on a given mass scale \( \sigma(M) \), because this is the relevant quantity in the calculations that follow. The COBE normalization amounts to choosing a suitable combination of \( V_0 \) and \( c \) and the effective spectral index
is insensitive to this choice. Normalized to COBE, this model predicts a perfectly acceptable value of $\sigma_8 \simeq 0.83$.

Before proceeding, I would like to mention briefly that the particle physics motivation for this type of potential may be somewhat dubious. In particular, if $p = 4$, COBE normalization requires the dimensionless $\phi^4$ coupling constant to be of order $\sim 10^{-14}$; however, this fine-tuning problem may be obviated by considering the coupling to be a parameter of an effective field theory rather than a fundamental parameter [193]. Nevertheless, the IPL model serves as a good, illustrative example because it is simple and has the general behavior $|n(k_*) - 1| \simeq \mathcal{O}(1)/N_e$ and $|dn/d\ln k| \ll 0.01$ that is exhibited by a wide variety of models, including many specific incarnations of new inflation [282], hybrid inflation and mutated, hybrid inflation [283], as well as scalar-tensor models of gravity that have a variable gravitational constant [255]. In addition, this potential mimics the potential encountered in a particular variation of mutated hybrid inflation known as “smooth” hybrid inflation [182].

5.2.2 Running-mass Inflation

Stewart has proposed a model in which the need to fine-tune the inflaton mass in order for the slow roll conditions to be met and sufficient inflation to occur in the context of supergravity is eliminated by a “flattening” of the effective inflaton potential due to loop corrections [284, 285]. This provides a natural mechanism for generating a potential that gives rise to inflation. Interestingly, the resulting effective potential can lead to a spectral index considerably different from $n = 1$ and with a significant scale dependence. In this model it is assumed that in the

7The length scale 8 $h^{-1}$ Mpc corresponds to a mass scale of $M \simeq 1.8 \times 10^{14}$ $h^{-1}$ $M_\odot$ under the assumption that $\Omega_M = 0.3$.  

122
sector of the inflaton field, supersymmetry is broken explicitly during inflation and the scalar fields have “so-called” soft supersymmetry-breaking mass terms as would generally be the case (for a review of these issues see [17]). Through couplings to fields with soft supersymmetry-breaking masses, the scalar field masses may get important renormalization corrections. The one-loop correction to the inflaton potential then gives an effective potential with a running inflaton mass,

\[ V(\phi) \simeq V_0 + \frac{1}{2} m^2(\phi) \phi^2 + \ldots, \]  

(5.7)

where the ellipsis represents nonrenormalizable terms that are suppressed by inverse powers of the Planck mass and become important at the Planck scale (assuming a cutoff at the Planck scale and an effective SUSY GUT at lower energy scales). The value of \( V_0 \) is tied to the scale of the supersymmetry breaking during inflation, \( M_S \sim V_0^{1/4} \).

This type of model has been discussed extensively in the literature by Covi, Lyth and Roszkowski [75], Covi and Lyth [73], and Covi, Lyth, and Melchiorri [74] who derived cosmic microwave background constraints on such models. Inflation occurs in the vicinity of an extremum of the potential and the established blueprint for analyzing inflation in the context of this model is to assume that \( m^2(\phi) \) can be approximated by a function that is linear in \( \ln(\phi) \) while cosmological scales are leaving the horizon [284, 75, 73]. In the notation of Refs. [75, 73], the effective potential is then written as in Eq. (5.7) with \( m^2(\phi) \simeq -(V_0/m^2_p)(c/2 - c\ln(\phi/\phi_0)) \) and the spectral index in this case is given by

\[ n(k_s) \simeq 1 + 2\sigma e^{-cN_c} - 2c \]  

(5.8)
where $\phi_0$ is chosen such that $V'(\phi_0) = 0$. The quantities $c$ and $\sigma$ are parameters that may be either positive or negative and we generally expect that $|c| \lesssim |\sigma| \lesssim 1$ [75, 73] because $c$ contains loop suppression factors and $\sigma \equiv c \ln(\phi_*/\phi_{\text{end}})$, where $\phi_*$ is the value of the field when $k_*$ leaves the horizon and $\phi_{\text{end}}$ is the value of the field when inflation ends. With $c > 0$ and $\sigma < 0$, we have the particularly interesting case that $n < 1$ and decreases with increasing $k$. With $c < 0$ and $\sigma > 0$, $n > 1$ and decreases with scale. I employ such a model in several places to illustrate the predictions of a model with a primordial power spectrum that has $n > 1$. For reasonable parameter choices there may be a significant tilt and the scale dependence may be as strong as $|dn(k)/d\ln(k)| \simeq 0.01$ on cosmologically interesting scales. The significant running of the tilt is not surprising. As I stated earlier, in the context of slow roll inflation, models with significant tilts typically exhibit strong variation of $n$ with scale. The COBE normalization is related to the parameters $V_0$ and $\tau \equiv |c| \ln(m_{\text{pl}}/\phi_0)$. In each case I make physically reasonable choices of these parameters to enforce the COBE normalization (see Ref. [73]). The shape of the spectrum is relatively insensitive to these choices. I choose this model as a simple example of an inflationary model that may naturally predict significant deviations from $n = 1$ and potentially detectable running of the spectral index.

In Figure 5.1, I show present-day linear power spectra for two particular choices of the aforementioned parameters. Adopting the parameter choices $\sigma = -0.31$ and $c = 0.04$ results in a model with a tilt $n < 1$ and significant running. For this model, $n(k_*) \simeq 0.84$ and $dn/d\ln k \simeq -0.008$ which is consistent with constraints on tilt from various analyses of CMB, large-scale structure and Ly$\alpha$ forest data [320, 128]. For this model $\sigma_8 \simeq 0.65$ which is on the low side of the observationally acceptable
range. In what follows, I denote this model by \( \sigma_8 = 0.65 \). I have also chosen a model with yields a less significant tilt, \( n \approx 0.90 \), and a small amount of running \( \frac{dn}{d\ln k} \approx -0.002 \). This model, subject to the COBE normalization, results in \( \sigma_8 \approx 0.75 \).

5.2.3 Broken Scale-Invariance in the Primordial Power Spectrum

In the models of the previous sections, the inflaton rolled smoothly down its potential resulting in a smooth, slowly varying power spectrum. In contrast with the above models, there may be a feature at some scale in the inflationary potential that causes the power to drop abruptly. For instance, the derivative of the effective inflaton potential may suddenly increase [cf., Equation (5.1)] due to, perhaps, a phase transition. This possibility leads one to consider models with so-called “broken scale-invariant” (BSI) spectra. In these models, there is a critical scale \( k_c \), and for \( k \gg k_c \) and \( k \ll k_c \) the primordial power spectrum has an effective power law index \( n \approx 1 \). However, on scales near the critical scale the amplitude of the initial density perturbations changes abruptly so that the power on scales \( k > k_c \) can be significantly less than that on scales \( k < k_c \). This type of spectrum may arise in models in which more than one field plays an important role in inflation while cosmological scales are leaving the horizon and sharp transitions in the power spectrum may evince a phase transition during the inflationary epoch [168, 267, 255, 242, 3]. In all such models, however, placing the scale \( k_c \) in an observationally interesting range usually introduces a fine-tuning issue.
As an idealized case of BSI, Starobinsky derived an analytic expression for the primordial power spectrum in a model where the inflaton potential has a step discontinuity in its first derivative [277]. This is a useful model to study for several reasons. First, the primordial power spectrum can be written in a relatively simple closed form. Second, and more material, this model exhibits the most rapid drop in power possible in a single field model of inflation [152]. Third, the two parameters of this model form a useful set with which to parameterize any such feature in the observed power spectrum in a manner that is useful to theorists and observers. Lesgourgues, Polarski and Starobinsky [183] investigated using primordial power spectra of this type to explain a feature on scales of about 125 h\(^{-1}\) Mpc in the galaxy power spectrum measured by the Automatic Plate-measuring Machine survey [113] (such a feature is not present in more recent determinations of the galaxy power spectrum [68]) while Kamionkowski and Liddle [152] explored the effects of such a primordial spectrum on the typical abundance of dwarf satellites.

In this scenario, the power spectrum of density perturbations \( \Delta_p^2(k) \), prior to being modified by causal physical processes, is given by the following exact relation [277]:

\[
\Delta_p^2(k) \propto y^4 \left[ 1 - 3(p - 1) \frac{1}{y} \left( f_- \sin(2y) + \frac{2}{y} \cos(2y) \right) \right. \\
\left. + \frac{9}{2} (p - 1)^2 \frac{1}{y^2} f_+ \left( f_+ + f_- \cos(2y) - \frac{2}{y} \sin(2y) \right) \right]
\]  

(5.9)

where \( y \equiv k/k_c, f_\pm \equiv 1 \pm 1/y^2 \), and \( p \) is the ratio of the amplitude of fluctuations on scales \( k < k_c \) to the amplitude of fluctuations on scales \( k > k_c \). In this model, one is free to choose the amplitude of primordial tensor perturbations and I assume that they are negligible. The normalization is according to the COBE data. Inspired by
the work of Kamionkowski and Liddle [152], I choose \( p = 10 \) and \( k_c = 0.9 \) h Mpc\(^{-1}\) in order to suppress power on mass scales \( M \lesssim 10^{10} \) h\(^{-1}\) M\(_\odot\) and thereby alleviate the dwarf satellite problem (note that I have chosen a different \( k_c \) than Kamionkowski and Liddle [152], partly because I assume a different cosmological model). The dotted line in Fig. 5.1 shows the \( z = 0 \), linear power spectrum predicted by this model. As there is a rise in power prior to the cutoff at \( M \approx 10^{10} \) h\(^{-1}\) M\(_\odot\) (see Fig. 5.1), we find that \( \sigma_8 \approx 0.97 \) which is slightly larger than in the standard scale-invariant model.

### 5.2.4 The Power Spectrum of WMAP

Finally, as we saw in Section 2.5, the WMAP team has recently performed a likelihood analysis on microwave anisotropy data from several experiments, galaxy clustering measurements, and estimates of the power spectrum using Ly-\( \alpha \) forest data. Their best-fit parameters are shown in Table 2.1. Interestingly, the WMAP team find a best-fit primordial power spectrum with \( n(k = 0.05 \) Mpc\(^{-1}\) = 0.93\(^8\) and \( \frac{dn}{d \ln k} = -0.031 \) normalized such that \( \sigma_8 = 0.84 \). Although this result has yet to be confirmed by other groups, such a model seems worth investigating, especially in light of the small-scale difficulties that it may help to alleviate. As a final, observationally-motivated example, I also consider the effect of the WMAP best-fit power spectrum with a running index (RI) on the properties of galaxies. All of the power spectra described in this Section are depicted in Figure 5.1.

\(^8\)The WMAP team report their results on the scale \( k = 0.05 \) Mpc\(^{-1}\). Evaluating the WMAP tilt at \( k_* \) and taking note of the running of the spectral index yields \( n(k_*) \approx 1.03 \). This is the value that should be compared to the other models discussed in this Chapter.
Figure 5.1: The $z = 0$, rms overdensity as a function of mass scale for several of my adopted power spectra (see Section 5.2). The models shown are the standard $n = 1, \sigma_8 = 0.95$ scale-invariant model normalized to COBE (solid), the broken scale-invariant (BSI) model (dotted), the IPL4 model with $\sigma_8 = 0.83$ (short-dashed), the running-mass model with $\sigma_8 = 0.65$ (long-dashed), the running-mass model with $n \simeq 0.9$ and $\sigma_8 \simeq 0.75$ (dot-long-dashed), and the WMAP best fit model with $\frac{dn}{d\ln k} = -0.03$ (dot-short-dashed).

5.3 Massive Neutrinos and Hot Dark Matter

A preponderance of evidence from solar and atmospheric neutrino oscillation experiments like Super-Kamiokande [110], the Sudbury Neutrino Observatory [4], the Russian-American Gallium Experiment [2], the Gallium Neutrino Observatory [9], the Gallium Experiment [126], and the Soudan Experiment [7] seems to imply that neutrinos are indeed massive. Yet these experiments cannot determine the absolute magnitude of the neutrino masses and it may be that the masses are large enough to
have significant cosmological implications. If massive neutrinos (or other “hot dark matter” particles) make up a non-negligible portion of the dark matter, the effect of their free-streaming will be to reduce power relative to the standard model on small length scales. This situation is commonly referred to as the cold+hot dark matter scenario [244].

It is easy to estimate the scale at which this effect becomes important. Massive neutrinos will move at a speed over order $c$ until they become nonrelativistic when $m_\nu \sim 3T_\nu$ which occurs at a redshift of $z_{NR} \simeq 2 \times 10^3 (m_\nu / \text{eV})$. We expect power to be suppressed on scales smaller than the horizon scale at redshift $z_{NR}$, because the neutrinos cannot be trapped by gravitational wells until they become nonrelativistic. As such, a rough estimate is that power will be damped on all scales $k \lesssim k_{FS}$ where

$$k_{FS} \simeq 0.03 \Omega_M^{1/2} \left( \frac{m_\nu}{\text{eV}} h \right) \text{Mpc}^{-1}.$$  

This corresponds to suppression of power on mass scales $M \lesssim M_{FS} \simeq 3 \times 10^{18} \Omega_M^{-3/2} (m_\nu / \text{eV})^{-3/2} h^{-1} \text{M}_\odot$. The contribution of $N_\nu$ massive, light ($m_\nu \ll 1$ MeV, so that they decouple while still relativistic, see Section 3.2.1) neutrinos to the mean matter density, relative to the critical density, is $\Omega_\nu \simeq N_\nu (m_\nu / \text{eV}) h^{-2} / 91.5$. On scales $k >> k_{FS}$, the fractional suppression of power due to massive neutrinos relative to pure CDM model approaches the asymptotic value of $\sim (1 + 8 \Omega_\nu / \Omega_M)^{-1}$ [195, 95].

This modification to the power spectrum on small scales has been studied in detail by many authors (e.g., Refs. [195, 95, 101]). In fact, the current best bounds on neutrino masses come from demanding consistency of the power spectrum on scales probed by COBE and the smaller scales probed by clusters [112] and the Lyα forest [76] or from the shape of the observed galaxy power spectrum on scales $0.01 \lesssim k \lesssim 0.2$ [104]. Roughly speaking, these cosmological bounds dictate that three
nearly mass-degenerate neutrinos must have \( m_\nu \lesssim 1 \) eV, while direct bounds on the electron neutrino mass from tritium decay experiments give \( m_\nu \lesssim 2.7 \) eV [325, 192].

In what follows, I study the effect of the suppression of small-scale power by three neutrinos with effectively degenerate masses \( m_\nu \lesssim 1 \) eV on the central densities of dark matter halos. As was pointed out by Fukugita, Liu, and Sugiyama [112], excessively large neutrino masses lead to unacceptably low values of \( \sigma_8 \). The strategy that I adopt here is to fix \( \Omega_{M,0} = 1 - \Omega_{\Lambda,0} = 0.3 \) and \( n = 1 \) and to ascertain whether or not a neutrino mass that saturates the lower limit of \( \sigma_8 > 0.55 \) can alleviate the central densities problem associated with the CDM paradigm. For this cosmology, the lower limit on \( \sigma_8 \) is saturated by a neutrino with \( m_\nu = 0.65 \) eV\(^9\). For comparison, I also report results for a model with \( m_\nu = 0.5 \) eV which has \( \sigma_8 \simeq 0.64 \). I show in Fig. 5.2, the present-day linear power spectra of the scale-invariant reference model and the two models with massive neutrinos. Notice the suppression of power on all mass scales relevant to galactic properties, \( M \lesssim 10^{13} \) M\(_\odot\).

5.4 The Central Densities of Dark Matter Halos

I describe halo structure and densities using the quantities defined in Section 4.2. To find the central densities of dark matter halos predicted by the aforementioned inflationary models, I make use of the semi-analytic concentration model of Bullock et al. [48] who were stimulated by the previous successes of NFW [218, 219, 220]. This model has been calibrated to the results of high-resolution cosmological N-body simulations. The Bullock et al. model was shown to work well in predicting the redshift and mass dependence of halo profiles for a standard ΛCDM model, and also

\(^9\)Incidentally, this lower limit reinforces my earlier statement that cosmological observations demand that \( m_\nu \lesssim 1 \) eV
Figure 5.2: Power spectra with massive neutrinos compared to the standard, scale-invariant power spectrum with no massive neutrinos. The different neutrino masses are labeled in the Figure.

reproduces the \( z = 0 \) results presented by NFW for standard CDM, open CDM, \( \Lambda \)CDM, and several power-law models [48]. This model also correctly predicts halo density structure in so-called quintessence cosmologies in which the vacuum energy is not in the form of a cosmological constant, but the potential of a scalar field [179]. Although the model was developed in the context of scale invariant CDM power spectra, it has also been shown to work remarkably well in predicting the results of an LCDM simulation with significant \( n = 0.9 \), as discussed in Ref. [5]. Bullock et al. model represents an improvement over the previous NFW model because it reproduces the relationship between \( c_{\text{vir}} \) and \( M_{\text{vir}} \) observed in N-body simulations.
as a function of redshift whereas the NFW model fails at \( z > 0 \). Nevertheless, it is important to realize that the following treatment is simplified and untested over the full range of power spectra I apply it to. Specifically, this model has not been tested against simulations with running spectral indices nor has it been tested against simulations with a significant hot dark matter component. Note, however, that halos formed within hot + cold dark matter simulations do seem to follow an NFW profile down to \( \sim 2\% \) of the halo virial radius [178]. Ideally, the results presented here will motivate future work using N-body simulations that will confirm or refute these preliminary conclusions.

Briefly, the model of Bullock et al. [48] (as well as the model of NFW) embodies the fact that we expect the central densities of dark matter halos to reflect the mean density of the Universe at a time when the central region of the halo was accreting matter at a high rate [321]. Therefore, we expect halos with central regions that collapsed earlier to be denser than their late-forming counterparts. Accordingly, the first step in the model is to assign an epoch of collapse to a halo via the prescription that, at the collapse epoch \( z_c \), the typical collapsing mass, \( M_\star(z_c) \), is equal to some fixed fraction \( F \), of the halo's virial mass. Explicitly, I define

\[
M_\star(z_c) \equiv FM_{\text{vir}}. \tag{5.11}
\]

\( M_\star(z_c) \) is the mass scale at which the rms density fluctuation is equal to the equivalent linear overdensity at collapse, \( \delta_c \simeq 1.69^{10} \). If \( \sigma(M, z) \) is the rms overdensity on mass scale \( M \) at redshift \( z \) [I use \( \sigma(M) \) with no redshift argument to denote \( \sigma(M, z = 0) \)]

\(^{10}\)The equivalent linear overdensity at collapse is a function of cosmology. NFW have found that a good approximation to the equivalent linear overdensity at collapse is \( \delta_c(z) \simeq 1.686\Omega_M^{-0.0055}(z) \) [220]. The matter density parameter evolves according to \( \Omega_M(z) = \Omega_{M,0}(1 + z)^3/[\Omega_{M,0}(1 + z)^3 + \Omega_{\Lambda,0}] \) in flat \( \Lambda \)CDM cosmologies.
as usual], then the collapse criterion can be written as $\sigma(M_*, z_*) = \delta_c$. Notice that 
this definition of the epoch of collapse differs from that of NFW who defined the 
collapse epoch using the extended Press-Schechter formalism [181]. This is the key 
difference that gives the Bullock et al. improved model the ability to trace the redshift 
dependence of the $M_{\text{vir}} - c_{\text{vir}}$ relationship. $F$ is a free parameter that describes the 
mass of the typical objects that are being accreted by the halo at $z_c$ and Bullock et al. 
found that the model is in good agreement with the results of N-body simulations if 
$F = 0.01$ [48]. The small value of the parameter $F$ is not surprising. The densities that 
characterize the very central regions of halos are determined by the power on scales 
much smaller than the size of the halo, scales that broke away from the expansion at a much earlier time than the mass scale $M_{\text{vir}}$. I will elaborate on this in the next 
Chapter when I discuss alternative models.

It is already evident that the central densities of dark matter halos are very sensi-
tive to $\sigma(M)$ on small scales and hence, to the slope of the primordial power spectrum 
or the presence of hot dark matter. At early times, well before the recent epoch of 
vacuum domination, $\Omega_M(z) \sim 1$, and $\sigma(M, z) \propto (1 + z)^{-1}$. Thus the epoch of collapse 
varyes approximately as $(1 + z_c) \propto \sigma(FM_{\text{vir}})$. If the central densities do, in fact, reflect 
the mean density of the Universe at the epoch of collapse then, roughly speaking, we 
expect $\Delta_{V/2} \propto \sigma^3(FM_{\text{vir}})$ so that a change in power by a factor of 2 leads to a change 
in central density by a factor $\sim 8$ [cf., Equation (4.15)].

The second step in the model is to relate the mass density of the Universe at $z_c$ 
to a characteristic halo density. Bullock et al. [48] chose to use $\bar{\rho}_s$ defined by 

$$M_{\text{vir}} \equiv \frac{4\pi}{3} r_s^3 \bar{\rho}_s.$$  \hspace{1cm} (5.12)
For an NFW profile, \( \tilde{\rho}_s = 3\rho_s f(c_{\text{vir}}) \). Introducing the free parameter \( K \), we associate \( \tilde{\rho}_s \) with the universal density at collapse via

\[
\tilde{\rho}_s = K^3 \rho_{\text{crit}} \Delta_{\text{vir}}(z) \Omega_M (1 + z_c)^3.
\]  

(5.13)

Solving Eqs. (5.12) and (5.13) for \( c_{\text{vir}} \) gives

\[
c_{\text{vir}}(M_{\text{vir}}) = K (1 + z_c(M_{\text{vir}})).
\]  

(5.14)

Agreement with N-body simulations fixes \( K = 4.0 \). Bullock et al. [48] and Wechsler et al. [321] claim that the statistical variance in the \( c_{\text{vir}} - M_{\text{vir}} \) relation among halos of the same mass is roughly \( \Delta \log(c_{\text{vir}}) \simeq 0.14 \) while Jing has argued for a somewhat smaller variance given by \( \Delta \log(c_{\text{vir}}) \simeq 0.08 \) [149]. With this model in place, I can use the linear power spectra of the previous two Sections to predict \( c_{\text{vir}}(M_{\text{vir}}) \) and, more practically, \( \Delta v_{/2} \) and compare these predictions with observations of dark matter-dominated dwarf and LSB galaxies.

5.5 Results

In this section I compare the predictions of the previous models with data on the rotation curves of dwarf and LSB galaxies. It is useful to concentrate our discussion on galaxies with both HI and H\( \alpha \) data or HI data that has been corrected for the effects of beam-smearing. The data I use are taken from the recent works of Swaters [291] (mass-modeling of these galaxies has been performed by van den Bosch and Swaters [311]) de Blok, McGaugh and Rubin [88] and de Blok and Bosma [86] who combined the existing HI measurements of Refs. [291, 312, 89, 287] with their high-resolution H\( \alpha \) rotation curve measurements. We use these data to derive observational estimates
of $\Delta_{V/2}$ for comparison with the theoretical predictions. The raw data of de Blok and Bosma [86] are currently not publicly available. Consequently I use their best fitting model for the dark matter distribution of each galaxy in the absence of baryons to derive estimates for $\Delta_{V/2}$. For the data of Swaters [291] and de Blok, McGaugh, and Rubin [88], I fit the raw data to the velocity profile proposed by Kravtsov et al. [178],

$$V_e(r) = V_e^0 \frac{(r/r_k)^\gamma}{[1 + (r/r_k)^\alpha]^{(\gamma+\beta)/\alpha}},$$

(5.15)

and use the best fitting models to estimate $\Delta_{V/2}$. Observed rotation curves generally do not robustly constrain all five of the parameters of the velocity profile of Equation (5.15), so in order to prevent the fitting routines from meandering through degenerate valleys in parameter space, I fix $\beta = 0.5$, which is the NFW velocity profile power law at very large radii. The quantities that I am exploring are defined on small scales so this pragmatic choice has little effect on the results presented here.

The profile in Equation (5.15) has the practical advantage that it parameterizes the sharpness of the transition between the two power laws at large and small radii through the quantity $\alpha$. Hence, the fitted value of the effective power law index at small radii is to some degree decoupled from the details of the rotation curve at $r \gtrsim r_k$. This added versatility makes it a very useful and accurate formula for describing observed rotation curves at small radii.

The estimates of $\Delta_{V/2}$ that I procure from fits to Equation (5.15) are robust in that for most galaxies in the aforementioned samples, the inferred values of $\Delta_{V/2}$ change by less than 40% if I instead fit the data with NFW, pseudo-isothermal or Burkert [54] density profiles. It is also interesting to note that there is no systematic difference in the derived $\Delta_{V/2}$ from one profile to the next. However, Moore profiles tend to fit the data more poorly, and give larger variation in the implied $\Delta_{V/2}$. This is similar
to the result found by van den Bosch and Swaters [311]. As I discussed above, the robustness of the central density parameter is a consequence of it being an integrated measure of density so any profile that faithfully represents observed rotation curves on scales of a few kpc can be used to estimate $\Delta_{V/2}$. This is yet another advantage of using $\Delta_{V/2}$ as a diagnostic of the central densities of dark matter halos. Nevertheless, the inferred values of $\Delta_{V/2}$ must be considered uncertain at the $\sim 30\%$ level.

Any comparison of the predictions of N-body simulations or semi-analytic calculations that model the behavior of CDM with data rests on some assumptions about the physics of baryon infall. Baryonic infall tends to make host halos more centrally concentrated as the baryons act to drag the dark matter towards the center of the halo [36, 35]. The consequence of this is that measured halos have higher $V_{\text{max}}$, smaller $r_{\text{max}}$, and therefore larger $\Delta_{V/2}$ [see Equation (4.13)], than commensurate halos prior to baryonic contraction or the pristine halos described by N-body simulations. In making this comparison, I believe that my methods are conservative in the sense that I likely overestimate the value of $\Delta_{V/2}$ that should be mapped onto the dark matter halo based on the observational data. This is conservative in the sense that I give the data every opportunity to match theoretical predictions (including the scale invariant "standard model"). First, I restrict the discussion to dwarf and LSB galaxies which are generally believed to be dominated by their dark matter components [291, 87]. In so doing, I anticipate that any effects of baryonic infall are mitigated but recognize the fact that I may be introducing a heretofore unappreciated selection effect. Second, I calculate $\Delta_{V/2}$ based on the raw rotation curve data without mass modeling or estimating baryon subtraction. I therefore overestimate the central density of the primordial dark matter halo because the cooling and contraction of the baryons likely
lead to contraction of the dark matter component as well (see Ref. [35], although see Ref. [323] for a competing effect). Third, the measured rotation curves of about 20\% of the galaxies in the sample may not extend out to large enough radii for an accurate determination of $V_{\text{max}}$ and consequently $V_{\text{max}}$ may be significantly underestimated for several galaxies. In these cases, I simply take the last point in the rotation curve as an estimate of $V_{\text{max}}$. By examining Eq. (4.13) it is easy to see that if $V_c(r) \propto r^\gamma$ with $\gamma \leq 1$ at small radii, an underestimation of $V_{\text{max}}$ by a factor $f_{V_{\text{max}}}$ leads to an overestimation of $\Delta V/2$ by a factor of $f_{V_{\text{max}}}^{2(1-1/\gamma)}$ (clearly, for $\gamma = 1$, corresponding to a constant density core, the error cancels exactly). In other words, the error introduced has the net effect of bringing theory and observation closer together.

In Fig. 5.3, I show the theoretical predictions for the concentration parameter $c_{\text{vir}}$ in the context of our inflationary models. Figure 5.4 shows the predictions in scenarios with massive neutrinos. Notice the wide swath of the $c_{\text{vir}} - M_{\text{vir}}$ plane that is carved out by the various models and, in particular, that $c_{\text{vir}}$ can be reduced by a factor of two or more by adopting primordial power spectra predicted by reasonable models of inflation or by adding neutrino masses that are not ruled out by any observation or experiment. Dark matter halos may be significantly less concentrated than standard LCDM plus scale-invariance predicts.

This can be quantified in another way that more directly confronts the so-called concentration problem. In Figures 5.6, 5.5, and 5.7, I show the same histogram of fitted $c_{\text{vir}}$ values as shown in Figure 4.4 in the previous chapter. The data used to construct this histogram are the same as those described here. In addition, I show the expected distribution of $c_{\text{vir}}$ values in the models with different primordial power spectra, the $\sigma_8 = 0.75$ tilted model, the $dn/d\ln k = -0.03$ running index
Figure 5.3: The median $c_{\text{vir}} - M_{\text{vir}}$ relation predicted by several different primordial power spectra. The predictions corresponding to the different primordial power spectra are labeled in the same fashion as in Fig. 5.1. The labeling scheme is explicitly shown in the key. Bullock et al. [48] estimate the 1σ scatter to be $\Delta \log(c_{\text{vir}}) \simeq 0.14$ while Jing argues for a smaller scatter of $\Delta \log(c_{\text{vir}}) \simeq 0.08$ [149]. These estimates for the 1σ scatter are illustrated by the error bars in the upper right corner.

model, and the broken scale-invariant (BSI) model. It is evident that the distributions predicted by these two alternative models more closely matches the apparent observed distribution. That the observed distribution has a larger scatter is not surprising: environmental effects as well as the effects of baryonic infall have been neglected in the theoretical modeling and these effects are likely to increase the scatter of the observed relation. Nevertheless, the low power models go a long way toward mitigating the concentration problem.

Unfortunately, the $c_{\text{vir}} - M_{\text{vir}}$ relation is not directly observable and, what is more, it is defined in terms of a particular density profile. To connect theory with
Figure 5.4: The median $c_{\text{vir}} - M_{\text{vir}}$ relation in models with massive neutrinos. The different lines are described in the key and the errorbars are as in Figure 5.3.

observations, it is better to compare the quantity $\Delta V/2$, as a measure of inner halo concentration, to $V_{\text{max}}$ as a measure of the absolute size of the halo. For an NFW profile, $V_{\text{max}}$ is related to $M_{\text{vir}}$ through Eqs. (4.1) and (4.8).

The results of this comparison are shown in Figs. 5.8 and 5.9. First, consider the predictions of the various models of inflation. Although the agreement or disagreement of a particular model with the data is hard to quantify, it is not surprising that inflationary models with $n > 1$ are effectively ruled out by the data. More interestingly, I find that, in agreement with previous studies [5], the $n = 1$ “standard” scale-invariant spectrum also has difficulty reproducing the observed galactic central densities. This is simply a restatement of the central density problem: if I am not preferentially selecting low density galaxies by restricting attention to low surface
Figure 5.5: The broken line represents a histogram of the best-fit values of $c_{\text{vir}}$ for sixty measured rotation curve. The heavy dashed line shows the expected distribution in the $\sigma_8 = 0.75$ model (neglecting any additional scatter due to baryonic or other effects). The light dotted line represents the expected distribution in the standard, $n = 1$, $\Lambda$CDM cosmology.

*If not enough galaxies, then some additional physics is needed to reconcile the standard model of CDM plus scale-invariant primordial spectrum with the observed central densities of dark matter-dominated galaxies.* The IPL4, $\sigma_8 = 0.83$ model does a somewhat better job of matching the data but the moderate tilt and spectral index running in this model are likely not sufficient to bring theory and observation together. For BSI, the agreement is much better but note that it is difficult to lower the theoretical $\Delta_{V/2}$ values further by adjusting the parameters of the model. Increasing $p$, the ratio of power on scales $k << k_c$ to power on scales $k >> k_c$, does not do much to help the BSI model come closer to matching the data because the
Figure 5.6: The broken line represents a histogram of the best-fit values of $c_{\text{vir}}$ for sixty measured rotation curve. The heavy dashed line shows the expected distribution in the $dn/d\ln k = -0.03$ RI model (neglecting any additional scatter due to baryonic or other effects). The light dotted line represents the expected distribution in the standard, $n = 1$, $\Lambda$CDM cosmology.

Fluctuation amplitude cannot drop quickly enough to produce a significant decrease in $\sigma(M)$ on the scales of interest. Meanwhile, $k_c$ cannot be increased much further without threatening the success of the standard model on large scales.

Notice that the running-mass model with $n < 1$ ($\sigma_8 = 0.65$) is a relatively good match to the median of the data in the $V_{\text{max}} - \Delta V/2$ plane (perhaps even undershooting the median). It is worth noting that this agreement has come without the need to saturate the lower bounds on spectral tilt from CMB and large scale structure ($n \approx 0.9 \pm 0.1$, see Refs. [128, 320]) or the lower limit on $\sigma_8$ that I adopted in Section 5.1. The central densities of dark matter halos are very sensitive to the initial power
Figure 5.7: The heavy dashed line shows the expected distribution of $c_{\text{vir}}$ in the BSI model (neglecting any additional scatter due to baryonic or other effects and with arbitrary normalization). Other lines are the same as in Figure 5.6.

spectrum and it seems as though the predicted central densities of dark matter halos in a CDM or ΛCDM cosmology may be reduced to acceptable levels by invoking simple and well-motivated models of inflation with $n < 1$ and/or a running spectral index.

Likewise, in the case of massive neutrinos, it is evident that by saturating the lower bound on $\sigma_8$, one may reduce the predicted median value of $\Delta_{V/2}$ to observationally acceptable levels. It seems that three massive neutrinos with $0.5 \text{ eV} \lesssim m_\nu \lesssim 0.65 \text{ eV}$ can decrease small-scale power enough to provide a relatively good match to the values of $\Delta_{V/2}$ inferred from rotation curve data.
Figure 5.8: $V_{\text{max}}$ vs. $\Delta v/2$ predictions compared with data. The symbols show the values of $\Delta v/2$ inferred from the rotation curve data. The data are taken from de Blok, McGaugh, and Rubin [88] (triangles and hexagons), de Blok and Bosma [86] (squares), and Swaters [291] (pentagons). The lines show the theoretical expectations for the different power spectra described in the text. The error bars in the upper right corner show the expected 1σ scatter in the theoretical predictions. The smaller range corresponds to the Jing [149] estimate and the larger range corresponds to the estimate of Bullock et al., [48].

5.6 Central Densities and the Power Spectrum: Conclusions and Discussion

The central density problem is one of several difficulties confronting the standard paradigm of structure formation that I reviewed in Chapter 4. In this Chapter, I explored potential resolutions that do not invoke uncertain baryonic physics while preserving the cold and collisionless properties of the dark matter particles. In Section
Figure 5.9: $V_{\text{max}}$ vs. $\Delta v/2$ predictions in models with massive neutrinos compared with data. The data points and the error bars in the upper right corner are the same as in Figure 5.8.

5.5, I showed that models of inflation that predict moderate, yet observationally acceptable, tilts $0.8 \lesssim n \lesssim 0.9$, may provide an acceptable solution to the central density problem. These tilts are consistent with the latest constraints from joint analyses of CMB anisotropy, large-scale structure and Ly$\alpha$ forest data [128, 320, 77, 274]. Moreover, these tilts can be produced in well-motivated models of inflation. In fact, I worked in the context of specific models throughout this Chapter and in so doing, I was able to take into account the important effect of the running of the spectral index. To illustrate the importance of the running, I also considered a “tilted” model with no spectral running and $n \simeq 0.84$ (the effective tilt of the RM $n < 1$ model on
the COBE scale) and found that this model predicts central densities that are more than 40% larger than the those predicted by the RM $n < 1$ model over the range $30 \text{ kms}^{-1} \leq V_{\text{max}} \leq 200 \text{ kms}^{-1}$. The effect of the very strong running the best-fit WMAP RI model is evident in that the slopes of the $c_{\text{vir}}-M_{\text{vir}}$ and $\Delta V/2-V_{\text{max}}$ relations are significantly different. Furthermore, the RI model with $dn/d\ln k = -0.03$ has an effective tilt of $n \approx 1.03$ on the COBE pivot scale, $k_*$, yet due to the running it predicts lower central densities IPL4 model with $n(k_*) \approx 0.94$. This is a striking example of the large lever arm that galactic and sub-galactic scale properties have for constraining the tilt and running (or perhaps more dramatic features) of the primordial power spectrum. Quantitatively, observations on galactic and sub-galactic scales almost double the effective range in scales that can be used to constrain $n(k)$ to $\Delta \ln k \sim 12$, from $\Delta \ln k \sim 6$ probed by the CMB and large-scale structure.

Given that precise measurements of the tilt of the power spectrum and the running of the spectral index using the data from the Sloan Digital Sky Survey [332, 268], future data releases from the MAP mission [24, 331], and the upcoming PLANCK CMB anisotropy mission [294] are on the horizon [102], it may be useful to adopt a purely empirical stance and consider the maximum tilt and running that are acceptable with respect to galactic central densities without linking the tilt and running through a particular inflationary model. As it is difficult to quantify the agreement or disagreement of a particular parameter choice with the data and because the current data certainly do not constrain the slope of the relationship between $V_{\text{max}}$ and $\Delta V/2$, I adopt the somewhat arbitrary, but sensible, criterion that a model predicts unacceptably diffuse galaxies if $\Delta V/2 \leq 10^5$ at $V_{\text{max}} = 100 \text{ kms}^{-1}$. Using the this criterion, I find that a lower limit on $n(k_*)$ allowed as a function of $dn(k_*)/d\ln k$ can
be approximated as
\[ n(k_*) + 6.77dn(k_*)/d \ln k \gtrsim 0.77. \]  \hspace{1cm} (5.16)

These maximally tilted models have \( \sigma_8 > 0.55 \) for \( n(k_*) \gtrsim 0.75 \). Adopting the criterion that a “good” fit to the data has \( \Delta V/2 \approx 3 \times 10^5 \) at \( V_{\text{max}} = 100 \text{ km s}^{-1} \), then a good fit to the data is given approximately by
\[ n(k_*) + 6.97dn(k_*)/d \ln k \approx 0.87. \]  \hspace{1cm} (5.17)

Notice that a “good” fit with \( dn(k_*)/d \ln k = -0.03 \) as in the RI model, gives \( n(k_*) \approx 1.06 \), close to but higher than the WMAP prediction for the effective tilt on the scale \( k_* \).

I also showed that massive neutrinos with \( 0.5 \text{ eV} \lesssim m_\nu \lesssim 0.65 \text{ eV} \) may provide an alternative solution to the central density problem; however, this solution seems to be rather less attractive. In order for neutrinos to solve the central density problem, it is necessary to nearly saturate the observational lower limit on \( \sigma_8 \) because, relative to the standard scale-invariant model, the power spectrum is damped by a factor \( \sim (1 + 8\Omega_\nu/\Omega_{\text{M,pl}})^{-1} \) on scales smaller than \( \sim 10^{16} \text{ M}_\odot \) (corresponding, roughly speaking, to \( k \approx k_{\text{FS}} \)) whereas in the inflationary models, power falls off continuously with increasing wave number. Again, a tilted spectrum can take advantage of the large lever arm from the Sachs-Wolfe plateau to the scales of galaxies. The range of neutrino masses allowed by the above criterion that the dark matter halos not be too diffuse is \( m_\nu \lesssim 0.9 \text{ eV} \), but as I mentioned earlier a neutrino mass greater than \( \sim 0.65 \text{ eV} \) leads to unacceptably small values of \( \sigma_8 \). A neutrino mass of \( m_\nu = 0.9 \text{ eV} \) implies that \( \sigma_8 \approx 0.46 \). 

146
I did not deal directly with the issue of central slopes in this Chapter. The problematic issue here is that cold (and warm) dark matter halo densities diverge at small radii whereas galactic rotation curves seem to be fit better with constant density cores [107, 212, 54, 34, 214, 88, 86]. While this is a worrisome situation, as I discussed above and at length in Chapter 4, it is difficult to tell the degree to which this is a serious challenge to ΛCDM. Recall that all observational errors favor constant density cores over cusps. Second, while it has been observed that pseudo-isothermal density profiles with constant density cores often fit galactic rotation curves better than NFW profiles [88, 86], the conclusion that observations indicate halos have cores rather than cusps is a non sequitur. First, all points on the curve contribute to the fit. Rotation curve fits are often largely determined by the transition region between the two power laws of the profile, and may not be faithful representations of the observed rotation curves at small radii (where there are relatively few data points). In addition, van den Bosch and Swaters showed that most rotation curves can be acceptably fit by divergent density profiles as long as the galaxies are much less centrally concentrated than standard LCDM predicts [311]. Lastly, the constant density cores of Burkert [54] or pseudo-isothermal profiles are not motivated by any theory and as NFW profiles are acceptable fits, there is no reason to abandon the NFW hypothesis until much more precise data can be brought to bear. From a practical standpoint, it is probably more useful to use a fitting form similar to Eq. (5.15) to address the issue of the observed value of the inner slopes of rotation curves, despite the fact that it is not inspired by a theoretical model, because it “decouples” the two power laws of the model rotation curve. The solution to the central density problem presented here likely cannot solve the cuspy halo issue by itself because central cusps are more a
reflection of the cold and collisionless properties of the dark matter than the amount
of small-scale power \(e.g., [92]\)). For that matter, however, warm dark matter models
likely cannot solve the cuspy halo issue either unless baryonic processes can be shown
to wipe out central cusps in large spirals \(e.g., [323]\). Nonetheless, the cuspy halo
issue is to some degree degenerate with the central density problem and it may be
that solving the latter problem may go a long way toward resolving the former.

A third problem associated with the standard \(\Lambda\)CDM paradigm is the dwarf satel-
line problem \([163, 213]\). In essence, this problem can be stated in the following way:
standard LCDM overpredicts the number of satellite halos with 10 \(\text{km s}^{-1}\) \(\lesssim\)
\(V_{\text{max}} \lesssim 40\)
\(\text{km s}^{-1}\) by as much as an order of magnitude relative to the number of observed satel-
lite galaxies in the local group. As I mentioned earlier, Kamionkowski and Liddle
\([152]\) investigated solving this problem with BSI initial power spectra. It is probable
that at least part of the solution lies in a feedback mechanism, like reionization sup-
pression \([49]\). However, the degree of feedback needed will depend crucially on the
input power spectrum. I will examine the subhalo issue in the context of inflation in
detail in the following Chapter. Briefly, I find that the discord between theory and
observation can be greatly allayed by considering models similar to those studied here
and thus, the feedback needed to meet observations can be greatly reduced or even
eliminated \([339]\).

Related to the dwarf satellite problem is the recent result of Dalal and Kochanek
\([81]\). The perturbing effect of substructure in strong gravitational lenses allowed
them to constrain the fraction of the host halo mass bound up in substructure to be
\(0.006 \leq f_{\text{smt}} \leq 0.07\) (90\% confidence). They then applied this result to limit the tilt
of the primordial spectrum and put constraints on the neutrino mass. They obtained
$n \geq 0.94$ and $m_\nu \leq 0.74$ eV at 95\% confidence [82]. The results on substructure that I present in the following Chapter differ significantly from those of Dalal and Kochanek [82]. I find that for a host halo of the relevant mass, the total mass fraction in subhalos is typically larger than the lower limit of Dalal and Kochanek ($f_{\text{sat}} \geq 0.006$) even with significantly tilted primordial spectra, $n \lesssim 0.8$. Thus the tilt of the primordial power spectrum may not yet be significantly constrained by strong lensing results [339]. However, as I have demonstrated here, the long “lever arm” from COBE scales to the sub-galactic regime offers a potentially useful avenue for constraining models of inflation and hints at the intriguing possibility that galaxy rotation curves may be telling us something fundamental about the early Universe.
CHAPTER 6

DARK MATTER HALO SUBSTRUCTURE AND THE PRIMORDIAL POWER SPECTRUM

6.1 Halo Substructure: Introduction and Background

In the previous two Chapters I discussed the impressive successes of the standard ΛCDM cosmology on large scales and the challenges to ΛCDM on small scales. In Chapter 5, I discussed the fact that dark matter halos seem to be significantly less dense than the standard ΛCDM model with scale-invariant primordial power spectrum would predict. I explicitly showed that the central densities of ΛCDM dark matter halos can be brought into reasonable agreement with the rotation curves of dark matter-dominated galaxies by simply reducing galactic-scale fluctuations in the initial power spectrum relative to horizon-scale fluctuations [337] (see also Refs. [5, 205, 310] which reach similar conclusions). In this chapter, I build directly upon this work.

In Section 4.5, I discussed the fact that one of the problems with the standard cosmology is the mismatch between the observed satellite population of the Milky Way and Andromeda, and the number of small satellite halos observed around Galaxy-sized halos in numerical simulations. Inspired by the successes of the previous Chapter, in
this Chapter I explore how changes in the initial power spectrum affect the substructure content of $\Lambda$CDM halos. This extension is important for several reasons. First, there currently exists no extensive study of substructure populations as a function of the input power spectrum. The effects of changes in the primordial power spectrum are unknown. Second, it is imperative that any proposed solution to the central density problem in the very least not exacerbate other known problems, such as the dwarf satellite problem. It is also interesting to consider that a single modification may alleviate all of the small-scale woes of the standard cosmology, leading to a remarkably simple, consistent, and unified model, instead of introducing a set of piecemeal additions that solve problems individually. This drive for consistency is a major motivator of this work. Furthermore, there have been concerted efforts recently to confront the dwarf satellite problem in another way, by detecting dark matter halos that are truly dark and do not host luminous galaxies [151, 144, 143, 203, 81, 158, 215] and in so doing, find the “missing” Galactic satellites. A particularly promising avenue for directly detecting dark matter substructure in galaxy-sized halos is through gravitational lensing [158, 215] and some studies already have claimed to find large amounts of substructure. As I discuss below, the discovery of large amounts of substructure may fundamentally change the nature of the dwarf satellite problem: instead of a problem with the number of subhalos, the problem will become one of galaxy formation and understanding why it is some halos host galaxies while others do not “light up.” Therefore, it is important to understand the cosmological predictions for substructure populations in order to appreciate the nature of the dwarf satellite problem and to interpret independent measures of substructure as tests of and constraints on the standard model.
It is straightforward to grasp why a fundamental prediction of the CDM paradigm is that galactic halos play host to a large number of distinct, bound substructures, or “subhalos.” Substructure is a natural outcome of the modern picture of hierarchical structure formation in which low mass systems collapse early and merge to form larger systems over time [330, 36, 156]. Small halos collapse at high redshift when the universe is very dense, so their central densities are correspondingly high. This observation forms the basis of the model that I discussed in the previous chapter. When these small halos merge into larger hosts, their high densities allow them to resist the strong tidal forces that act to destroy them. While gravitational interactions do serve to unbind most of the mass associated with merged progenitors, a significant fraction of these small halos survive as distinct substructure.

Our understanding of this process has increased dramatically in the last five years thanks to remarkable advances in N-body techniques that allow the high force and mass resolution necessary to model halo substructure in detail (see Refs. [114, 115, 176, 163, 162, 170, 213, 214, 108, 288]). For the halos that have been simulated at sufficiently high resolution in standard $n = 1$, $\Lambda$CDM cosmological simulations, the total mass fraction bound up in substructure is measured at $f \sim 5-15\%$ [114, 162], with a significant portion contributed by the most massive subsystems, $dN/dM \propto M^{-\gamma}$, with $\gamma \approx 1.6 - 2.0$. The substructure content of halos seems to be roughly self-similar when the subhalo mass is scaled by the host halo mass [213] and the subhalo count is observed to decline at the host halo center, where tidal forces are strongest [114, 66, 61].

Unfortunately, issues of numerical resolution are endemic to the study of halo substructure using N-body simulations. Simulations with the capability to resolve
substructure are extremely computationally expensive, requiring months of supercomputer time to obtain information on substructure in a fairly small number of halos. These tools cannot be used to study the implications of many unknown input parameters and cannot attain both the resolution and the statistic samples needed to confront observational data on substructure that appear to be on the horizon simultaneously. Moreover, even the state of the art simulations face difficulties in the centers of halos where “overmerging” becomes more of a problem (see the discussions in [61] and [162]). As I discuss below, attempts to measure the substructure fraction via lensing are highly sensitive to these uncertain, central regions.

The goal of this Chapter is to present and apply a semi-analytic model (as opposed to an N-body simulation) that suffers from no inherent resolution effects and that is based on the processes that were observed to govern substructure populations in past N-body simulations. Having discussed the drawbacks associated with N-body work, the advantages of this type of study are manifest. This kind of model can generate predictions for hundreds of halos, therefore the predictions are statistically significant. Moreover, these predictions can be calculated very quickly and with limited computing resources. This semi-analytic model can therefore be used to study the dependence of substructure on a variety of inputs and these predictions can be used to guide expectations for the next generation of N-body simulations. Conversely, the model that I present here represents in many ways an audacious extrapolation of N-body results into unexplored domains and it is imperative that these results be tested by future numerical studies. At present, my objective is to explore how assumptions about the power spectrum affect the population of surviving subhalos in

153
models like the standard ΛCDM cosmology; however, in principle these methods are suitable for testing substructure ramifications for a variety of cosmological inputs.

As I alluded to previously, one of the primary motivations for this work comes from simulation results that indicate that Galaxy-sized ΛCDM halos play host to hundreds of subhalos with maximum circular velocities in the range $10 \text{ km s}^{-1} \lesssim V_{\text{max}} \lesssim 30 \text{ km s}^{-1}$. The Milky Way, as a comparative example, hosts only 11 dwarf satellites of similar size within $\sim 300$ kpc of the Galactic center. This is the dwarf satellite problem and it specifically refers to the apparent discordance between the predicted number of ΛCDM subhalos and the number count of satellite galaxies in the Local Group [163, 213, 108]. Kauffmann et al. [156], also intimated that there may be a problem using simple, analytic arguments. The dwarf satellite problem and other small-scale issues led many authors to consider modifications to the standard framework. If the dark matter were “warm” [232, 67, 69, 137, 37, 165] sub-galactic scale problems with CDM may be allayed without vitiating the overall successes of ΛCDM on large scales as discussed in the previous Chapter. Unfortunately, it seems that the central density problem cannot be cured by a introducing a WDM particle that is a thermal relic (see Section 3.3) because the mass of the particle needed to solve the problem would violate previous bounds, $m_{\text{W}} \gtrsim 1 \text{ keV}$ [5, 216, 21]. Nevertheless, the dark matter may be effectively warm if they are not thermal relics; it is possible to endow the dark matter particles with significant velocity dispersions by allowing them to be the daughters of massive relic decays [187]. If the primordial power spectrum were sharply truncated on small scales then sub-galactic-scale problems may likewise be remedied [277, 152]. Another possibility is that CDM substructure is abundant in all galaxy halos, but that most of the low-mass systems are simply devoid of stars.
An intermediate solution may involve a simple modification of the assumed primordial spectrum of density perturbations that lowers power gradually on galactic scales relative to the horizon, e.g., via tilting the power spectrum.

Some issues with small-scale structure may be resolved by observational improvements so it is important to emphasize that probing models with low galactic-scale power is motivated not solely by the small-scale crises facing standard ΛCDM but also by more direct probes of the power spectrum. While many analyses continue to measure "high" values for $\sigma_8 \sim 0.9 - 1.0$ [314, 173, 14] (as I stated earlier, $\sigma_8$ is the linear, rms fluctuation amplitude on a length scale of $8 \, h^{-1} \text{Mpc}$, see Section 2.3), an ever-growing number of recent studies, relying on similar techniques, advocate what would be considered "low" values with $\sigma_8 \sim 0.7 - 0.8$ [146, 16, 260, 317, 44, 6, 125, 241, 208, 40]. Similarly, the Ly-α forest measurements of the power spectrum by Croft et al. [78, 77] and McDonald et al. [204] are consistent with reduced galactic-scale power. Set against the normalization of fluctuations on large scales implied by the Cosmic Background Explorer measurements of cosmic microwave background radiation anisotropy [23, 52, 53], these measurements suggest that the primeval power spectrum from inflation may be tilted to favor large scales, with $n < 1$, may have a non-negligible running, $dn/d\ln k < 0$, or something even stranger.

The recent analysis of the Wilkinson Microwave Anisotropy Probe measurements of CMB anisotropy presented by Spergel et al. [274, 316, 237] (see Section 2.5 and particularly Table 2.1) returns a best-fit spectral index to a pure power law primordial spectrum of $n = 0.99 \pm 0.04$ when only the WMAP data are considered. However, when data from smaller scale CMB measurements, the 2dF Galaxy Redshift Survey, and the Ly-α forest are included, the analysis favors a mild tilt, $n = 0.96 \pm 0.02$. 

155
Interestingly, all of the data sets together yield a better fit if the index is allowed to run. The WMAP team find what may prove to be the first indications of spectral index running (discounting what the small-scale structure crises may have been hinting at all along), \( \frac{dn}{d\ln k} = -0.031^{+0.016}_{-0.017} \). This result is consistent with no running at \( \sim 2\sigma \) and the statistical significance of this result is further weakened when additional uncertainties in the mean flux decrement in the Ly-\( \alpha \) forest are considered [261], yet such a possibility certainly seems worth investigating. This is especially true in light of the fact that it may temper the wind to the shorn lamb as far as small-scale difficulties with CDM are concerned.

I explicitly showed how models with reduced small-scale power help to remedy the halo density problem in Chapter 5, particularly in Figures 5.5, 5.6, 5.7, 5.8, and 5.9. Clearly, the data favor low small-scale power relative to the standard \( n = 1 \) case. The running index (RI) model that is based on the WMAP measurements does quite well, for instance. In this Chapter, I will consider many of the same power spectra.

The possibility of discriminating between standard \( n = 1 \), \( \Lambda \)CDM and models with reduced power on small scales has inspired significant efforts to measure and quantify the substructure content of galactic halos. Some methods are direct and focus on observations within the Milky Way. One potentially useful example relies on studying the tidal tails associated with known Galactic satellites [151, 144, 143, 203, 108]. Subhalos passing through cold tidal streams tend to scatter stars away from their original orbits, and the signatures of this debris from tidal stripping may be detectable in the velocity data that will be provided by future astrometric missions and several deep halo surveys that will soon be completed. These methods, however, can only be applied to the Milky Way.
Of particular interest for obtaining measurements in large numbers of distant galaxies are studies that aim to detect substructure via flux ratio anomalies in strong gravitational lenses [210, 211, 158, 42]. Using a sample of seven four-image radio lenses, Dalal & Kochanek [81] estimated a mass fraction in substructure of \( f = 0.006 - 0.07 \) (90% confidence level) bound up in substructure less massive than \( \sim 10^8 - 10^{10} \) \( M_\odot \), in line with the rough expectations of CDM. (In their original paper, Dalal & Kochanek (2002) quoted an estimated upper mass limit of \( 10^6 - 10^9 \) \( M_\odot \). They have since concluded that an upper limit of \( \sim 10^8 - 10^{10} \) \( M_\odot \) is more appropriate [80].)

While measurements of this kind are susceptible to potential degeneracies with the adopted smooth lens model and other uncertainties, they are encouraging and serve as prime motivators for this work (see Ref. [167]). In addition, new observational techniques that focus on astrometric features [209], and particularly spectroscopic studies of strong lensing systems like that of Moustakas and Metcalf [215], promise to hone in on the masses of the subclumps responsible for the lensing signals.

If the results of the first lensing studies are confirmed and the Milky Way really is surrounded by a large number of dark subhalos, the dwarf satellite problem serves as a conspicuous reminder that feedback mechanisms must play an important role in hierarchical galaxy formation. Of course, the need for feedback in small systems has been generally recognized since the pioneering work of White and Rees [330]. Supernova blow-out likely plays a key role in regulating star formation if CDM is the correct theory [93, 156, 63, 272]; however, supernova winds do not naturally suggest a natural feedback scale at \( V_{\text{max}} \sim 30 \) km s\(^{-1}\), nor does supernova blow-out explain why some halos of this size should have stars while most have none at all. It seems more
likely that supernovae play an important role in setting scaling relations in slightly larger galaxies \( V_{\text{max}} \sim 100 \text{ km s}^{-1} \) [94, 197].

Perhaps a more natural feedback source on satellite galaxy scales is the ionizing background, which should suppress galaxy formation in halos with \( V_{\text{max}} < 30 \text{ km s}^{-1} \) [250, 263, 302, 246, 116]. Bullock, Kravtsov, & Weinberg (BKW) [49] suggested that dwarf galaxies should be associated with small halos that collapsed before the epoch of reionization, \( z_{\text{rec}} \). BKW estimated that this would work in standard \( \Lambda \text{CDM} \) if \( z_{\text{rec}} \approx 6.5 - 12 \). The general idea of feedback mechanisms and the efficacy of the reionization feedback scenario are depicted in Figure 6.1, which shows an example of the BKW proposal. For the example is shown, it is assumed that for all halos with \( V_{\text{max}} < 30 \text{ km s}^{-1} \), a halo must have had > 30\% of its mass in place prior to reionization. The threshold of 30 km s\(^{-1}\) is based on the result of Thoul and Weinberg [302] and reionization is assumed to occur at a redshift of \( z_{\text{rec}} = 8 \). Notice that this is a differential feedback mechanism in that \( \sim 1/10 \) of all halos below the threshold go on to host observable galaxies. Further, observe that the Milky Way number count is better than a 1\( \sigma \) match to the theoretical predictions in this case. However, it must be borne in mind that this is only an example, illustrating the role of feedback in the dwarf satellite problem. Though the method used by BKW to estimate dwarf luminosities was crude, more sophisticated models have since led to similar conclusions [62, 269, 26]. For the very smallest systems, \( V_{\text{max}} \lesssim 10 \text{ km s}^{-1} \), the ionizing background likely prevents star formation altogether by photo-evaporating gas in halos, even after they have collapsed [22].
Figure 6.1: This Figure demonstrates the ability for reionization feedback to reconcile the predictions of ΛCDM with the observed number of galactic satellites. The solid line represents the predictions of standard ΛCDM. The observed velocity function of the Milky Way satellites is shown by the line marked with squares. These lines are the same as in Figure 4.6. The dashed line represents the predictions of a model of reionization that requires that a halo must have had $\geq 30\%$ of its mass in place by reionization at a redshift of $z_{re} = 8$ in order to host a visible galaxy (see [49] for details).

Precisely what can be learned about galaxy formation and/or cosmology by counting dwarf satellites depends sensitively on one’s expectations for the density and velocity profiles of their host halos. To count satellites of a given maximum circular velocity, we must infer a halo $V_{\text{max}}$ using the observed central velocity dispersion $\sigma_*$, and the mapping between these two quantities depends sensitively on the precise nature of each satellite’s dark matter halo [328]. If a dwarf galaxy sits in a dark halo with a very slowly rising rotation curve that reaches a maximum at $r_{\text{max}} \gg \text{kpc}$, then the multiplicative factor that converts a $\sim 1$ kpc velocity dispersion measurement to
the halo’s maximum circular velocity can be quite large because the velocity dispersion information comes from only the inner \( \sim \) kpc, where the stars are. This is shown graphically in Figure 6.2 where I show the measured stellar profile of Carina [200] along with two rotation curves that both lead to the same observed central velocity dispersion. Notice that Carina can be associated with dark matter halos that span a wide range of \( V_{\text{max}} \) and still be consistent with observational data. I will discuss this issue in detail in Section 6.8. Standard ΛCDM halos of \( \sim 30 \text{ km s}^{-1} \) in size are expected to be very concentrated as I mentioned in Chapter 5 [67, 48], with \( r_{\text{max}} \sim 1 \) kpc, so the multiplicative factor that maps velocity dispersion to maximum circular velocity is rather modest: \( V_{\text{max}} \sim \sqrt{3} \sigma_* \). Klypin et al. [163] adopted this conversion when they first identified the dwarf satellite problem. However, as I discuss below, the appropriate conversion is cosmology-dependent because models with later structure formation tend to produce halos with more slowly rising rotation curves, implying a larger multiplicative correction [339, 338, 51]. Shifts of this kind in the “observed” velocity function change the implied velocity (or halo mass) scale of discrepancy, and perhaps influence our ideas about the type of feedback that gives rise to the mismatch.

Hayashi et al. [135] and Stoehr et al. [288] have suggested that substructure halos experience significant mass redistribution in their centers as a result of tidal interactions and that they are therefore less concentrated than comparable halos in the field. They argue that when this is taken into account, the dwarf satellite mismatch sets in at \( V_{\text{max}} \sim 20 \text{ km s}^{-1} \), and that the transition is sudden — that below this scale all halos are completely devoid of observable galaxies. While these conclusions have yet to be confirmed and are dependent upon subhalo merger histories and the isotropy of dwarf velocity dispersions, they highlight the need to refine predictions about halo
Figure 6.2: Two dark matter rotation curves that lead to the same central line-of-sight velocity dispersions for Carina. The dashed line represents the observed stellar density profile for Carina. The solid and dot-dashed lines represent two rotation curves that, when coupled with the observed stellar profile of Carina, lead to the same prediction for the central velocity dispersion. Notice that these halos differ in $V_{\text{max}}$ by a factor of $\sim 2.5$. The units on the ordinal axis are arbitrary.

substructure. They also motivate us to explore how minor changes in cosmological parameters can influence our interpretation of the dwarf satellite problem.

In the remainder of this Chapter I present a detailed study of halo substructure. In Section 6.2, I describe a semi-analytic model for halo substructure, provide some illustrative examples, and compare the results of this model for standard $\Lambda$CDM to previous N-body results. In Section 6.3, I briefly review the input power spectra that serve as the basis for this study. I present the first results regarding accretion histories in Section 6.4. In Section 6.5, I present results on subhalo mass functions and velocity functions. I make predictions aimed at measuring the projected mass fraction

161
of halos via gravitational lensing in Section 6.6. In Section 6.7, I discuss the warm dark matter (WDM) alternative to CDM and present results aimed at testing the WDM model with gravitational lensing. In Section 6.8, I address the dwarf satellite problem in light of some of the findings of this work. In Section 6.10 I discuss some shortcomings of the model presented here and how they might be improved in future work. In Section 6.12 I summarize this study and draw conclusions from these results. Although this substructure recipe is more generally applicable, in this study, I vary only the primordial power spectrum and work within the context of the so-called “concordance” cosmological model with $\Omega_{M,0} = 0.3$, $\Omega_{\Lambda,0} = 0.7$, $h = 0.72$, and $\Omega_{b,0} h^2 = 0.02$ [274, 307].

6.2 Modeling Halo Substructure

6.2.1 Overview

In order to determine the substructure properties of a dark matter halo one must model its mass accretion history as well as the orbital evolution of the subsystems once they are accreted. For the first step, I rely on the the well-studied extended Press-Schechter (EPS) formalism to create merger histories for each host system. The EPS formalism is statistical and reflects the fact that halos of a given mass today may have had significantly different mass assembly histories. Using EPS one can account for the statistical halo-to-halo variation in any given cosmology by generating accretion histories for a significant number of halos. I give a very brief description of EPS merger trees in Section 6.2.2. In Section 6.2.3 I review models of the density structure of accreted halos and the host system and in Section 6.2.4. I describe my method for
following the orbital evolution of each merged system. I show tests and examples of
this model in Section 6.2.5.

6.2.2 Extended Press-Schechter Theory and Hierarchical Halo
Merger Histories

The Press-Schechter Halo Multiplicity Function

The costliness of N-body simulations makes it important to develop analytic meth-
ods for studying the clustering of matter. The most popular method is that devel-
oped by Press and Schechter based on the statistics of the perturbation field. They
assumed the density field to be Gaussian, which is a prediction of the simplest infla-
tionary models in which the Fourier modes have uncorrelated random phases. They
considered the statistics of fluctuations as a function of mass scale by examining the
density field smoothed on the window scale \( R_W \)

\[
\delta(\vec{x}, M) = \frac{\int \text{d}^3 y \ W(|\vec{y} - \vec{x}|) \delta(\vec{y})}{\int \text{d}^3 y W(\vec{y})},
\]

(6.1)

where \( W(r) \) is a spherically symmetric smoothing window of approximate radial ex-
tent \( \sim R_W \), and is mapped onto a mass scale by

\[
M = \rho_M \int \text{d}^3 x W(r),
\]

(6.2)

where \( \rho_M \) is the mean matter density. The window volume is given by \( V_W \equiv \int \text{d}^3 x W(\vec{x}) \). It is common practice to scale the window function such that the volume
is unity. Adopting this convention, the most popular window functions are: the real
space top hat,
\[ W(r) = \frac{3}{4\pi R_w^3} \text{ if } r < R_w \]
\[ = 0 \text{ otherwise,} \quad (6.3) \]

with Fourier transform

\[ W(k) = 3V^{-1} \frac{\sin(kR_w) - kR_w \cos(kR_w)}{(kR_w)^3}; \quad (6.4) \]

the Gaussian window

\[ W(r) = \frac{1}{(2\pi R_w^3)^{3/2}} \exp \left( -\frac{r^2}{2R_w^2} \right), \quad (6.5) \]

with

\[ W(k) = V^{-1} \exp \left( -\frac{R_w^2 k^2}{2} \right); \quad (6.6) \]

and the k-space top hat,

\[ W(r) = \frac{9}{4\pi R_w^3} \frac{(\sin(r/R_w) - (r/R_w) \cos(r/R_w))}{(r/R_w)^3}, \quad (6.7) \]

with the obvious Fourier transform,

\[ W(k) = V^{-1} \text{ if } kR_w < 1 \]
\[ = 0 \text{ otherwise.} \quad (6.8) \]

The factors of fiducial volume, \( V \), arise from the Fourier conventions of Equations (2.8) and (2.9), but all observables are independent of \( V \).
At a given point in space, the probability of an overdensity $\delta$ on mass scale $M$, is then

$$P(\delta) d\delta = \frac{1}{\sqrt{2\pi}\sigma(M)} \exp \left( -\frac{\delta^2}{2\sigma^2(M)} \right),$$

where $\sigma(M)$ is the variance given in Equation 2.10 and is a monotonically decreasing function of mass in all practical cases. The next step is to assume that if the overdensity $\delta(M)$ at some point exceeds some threshold for collapse $\delta_c$, that this corresponds to a collapsed halo of mass $> M$, because if $\delta(M)$, is greater than $\delta_c$, then $\delta(M' > M)$ will eventually become equal to $\delta_c$ on some larger mass scale. The critical density for collapse is a weak function of cosmology and can be estimated using the approximation for the spherically symmetric collapse of a uniform top hat overdensity. This yields $\delta_c \simeq 1.686\Omega_M^{0.0055}$ in flat $\Lambda$CDM cosmologies [217]. The probability that a point lies within a virialized structure of mass $> M$ is then

$$F(> M) = \int_{\delta_c}^{\infty} P(\delta) d\delta = \frac{1}{2} \text{erfc} \left( -\frac{\delta_c}{\sqrt{2}\sigma(M)} \right).$$

Press and Schechter realized that this reasoning had no provision for dealing with underdensities and therefore only counted $1/2$ of the mass of the Universe. They therefore multiplied by a factor of 2 and obtained an expression for the number density of halos as a function of $M$ by differentiating Equation 6.10 and multiplying by $\rho_M/M$ to convert from mass to number, yielding

$$n(M) dM = \left( \frac{2}{\pi} \right)^{1/2} \frac{\rho_M}{M^2 \sigma(M)} \delta_c \exp \left( -\frac{\delta_c^2}{2\sigma^2(M)} \right) \left| \frac{d\ln \sigma(M)}{d\ln M} \right| dM,$$
which is in fairly good agreement with the results of N-body simulations. See Sheth and Tormen [264] and Jenkins et al. [147] for detailed comparisons with N-body codes and fits the the Press-Schechter functional form.

**Excursion Sets of the Density Field**

Bond et al. [38] built a more rigorous foundation for the Press-Schechter theory and developed an interpretation that made the theory more flexible by examining the excursion sets of the density field. The details are in Bond et al. [38]. I present a very brief summary here. Bond et al. considered the trajectories of the density field at fixed position $\delta(S)$, where $S \equiv \sigma^2(M)$ relates to the filter mass. At extremely large mass, $S \rightarrow 0$ and $\delta(S) \rightarrow 0$. They chose to identify bound objects by the largest mass, or smallest $S$, at which a particular realization of the density field first had $\delta(S) > \delta_c$.

In the case of the k-space tophat filter, this analysis is the simplest. In this case, each incremental increase in $S$ corresponds to the addition of uncorrelated extra k-modes so that each step is a step in a random walk process. The distribution of $\delta(S)$ at a given $S$ is then given by

$$P(\delta(S))d\delta = \frac{1}{\sqrt{2\pi S}} \exp\left(-\frac{\delta^2(S)}{2S}\right).$$

The probability distribution for all paths that have never meet the threshold prior to $S$ is then

$$Q = \frac{1}{\sqrt{2\pi S}} \left[ \exp\left(-\frac{\delta^2(S)}{2S}\right) - \exp\left(-\frac{(2\delta_c - \delta(S))^2}{2S}\right) \right]$$

166
for $\delta(S) < \delta_c$. Clearly, this probability goes to zero for all trajectories with $\delta(S) > \delta_c$ at “time” $S$. The number of trajectories that achieve $\delta(S) > \delta_c$ prior to $S$ is then

$$F(> M) = 1 - \int_{-\infty}^{\delta_c} Q d\delta. \quad (6.14)$$

Using the machinery of Brownian random walks, it can be shown that $Q$ satisfies a diffusion equation of the form $\partial Q / \partial S = 1/2 \partial^2 Q / \partial \delta^2$, and it is easy to show that Equation (6.13) is a particular solution to this equation. Differentiating (6.13) and using the diffusion equation results in the fraction of trajectories that first cross the threshold in the interval $S$ to $S + dS$:

$$f(S, \delta_c) dS = \frac{1}{\sqrt{2\pi \delta_c^3}} \exp\left(-\frac{\delta_c^2}{2S}\right) dS. \quad (6.15)$$

This results in the exact same mass function as the Press-Schechter analysis, Equation (6.11), including the factor of two that was added by Press and Schechter arbitrarily. In the language of the excursion sets, the factor of two comes from the trajectories that crossed above the threshold at a larger mass (smaller $S$) and then crossed back below. Press and Schechter ignored these so that an underdensity engulfed in a larger overdensity was not counted at the larger mass threshold. The excursion set formalism explicitly counts these as part of the larger mass object. This is the second term in Equation (6.13).

**Constructing Merger Trees**

The formulation of the hierarchical clustering of matter in terms of the excursion sets of the trajectories of the density field provides a basis for interpretation that allows one to calculate more complicated quantities. For instance, consider a fixed
density field with the evolution of time embodied in a continuously decreasing threshold, \( w(t) = \delta_c/D(t) \), where \( D(t) \) is the linear growth function. Lowering \( w(t) \) as \( D(t) \) increases corresponds to moving forward in time because overdensities that collapsed early in the history of the Universe would have a very large equivalent linear overdensity today. Conversely, raising the threshold corresponds to stepping backward in time.

Consider the conditional probability of a first up-crossing of threshold \( w_1 \) at time \( t_1 \) in the interval \( S_1 \) to \( S_1 + dS \), given that there is a first up-crossing of a barrier of height \( w_2 \) in the interval \( S_2 \) to \( S_2 + dS \), with \( w_1 > w_2 \), \( S_1 > S_2 \), and \( t_1 < t_2 \). In the language of trajectories of the density field as a function of \( S \), this corresponds to the conditional probability that a halo of mass \( M_2 \) (corresponding to \( S_2 \)) at time \( t_2 \) (corresponding to threshold \( w_2 \)) was formed from a progenitor halo of mass \( M_1 \) at time \( t_1 \). The method of solution is the same as before with the only difference being that the beginning of the trajectory \( \delta(S) \) is shifted from the point \( (S = 0, \delta = 0) \) to the point \( (S = S_1, \delta = w_1) \). The resulting conditional probability is [38, 181]

\[
f(S_1, w_1 | S_2, w_2) d(\Delta S) = \frac{\delta w}{\sqrt{2\pi}\Delta S^{3/2}} \exp \left( -\frac{(\delta w)^2}{2\Delta S} \right) d(\Delta S), \tag{6.16}
\]

where \( \Delta S \equiv S_1 - S_2 \) and \( \delta w \equiv w_1 - w_2 \). Repeated application of this equation can be used to determine the progenitor populations of halos of a given final mass at a given final time [181]. In this way, a halo “merger tree” that reproduces many of the results of N-body simulations can be constructed using only the linear power spectrum of density fluctuations.

I track diffuse mass accretion and satellite halo acquisition of host systems by constructing merger histories using the EPS method [38, 181]. There are several
approximate methods for building merger trees that differ in their details and there is no method that exactly reproduces the mean progenitor distribution predicted by the Press-Schechter mass functions while exactly conserving mass. In particular, I employ the merger tree algorithm proposed by Somerville and Kolatt that explicitly conserves mass at each time-step [271]. This allows one to generate a list of the masses and accretion redshifts of all subhalos greater than a given threshold mass that merged to form the host halo. I describe the method briefly here, and encourage the interested reader to consult the papers of Lacey and Cole [181] and Somerville and Kolatt [271] for further details.

The probability that a halo of mass \( M \), at time \( t \), accreted an amount of mass associated with a step of \( \Delta S \), in a given time step implied by \( \delta w \) is given by Equation (6.16). Merger histories are constructed by starting at a chosen redshift and halo mass and stepping back in time with an appropriate time step. If the minimum mass of a progenitor that we wish to track is \( M_{\text{min}} \), then Somerville and Kolatt tell us that each time step must be small in order to reproduce the conditional mass functions of EPS theory: \( \delta w \lesssim \sqrt{M_{\text{min}}(dS(M)/dM)} \) [271].

In this study, I build merger trees by selecting progenitors at each time step according to Equation (6.16) and treating events with \( \Delta M < M_{\text{min}} \) as diffuse mass accretion. At each time step, I identify the most massive progenitor with the host halo and the less massive progenitors with accreted subhalos and continue this process until the host mass falls below \( M_{\text{min}} \). In practice, I use a slightly modified version of the Somerville and Kolatt scheme. At each stage I demand that the number of progenitor halos in the mass range we consider be close to the mean value. As discussed by BKW, this modification considerably improves the agreement between the
analytically predicted progenitor distribution and the numerically generated progenitor distribution [49]. In the following, I set \( M_{\text{min}} = 10^5 \, M_\odot \). I take a fiducial, \( z=0 \) host mass of \( 1.4 \times 10^{12} \, M_\odot \), but I vary these choices in order to test the sensitivity of various results to the host mass and redshift of observation, as described below. I refer to each individual mass accretion history that I build from repeated application of Equation (6.16) as a realization of a merger history. In what follows, I sample the distribution of mass accretion histories by generating ensembles of between 50 and 200 realizations for each choice of input parameters.

### 6.2.3 Halo Density Structure

Whether a merged system survives or is destroyed depends on the density structure of the subhalo and on the gravitational potential of the host system. Therefore, it is worthwhile to describe all assumptions about CDM density profiles in some detail. I quantify the absolute size of a virialized dark matter halo in terms of its virial mass \( M_{\text{vir}} \), or equivalently its virial radius \( R_{\text{vir}} \), or virial velocity \( V_{\text{vir}}^2 \equiv GM_{\text{vir}}/R_{\text{vir}} \) as in Section 4.2. In the interest of simplicity and in light of the discussion of Section 4.2, I choose to model all halos with the density profile proposed by Navarro, Frenk, & White [218, 219, 220] and given explicitly by Equation (4.3).

As I discussed in Section 4.2, the results of the study by Wechsler et al. [321] and several notable precursors (e.g., Refs. [333, 219, 220, 48, 46, 12]) teach us that dark matter halo concentrations are determined almost exclusively by their mass assembly histories. The gross picture advocated by W02 is that the rate at which a halo accretes mass essentially determines how close to the center of the host halo the accreted mass is deposited. When the mass accretion rate is high, near equal mass mergers are very
likely, and dynamical friction acts to deposit mass deep into the interior of the host. After an early period of rapid mass accretion, the central densities of halos remain constant at a value proportional to the mean density of the Universe at the so-called “formation epoch” \( z_c \), defined as the redshift when the relative mass accretion rate was similar to the rate of universal expansion (see [321] for details). For typical halos, this formation epoch occurred at a time when halos were roughly \( \sim 10 - 20\% \) of their final masses. Additionally, W02 found that the scale radius and central density of the best fit NFW profile remain practically constant after the initial phase of rapid accretion. After this time, the mass increase is slow, and as the virial radius of the halo grows, its concentration decreases as \( c_{\text{vir}} \propto (1 + z)^{-1} \).

The results of Wechsler et al. lend support to Bullock et al. (B01) [48], who explained the observed trends with halo mass and redshift using a simple, semi-analytic model that I used in Chapter 5 and that I also adopt in this study. In the B01 model, halo concentrations \( c_{\text{vir}}(M, z) \), depend only on the value of \( \sigma(M) \) and the evolution of linear perturbations, \( \delta(z)/\delta(z = 0) \). Specifically, the density of a halo of mass \( M \) is set by the density of the universe at the time when systems of mass \( \sim 0.01M \) were typically collapsing. The collapse epoch \( z_c \), is defined by \( \sigma(0.01M) \equiv \delta_c(z_c) \). Again, \( \delta_c(z) \) is the linear overdensity for collapse at redshift \( z \). Central densities determined in this manner connect well to the findings of Wechsler et al. Halo density structure is set at a time of rapid accretion, when progenitor masses typically were \( M_{\text{prog}} \sim 0.1M \). Most of the mass in a halo at any given time is set by accretion events with subhalos of mass \( \sim 1/10 \) the host halo mass (cf., §6.5). Thus the period of rapid mass accretion involves objects of mass \( \sim 0.1M_{\text{prog}} \sim 0.01M \), and
it is the collapse times and densities of these constituents that set the central density of the mass $M$ halo.

As I mentioned in the preceding Chapter, the B01 model reproduces N-body results for $n = 1$, $n = 0.9$, power-law CDM, and quintessence+CDM models [66, 48, 179]. Likewise, this model also matches well the warm dark matter (WDM) simulations of Colín et al. [67] and Avila-Reese et al. [11]. Nevertheless, it is important to stress that most N-body tests were restricted to the mass range $\sim 10^9 - 10^{14} M_\odot$ because of the limited dynamic range of numerical experiments (see [46]). In what follows, I nevertheless use the B01 model to compute concentrations for halos with masses $\ll 10^9 M_\odot$. The means that these results for $M \lesssim 10^9 M_\odot$ can then be regarded as a “best-guess” extrapolation of N-body results. However, a new set of N-body simulations seems to confirm the validity of the B01 model all the way down to masses as small as $\sim 10^7 M_\odot$ [65], which is a promising, yet unconfirmed result.

Before proceeding, important to mention an alternate prescription for assigning $c_{\text{vir}}(M, z)$ proposed by Eke, Navarro, & Steinmetz (ENS hereafter) [103]. ENS investigated the power spectrum dependence of the $c_{\text{vir}} - M_{\text{vir}}$ relation for several ΛCDM and WDM models. While the B01 and ENS recipes for $c_{\text{vir}}(M, z)$ nearly coincide for ΛCDM, the B01 model failed to reproduce the mass dependence seen in simulations by ENS for WDM halos with masses smaller than the “free-streaming” mass (see §6.3). The four WDM halos simulated by ENS with masses small enough to be appreciably affected by the free-streaming scale all had $c_{\text{vir}}$ values that were $\sim 2\sigma$ lower than the B01 model results for the median relation and statistical scatter. Based on these data, ENS proposed a model in which halo collapse time depends not on the
amplitude of the power spectrum $\sigma(M)$, but on an effective overdensity amplitude given by the product $\sigma_{\text{eff}} \equiv -\sigma(M)\ln \sigma(M)/\ln M = Md\sigma(M)/dM$. This results in a $c_{\text{vir}}(M)$ relation that increases with mass for masses smaller than the truncation scale and decreases at larger masses as in $\Lambda$CDM. By defining an effective overdensity in this way, ENS were able to account for the low $c_{\text{vir}}$ values observed in their WDM simulations and still reproduce the redshift and mass dependence seen in $\Lambda$CDM simulations. The slope of the $c_{\text{vir}}-M_{\text{vir}}$ of ENS is shallower than the slope predicted by the B01 relation, therefore the ENS model also leads to less concentrated halos at small mass ($M < 10^{10} \ M_\odot$) even for identical input power spectra. This disparity grows larger when tilted and/or running spectra are considered. I compare these disparate results in Figure 6.3. Notice that on group and cluster scales ($M \gtrsim 10^{13} \ M_\odot$), where these models have been most reliably tested, the two prescriptions agree. On much smaller scales, the two prescriptions yield significantly different concentrations, exceeding 30% at $M \sim 10^8 \ M_\odot$.

Unfortunately, the ENS model cannot be applied in the WDM cases we explore because in these models $\sigma(M)$ is very flat on scales smaller than the free-streaming mass and the ENS model breaks down when $d\sigma(M)/dM$ becomes very small. In the ENS model, WDM halos smaller than $\sim 1\%$ of the free-streaming mass never collapse because $\sigma_{\text{eff}} \approx 0$. In addition to this practical problem with utilizing this model, the ENS predictions are not supported by the results of Avila-Reese et al. [11] and Colin et al. [67]. Using $\sim 25$ halos, Avila-Reese et al. found WDM halo concentrations to be roughly constant with mass down to several orders of magnitude below the free-streaming scale, in accordance with the B01 model predictions. In light of these difficulties and the discordant results of different N-body studies, we have not
Figure 6.3: This Figure illustrates the disparate results of the ENS model and the B01 model for dark matter halo concentrations. In the left panel, I show a comparison for the standard $n = 1$ initial power spectrum. The solid line shows the median value of the B01 relation. The dotted lines show the expected $1\sigma$ scatter about the median value. The dashed line shows the median value of the ENS relation. In the right panel I show a similar comparison for the BSI initial power spectrum.

explored the implications of the ENS model in this work. This is not an indictment of the ENS model. Rather, the results of ENS highlight the uncertainty in assigning halo concentrations to low-mass systems, especially with power spectra that vary rapidly with scale. Our choice of the B01 relation is a matter of pragmatism and represents a conservative choice in that halos are assigned the higher of the two predictions of $c_{\text{vir}}$ at small mass. Lower $c_{\text{vir}}$ values (in line with ENS expectations) would result in less substructure and larger deviations from standard the $\Lambda$CDM model than the predictions in the following Sections.
6.2.4 Orbital Evolution

With the accretion history of the host halo in place, and with a recipe in hand that fixes the density structure of host and satellite halos, the next step is to track the orbital evolution of the accreted systems. This is necessary in order to account for the effects of dynamical friction and mass loss due to tidal forces. These processes will cause many of the accreted subhalos either to sink to the center of the host halo and become “centrally merged,” or to lose most of their mass and be “tidally disrupted” and no longer identifiable as distinct substructure. We model these effects using an improved version of the Bullock, Kravtsov, and Weinberg [49, 50] technique, borrowing heavily from the dynamical evolution model proposed by Taylor and Babul [295, 297] (TB01 hereafter) and the dynamical friction studies of Hashimoto, Funato, & Makino [133] (HFM02 hereafter) and Valenzuela & Klypin [308, 309].

Let the mass of an accreted subhalo be \( M_{\text{sat}} \), its outer radius as \( R_{\text{sat}} \), and the accretion redshift as \( z_{\text{acc}} \). I set the subhalo concentration to the median value given by the B01 model for this mass and redshift. Although initially set by the virial mass and radius of the in-falling halo, \( M_{\text{sat}} \) and \( R_{\text{sat}} \) are allowed to evolve with time, as described in more detail below. I then track the orbit of each subhalo in the potential of its host from the time of accretion, \( t_{\text{acc}} \), until today (\( t_0 \approx 13.5 \) Gyr in the cosmology we have chosen) or until it is destroyed. The mass accretion history also gives us the host halo mass at each time step. I fix the density profile of the host at each accretion time using the median B01 expectation for a halo of that mass; however, as I mentioned earlier, the scale radius of the host remains approximately constant.
For the purpose of tracking each subhalo orbit, I assume the host potential to be both spherically symmetric and static. I update the host halo profile using the B01 expectation at each accretion event, but hold it fixed while each orbit is integrated. While the approximation of a static host potential for each orbit is not ideal, it allows for an extremely simple prescription that significantly reduces the computational expense of our study. Moreover, this approximation is not bereft of physical motivation. As I discussed above, halos observed in numerical simulations appear to form dense central regions early in their evolution after which their scale radii and central densities remain roughly fixed with time.\textsuperscript{11} Additionally, I have run test examples that include an evolving halo potential (set by the results of W02) and find that this addition has a negligible effect on the statistical properties of substructure that I am concerned with here.

Upon accretion onto the host each halo must be assigned an initial position and an initial angular momentum which serve to specify its initial orbit\textsuperscript{12}. To each incoming halo, I assign an initial orbital energy based on the range of binding energies observed in the numerical simulations of Klypin et al. [163]. I place each satellite halo on an initial orbit of energy equal to the energy of a circular orbit of radius \( R_{\text{circ}} = \eta R_{\text{vir}} \), where \( R_{\text{vir}} \) is the virial radius of the host at the time of accretion and \( \eta \) is drawn randomly from a uniform distribution on the interval \([0.4, 0.75]\). I then apportion to each satellite an initial specific angular momentum \( J = \epsilon J_{\text{circ}} \), where \( J_{\text{circ}} \) is the specific angular momentum of the aforementioned circular orbit and \( \epsilon \) is known as

\textsuperscript{11}The exception to this is the case of a late-time merger of halos of comparable mass. In such a major merger, the central densities and scale radii of the participating halos may change considerably (see Wechsler et al. [321]).

\textsuperscript{12}I use the term initial orbit because the orbits are not fixed but slowly decay due to the effect of dynamical friction as described below.
the “orbital circularity.” Past studies have drawn $\epsilon$, from a uniform distribution on the interval $[0.1, 1]$ (e.g., Bullock, Kravtsov, and Weinberg [49]). This distribution was chosen roughly to match the circularity distributions reported by Ghigna et al. [114] for their population of surviving subhalos at $z = 0$. However, it is important to account for the fact that the orbits of surviving halos are biased relative to the orbits of all accreted subsystems because subhalos on radial orbits pass closer to the center of the host halo, where dynamical friction and tidal forces are the strongest, and so they are preferentially destroyed. I find that a better match the Ghigna et al. result for surviving satellites can be achieved if I draw the initial $\epsilon$ distribution from the simple, piecewise-linear distribution depicted in Figure 6.4. I set the initial radial position of each satellite halo is set to $R_{\text{init}} = R_{\text{circ}}$ and for all non-circular orbits, we set the subhalo to be initially in-falling, that is $dR/dt < 0$. I assume the rate of mergers to be spherically symmetric.

To calculate the trajectories of subhalos, I treat them as point masses under the influence of the NFW gravitational potential of the host halo. I model orbital decay by dynamical friction using the formula of Chandrasekhar [60]. The Chandrasekhar formula was derived in the context of a highly idealized situation. In particular, Chandrasekhar assumed a massive body to be moving through an infinite, homogeneous sea of background particles (these are the dark matter particles in this case) described by a Maxwellian phase space distribution. However, numerical studies indicate that this approximate relation can be applied more generally (e.g., Valenzuela & Klypin [308] have performed a new test that validates the use of this approximation). Using the Chandrasekhar approximation, there is a frictional force exerted by the live host
Figure 6.4: Input orbital circularity distribution of initially in-falling substructure (dashed) shown along with the circularity distribution of the final surviving population of \((n = 1)\) LCDM subhalos at \(z = 0\) (solid). For reference, the thin dotted line shows the circularity distribution of surviving substructure measured by Ghigna et al. [114] in their N-body simulations.

A halo that points opposite to the subhalo velocity with magnitude

\[
F_{\text{DF}} \sim \frac{4\pi \ln(\Lambda)G^2M^2_{\text{sat}}\rho(r)}{V_{\text{orb}}^2}\left[\text{erf}(X) - \frac{2X}{\sqrt{\pi}}\exp(-X^2)\right].
\]  

(6.17)

In equation (6.17), \(\ln(\Lambda)\) is the Coulomb logarithm, \(r\) is the radial position of the orbiting satellite, and \(\rho(r)\) is the density of the host halo at the satellite radius. The quantity \(V_{\text{orb}}\) is the orbital speed of the satellite halo and \(X\) is defined by \(X \equiv V_{\text{orb}}/\sqrt{2\sigma^2}\), where \(\sigma\) is the one-dimensional velocity dispersion of particles in the host halo. For an NFW profile, the one-dimensional velocity dispersion can be determined
using the Jeans equation. Assuming an isotropic velocity dispersion tensor,

\[
\sigma^2(x = r/r_s) = \frac{GM_{\text{vir}}}{R_{\text{vir}}} \frac{c_{\text{vir}}}{g(c_{\text{vir}})} x (1 + x)^2 \int_x^\infty \frac{g(x')}{x'^3 (1 + x')^2} dx'.
\]

The following approximation is useful and accurate to 1% for \( x = 0.01 - 100 \):

\[
\sigma(x) \simeq V_{\text{max}} \frac{1.4393 x^{0.354}}{1 + 1.1756 x^{0.725}}.
\]

There has been much debate on the appropriate way in which to assign the Coulomb logarithm in Eq. (6.17). Dynamical friction is caused by the scattering of background particles into an overdense “wake” that trails the orbiting body and tugs back on the scatterer. The Coulomb logarithm is interpreted as \( \ln(b_{\text{max}}/b_{\text{min}}) \), where \( b_{\text{max}} \) is the maximum relevant impact parameter at which background particles are scattered into the wake and \( b_{\text{min}} \) is the minimum relevant impact parameter. Many authors parameterize the deviation of realistic physical situations from the idealization of Chandrasekhar by absorbing the uncertainty into the choice of the value of the Coulomb logarithm as the definition of the minimum and maximum relevant impact parameters is somewhat ambiguous. A rather common approach is to choose a constant value of the Coulomb logarithm (perhaps by calibrating the analytic expression to the results of numerical experiments as in TB01), but several studies indicate that this approach significantly underestimates the dynamical friction timescale when tested against N-body simulations (e.g., Refs. [70, 133]). Motivated by the results of HFM02 and the study of Valenzuela & Klypin [308, 309], I allow the Coulomb logarithm to evolve as the radius evolves and set \( b_{\text{max}} = r(t) \), where \( r \) is the radial position of the orbiting subhalo. In accordance with HFM02, I assign the minimum impact parameter according to the prescription of White [327] and integrate the effect of encounters with background particles over the density profile of the subhalo.
Repeating this calculation for an NFW halo furnishes an expression for the Coulomb logarithm that can be written

\[
\ln(A) = \ln\left(\frac{r}{R_{\text{sat}}}\right) + \frac{1}{g^2(x_{\text{sat}})} I(x_{\text{sat}})
\]  
(6.20)

where

\[
I(x_{\text{sat}}) \equiv \int_0^{x_{\text{sat}}} x_b^3 \left[ \int_{x_b}^{\infty} \frac{g(x)}{x^2 \sqrt{x^2 - x_b^2}} \, dx \right]^2 \, dx_b,
\]  
(6.21)

\[x_{\text{sat}} \equiv R_{\text{sat}}/r_s^{\text{sat}},\] and \(r_s^{\text{sat}}\) is the NFW scale radius of the satellite. The integral \(I(y)\) is well-approximated by the following function, which is accurate to 1% for \(0.1 \leq y \leq 100\):

\[
I(y) \simeq \frac{0.10947y^{3.989}}{[1 + 0.90055y^{1.099} + 0.03568y^{1.189} + 0.06403y^{1.989}]}.
\]  
(6.22)

As the satellite orbits within the potential of the host, it is stripped of mass as a result of the tidal forces it experiences. The next step in tracking the dynamical evolution of the subhalo is to estimate the rate at which mass is removed from the outer regions of the halo by the tides. I estimate the instantaneous tidal radius of the subhalo \(r_1\), at each point along its orbit. In the limit that the satellite is much smaller than the host, the tidal radius is given by the solution to the equation (see [318, 160])

\[
r_1^3 \simeq \frac{M_{\text{sat}}(< r_1)/M_{\text{host}}(< r)}{2 + \omega^2 R^3/G M_{\text{host}}(< r) - \partial \ln M_{\text{host}}(< r)/\partial \ln r^3},
\]  
(6.23)

where \(r\) is the radial position of the satellite, \(M_{\text{host}}(< r)\) is the host’s mass contained within this radius [cf., Eq. (4.4)], \(M_{\text{sat}}(< r_1)\) is the satellite’s mass contained within \(r_1\), and \(\omega\) is the instantaneous angular speed of the satellite. Equation (6.23) is merely
an estimate of the satellite’s tidal limit. For a satellite on a circular orbit, the value of \( r_1 \) derived from Equation (6.23) represents the distance from the satellite center to the point along the line connecting the satellite and the host halo center where the tidal force on a test particle just balances the attractive force of the satellite. In reality, the situation is clearly much more complicated. The tidal limit of a satellite cannot be represented by a spherical surface: some particles within \( r_1 \) will be unbound while others without \( r_1 \) may be bound. Nevertheless, TB01 showed that this can serve as a very useful approximation [295].

As the tidal radius shrinks, unbound mass in the periphery is stripped. Clearly, the tidal force is the strongest, and \( r_1 \) the smallest, when the orbit reaches pericenter; however, all of the mass outside of \( r_1 \) is not stripped instantaneously at each pericenter passage. Rather, mass is gradually lost from the satellite on a timescale set by the orbital energy of the liberated particles. Johnston (1998), found that the typical energy scale of tidally stripped debris is set by the change in the host halo potential on the length scale of the orbiting satellite,

\[
\epsilon \approx r_1 \frac{d\Phi_{\text{host}}(r)}{dr}.
\]  

(6.24)

Particles on circular orbits of energy \( E \) and \( E \pm \epsilon \) move a distance \( r_1 \), relative to each other on a timescale of order the orbital period, \( T \). As such, one may expect \( T \) to be the relevant timescale for tidal mass stripping. Possibly motivated by the same reasoning, Taylor and Babul [295] used this timescale in their model to reproduce the results of several idealized N-body experiments. Encouraged by this success, I follow Taylor and Babul and model satellite mass loss by dividing the orbit into discrete
time steps of size $\delta t \ll T$. At each step, I remove an amount of mass

$$\delta m = M_{\text{sat}}(> r_1) \frac{\delta t}{T},$$

(6.25)

where $T = 2\pi/\omega$ and $M_{\text{sat}}(> r_1)$ is the satellite’s mass exterior to its tidal radius, $r_1$.

I make a further assumption regarding the distribution of mass within each subhalo. As a subhalo loses mass due to tidal stripping, I assume that its inner density profile is unmodified within its outer radius $R_{\text{sat}}$. Rather than identify $R_{\text{sat}}$ with the tidal radius, which does not change monotonically with time, I set its value by determining the radius within which the mass profile retains the appropriate bound mass [cf., Eq. (4.4)]:

$$g(x_{\text{sat}}) \equiv \frac{M_{\text{sat}}(t)}{M_{\text{sat}}(t_{\text{acc}})} g(c_{\text{sat}}).$$

(6.26)

The scale radius of the subhalo $r_{\text{sat}}^{\text{sat}}$, stays fixed at the value defined at the epoch of accretion.

The aim of adopting these approximations is to include the relevant physical scalings that determine the properties of halo substructure. These approximations for dynamical friction and tidal stripping are the least accurate when the mass of the satellites are not very small compared to the mass of the host. However, as $F_{\text{DF}} \propto M_{\text{sat}}^2$, it is in precisely these cases that one would expect the satellite to quickly merge with the host and no longer be identifiable as distinct substructure. Dynamical friction causes large halos to sink to the center of the host on a timescale that varies roughly as $\tau_{\text{DF}} \propto (M_{\text{host}}/M_{\text{sat}})$. As such, the precise dynamics should not have a significant effect upon the results in these cases because almost all such halos will be erased by dynamical friction. These worries are also less important for this work, because the main predictions of this section involve low-mass substructure.
However, more detailed modeling will be important for investigations that focus on more massive substructures, for example, explorations that use disk thickening as a test of the \( \Lambda \)CDM cosmological model or to constrain the power spectrum of primordial fluctuations.

The final ingredients for this semi-analytic model of halo substructure are the criteria for declaring subhalos to be tidally disrupted and centrally merged. Let \( r_{\text{max}}^\text{sat} \) be the radius at which the subhalo's initial velocity profile attains its maximum, and \( M_{\text{sat}}(< r_{\text{max}}^\text{sat}) \) be the mass of the satellite originally contained within the radius \( r_{\text{max}}^\text{sat} \). In this model, I declare a subhalo to be centrally merged with the host if its radial position relative to the center of the host becomes smaller than \( r_{\text{max}}^\text{sat} \). I declare a satellite to be tidally disrupted if the mass of the satellite becomes less than \( M_{\text{sat}}(< r_{\text{max}}^\text{sat}) \). This criterion is partially motivated by the numerical results of Hayashi et al. [135] (H03 hereafter) and Klypin et al. [163, 177], who find that NFW subhalos are completely tidally destroyed shortly after \( r_t \) becomes less than \( r_{\text{max}}^\text{sat} \).

Of course the distinction between centrally merged and tidally destroyed satellites is somewhat arbitrary as subhalos are typically severely tidally disrupted as they approach the center of the host potential. It is fortunate that for the issues that I explore here, the precise nature of a satellite halo’s destruction is not important. This issue will be discussed further in a forthcoming extension of this work [336].

In reporting results concerning the velocity function of substructure, I invoke one further modification. Hayashi et al. [135] noted that subhalos that experienced significant tidal stripping suffered not only mass loss at radii \( \gtrsim r_t \), but mass redistribution in their central regions, at radii smaller than \( r_t \). To account for this mass redistribution in severely stripped objects, I determine whether or not the tidal radius of
each surviving subhalo was ever less than \( r_{\text{sat}}^{\text{max}} \). If so, I then follow the prescription of Hayashi et al. [135] to account for mass redistribution and scale the maximum circular velocity of the satellite via

\[
V_{\text{max}}^{\text{final}} = \left( \frac{M_{\text{sat}}^{\text{final}}}{M_{\text{sat}}^{\text{initial}}} \right)^{1/3} V_{\text{max}}^{\text{initial}},
\]

(6.27)

where \( V_{\text{max}}^{\text{initial}} \) is the maximum circular velocity of the satellite according to its initial density profile, \( M_{\text{sat}}^{\text{final}} \) is its final mass, and \( M_{\text{sat}}^{\text{initial}} \) is its initial mass before being tidally stripped. In practice, this rescaling has a fairly small effect on the resultant velocity functions. Roughly \( \sim 30\% \) of surviving halos meet this condition for \( r_{\text{t}} \). For those halos that do experience this kind of mass loss, the typical reduction in \( V_{\text{max}} \) is \( \lesssim 25\% \).

Currently, a significant effort is being put into checking this model against idealized N-body experiments designed to mimic the type of orbital histories that I model here [47]. Preliminary results show promising agreement.

### 6.2.5 Preliminary Tests and Examples

Figure 6.5 shows three example calculations of subhalo trajectories aimed at demonstrating how various factors affect the orbital evolution of a satellite system. Each satellite system was started with the same initial orbit, \( \epsilon = \eta = 0.5 \), but the satellite properties were varied: \( M_{\text{sat}}^{0} = 10^{8} \, M_{\odot} \), \( c_{\text{vir}} = 15 \) (solid); \( M_{\text{sat}}^{0} = 10^{8} \, M_{\odot} \), \( c_{\text{vir}} = 7.5 \) (dashed); and \( M_{\text{sat}}^{0} = 5 \times 10^{9} \, M_{\odot} \), \( c_{\text{vir}} = 15 \) (short-dashed). The superscript “0” indicates that the mass given is the initial mass which must be distinguished from the bound mass in general because the total mass that is bound to the system varies with time. The upper and lower panels depict the evolution of orbital radius.
Figure 6.5: Orbital evolution for three sets of subhalo input parameters: $M_{\text{sat}}^0 = 10^8 \, M_\odot$, $c_{\text{vir}} = 15$ (solid); $M_{\text{sat}}^0 = 10^8 \, M_\odot$, $c_{\text{vir}} = 7.5$ (dashed); and $M_{\text{sat}}^0 = 5 \times 10^9 \, M_\odot$, $c_{\text{vir}} = 15$ (short-dashed). The superscript “0” indicates that this is the initial mass of the host. The total bound mass varies with time as the satellite is tidally stripped. Initial orbital parameters and host mass properties are fixed, as described in the text. The top panel shows the radial evolution in units of the initial radius as a function of time. The bottom panel shows the mass of each system as a function of time. The less concentrated subhalo is tidally destroyed after only a few pericenter passages, and the more massive system is similarly destroyed because dynamical friction causes the orbit to decay so that is remains closer to the center of the host and experiences more pericenter passes.

and mass of the subhalo respectively. I set the accretion time to 8 Gyr in the past for these examples. This corresponds to a scale factor $a = (1 + z)^{-1} \approx 0.45$ for this cosmology. The host halo parameters were chosen to match reasonable expectations for a Milky Way-sized progenitor at that time: $M_{\text{host}} = 5 \times 10^{11} \, M_\odot$ ($R_{\text{vir}} \approx 110\, \text{kpc}$) and $c_{\text{vir}} = 6$. While the subhalo represented by the solid line experiences gradual tidal mass loss and slight orbital decay as a result of dynamical friction, its core survives
Figure 6.6: The velocity functions of progenitor and surviving subhalo populations derived using the fiducial ΛCDM ($n = 1, \sigma_8 = 0.95$) cosmology and a 200-halo ensemble of $1.4 \times 10^{12}$ M$_\odot$ systems at $z = 0$. Shown are all accreted halos (dashed), and the fraction of those that are tidally destroyed (short-dashed) and centrally merged (dotted). The solid line shows the surviving population of subhalos at $z = 0$ and, for comparison, the thin dashed line shows the surviving population derived by Klypin et al. (1999) using N-body simulations. The error bars represent the sample variance.

for the full time period. The less concentrated subhalo (dashed) is more strongly affected by tides, and is completely disrupted $\sim 3.5$ Gyr after being incorporated into the host. (Although not shown, a similar effect is seen if the host halo concentration is increased and the subhalo concentration is held fixed.) In the case of the massive subhalo (dashed line), dynamical friction causes the orbit to decay more quickly. As a result, the massive subhalo does not range as far from the center of the host in this case and the massive subhalo experiences more frequent pericenter passages, where tidal forces are the strongest. As a consequence, disruption occurs $\sim 6$ Gyr
after accretion. Notice that the stripping process is gradual (unless orbits are very radial) and the orbital timescales involved are of order $\sim$ Gyr so the accretion time is also important in determining survival probability. If any of these orbits were made more recent, their chances of survival to the present day would increase accordingly. The combination of factors illustrated here — accretion times, satellite mass, and the relative concentrations of host and satellite — will be important in later sections for understanding what sets the characteristics of the subhalo population from one cosmology to the next.

Figure 6.6 shows ensemble-averaged, cumulative velocity functions for the progenitors of 200 Milky Way-like host halos computed for our standard $\Lambda$CDM cosmology. The host properties at $z = 0$ are $M_{\text{vir}} = 1.4 \times 10^{12} \, M_\odot$, $c_{\text{vir}} \simeq 13.9$, and $V_{\text{max}} \simeq 187 \, \text{km s}^{-1}$. The lines represent the means of 200 merger tree realizations, and the error bars represent the sample variances over these realizations. The long-dashed line represents all subhalos accreted over the merger history of the halo. The dotted line shows all accreted subhalos that centrally-merged with the host, the short-dashed line shows the subhalos that were tidally disrupted, and the solid line shows the surviving subhalo population at $z = 0$. For comparison, the thin dashed line is the best-fit velocity function reported by Klypin et al. [163] based on an analysis of substructure in $\Lambda$CDM halos. The line is plotted over the range that their resolution and sample size allowed them to probe. The apparent agreement between the semi-analytic model presented in this section and the N-body result is excellent, and lends confidence in our ability to apply this model to different power spectra.

The radial density distribution of substructure at $z = 0$ for the same ensemble of halos is shown in Figure 6.7. Open circles show the differential number density profile
Figure 6.7: The radial number density profile of substructure derived from 200 model realizations of a $M_{\text{vir}} = 1.4 \times 10^{12} \, M_\odot$ host halo at $z = 0$ in our fiducial ($n = 1$, $\sigma_8 = 0.95$) $\Lambda$CDM cosmology. The open circles show the number density of subhalos with $M_{\text{sat}} > 10^6 \, M_\odot$ divided by the average number density of systems meeting this mass threshold within the virial radius of the host system. The points reflect the radial profile averaged over all realizations, and the error bars reflect the sample variance. Solid pentagons show the same result for $M_{\text{sat}} > 6 \times 10^8 \, M_\odot$ subhalos. The variance (not shown) is significantly larger for the higher mass threshold because there are significantly fewer such systems in each host. For reference, the solid line shows NFW density profile of the host at $z = 0$. The virial radius for a host halo of this size is $R_{\text{vir}} \approx 287 \, \text{kpc}$ and the typical NFW scale radius is $r_s \approx 20 \, \text{kpc}$.

of subhalos with $M_{\text{sat}} > 10^6 \, M_\odot$ normalized relative to the total, volume-averaged number density of subhalos within $R_{\text{vir}}$ that meet the same mass requirement. The solid pentagons show the same quantity for more massive subhalos, $M_{\text{sat}} > 6 \times 10^8 \, M_\odot$. The line shows the NFW dark matter profile for the host system normalized relative to the average (virial) density within the halo. Observe that the subhalo profile traces the dark matter profile at large radius, but flattens towards the center.
as a consequence of tidal disruption. This result agrees remarkably well with that presented in Figure 3 of the study by Colín et al. [66]. Using an N-body analysis of a cluster-sized host, Colín et al. showed that the number density of systems with \( M_{\text{sat}} \) greater than 0.04\% of the host mass traces the background halo profile at large radius, begins to flatten at \( r \sim 0.2 R_{\text{vir}} \), and is roughly a factor of 5 below the background at \( r \sim 0.07 R_{\text{vir}} \) (their innermost point).\(^{13}\) The solid pentagons in Figure 6.7 correspond to the same mass fraction relative to the host. Notice that at \( r = 0.07 R_{\text{vir}} \simeq 20 \) kpc, the factor of \( \sim 5 \) mismatch is reproduced. Ghigna et al. [114] observed the same qualitative behavior for subhalos in a standard CDM simulation of a cluster-size halo. Chen et al. [61] have measured the substructure profile using a high-resolution Galaxy-sized halo with \( M_{\text{sat}} \gtrsim 0.0015\% M_{\text{host}} \), corresponding to subhalos intermediate in mass between those represented by the open circles and solid pentagons in Figure 6.7. Chen et al. [61] similarly find core behavior setting in at a radius of \( \sim 30 \) kpc, but also find a stronger overall suppression in substructure counts within \( r \lesssim 70 \) kpc. These preliminary results may suggest that some of the observed suppression may be caused by overmerging in their simulated halo’s central region. Of course, only the next generation of numerical simulations can reliably test this.

That this produces a reasonable approximation to the expected central flattening of the number density profile of substructure is an indication that the disruption model is sound, but the ability of the model to reproduce the outer profile is perhaps more striking. I stress that the background dark halo profile was chosen by hand to match theoretical expectation, but the final substructure profile is subject to all of the inputs of the model, including the initial radius and accretion time of each halo, as

\(^{13}\)Results are quoted relative to \( R_{\text{vir}} \) and \( M_{\text{host}} \) because the host halo in Colín et al. (1999) [66] is significantly more massive than the halos considered here.
well as the orbital evolution. Thus, the fact that I obtain an outer profile slope close
to \( \rho \propto r^{-3} \) is an encouraging achievement. This finding lends support to the idea
that the outer profiles of dark matter halos are set by late-time, quiescent accretion.
Furthermore, because the flattening at small radius is caused by disruption, it is
tempting to speculate that these destructive processes give rise to the characteristic
bend, and inner-slopes of dark matter halos. Although an exploration of these ideas
is well beyond the scope of this paper, the framework I have put forward could serve
as a useful tool in such an investigation. Maller & Dekel [198] have already used a
similar, merger-tree-based approach to explain the angular momentum structure of
halos, and some interesting ideas towards explaining the density structure of halos
within a similar framework were discussed by Dekel et al. [91], but not within a
cosmological context.

6.3 Model Power Spectra

In the previous Chapter, I discussed halo central densities as a function of the
power spectrum. I used specific models of inflation as a conceptual device to demon-
strate the unity of the cosmological picture. In this Chapter I will consider many
of the same power spectra along with some additional examples and so I provide a
brief summary of the power spectra that I explore for convenience. The approach
of this Chapter is fundamentally different though. I adopt an empirical approach
to the study of substructure as a function of the primordial power spectrum and
the spectra that I consider may be viewed simply as a representative set that span
the observationally viable range of initial power spectra. Recall that initial power
spectrum of density fluctuations is conventionally written as an approximate power

190
law in wavenumber $k$, $P(k) \propto k^n$, corresponding to a variance per logarithmic interval in wavenumber of $\Delta(k) = \sqrt{k^3 P(k)/2\pi^2}$. If the fluctuations were seeded during an early inflationary stage, as is commonly supposed, then the initial spectrum is likely to be nearly scale-invariant, with $n \approx 1$. Any deviation from power law behavior, or “running” of the power law index with scale is usually expected to be small, $|dn/d\ln k| < 0.01$. In addition to these theoretical prejudices, large-scale observations of galaxy clustering and CMB anisotropy seem to favor nearly scale-invariant models that can be parameterized in this way. In this Chapter, I explore the effects on halo substructure of taking $n \neq 1$ and allowing for scale-dependence in the power law index and more dramatic features in the power spectrum.

Table 6.1 summarizes the relevant features of the example power spectra. The second and third columns list the primordial spectral index evaluated at the pivot scale of the COBE measurements $k_{\text{COBE}} \approx 0.0023 \text{ h Mpc}^{-1}$, and the running of the spectral index. I have neglected any variation in the running with scale. Explicitly, we assume $d^2 n(k)/d(\ln k)^2 = 0$. The fourth column gives the implied, $z = 0$, linear, rms fluctuation amplitude on a length scale of $8 \text{ h}^{-1} \text{ Mpc}$, $\sigma_8$. Except for the running index (RI) case, I have normalized all models to the COBE measurements of the CMB anisotropy using the fitting formulae of Bunn, Liddle, & White (1996; also Bunn & White 1997). I calculate spectra using the transfer functions of Eisenstein & Hu (1999). In Figure 6.8 I illustrate the implied $\sigma(M)$ for these models.

The models span the extrema of observationally acceptable values of tilt and $\sigma_8$, spanning from $n \approx 0.84$ with $\sigma_8 = 0.65$ to $n = 1$ and $\sigma_8 = 0.95$. The model with $\sigma_8 = 0.75$ was specifically chosen to match galaxy central densities, as described in Chapter 5. I also explore the best-fit, running-index model of the WMAP team
<table>
<thead>
<tr>
<th>Model Description</th>
<th>$n(k_{\text{COBE}})$</th>
<th>$dn(k)/d\ln k$</th>
<th>$\sigma_8$</th>
<th>comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale-invariant</td>
<td>1.00</td>
<td>0.000</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>Inverted power law</td>
<td>0.94</td>
<td>-0.002</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>Running-mass I</td>
<td>0.84</td>
<td>-0.008</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>Running-mass II</td>
<td>0.90</td>
<td>-0.002</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Running index (RI) model</td>
<td>1.03</td>
<td>-0.03</td>
<td>0.84</td>
<td>WMAP best fit [274]</td>
</tr>
<tr>
<td>Broken scale-invariant (BSI)</td>
<td>1.00</td>
<td>0.000</td>
<td>0.97</td>
<td>cutoff at $k \gtrsim 1 \text{ h Mpc}^{-1}$,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>or $M \lesssim 10^{10} \text{ M}_{\odot}$</td>
</tr>
</tbody>
</table>

WDM, $m_w = 3.0$ keV     | 1.00                | 0.000          | 0.95      | $M_f \simeq 8.3 \times 10^8 \text{ M}_{\odot}$|
WDM, $m_w = 1.5$ keV     | 1.00                | 0.000          | 0.95      | $M_f \simeq 1.3 \times 10^{10} \text{ M}_{\odot}$|
WDM, $m_w = 0.75$ keV    | 1.00                | 0.000          | 0.94      | $M_f \simeq 2.1 \times 10^{11} \text{ M}_{\odot}$|

Table 6.1: Column (1) gives a brief description of the inflation or warm dark matter model used to predict the power spectrum. In the text, I distinguish the first five models by their tilts and/or their values of $\sigma_8$. I label the warm dark matter models by the warm particle mass. Columns (2) and (3) give the tilt $n(k_{\text{COBE}})$ on the pivot scale of the COBE data $k_{\text{COBE}} \approx 0.0023 \text{ h Mpc}^{-1}$, and the running of the spectral index $dn(k)/d\ln k$, respectively. I have explicitly assumed the “running-of-running” to be small and taken $d^2n(k)/d(lnk)^2 = 0$. Column (4) contains the values of $\sigma_8$ implied by the tilt or warm particle mass, the COBE normalization, and the fiducial cosmological parameters except in the case of the WMAP best-fit running index (RI) model, in which case the value of $\sigma_8$ reflects their best-fit normalization [274].
Figure 6.8: The $z = 0$ rms overdensity as a function of mass scale for several of the power spectra that I consider in this Chapter (see Section 6.3 and Table 6.1 for more details). I exhibit spectra that deviate from the standard $n = 1$ scale-invariant model. The models shown in this panel are standard $n = 1$ (solid), a broken scale-invariant model (dotted), $\sigma_8 = 0.84$ & $n \simeq 0.94$ (short-dashed), $\sigma_8 = 0.65$ & $n \simeq 0.84$ (long-dashed), $\sigma_8 = 0.75$ & $n \simeq 0.90$ (dot-long-dashed), and a model based on the results of the WMAP team with $n \simeq 1.03$, $dn/d\ln k = -0.03$, and $\sigma_8 = 0.84$ (dot-short-dashed).

(Spergel et al. 2003), with $dn/d\ln k = -0.03$ (see Section 2.5). I refer to this as the “running index model” or “RI model.” Note that Spergel et al. (2003) quote a value of $n = 0.93$, evaluated at $k = 0.05$ Mpc$^{-1}$. The value listed in Table 1 is larger because I quote it at a smaller wavenumber, $k = k_{\text{COBE}}$.

I consider power spectra computed in the context of a WDM scenario in which the primordial power spectrum is scale-invariant but small-scale fluctuations are filtered by free-streaming. The free-streaming scale is set by the primordial velocity dispersion of the warm particles. In the canonical case of a “neutrino-like,” thermal relic with
Figure 6.9: The $z = 0$ rms overdensity as a function of mass scale for the warm dark matter (WDM) power spectra that I consider in this Chapter (see Section 6.3 and Table 6.1 for more details). In this Figure, I show several warm dark matter power spectra. I depict spectra implied by warm particle masses of $m_W = 3.0 \text{ keV}$ (short-dashed), $m_W = 1.5 \text{ keV}$ (long-dashed), and $m_W = 0.75 \text{ keV}$ (dotted) along side the standard $n = 1$, $\Lambda$CDM spectrum (solid).

Two internal degrees of freedom, the free-streaming scale can be expressed in terms of the warm particle mass $m_W$ and WDM relic abundance, $\Omega_W h^2$:

$$R_f \simeq 0.11 \left[ \frac{\Omega_{WDM} h^2}{0.15} \right]^{1/3} \left[ \frac{m_W}{\text{keV}} \right]^{-4/3} \text{ Mpc.}$$  \hspace{1cm} (6.28)

I calculate WDM spectra assuming the same flat cosmology with $\Omega_M = \Omega_{WDM} + \Omega_B = 0.3$, and use the approximate WDM transfer function given by Bardeen et al. (1986),

$$P(k) = \exp[-kR_f - (kR_f)^2]P_{CDM}(k).$$  \hspace{1cm} (6.29)
I discuss clustering and halo properties in WDM cosmologies in more detail in Section 6.7.

Several studies have placed approximate constraints on WDM masses based on either the argument that there must be enough power on small scales to re-ionize the Universe at sufficiently high redshift \( z_{\text{re}} \gtrsim 6 \) or by probing the power spectrum on small scales directly with the Lyman-\( \alpha \) forest [21, 216]. These authors essentially find that \( m_W \gtrsim 0.75 \text{ keV} \) assuming a neutrino-like thermal relic; however, it is important to note that the WDM constraint may be significantly more restrictive if measurements of \( z_{\text{re}} \sim 17 \) by the WMAP collaboration [169, 274] are confirmed [270]. As such, I consider three illustrative examples in what follows, \( m_W = 0.75 \text{ keV}, 1.5 \text{ keV}, \) and \( 3.0 \text{ keV} \). The corresponding “free-streaming” masses, below which the fluctuation amplitudes are suppressed, are listed in Table 6.1. The power spectra are shown in Figure 6.9.

### 6.4 Results: Accretion Histories

The first results that I present concern the merger histories of halos that are approximately Milky Way-sized, with \( M_{\text{vir}} = 1.4 \times 10^{12} \, M_\odot \) at \( z = 0 \). For the \( n = 1 \), \( \Lambda \)CDM model, I present results based on 200 realizations. For all other models, these findings are based on 50 model realizations.

Figure 6.10 focuses on the distribution of accreted halos with mass, integrated over the entire merger history of the host. I plot \( df/d\log(M_{\text{sat}}) \), the fraction of mass in the final halo that was accreted in subhalos of a given mass per logarithmic interval in subhalo mass. Observe that the mass fraction accreted in subhalos of a given mass is relatively insensitive to the shape of the power spectrum. Although similarity from
Figure 6.10: Fraction of final host mass accreted in subhalos of mass $M_{\text{sat}}$ as a function of $M_{\text{sat}}$. The final host mass is $1.4 \times 10^{12} \, M_\odot$. The results for several input power spectra are shown. The mass fraction accreted in substructure of a given mass is relatively insensitive to the initial power spectrum.

model to model may be somewhat surprising at first, it follows directly from repeated application of Equation (6.16). In particular, the shape of the progenitor distribution for $M_{\text{sat}} \ll M_{\text{host}}$ must follow $df/d\log(M_{\text{sat}}) \propto M_{\text{sat}}^{1/2}$, and the turnover occurs because mass conservation suppresses the number of major mergers. The specific shape shown in Figure 6.10 and its insensitivity to the power spectrum is discussed in detail by Lacey and Cole [181].

While the total mass function of accreted substructure is relatively independent of the primordial spectrum, the merger histories themselves are not. In models with less power on galaxy scales, halos assemble their mass later and experience more
recent mergers and disruption events. An example of this shift in mass accretion
time is depicted in Figure 6.11. Here I plot the average accretion rate of subhalos
with $M_{\text{sat}} > 10^8 M_\odot$ for host halos in the standard $n = 1$, ΛCDM model (top
panel), the RI model (middle panel), and the lowest normalization case ($n = 0.84,$
$\sigma_8 = 0.65$; bottom panel). The total accretion rate is divided in two pieces: dashed
lines show those subhalos that are eventually destroyed, either by central merging
or tidal disruption, and solid lines show the accretion times of subhalos that survive
until the present day. For the standard ($n = 1, \sigma_8 = 0.95$) case, the event rate peaks
sharply about $\sim 12$ Gyr in the past, while the low-normalization case has a broader
distribution, peaking later at $\sim 9$ Gyr ago, and with a long tail of accretion events
extending towards the present day.

The shift in accretion times in models with less power plays a role in regulating
the number of surviving subhalos. As discussed in relation to Figure 6.5, a finite
amount of time is required for an orbit to decay or for a system to become tidally
unbound and in many cases the longer a subhalo orbits in the background potential,
the more probable its disruption becomes. The later accretion times in models with
less power partially compensate for the fact that subhalos in these models are less
centrally concentrated and more susceptible to disruption at each pericenter passage.
Particular results for substructure populations in each model are given in the following
subsections.

That a characteristic merger/disruption phase in each halo’s past is expected is
an intriguing result. It turns out that this phase is approximately coincident with the

$^{14}$LC93 showed that $z_{1/2}$, the typical redshift, at which a halo attains half of the mass that it has
at $z_0$ is roughly set by $\delta_c(z_{1/2}) \approx \delta_c(z_0) + \mathcal{O} \sqrt{\sigma^2(M_0/2) - \sigma^2(M_0)}$, where $\mathcal{O}$ is of order unity. A
lower normalization for $\sigma(M)$ implies a smaller value of $\delta_c(z_{1/2})$ and, therefore, a lower value for
$z_{1/2}$ for a fixed cosmology.
Figure 6.11: Accretion rate averages, $dN/dt$ (Gyr$^{-1}$), for merged halos more massive than $10^8$ M$_\odot$ and host halos of mass $1.4 \times 10^{12}$ M$_\odot$ at $z = 0$. Three different power spectra are shown: $n = 1$ (upper); RI model (middle); and $n = 0.84$, $\sigma_8 = 0.65$ (bottom). Dashed lines show objects that are destined to be destroyed, either by tidal disruption or central merging, and solid lines show subhalos that survive until $z = 0$. The $n = 1$ results are based on 200 EPS realizations while the others are based on 50 realizations.

estimated ages of galactic thick disks, $t_{td} \sim 8 - 10$ Gyr (e.g., see Quillen and Garnett [245] for the Milky Way), which seem to be ubiquitous and roughly coeval [83]. In this context, the age distributions of thick disks might serve as a test of this characteristic accretion time, which varies as a function of normalization and cosmology. I stress that the look-back times shown for the dashed lines in Figure 6.11 are the times that subhalos were accreted, not the times at which the tidal disruptions or central mergers occurred. If instead I were to plot the central merger rates or destruction rates, the distributions would peak at slightly more recent times but their widths would be
broader, with longer tails towards the present epoch. A more detailed investigation into these issues will be presented elsewhere.

It is interesting to note that the surviving halos in Figure 6.11 represent a distinctly different population of objects than the destroyed systems — they tend to have been accreted more recently. This points toward several interesting avenues of research. One may be inclined to speculate that the star formation histories of galaxies that were destroyed after being accreted could be distinctly different from those of the surviving (satellite dwarf) galaxies as well. This may have interesting implications for understanding whether the stellar halo of our Galaxy formed from disrupted dwarfs or some other process. While the global structure of the stellar halo seems consistent with the disruption theory [50], the element ratios of stellar halo stars and stars in dwarf galaxies are not consistent with a common history of chemical evolution [265]. The results shown in Figure 6.11 provide general motivation to model dwarf galaxy evolution and Milky Way formation in a cosmological context. Work in this direction is underway.

6.5 Results: Mass and Velocity Functions

I now present results on surviving halo substructure beginning with the abundance of satellites in Milky Way-like galaxies. I plot the mass function of subhalos \( N(> M_{\text{sat}}) \), or the number of subhalos with mass greater than \( M_{\text{sat}} \) as a function of \( M_{\text{sat}} \), for each model in Figure 6.12. The host halo mass is again fixed at \( 1.4 \times 10^{12} \) \( M_\odot \) at \( z = 0 \). From this figure, one can see that even in the significantly tilted, low-normalization model (\( \sigma_s \approx 0.65 \)), the number of satellite halos with mass greater than \( 10^6 \) \( M_\odot \) is roughly equal to that in the standard \( n = 1 \) model. The systematic
Figure 6.12: The cumulative mass function $N(> M)$, of surviving subhalos computed for an ensemble of host halos of mass $M_{\text{host}} = 1.4 \times 10^{12} M_\odot$ at $z = 0$. The lines show the means computed over all realizations. The different line types relate to different models following the convention in the left panel of Figure 6. The $n = 1$ results are derived from 200 realizations and the results for the other models are based on 50 realizations. The error bars show the variance over the $n = 1$ realizations (upper) and BSI realizations (lower).

differences between the models are small compared to the scatter. The suppression is not stronger because several competing effects countermand the effect of the reduced subhalo concentrations. As discussed in the previous section, subhalos are accreted later in models with less galactic-scale power. In addition, the host halos are less concentrated and less capable of disrupting their satellites.

The BSI model, on the other hand, shows a substantial decrease (a factor of $\sim 3$) in the number of surviving satellite halos at fixed mass. The reason for the dramatic reduction in this case is easy to understand. First, power is reduced only
on scales smaller than a critical scale around $\sim 10^{10} \, M_\odot$ (cf., Figure 6.8) and so, the concentration and accretion history of the $\sim 10^{12} \, M_\odot$ host halo are minimally altered while the concentrations of the small subhalos are drastically reduced (see Chapter 5). In other words, the host halo has a similar density structure as in the $n = 1$ model and is just as capable of tidally disrupting satellites, but the satellites are significantly more susceptible to disruption. A second difference is that Galaxy-size halos in the BSI model, in contrast to the tilted case, accrete $\sim 40\%$ fewer low-mass ($\lesssim 10^7 \, M_\odot$) halos over their lifetimes, and this further widens the disparity between the BSI and CDM-type models.
It is conventional to discuss the substructure population of Milky Way-like halos in terms of the velocity function because this quantity is somewhat more closely related to observations. The velocity function is defined in a manner similar to the mass function. The maximum circular velocity of the rotation curve of the subhalo is used as a measure of subhalo size. The cumulative velocity function $N(> V_{\text{max}})$ is defined as the number of subhalos with maximum circular velocities greater than $V_{\text{max}}$ as a function of $V_{\text{max}}$. In Figure 6.13, I show the model results for the cumulative velocity functions of subhalos, again for a fixed host mass of $M_{\text{host}} = 1.4 \times 10^{12} \, M_\odot$. Notice that the velocity functions reveal a stronger trend with power spectrum than the mass functions (Figure 6.12), but the effect is still rather modest. For the most extreme tilted model, the total number of subhalos with $V_{\text{max}} \gtrsim 10 \, \text{km s}^{-1}$ is only a factor of $\sim 2$ lower than in the standard, scale-invariant case. In the case of the tilted models, the reduction in the velocity function is largely due to the fact that the subhalos are less concentrated, so the $V_{\text{max}}$ values are correspondingly smaller for fixed halo masses [cf., Eq. (4.4) and the discussion that follows].

This effect is illustrated explicitly in Figure 6.14, where, rather than fixing the host mass at $z = 0$, I have fixed its maximum circular velocity at $V_{\text{max}} = 187 \, \text{km s}^{-1}$, the value of a typical $n = 1$, $M_{\text{host}} = 1.4 \times 10^{12} \, M_\odot$ halo at $z = 0$. Normalizing the model halos by maximum velocity rather than halo mass is perhaps a more reasonable choice because $V_{\text{max}}$ is more closely related to observations. The maximum rotation velocity for the Milky Way, for example, is $V_{\text{max, MW}} \approx 220 \, \text{km s}^{-1}$; however, the Milky Way is a large spiral galaxy in which baryons play an important role gravitationally and $V_{\text{max, MW}} \approx 187 \, \text{km s}^{-1}$ is in line with expectations for the primeval dark matter halo of the Milky Way once the effects of baryon contraction have been included.
Figure 6.14: The cumulative velocity function of subhalos in a host of fixed maximum circular velocity $V_{\text{max}} \simeq 187 \text{ km s}^{-1}$. The lines represent averages over 200 merger history realizations for $n = 1$ and 50 realizations for all other models. The error bars represent the dispersion in these realizations. The different models are as in Figure 6.13.

[164]. Models with less galactic-scale power require a more massive host in order to obtain the same value of $V_{\text{max}}$, and their velocity functions shift correspondingly (see Equation (4.8). For example, a host with $V_{\text{max}} = 187 \text{ km s}^{-1}$ in the $\sigma_8 = 0.65$ model requires $M_{\text{host}} \approx 2.2 \times 10^{12} \, M_\odot$. With this adjustment, the velocity functions of the various titled models are now very similar. Again, the BSI case is different from the tilted models because the relative shift in the $V_{\text{max}} - M_{\text{vir}}$ relation changes with mass scale. It is also encouraging that the model BSI velocity function agrees quite well with the N-body results of Colín et al. [67] for a similar type of truncated power spectrum (see their $R_f = 0.1 \, \text{Mpc}$ model, Fig. 2).
Figure 6.15: The average differential mass fraction, $df/dM_{\text{sat}}$, normalized relative to host mass and satellite mass. The upper set of (bold) curves were computed for the $n = 1$ cosmology with $M_{\text{host}} = 1.4 \times 10^{12} \, M_\odot$ at $z = 0$ (solid) and $M_{\text{host}} = 10^{11} \, M_\odot$ (dotted), $3 \times 10^{12} \, M_\odot$ (long-dash) and $10^{13} \, M_\odot$ (dot-dash) all at $z = 0.6$. The lower set of thin curves correspond to BSI halos of $M_{\text{host}} = 1.4 \times 10^{12} \, M_\odot$ at $z = 0$ (solid) and $M_{\text{host}} = 3 \times 10^{12} \, M_\odot$ at $z = 0.6$ (long-dash). The crosses reflect an analytic fit to the $n = 1$ results, as discussed in the text.

Another convenient way of describing the substructure content of halos is in terms of the mass fraction in substructure bound up in subhalos less massive than $M_{\text{sat}}$: $f(< M_{\text{sat}})$. Figure 6.15 shows the differential fraction, $df/dM_{\text{sat}}$, normalized relative to the host mass for several different host masses and redshifts. The top set of lines show ensemble averages for several different host masses and redshifts (see caption) for the $n = 1$ model. The mass fractions are approximately self-similar with respect to the host mass, and can be well-represented by the analytic form,

$$
\frac{df}{dx} = \left(\frac{x}{x_0}\right)^{-a} \exp \left(- \frac{x}{x_0}\right),
$$

(6.30)
with \( x = M_{\text{sat}}/M_{\text{host}}, \alpha = 0.6 \) and \( x_0 = 0.07 \pm 0.05 \). The quoted range in \( x_0 \) characterizes well the rms scatter from realization to realization (not shown). This function (with \( x_0 = 0.07 \)) is shown as the set of bold crosses in Figure 6.15. The lower set of lines correspond to the BSI model for two different halo masses (see caption). As expected, the mass fractions are somewhat lower in these cases. The other CDM-type models all yield differential mass functions similar to those shown for the \( n = 1 \) case. While in the next section I present results for a particular choice of host mass as a function of \( M_{\text{sat}} \), the self-similarity demonstrated here implies that the results at a fixed satellite mass ratio \( x \), can be scaled appropriately in order to apply these results to any value of \( M_{\text{host}} \).

### 6.6 Results: Mass Fractions and Gravitational Lensing

Dalal & Kochanek [81] made a significant advance when they employed a statistical technique to use flux ratios in multiply-imaged quasars to constrain the substructure content of galactic, dark matter halos. They found the fraction of mass in the lens halos bound up in substructure to be \( f = 0.006 - 0.07 \) (at 90% confidence) for \( M_{\text{sat}} \lesssim 10^8 - 10^{10} \, M_\odot \). In their sample of lens systems, the lens redshifts span the range \( 0.31 \lesssim z_l \lesssim 0.97 \) with a median lens redshift of \( z_l \simeq 0.6 \). The primary goal of this section is to present theoretical predictions aimed at lensing studies. Consequently, I have chosen to present results for host systems at \( z = 0.6 \), and with \( M_{\text{host}} = 3 \times 10^{12} \, M_\odot \), which was taken as a typical lens mass by Dalal & Kochanek.

In Figure 6.16, we show results for each of the primordial power spectrum models for host halos at \( z = 0.6 \). Here, \( f(10^6 \, M_\odot < M < M_{\text{sat}}) \) is the cumulative fraction of host halo mass that is bound up in substructure with masses larger than \( 10^6 \, M_\odot \) and
Figure 6.16: The fraction of the parent halo mass that is bound up in substructure in the mass range between $10^6 \ M_\odot$ and $M_{\text{sat}}$ as a function of $M_{\text{sat}}$. The host halo in each case has $M = 3 \times 10^{12} \ M_\odot$ at $z = 0.6$. Lines reflect the mean over all realizations, and results are shown for the $n = 1$ model (solid), RI model (dot-short-dash), $\sigma_8 = 0.75$ (dashed), $\sigma_8 = 0.65$ (dot-long-dash), and BSI (dotted). The error bars on the top set of lines reflect the 90 percentile range determined using 200 merger tree realizations for the $n = 1$ case (the other models in the top set of lines have very similar scatter). The bottom set of errors reflect the same range determined using 50 realizations of the BSI model.

less than $M_{\text{sat}}$. As expected from the prior discussion in Section 6.5, the mass fraction in substructure is not a strong function of the tilt of the primordial power spectrum, although it is sensitive to a sudden break in power at small scales. Specifically, the subhalo mass fraction in the BSI model is roughly a factor of $\sim 3$ below that seen for the CDM-type spectra in this mass range. The top set of error bars reflect the
Figure 6.17: The fraction of the mass in substructure in a central, cylindrical projection of radius 10 kpc, computed for the same set of halos as shown in Figure 6.16. The line types are the same as those in Figure 6.16, and again represent the mean fraction computed over all realizations. The large and small error bars represent the 90 and 64 percentile ranges, respectively. A down-arrow is plotted instead of a lower, large error tick if at least 5% of the realizations had \( f = 0 \) in that bin. A down-arrow with no accompanying lower error bar means that at least 18% of the realizations were without projected substructure in that bin.

90 percentile range derived using 200 realizations for the \( n = 1 \) model (other CDM-type models show similar scatter) and the bottom set of errors reflect the same range determined from 50 realizations of the BSI spectrum.

Rather than the total mass fraction, lensing measurements are sensitive to the mass fraction in substructure projected onto the plane of the lens at a halo-centric distance of order the Einstein radius, \( R_E \sim 5-10 \) kpc. In Figure 6.17 I show \( f_{\text{sat}}(> 10^6 \, M_\odot) \) projected through a cylinder of radius 10 kpc centered on the host halo for the
Figure 6.18: The cumulative mass fraction of substructure with $10^6 \, M_\odot < M_{\text{sat}} < 10^9 \, M_\odot$ shown spherically averaged as a function of radius $r$ (solid line) and in projection, as a function of the projected radius $\rho$ (dashed line). The averages (lines) and 64% percent ranges (error-bars) were determined using 200 realizations of $M = 3 \times 10^{12} \, M_\odot$ halos $z = 0.6$ for our $n = 1$ model. Down-arrows indicate that more than 18% of the realizations had $f = 0$ in the corresponding radial bin.

The same set of halos shown in Figure 6.16. The large and small error bars reflect the 90 and 64 percentile ranges, respectively, in measured projected mass fractions derived using 200 $n = 1$ realizations (top set) and 50 BSI realizations (bottom set). A down-arrow is plotted instead of a lower, large error tick if at least 5% of the realizations had $f = 0$ in that bin. A down-arrow with no accompanying lower error bar indicates that at least 18% of the realizations were without projected substructure in that bin.

The central projected mass fractions are not as severely suppressed relative to the volume-averaged mass fractions as one might have expected given that tidal forces act to systematically destroy substructure near host halo centers (see, e.g., Figure 6.5).
Figure 6.19: Mass fraction in substructure in cylindrical projection of radius $\rho$ for the same set of $n = 1$ halos described in Figure 6.18. The upper left, upper right, and lower left panels show the mass fraction profiles in subhalos larger than $10^6 \, M_\odot$ and less than $10^9$, $10^8$ and $10^7 \, M_\odot$ respectively. Error bars reflect the same percentile ranges as do those in Figure 6.18. The bottom right panel illustrates how the mean projected mass fraction profiles vary as a function of the maximum subhalo mass considered: $10^{6.3}$, $10^7$, $10^8$, $10^9$, and $10^{10} \, M_\odot$ from bottom to top.

The reason is that in examining substructure in a cylindrical volume, one picks up many subhalos subhalos in projection that have large halo-centric radii. I illustrate demonstrate effect in Figure 6.18, where I compare the mass fraction in cylindrical, projection radius $\rho$ with the mass fraction in spherical shells with the same value of spherical radius $r$. Notice that the mass fraction in spherical regions is significantly reduced in the center (as a result of tides), while the projected mass fraction is less severely affected. As expected, the mass fraction approaches the global value at large radii. Figures 6.19 and 6.20 demonstrate how the mass fractions change as a function
of projection radius for various subhalo mass cuts for the $n = 1$ model and BSI models respectively. It is interesting that the relative drop in mass fraction as a function of projected radius is slightly stronger in the BSI model than in the $n = 1$ case. This reflects both the fact that tidal disruption is more important in the BSI case and core-like behavior of the subhalo radial distribution sets in at a larger radius in this model and the fact that individual halos are less concentrated and contribute less mass along the line-of-sight when partially contained within a cylinder.

### 6.7 Warm Dark Matter and Gravitational Lensing

As I mentioned in Section 6.1, another alternative to the CDM paradigm that may have the ability to ameliorate some of the small-scale problems with the standard
cosmology is to allow the dark matter particles to be “warm.” By “warm” what is meant is that the particles have a primordial velocity dispersion that is large enough so that they resist gravitational clustering on relevant scales. In practice, however, the mass and/or velocity dispersion has to be tuned to a value that leads to an effect on the particular property that one chooses to examine. Warm dark matter (WDM) particles can be produced via many mechanisms, such as the decay of heavy parent particles; however, the most common practice is to consider the case of fermionic, thermal relics that decoupled from the thermal bath while still relativistic (Ref. [69] discusses several mechanisms for WDM particle production). In the canonical case of a thermal relic, it is a simple matter to calculate the relic abundance of such particles. The technique involves finding solutions to Equation (3.22) and yields [257]

$$\Omega_{\text{WDM}} \simeq 1.14 \left( \frac{g_{\text{WDM}}}{2} \right) \left( \frac{100}{g_{*,D}} \right) \left( \frac{m_{W}}{\text{keV}} \right).$$  \hspace{1cm} (6.31)

In Equation (6.31), $g_{\text{WDM}}$ is the number of effective fermionic degrees of freedom of the WDM species and $g_{*,D}$ is the number of relativistic degrees of freedom in thermal equilibrium at the time that the WDM particle decouples. It is conventional to choose a fiducial model of “neutrino-like” WDM in which case $g_{\text{WDM}} = 2$ and I will follow this convention here. This dependence should have been expected from the discussion of neutrino decoupling in Section (3.2.1). The primordial velocity dispersion (as distinguished from the velocity dispersion of WDM particles in a virialized halo) can be expressed as

$$< v^2 > \simeq 0.08 (1 + z) \left( \frac{100}{g_{*,D}} \right)^{1/3} \left( \frac{\text{keV}}{m_W} \right) \text{km} \text{s}^{-1}. \hspace{1cm} (6.32)$$
The warm particles “free-stream” out of potential wells while they are relativistic and so they wipe out the primordial fluctuations on all scales smaller than the comoving horizon scale when the WDM becomes non-relativistic. Utilizing the above relations and eliminating $g_{*,D}$ among them gives a characteristic free-streaming length scale

$$R_f \simeq 0.2 (\Omega_{\text{WDM}} h^2)^{1/3} \left( \frac{\text{keV}}{m_{\text{WDM}}} \right)^{4/3} .$$

(6.33)

A detailed fit to a full Einstein-Boltzmann integration yields a power spectrum that is filtered on this scale, and can be expressed in terms of the standard CDM power spectrum as [19],

$$P(k) \simeq \exp(-kR_f - k^2 R_f^2)P_{\text{CDM}}(k) .$$

(6.34)

In terms of a mass scale, this corresponds to reduced power on mass scales smaller than the filter mass,

$$M_f \simeq 6.7 \times 10^{10} \left( \frac{\Omega_{\text{WDM}} h^2}{0.15} \right)^2 \left( \frac{\text{keV}}{m_{\text{WDM}}} \right)^4 .$$

(6.35)

It is this filtering of small-scale power that yields to a reduction in galactic central densities and may yield a resolution to the dwarf satellite problem. The WDM power spectra that I consider here are depicted in Figure 6.8 and the particle masses and filtering masses are given in Table 6.1.

In the previous section I demonstrated that the substructure mass fraction is sensitive to abrupt changes in the mass power spectrum, and used the BSI model as an example of such a power spectrum. In this section I investigate these differences in the context of WDM. I will label the different WDM models by the warm particle mass and assume that the particle is a thermal relic in which case the mass maps onto
velocity dispersion in a unique way such that there is no ambiguity in the clustering properties of the particles. Throughout this section, I also assume the canonical case of “neutrino-like” WDM and take the particle to have two internal degrees of freedom, $g_w = 2$.

Figure 6.21 shows the total mass fraction of $3 \times 10^{12}$ M$_\odot$ host halos at $z = 0.6$ as a function of $M_{\text{sat}}$ implied by our three WDM model power spectra compared to our standard $\Lambda$CDM case. For substructure smaller than $\sim 10^7$ M$_\odot$, the differences between the models are as large as an order of magnitude, and even the largest WDM particle mass (3 keV) provides a potentially measurable suppression of substructure. Figure 6.22 shows the mass fraction in projected cylinders of radius 10 kpc.

The differences in mass fractions seen for the different models in Figures 6.21 and 6.22 come about because subhalos become less concentrated relative to their host halos as the WDM particle mass is decreased and power is suppressed on larger scales, essentially the same effect as in the BSI case. However, in true WDM models there are other processes that, in principle, can alter the formation and density structure of dark matter halos. In Figures 6.21 and 6.22, I have only accounted for the effect of the power spectrum on substructure mass fractions and assumed that the density structure of WDM halos is identical to that for CDM halos. For high mass systems, this is a sensible approximation (see [67, 11]); however, this approximation should break down at small masses and lead to further suppression of substructure.

One consequence of a WDM particle with non-negligible velocity dispersion is that gravitational clustering is resisted for structures below the effective Jeans mass of the warm particles [137, 37]:

213
\[ M_J \approx 6 \times 10^3 \left( \frac{\text{keV}}{m_W} \right)^4 \left( \frac{\Omega_{\text{WDM}} h^2}{0.15} \right)^{1/2} \left( \frac{2}{g_W} \right) (1 + z)^{3/2} M_\odot. \] (6.36)

For both the \( m_W = 1.5 \) keV and \( m_W = 3.0 \) keV models, \( M_J \ll 10^5 M_\odot \) when \( z \lessgtr 10 \), so all halos of interest in this context are minimally affected. The situation is somewhat more complicated in the \( m_W = 0.75 \) keV model, where \( M_J \gtrsim 10^5 M_\odot \) for redshifts \( z \gtrsim 2 \). We therefore expect that the formation of these halos should be suppressed compared to the predictions of the EPS formalism. This suppression should only have a minor effect on the predictions of relevance here because I restrict myself to satellite masses \( \gtrsim 10^5 M_\odot \) and most surviving subhalos are accreted at \( z \lesssim 2 \). In the interest of simplicity, I chose to ignore this effect here. As a result, I may significantly over-predict substructure mass fractions at low \( M_{\text{sat}} \) in these cases. In the context of this study, this is a conservative approach because the true mass fraction would be reduced by these effects, bringing it further away from the measured substructure mass fractions.

In addition to the effective Jeans suppression due to non-negligible primordial velocity dispersions, WDM halos, unlike their CDM counterparts, cannot achieve extremely high densities in their centers due to phase space constraints [303]. In the early, hot, homogeneous Universe, the primordial phase space distribution of the WDM particles is a Fermi-Dirac distribution which attains a maximum of \( f_{\text{max}} = g_W/h_{\text{PL}}^3 \) at low energies (\( h_{\text{PL}} \) is Planck’s constant not to be confused with the Hubble parameter, \( h \)). For a collisionless species, the phase space density is conserved during gravitational collapse and this maximum phase space density may not be exceeded within WDM halos. Defining the so-called phase density as \( Q = \rho / (2\pi \sigma^2)^{3/2} \), then the maximum allowed phase density is

214
\[ Q_{\text{max}} \simeq 5.2 \times 10^{-4} \left( \frac{m_{\text{W}}}{\text{keV}} \right)^4 \left( \frac{g_{\text{W}}}{2} \right) \frac{M_{\odot} / \text{pc}^2}{(\text{km/s})^3}. \]  

(6.37)

This phase space limit means that WDM halos cannot achieve the central density cusps of the kind observed in simulated CDM halos. Instead, one may expect a core in the density profile. For viable WDM models, the phase space core is expected to be dynamically unimportant for halos massive enough to host visible galaxies [5]. Nevertheless, for the the lowest-mass subhalos \((M \lesssim 10^7 M_\odot)\) the presence of phase space-limited cores may be important because halos with large cores are less resistant to tidal forces than cuspy halos.

I have attempted to make a crude estimate of the way in which the phase space limit affects the substructure population of WDM halos by adopting the standard model of halo accretion and orbital evolution, but allowing the density structure of the appropriately small subhalos to be set by the phase space limit. For these calculations I used the phenomenological density profile of Burkert [54],

\[ \rho_B(r) = \frac{\rho_0}{[1 + r/r_B][1 + (r/r_B)^2]}. \]  

(6.38)

The Burkert profile resembles the NFW form for large radius, but has the beneficial feature of a constant density core at its center, and thus a velocity dispersion that approaches a constant at small radius: \(\sigma_0^B \simeq 0.55 V_{\text{max}}\). For Burkert profiles, \(V_{\text{max}} \simeq 0.86 V_{\text{vir}} c_B / g_B(c_B)\) at a radius \(r_{\text{max}} \simeq 3.24 r_B\), where \(c_B \equiv R_{\text{vir}}/r_B\) is the Burkert concentration and \(g_B(y) \equiv \ln(1 + y^2) + 2 \ln(1 + y) - 2 \tan^{-1}(y)\). Solving for the phase density in the core \((r \ll r_B)\) gives
Figure 6.21: Total cumulative mass fractions in substructure more massive than $10^5 \text{M}_\odot$ for the $n = 1$ and 3 WDM models. The models are $\Lambda$CDM (solid), $m_W = 3.0$ keV (dashed), $m_W = 1.5$ keV (dash-dot), and $m_W = 0.75$ keV (dotted). For clarity, we show error bars only for the $\Lambda$CDM and $m_W = 1.5$ keV models. The error bars and arrows have the same meaning as in Figure 6.17.

$$\frac{Q_{B,0}}{\left(\frac{\text{M}_\odot}{\text{pc}^2}\right)\left(\frac{\text{km} \text{ s}^{-1}}{\text{km} \text{ s}^{-1}}\right)^3} \simeq 4.65 \times 10^{-6} \left(\frac{\Omega_{\text{WDM}} h^2}{0.15}\right) \left(\frac{\Delta_{\text{vir}}}{178}\right) \left(\frac{\text{km} \text{ s}^{-1}}{V_{\text{vir}}}\right)^3 (1 + z)^3 f_B^{1/2} c_B^{3/2}.$$  

(6.39)

Equating the phase density in the Burkert core with the maximum phase density of equation (6.37) yields the following relation for the maximum attainable value of $c_B$ [5, 339]:

$$c_B^{-3/2} g_B^{1/2} (c_B) \simeq 111 \left(\frac{0.15}{\Omega_{\text{WDM}} h^2}\right) \left(\frac{178}{\Delta_{\text{vir}}}{\text{km/s}}\right)^3 V_{\text{vir}}^{-3}$$
\[ \times \left( \frac{g_{\text{w}}}{2} \right) \left( \frac{m_{\text{w}}}{\text{keV}} \right) (1 + z)^{-3}. \] (6.40)

As with NFW profiles, it is important to account for the correlations between collapse time and the density structure of halos. I assigned the Burkert concentrations according to the following prescription. First, I computed NFW concentrations \( c_{\text{vir}} \), for each halo according to the B01 model. I converted from NFW concentration to Burkert concentration \( c_B \), by interpreting the B01 value of \( r_s \) as the radius at which logarithmic slope of the density profile is equal to \(-2\), explicitly \( \frac{\text{d} \ln \rho(r)}{\text{d} \ln r} \big|_{r=r_s} = -2 \). This implies that \( r_B \approx 0.66 r_s \) or \( c_B \approx 1.5 c_{\text{vir}} \). With this correspondence, the adopted Burkert profile achieves the maximum of its rotation curve at \( r_{\text{max}}^B \approx 0.986 r_{\text{max}} \), where \( r_{\text{max}} \) is the radius at which the corresponding NFW halo achieves \( V_{\text{max}} \). Similarly, the maximum circular velocity of the adopted Burkert profile is within 10\% of the corresponding NFW \( V_{\text{max}} \) for all relevant concentrations \((1 \leq c_{\text{vir}} \leq 25)\). Second, I computed the maximum value of \( c_B \) allowed by the phase space constraints using equation (6.40). I then took the smaller of these two values of \( c_B \) and assigned it to the accreted subhalo at the time of accretion. In this way, I guaranteed that the phase space constraint was met by all halos. I have checked that this prescription for Burkert halos does not yield any systematic bias in our results by applying it all of the CDM models. I found that it gave nearly identical results to that of the standard, NFW model which is not surprising in the context of the substructure model and disruption criteria.

The results for the cumulative mass fractions obtained by this calculation are presented in Figure 6.23 by the dashed lines. The upper and lower sets of lines correspond to the \( m_{\text{w}} = 0.75 \text{ keV} \) and \( 1.5 \text{ keV} \) models respectively. The solid lines show the cumulative mass fraction calculated without the phase-space constraint. It
Figure 6.22: Cumulative mass fractions in substructure more massive than $10^5$ $M_\odot$ within a projected radius of 10 kpc for the same models shown in Figure 6.21.

is clear, at least from this rough estimate, that the Tremaine-Gunn limit plays an important role only for the most extreme WDM models $m_W \lesssim 1$ keV and only the smallest halos $\lesssim 10^6$ $M_\odot$. However, it is important to emphasize that unlike my previous conclusions concerning power spectrum shapes, these new assumptions have not been tested with N-body simulations. N-body simulations have yet to examine the detailed density structure of halos that saturate the phase space bound and most studies to date have ignored the primordial velocity dispersion of the warm particles (see Refs. [11, 67, 165]), but the Burkert profile assumption seems plausible. With these precautions in mind, Figure 6.21 may be regarded as an approximate upper-limit on the substructure mass fraction for WDM halos. Any phase space bound or

218
Figure 6.23: Total cumulative mass fractions in substructure more massive than $10^5 \, M_\odot$ both with (dashed) and without (solid) estimating the effects of the phase space limit. The upper set of lines corresponds to the $m_W = 1.5 \, \text{keV}$ model and the lower set of lines corresponds to the $m_W = 0.75 \, \text{keV}$ model. The error bars and arrows are as in Figure 6.17.

the effects of primordial velocity dispersions on halo formation and density structure should only lead to enhanced disruption, resulting in lower mass fractions.

One physical process that might affect WDM (and BSI) models that I have not considered is top-down fragmentation [165]. It is possible that power can be transported from large scales to small in truncated models, resulting in a population of low-mass halos that would not be accounted for in Press-Schechter theory. While such a process could result in a higher substructure abundance than that estimated using our model, there are reasons to believe that the effects would be fairly small. Any systems that would form in this manner would collapse quite late, and their
density structure likely would be very diffuse compared to their hierarchically formed counterparts. Therefore, it is highly unlikely that systems formed via fragmentation could survive tidal disruption once incorporated into a galactic halo.

6.8 The Dwarf Satellite Problem

Comparisons between the expected subhalo population and the observed dwarf galaxy abundance are usually made by comparing counts as a function of maximum circular velocity, $V_{\text{max}}$ [163, 213]. Arguably, this is a robust method of comparison because it sidesteps the complicated issues of star formation and feedback in these poorly-understood galaxies. Yet, there are considerable uncertainties, even for this method of comparison and it is likely that efforts to compare predictions as a function of dwarf luminosity [26, 269] in tandem with velocity comparisons will be needed in order to fully understand the nature of this problem.

For most satellites, the quantity that is observed and used to infer the halo $V_{\text{max}}$ is the line-of-sight stellar velocity dispersion, $\sigma_*$. As discussed by Hayashi et al. [135] and Stoehr et al. [288], the mapping between $\sigma_*$ and $V_{\text{max}}$ depends upon the theoretical expectation for the density profile of the subhalo as well as on the stellar mass distribution of the galaxy. An additional complication concerns the unknown velocity anisotropy of the stars in the system.

A phenomenologically-motivated approximation for the stellar distribution in a dwarf galaxy is the spherically symmetric King profile (King 1962),

$$\rho_*(r) = \frac{k}{z^2} \left( \frac{\cos^{-1}(z)}{z} - \sqrt{1 - z^2} \right),$$  

where

$$z \equiv \frac{1 + (r/r_c)^2}{1 + (r/r_c)^2},$$  

\[ 220 \]
$r_c$ and $r_t$ are the core and tidal radii of the King profile, and $\rho_*(r > r_t) = 0$. The normalization is not important in what follows. As I’ve already mentioned, the Milky Way has eleven known satellite galaxies within 300 kpc of the Galactic center. Their names, distances, velocity dispersions and King profile parameters are listed in Table 6.2 along with the appropriate references. In Table 6.2, I also give the value of $V_{\text{max}}$ assigned to the halos that host the satellite galaxies by Klypin et al. [163] in their original study of the dwarf satellite problem. The Magellanic Clouds do not have well-determined central velocity dispersions or King profile parameters, but there is recent kinematic data on the clouds available in the studies of van der Marel et al. [313] and Stanimirović et al. [276] that I will discuss below.

If I assume that a stellar system described by Equation (6.41) is in equilibrium and embedded in a spherically symmetric dark matter potential characterized by the circular velocity profile $V_c(r)$, then the radial stellar velocity dispersion $\sigma_r(r)$ can be computed via the Jeans equation:

$$r \frac{d(\rho_\ast \sigma_\ast)}{dr} = -\rho_\ast(r) V_c^2(r) - 2\beta(r) \rho_\ast \sigma_\ast^2,$$

where the anisotropy parameter, $\beta \equiv 1 - \sigma^2_\ast / 2\sigma_r^2$, and $\sigma_r$ and $\sigma_\perp$ are the radial and tangential velocity dispersions, respectively. For the special case of an isotropic velocity dispersion tensor, $\beta = 0$ and the one-dimensional velocity dispersion profile can be written as a simple integral. A measured, line-of-sight velocity dispersion is determined by the projected velocity dispersion profile weighted by the luminosity distribution sampled along line-of-sight. For the isotropic case it is given by

$$\sigma_\ast^2 = \frac{\int_0^{r_t} \rho_\ast(r') V_c^2(r') dr'}{\int_0^{r_t} \rho_\ast(r') dr'},$$

(6.44)
<table>
<thead>
<tr>
<th>Satellite</th>
<th>$D$</th>
<th>$\sigma_*$</th>
<th>$r_c$</th>
<th>$r_t$</th>
<th>$V_{\text{K KV}}^\text{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kpc</td>
<td>km s$^{-1}$</td>
<td>kpc</td>
<td>kpc</td>
<td></td>
</tr>
<tr>
<td>Sagittarius</td>
<td>24 ± 2</td>
<td>11.7 ± 0.7</td>
<td>0.44</td>
<td>3.0</td>
<td>20</td>
</tr>
<tr>
<td>Fornax</td>
<td>138 ± 8</td>
<td>10.5 ± 1.5</td>
<td>0.46</td>
<td>2.3</td>
<td>18</td>
</tr>
<tr>
<td>Draco</td>
<td>82 ± 6</td>
<td>9.5 ± 1.6</td>
<td>0.18</td>
<td>0.93</td>
<td>17</td>
</tr>
<tr>
<td>Ursa Minor</td>
<td>66 ± 3</td>
<td>9.3 ± 1.8</td>
<td>0.20</td>
<td>0.64</td>
<td>16</td>
</tr>
<tr>
<td>Leo I</td>
<td>250 ± 30</td>
<td>8.8 ± 0.9</td>
<td>0.22</td>
<td>0.82</td>
<td>15</td>
</tr>
<tr>
<td>Carina</td>
<td>101 ± 5</td>
<td>6.8 ± 1.6</td>
<td>0.21</td>
<td>0.69</td>
<td>12</td>
</tr>
<tr>
<td>Leo II</td>
<td>205 ± 12</td>
<td>6.7 ± 1.1</td>
<td>0.16</td>
<td>0.48</td>
<td>12</td>
</tr>
<tr>
<td>Sculptor</td>
<td>79 ± 4</td>
<td>6.6 ± 0.7</td>
<td>0.11</td>
<td>1.5</td>
<td>11</td>
</tr>
<tr>
<td>Sextans</td>
<td>86 ± 4</td>
<td>6.6 ± 0.7</td>
<td>0.34</td>
<td>3.2</td>
<td>11</td>
</tr>
<tr>
<td>SMC</td>
<td>58 ± 2</td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>LMC</td>
<td>49 ± 4</td>
<td></td>
<td></td>
<td></td>
<td>50</td>
</tr>
</tbody>
</table>

Table 6.2: A summary of the properties of the satellite galaxies of the Milky Way. Column (1) gives the name of the galaxy. I abbreviate the names of the Small Magellanic Cloud and the Large Magellanic Cloud by “SMC” and “LMC” respectively. Column (2) gives the distance to the satellite in kpc. Column (3) gives the measured line-of-sight velocity dispersion of the galaxy in km s$^{-1}$. Columns (4) and (5) give the measured King profile core and tidal radii, $r_c$ and $r_t$, in kpc. The last column gives the value of $V_{\text{max}}$ assigned to the host halo of each satellite by Klypin, Kravtsov, Valenzuela, and Prada [163]. Except for Draco, all measured quantities are taken from the study of the Local Group by Mateo [200]. For Draco, I quote the parameters given by Odenkirchen et al. [222].
assuming a constant mass-to-light ratio. If a galaxy has a measured stellar profile (the
King profile parameters in this case) and measured value of $\sigma_*$, then Equation 6.44
[or equation (6.43) if $\beta \neq 0$] places only one constraint on the rotation curve of the
system, $V_c(r)$. The host halo velocity profile is expected to be at least a two-parameter
function (e.g., an NFW profile rotation curve or Burkert profile) so determining $V_{\text{max}}$
requires some theoretical input for the expected form of $V_c(r)$ in order to provide a
second constraint.

Motivated by dark matter models, I assume that the global rotation curve is set
by an NFW profile associated with the dwarf galaxy halo. The rotation curve for an
NFW halo is fully described by specifying two parameters and a natural pair is $V_{\text{max}}$
and $r_{\text{max}}$. For any given cosmology, the relation between $V_{\text{max}}$ and $r_{\text{max}}$ is expected to
be rather tight, and this provides a (theoretically-motivated) second constraint that
sets the $V_{\text{max}}-\sigma_*$ mapping implied by Eq. (6.43) [or Eq. (6.44)].

The $V_{\text{max}}-r_{\text{max}}$ relationships for subhalos in two of the substructure models are
shown in Figure 6.24. The lower set of points corresponds to the standard, $n = 1, \sigma_8 =
0.95$ model and the higher set of points is derived from the low-power $\sigma_8 = 0.65$ model.
In Figure 6.24, I plot one point for each surviving halo in ten merger tree realizations
in each case. The strong correlation, $r_{\text{max}} \propto V_{\text{max}}^\gamma, \gamma \approx 1.3$, follows directly from the
input correlations between $M_{\text{vir}}(z)$, and $c_{\text{vir}}$ (see Section 6.2.4 and Refs. [48, 321, 339]).
The normalizations and slopes are influenced by the cosmology, accretion times and
(mildly) by the orbital history of the subhalos. In reality, I expect the scatter in the
$V_{\text{max}}-r_{\text{max}}$ plane to be larger than that shown here because I have not included the
expected scatter in the input $c_{\text{vir}}-M_{\text{vir}}$ relation [149, 48, 321]. For $\sigma(\log c_{\text{vir}}) \approx 0.14,$
as estimated by Bullock et al. [48] and Wechsler et al. [321], the implied scatter is

223
Figure 6.24: The lower and upper sets of points show a scatter plot of \( V_{max} \) and \( r_{max} \) for all surviving halos produced in 10 merger tree realizations for the \( n = 1 \), and low-normalization, \( \sigma_8 = 0.65 \) models respectively. The thick solid line shows the locus of points in the \( V_{max}-r_{max} \) plane that corresponds to the central value of the measured velocity dispersion of Carina (\( \sigma_* = 6.8 \pm 1.6 \) km s\(^{-1}\)) given the measured King profile parameters of Carina (Table 6.2) and assuming NFW halos. The thin solid lines correspond to the \( \pm 1\sigma \) quoted errors in velocity dispersion for Carina. The dashed line corresponds to the locus of points that is consistent with the central value of the measured velocity dispersion of Draco (\( \sigma_* = 9.5 \) km s\(^{-1}\)). Consistency demands that Carina (Draco) resides in a halo with structural parameters that overlap with the solid (dashed) lines.

\[ \sigma(\log r_{max}) \approx 0.18 \] at fixed \( V_{max} \). For the smaller scatter \( \sigma(\log c_{vir}) \approx 0.08 \) advocated by Jing [149], the implication is \( \sigma(\log r_{max}) \approx 0.11 \).

The thick solid and dashed lines in Figure 6.24 show the locus of points in the \( V_{max}-r_{max} \) plane that correspond to the central values of the measured \( \sigma_* \) values for Carina and Draco respectively, given their measured King profile parameters. The \( \sigma_* \) values and King profile parameters that I have adopted are listed in Table 6.2 along with
appropriate references. The light solid lines illustrate how these contours expand when I include the \( \pm 1\sigma \) measurement errors in \( \sigma_* \) for Carina. A similar (although narrower) band exists for Draco, but I have omitted it for the sake of clarity. Consistency with the observed King parameters and velocity dispersions requires each dwarf to reside in a halo with structural parameters that lie within the region of overlap between the contours and the model points. For example, in the \( n = 1 \) model Carina is expected to reside in a halo with \( V_{\text{max}} \approx 11 \text{ km s}^{-1} \) and \( r_{\text{max}} \approx 1 \text{ kpc} \). For the \( \sigma_8 = 0.65 \) model, Carina is expected to sit in a much larger halo, with \( V_{\text{max}} \approx 29 \text{ km s}^{-1} \) and \( r_{\text{max}} \approx 10 \text{ kpc} \). Similar comparisons hold for Draco and all of the Local Group dwarf satellites and these comparisons can be made in a similar way for any cosmology. The point is that the maximum velocities that are assigned to satellite galaxies are cosmology-dependent. Therefore, “observed” velocity functions are also cosmology-dependent because theoretical inputs are used to convert from \( \sigma_* \) to \( V_{\text{max}} \). Reduced small-scale power demands systematically higher values of \( V_{\text{max}} \) because halos are less concentrated.

In Table 6.3 I show estimates for halo \( V_{\text{max}} \) values for the dwarf satellite galaxies of the Milky Way under the assumption that \( \beta = 0 \). In Table 6.4 I show a similar compilation of \( V_{\text{max}} \) values assuming that the stellar orbits are not isotropic within the satellite galaxies and that the anisotropy is given by an anisotropy parameter \( \beta = 0.15 \). I compute halo \( V_{\text{max}} \) estimates for six different power spectra, relying on the model-dependent \( r_{\text{max}}-V_{\text{max}} \) relationship for substructure in each case, and taking the central values of the measured velocity dispersions for each halo. Taking the quoted \( \pm 1\sigma \) range for the measured velocity dispersions typically leads to a shift in \( V_{\text{max}} \) of \( \sim 30\% \) which is a considerable effect compared to the inherent scatter.

225
<table>
<thead>
<tr>
<th>Satellite</th>
<th>$V_{\text{max}}$</th>
<th>$V_{\text{max}}$</th>
<th>$V_{\text{max}}$</th>
<th>$V_{\text{max}}$</th>
<th>$V_{\text{max}}$</th>
<th>$V_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 1$</td>
<td>$\sigma_8 = 0.83$</td>
<td>RI</td>
<td>$\sigma_8 = 0.75$</td>
<td>BSI</td>
<td>$\sigma_8 = 0.65$</td>
</tr>
<tr>
<td>Sagittarius</td>
<td>17</td>
<td>20</td>
<td>23</td>
<td>24</td>
<td>26</td>
<td>35</td>
</tr>
<tr>
<td>Fornax</td>
<td>15</td>
<td>18</td>
<td>20</td>
<td>21</td>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>Draco</td>
<td>21</td>
<td>31</td>
<td>38</td>
<td>41</td>
<td>41</td>
<td>67</td>
</tr>
<tr>
<td>Ursa Minor</td>
<td>21</td>
<td>31</td>
<td>37</td>
<td>41</td>
<td>42</td>
<td>69</td>
</tr>
<tr>
<td>Leo I</td>
<td>17</td>
<td>24</td>
<td>28</td>
<td>30</td>
<td>33</td>
<td>51</td>
</tr>
<tr>
<td>Carina</td>
<td>11</td>
<td>14</td>
<td>18</td>
<td>18</td>
<td>22</td>
<td>29</td>
</tr>
<tr>
<td>Leo II</td>
<td>13</td>
<td>19</td>
<td>23</td>
<td>24</td>
<td>27</td>
<td>40</td>
</tr>
<tr>
<td>Sculptor</td>
<td>13</td>
<td>18</td>
<td>21</td>
<td>23</td>
<td>26</td>
<td>38</td>
</tr>
<tr>
<td>Sextans</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>SMC</td>
<td>$&lt; 60^*$</td>
<td>$&lt; 60^*$</td>
<td>$&lt; 60^*$</td>
<td>$&lt; 60^*$</td>
<td>$&lt; 60^*$</td>
<td>$&lt; 60^*$</td>
</tr>
<tr>
<td>LMC</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50$^*$</td>
</tr>
</tbody>
</table>

Table 6.3: The maximum circular velocities of the halos of the Milky Way satellites assuming isotropic stellar velocities. The first column gives the name of the Milky Way satellite. Columns (2)-(7) give the inferred $V_{\text{max}}$ for their host halos in the context of each primordial power spectrum model that I consider. Each model is labeled at the head of the appropriate column. The SMC and LMC represent special cases. The quoted maximum velocities for the LMC and SMC are taken from the studies of van der Marel et al. [313] and Stanimirović et al. [276] respectively. The LMC rotation curve is observed to be flat from $r \sim 4$ kpc to greater than $r \sim 9$ kpc. For many of the lower-power models that I consider, the flat portion of the rotation curve (at $r \sim r_{\text{max}}$) is expected to be at somewhat larger radius. In order to explain this in the context of these models, it is necessary to assume that baryonic contraction plays an important role in setting the properties of the dark matter rotation curves [35]. In this case, the measured value of $V_{\text{max}}$ ($r_{\text{max}}$) is larger (smaller) than it would be for the pristine halo prior to baryonic infall. The SMC rotation curve is even more likely to be influenced by baryonic effects [276], and baryonic infall is likely to be of some importance for all cases. While not demanded by the data, the effects of baryonic infall may be important for all Milky Way satellites, thus the listed $V_{\text{max}}$ values should be considered lower limits. Lastly, the large value of $r_{\text{max}}$ associated with Draco in the $\sigma_8 = 0.65$ model may be difficult to reconcile with the kinematic data of Kleyna et al. [161], and may disfavor a model with such low small-scale power.
Figure 6.25: The satellite halo velocity functions for six of the models compared with the velocity functions of the Milky Way satellites after accounting for the cosmology-dependent mapping between $\sigma_\star$ and $V_{\text{max}}$. The squares represent the velocity function of Milky Way satellites based on the data in Table 6.3. The lines represent the means over all realizations and the error bars reflect the dispersion among these realizations.

expected in the $M_{\text{vir}} - c_{\text{vir}}$ relation. For reference, I have also included the adopted $V_{\text{max}}$ values from the original Klypin et al. [163] work on the dwarf satellite problem in Table 6.2. As expected, the implied $V_{\text{max}}$ values become larger as I explore models with less galactic-scale power. The estimates for the $n = 1$, $\beta = 0$ case are close to those of Klypin et al. [163] as expected.

The panels of Figure 6.25 show the Milky Way satellite counts for each model, assuming $\beta = 0$, along with the predicted velocity functions for each model. In addition to the satellites listed in upper portions Tables 6.3 and 6.4 and for which
<table>
<thead>
<tr>
<th>Satellite</th>
<th>$V_{\text{max}}$</th>
<th>$V_{\text{max}}$</th>
<th>$V_{\text{max}}$</th>
<th>$V_{\text{max}}$</th>
<th>$V_{\text{max}}$</th>
<th>$V_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 1$</td>
<td>$\sigma_S = 0.83$</td>
<td>RI</td>
<td>$\sigma_S = 0.75$</td>
<td>BSI</td>
<td>$\sigma_S = 0.65$</td>
</tr>
<tr>
<td>Sagittarius</td>
<td>13</td>
<td>15</td>
<td>17</td>
<td>17</td>
<td>19</td>
<td>25</td>
</tr>
<tr>
<td>Fornax</td>
<td>11</td>
<td>13</td>
<td>15</td>
<td>15</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>Draco</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>26</td>
<td>29</td>
<td>44</td>
</tr>
<tr>
<td>Ursa Minor</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>26</td>
<td>29</td>
<td>43</td>
</tr>
<tr>
<td>Leo I</td>
<td>12</td>
<td>16</td>
<td>19</td>
<td>20</td>
<td>23</td>
<td>32</td>
</tr>
<tr>
<td>Carina</td>
<td>8</td>
<td>10</td>
<td>13</td>
<td>13</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>Leo II</td>
<td>9</td>
<td>13</td>
<td>16</td>
<td>16</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>Sculptor</td>
<td>9</td>
<td>12</td>
<td>15</td>
<td>15</td>
<td>18</td>
<td>24</td>
</tr>
<tr>
<td>Sextans</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>SMC</td>
<td>$&lt; 60^*$</td>
<td>$&lt; 60^*$</td>
<td>$&lt; 60^*$</td>
<td>$&lt; 60^*$</td>
<td>$&lt; 60^*$</td>
<td>$&lt; 60^*$</td>
</tr>
<tr>
<td>LMC</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 6.4: This Table is formatted in the same way as Table 6.3. The difference in the content is that I have calculated halo $V_{\text{max}}$ assuming an anisotropy parameter of $\beta = 0.15$ for the orbits of the stars in the Milky Way satellites. This choice of $\beta$ is motivated by the results of [66]. Again, the SMC and LMC represent special cases (see Table 6.3.)
I have carried out the comparison described in this paragraph, I have also included
the Small Magellanic Cloud (SMC) and the Large Magellanic Cloud (LMC) in these
cumulative velocity functions. The Magellanic Clouds have very complex velocity
structures and the properties of the halo must be determined in a somewhat model-
dependent manner. Stanimirović et al. [276] estimate a circular velocity of \( V_c = 60 \) km s\(^{-1}\) at distance of \( r \sim r_{\text{max}} \) for the SMC and I will adopt this value as an
estimate of \( V_{\text{max}} \) for the SMC. The Large Magellanic Cloud (LMC) has \( V_c \approx 50 \) km s\(^{-1}\)
at radii \( r \sim 4 - 9 \) kpc as determined by van der Marel et al. [313]. I take \( V_{\text{max}} = 50 \) km s\(^{-1}\) as an estimate of the maximum circular velocity of the halo of the LMC
in my cumulative velocity functions. The precise choice of \( V_{\text{max}} \) for the Magellanic
Clouds has little affect on my conclusions because they are the largest Milky Way
satellites and there is no dwarf satellite mismatch for objects of this size.

For the standard case \( (n = 1) \) the discrepancy sets in at \( V_{\text{max}} \sim 30 \) km s\(^{-1}\),
as observed in previous studies [163, 213]. This requires roughly one-in-ten halos
at \( \sim 10 - 20 \) km s\(^{-1}\) to be the host of a luminous galaxy. The extremely tilted
model (with \( \sigma_s \approx 0.65 \)) actually under-predicts the dwarf count for large systems.
Interestingly, dwarfs in the \( \frac{d n}{d \ln k} = -0.03 \) RI model as well as the \( \sigma_s = 0.75 \) case
are consistent with inhabiting the \( \sim 10 \) most-massive subhalos, with only Sextans
standing as an outlier. The BSI model also looks to be in good agreement with the
data for \( V_{\text{max}} \gtrsim 12 \) km s\(^{-1}\). This is the model originally investigated by Kamionkowski
& Liddle [152] in order to solve the dwarf satellite problem (though the specific
truncation scale has been adjusted).

Nevertheless, it is important that I emphasize that the problem of Local Group
satellites is not “solved” in any of these models because the velocity function continues
Figure 6.26: The satellite halo velocity functions for six of the models compared with the velocity functions of the Milky Way satellites assuming a constant anisotropy parameter of $\beta = 0.15$ as discussed in the text. The squares represent the velocity function of Milky Way satellites based on the data in Table 6.4. The lines represent the means over all realizations and the error bars reflect the dispersion among these realizations.

to rise below velocity-scale of Sextans in all cases. What changes in the low-power models is the nature of the discrepancy. In one extreme, the mismatch sets in at $V_{\text{max}} \sim 30$ km s$^{-1}$ and gradually becomes worse for smaller systems. The mechanism that suppresses star/galaxy formation in the satellite halos must gradually become more effective as $V_{\text{max}}$ goes down. In the other extreme, the mismatch seems to imply a sharp threshold for dwarf galaxy formation at $\sim 10 - 20$ km s$^{-1}$.

Unfortunately, a detailed accounting for the mismatch is extremely difficult, even for a fixed cosmology and primordial power spectrum. The dwarf $V_{\text{max}}$ estimate is
very sensitive to the velocity anisotropy parameter, $\beta$. For example, when the values are recalculated with $\beta = 0.15$, the implied $V_{\text{max}}$ values are significantly lower than in the isotropic case because rotational support has been traded for pressure support. The velocity function comparisons with $\beta = 0.15$ are shown in the panels of Figure 6.26 and the $V_{\text{max}}$ values are listed in Table 6.4. In this case, only the $\sigma_8 \simeq 0.65$ case and the BSI model can easily account for the dwarf population without a differential feedback mechanism. The rest of the models over-predict the counts, with the greatest apparent discord in the $n = 1$ case. Of course, the specific choice of $\beta = 0.15$ serves mainly to illustrate the effect of a minor anisotropy. I chose this value because it is typical of what is seen in the central regions of simulated dark matter halos [66], and therefore it seems a reasonable possibility for the anisotropy parameter of particles in dwarf galaxies.

6.9 Another Model Application: What Can Be Learned From Neutralino Annihilation In Substructure?

It is also interesting to demonstrate how this type of modeling may be applied to other outstanding questions in cosmology. One of the most prominent remaining puzzles facing cosmology and particle physics is the nature of the dark matter. As we have seen, the preferred dark matter candidate seems to be a massive, weakly-interacting particle. Massive neutrinos ($m_\nu \sim 5 \text{ GeV}$) were formerly considered good candidates for dark matter because their relic abundance after decoupling from the thermal bath in the early Universe is appropriate for a dark matter candidate, $\Omega_{M,0}h^2 \approx 0.15$ or so (e.g., see [171]). However, the current bounds on neutrino mass rule out this scenario, at least for Standard Model neutrinos, and so a new candidate is needed. Supersymmetry (SUSY) has developed as the most natural extension to the Standard
Model with the capability of mitigating the hierarchy problem. Supersymmetry also provides electromagnetically neutral, massive particles that interact weakly so as to freeze-out at an abundance that is cosmologically acceptable, and may therefore fill the bill of a dark matter candidate, namely the neutralinos.

The fact that the standard CDM model of hierarchical structure formation predicts that the Milky Way halo should be permeated by small, dense subhalos and that these subhalos may be neutralinos has stimulated a great deal of work on direct detection of substructure through the photon by-products of neutralino annihilations in these dense environments [27, 57, 8, 293]. Such a detection could potentially confirm the existence of a significant number of “dark” subhalos in the Galactic halo and may be able to teach us about SUSY parameters. Most past studies have had to make assumptions about substructure properties in the Galactic halo. With the model of Section 6.2, however, it is possible to model substructure distributions and to perform a statistically meaningful study by analyzing a large number of realizations of Milky Way-like halos. In this section, I give a brief introduction to ongoing work related to these issues [175].

The signal expected from neutralino annihilations in subhalos depends upon four factors: (1) the distribution of substructure in the Milky Way halo which determines the typical distance to an object; (2) the sensitivity of the detector and the backgrounds; (3) the cross section for neutralino annihilation which, in turn, depends upon the SUSY parameters; and (4) the radius at which the inner profile of the halo achieves a constant core density, which may be due to the annihilations themselves. The substructure distribution can be determined statistically using the substructure model described earlier in this Chapter. The measured background is composed of
three components, electron and hadron cosmic rays [221] and the diffuse background 
from astrophysical sources [275]. The cross section for photon production also has two 
components. One is from direct annihilation into two photons or into a $Z^0$-photon 
final state via sfermion loops and the other is indirect via hadronization of annihila-
tion products. The goal of this study is to be as optimistic as possible and maximize 
the likelihood of neutralino detection and so we adopt the theoretical upper limits 
for both cross sections [29, 28] by adopting appropriate SUSY parameter values. We 
adopt the detector specifications of the VERITAS experiment [322]. The dark matter 
density must reach a maximum at some small radius, rather than diverging as in the 
NFW profile. Physically, this radius may be set by equating the rate for gravitational 
infall to the annihilation rate, which sets the core radius $r_c \approx 10^{-10} - 10^{-8} r_s$, but it 
is very likely set at a larger radius due to baryonic effects. Nevertheless, to maximize 
the prospects for neutralino detection, we assume the core radius to be small and 
vary it in the range $10^{-11} - 10^{-7} r_s$.

Taking all of these inputs into account it is then possible to estimate the likelihood 
of directly detecting neutralino subhalos via annihilations into photons. Taking the 
most optimistic case that the neutralino has mass $M_\chi = 300$ GeV, that we observe the 
continuum emission at $E \approx 100$ GeV, in order to maximize the observable signal given 
the specifications of VERITAS, and that the neutralino annihilation cross section to 
al channels is the maximum allowable given the available parameter space, one can 
calculate the average number of visible subhalos (at the $3\sigma$ level) in a Milky Way-like 
halo as a function of the uncertain core radius.

In Figure 6.27, I show the average number of detectable subhalos in a Galactic 
host after a month long exposure of VERITAS as a function of core radius. Notice
Figure 6.27: This figure shows the number of neutralino subhalos in the Milky Way that are detectable at 3σ via neutralino annihilation given the most optimistic choices for all SUSY and detector parameters. The upper line represents the result assuming the $n = 1$ standard model and the lower line assumes the WMAP running index power spectrum.

that even with all of the optimistic choices that were made in order to maximize the observed flux, on average there will be fewer than 7 observable subhalos in the Milky Way. Given the field of view of VERITAS, it would take 48 years to survey 1/7 of the sky and thus detect a single subhalo in this way.

Contrary to the results of prior studies, it seems unlikely that this strategy can be used to learn about SUSY parameters. The non-detection of gamma rays from neutralino annihilations seems likely even in optimistic cases. The detection of such a signal, however, may yield some information about the physics of subhalo disruption. It is unclear what happens to subhalos that become tidally disrupted. They are likely not the hosts of visible galaxies, but small cores from their very inner regions
may survive and N-body simulations are unable to track the evolution of these cores. Moreover, the very inner regions (many orders of magnitude below numerical resolution limits) of halos are precisely where all of the neutralino annihilation signal comes from. As such, a detection of such a signal may hint at the survival of the dense inner regions of subhalos after they have been stripped of almost all of their mass [175].

6.10 Caveats

The model I have built and employed in this study is a simple, semi-analytic model based on many previous studies [38, 181, 271, 49, 295, 48, 321, 133] and designed to produce large numbers of halo realizations with minimal computational effort. In developing this model, I have made many simplifying assumptions. In this section I draw attention to many of these shortcomings and discuss how they might affect my results and how these shortcomings may be improved upon in future work.

Among the most obvious omissions in this work is the neglect of any disk or bulge component in each halo. I have specifically chosen to ignore the effects of central galaxies because the physics of dark halo formation is relatively well-understood compared with that of galaxy formation. This also allows the model to be grounded in and calibrated against dissipationless N-body simulations. In order to include a galactic component, one is forced to adopt many poorly-constrained models and assumptions regarding gas accretion, cooling, angular momentum distributions, feedback, and the effects of merging substructure on the galaxy itself. Once a reliable framework for the dark matter has been developed, it will be possible to use this as a foundation for more speculative (yet quite interesting) explorations involving the baryonic components.
A central (disk) galaxy would add to the dynamical friction force experienced by subhalos orbiting near the plane of the disk and cause halos on highly inclined orbits to be tidally heated during rapid encounters with the disk potential (e.g., Gnedin & Ostriker 1999; Gnedin, Hernquist, & Ostriker 1999; TB01). These effects lead to enhanced satellite disruption. On the other hand, subsystems that are massive enough to contain galaxies might be rendered more resistant to tidal disruption as a result of the enhanced central component of cool baryons. For low-mass halos, the effect of including a central galaxy would likely act to reduce the substructure count, mainly at small radii. Although, even without including these effects, the model discussed here reveals that the substructure fraction drops significantly at small radii because of the dark matter potential, and that a large part of the projected central mass fraction comes from subhalos at large radii that are picked up in projection. Nevertheless, projected mass fractions are rather sensitive to the size of the core in the subhalo radial distribution [61], so if the core region were larger as a result of a central galaxy, the implied lensing signal would tend to be reduced relative to these estimates. As an extreme example, I find that eliminating all substructure within 20 kpc of the halo center, reduces the projected mass fractions in subhalos less massive than $10^{10} \, M_\odot$, $f_{10}$, by $\sim 30\%$.

In addition to the uncertainties associated with the process of galaxy formation, there are some potential shortcomings in the model that concern the physics of dark matter only. For example, I have allowed for only a mild redistribution of mass within the tidal radii of the orbiting subhalos up until the time the subhalo is totally tidally destroyed. The work of Hayashi et al. [135] and Stoehr et al. [288] suggest that this effect may be larger. However, it is important to regard these results with some
care. These results may have been compromised by limited numerical resolution or, in the case of Hayashi et al. [135], inappropriate assumptions having to do with initial orbits and/or accretion times of dwarf-sized subhalos. In this sense, the approach approach taken here represents a conservative extreme because I assume that the surviving subhalo density structure is typically very similar to that of halos in the field. In addition, I have adopted a halo concentration relation (B01) that has not been confirmed for $M \lesssim 10^9 \, M_\odot$. Similarly, the EPS merger tree calculations have yet to be tested in the low-mass regime. In light of these extrapolations, it is imperative that the results of this model be tested, and updated using the next generation of numerical studies.

Finally, this simple model does not self-consistently treat the substructure population of in-falling subhalos. I have neglected any subhalo-subhalo interactions which would serve to increase the internal heating of substructure and modify the rates of dynamical friction as orbital energy is exchanged between subhalos and traded for internal energy. I have adopted the approximation that all in-falling halos are “distinct” and have no subhalos of their own (see Taylor & Babul 2003 [296] for a recent study of merger tree “pruning”). However, the “tree-level” calculations of this Chapter suggest that the mass fraction in substructure is uniformly $\sim 10\%$ regardless of host mass, thus it seems sensible to expect the that this correction would typically affect our derived mass fractions by $\lesssim 10\%$. Considering the assumptions that have gone into these calculations, the additional caveats discussed above, and the current level of observational precision, this level of error is quite acceptable; yet, this model will likely need to be improved upon as observations zero in on the masses.
Figure 6.28: This figure shows a preliminary test of the semi-analytic orbit tracking code presented in Section 6.2. The mass (top panel) and radial (bottom panel) evolution for four our different orbits are shown. The solid lines represent the N-body results and the dashed lines represent the results of the analytic models. For the analytic lines, the lines that stop suddenly represent disruption events according to the criteria of the analytic model. For technical reasons, the N-body code is initialized with a subhalo of extent $R_{\text{sat}} = 35r_s$ which exceeds the virial radius of the halo. The mass is expressed in units of the initial mass at this radius $M_{35}$. The analytical model has no mass outside of $R_{\text{vir}}$ so that the initial mass of the analytically modeled halos are less than the N-body halos, but they quickly come together once tidal forces act to strip the exterior of the N-body halo. Figure courtesy J. S. Bullock and K. V. Johnston.

of the subclumps responsible for the lensing signals and the mass fractions in these subclumps.
6.11 Future Work: Testing and Refining the Analytic Model

The results presented in this Chapter are largely based on analytic approximations that make the calculations computationally feasible. As observations and theory (i.e., N-body techniques) improve, it will be increasingly necessary to test all analytic models using state of the art numerical codes. The most ill-understood aspect of the analytic modeling involves the physics of tidal disruption: How is mass stripped and when do halos become completely disrupted? Preliminary work is underway using idealized N-body simulations with two goals in mind [47]. The first is to build an understanding of halo disruption due to tidal forces. The second goal is to improve the semi-analytic model so that predictions of substructure characteristics and abundances can be made more accurate.

The N-body simulations being used in order to understand the disruption process are idealized. The host halo is modeled as a potential and thus there is no dynamical friction effect. The satellite is modeled as an NFW halo with $10^5$ particles and initial parameters set according to their cosmological expectations as in Section 6.2. Preliminary results are shown in figures 6.28 and 6.29. Notice that the overall agreement between the analytic and numerical results is rather promising, lending support to the model described in this Chapter. Yet, these figures do point out shortcomings of the analytic procedure. First, notice in Figure 6.28 that stripping and disruption are more gradual in the N-body results than in the analytic code in which stripping happens almost exclusively at pericenter. Also notice in Figure 6.29 that the analytic model tends to underestimate the amount of stripping at the first pericenter pass. In fact the analytic model never “recovers” from this and therefore tends to overestimate
Figure 6.29: This figure shows a possible shortcoming of the analytic model. The panels and lines are the same as in Figure 6.28 but for four more elliptical orbits. Notice that on elliptical orbits, the analytic model tends to underestimate the stripping at the first pericenter passage and never recovers.

the amount of remaining mass the satellite. Moreover the discrepancy becomes more pronounced with increasingly elliptical orbits.

6.12 Halo Substructure: Discussion and Conclusions

The substructure abundance in dark matter halos is determined through an ongoing competition between accretion and disruption. Accreted subhalos with dense cores are resistant to disruption, but over time they are more likely to be destroyed as their orbits decay and their mass is stripped away. The model I presented in this Chapter allows one to follow the complicated interplay between orbital evolution, accretion time, and survival probability in order to determine how changes in the power
spectrum affect the final substructure population in galaxy-sized halos. For a fixed set of cosmological parameters, changes in the power spectrum manifest themselves by changing collapse times for halos, where less power leads to later accretion times and lower densities. I have specifically focused on tilted models that help to relieve the central density crisis facing CDM (Ref. [339] and Fig. 5.8) and that may be favored by joint CMB and large-scale structure analyses [274]. The advantage of these models is that they preserve the success of the CDM paradigm on large scales and there is no need to invoke unmotivated and/or fine-tuned dark matter candidates. I have also considered a BSI inflation model and WDM models, where the power is sharply reduced on small scales.

For a large class of CDM-type models, including those with significant tilts and rapidly varying spectral indices, I find that the fraction of mass bound up in substructure $f$, for galaxy-mass halos is relatively insensitive to the slope of the primordial power spectrum. This result comes about because both the host halos and their accreted subsystems collapse later in these models, in a roughly self-similar way as the power is reduced. Note that this result would have been expected if one were to vary only the overall normalization of the power spectrum because the relative redshifts of collapse would be invariant (I assume host halos are small enough that they collapse before $z \sim 0$). The model of this Chapter suggests that this intuitive description holds even for tilted and running index models, at least over the parameter range that I have explored. All indications are that this insensitivity to the tilt of the power spectrum is a rather robust result, and should hold even if some unknown factor causes the overall normalization in predicted mass fractions to be in error (e.g., by the exclusion of central galaxies as discussed in Section 6.10). Interestingly, the shape of the mass
function, \( f(x \equiv M_{\text{sat}}/M_{\text{host}}) \), is also relatively insensitive to the mass of the host halo (see Eq. 6.30), and a similar shape holds for all of the tilted models that I explored.

The similarity in mass fractions breaks down significantly for models with sharp features in their power spectra, like the BSI case and WDM models. In these models, low-mass halo formation is delayed significantly relative to the formation time of their hosts. Consequently, fewer subhalos are dense enough to withstand the tidal field they experience upon accretion. I find that for the relevant WDM and BSI models, the mass fraction in substructure is reduced by a factor of \( \gtrsim 3 \) compared to the standard/tilted \( \Lambda \)CDM models.

Inspired by recent attempts to measure substructure mass fractions using multiply-imaged quasars, I applied this model to ensembles of host halos with \( M = 3 \times 10^{12} \) \( M_\odot \) at \( z = 0.6 \) in order to match the expectations for massive lens galaxies [81, 82]. For the \( \Lambda \)CDM/tilted cases, the expected substructure mass fractions within a 10 kpc projected radius in subsystems less massive than \( M = 10^8, 10^9, \) and \( 10^{10} M_\odot \) are \( f_8 \approx 0.2 - 0.4\%, f_9 \approx 0.4 - 1.5\%, \) and \( f_{10} \approx 0.6 - 2.5\% \) at the 64 percentile range. These estimates are consistent with, but on the low side, of first attempts to measure the substructure fraction using multiply-imaged quasars by Dalal & Kochanek [81], who obtain \( f \approx 0.6\% - 7\% \) at 90\% confidence, with an upper mass limit of \( 10^8 - 10^{10} M_\odot \) [80]. The lensing results disfavor the BSI model, which has a projected fractions of \( f_8 \approx 0.01 - 0.06\%, f_9 \approx 0.02 - 0.2\%, \) and \( f_{10} \approx 0.03 - 0.4\% \) (64 percentile). This is true unless the break scale in the power spectrum is pushed to such a small value that this model no longer has the attractive feature of alleviating the central density and dwarf satellite problems. A \( m_W = 0.75 \) keV WDM model is similarly disfavored, and even the highest mass WDM case that I considered in this study, \( m_W = 3 \) keV,
has a typical projected substructure mass fraction \( f_0 \sim 0.4\% \) that is fairly low compared to the Dalal & Kochanek constraint. Again, this indicates that if the warm particle is a thermal relic, the mass must be large enough that it no longer mitigates the small-scale problems of standard CDM. Yet, these results are interesting because they show how lensing effects may be used as one of the few probes of the dark matter particle mass in the range \( \gtrsim 1 \text{ keV} \) or a break in the primordial power spectrum at large wavenumber.

Of course, these conclusions must be regarded with some caution. In addition to the uncertainties of modeling discussed in Section 6.10, other issues make drawing definite conclusions difficult. For example, I have only accounted for the substructure within the virial radius of the host halo, yet the anomalous flux ratios of lensed images are sensitive to the presence of small halos along the line-of-sight to the lens. Keeton [158] showed that field halos can have a significant lensing effect even if they are separated from the lens by several tenths in redshift and in hierarchical, CDM-type models, small halos in the field are ubiquitous. For example, Chen et al. [61] have shown that the relative effect from halos outside the virial radius of the lens is typically a few percent, but it may be as large as 20 – 30\% of that from subhalos, depending upon assumptions about the subhalo population. What is more, as the mass fraction in substructure of a given mass depends on the mass of the host [Equation (6.30)], it may be important to constrain the host halo mass in order to fully exploit the ability of lensing measurements of substructure to probe cosmology and structure formation.

I compared the model predictions for the cumulative subhalo velocity function, \( N(> V_{\text{max}}) \), to the satellite galaxy count of the Milky Way. The approach here was to estimate the \( V_{\text{max}} \) value of each satellite galaxy’s dark matter halo based on
its observed line-of-sight velocity dispersion, $\sigma_*$. As I emphasized in Section 6.8, the mapping between $\sigma_*$ and $V_{\text{max}}$ is sensitive to theoretical prejudice regarding the density structure of the dwarf galaxy’s halo as well as the unknown velocity anisotropy parameter of the system, $\beta$. For a fixed value of $\beta$, less concentrated host halos imply larger values of $V_{\text{max}}$ because halo rotation curves are more slowly rising and stars probe only the inner $\sim 1$ kpc of the halo. Interestingly, this implies that tilted models and truncated models, do significantly better than $n = 1$, $\Lambda$CDM in reproducing apparent dwarf counts, even though their mass fractions are similar.

When I fix $\beta = 0$ as a representative case, the $dn/d\ln k = -0.03$ running index model, $\sigma_s = 0.75$ model, and the BSI case all do well in matching the known satellite population of the Milky Way for $V_{\text{max}} \gtrsim 20$ km s$^{-1}$. The lowest power model ($\sigma_s = 0.65$, $n \simeq 0.84$, and mild running) actually under-predicts the dwarf count for $V_{\text{max}} \gtrsim 30$ km s$^{-1}$. However, this result is achieved only for the optimistic assumption of isotropic velocities. If instead I adopt even a small level of anisotropy, $\beta = 0.15$, consistent with anisotropies in the centers of numerically-simulated dark matter halos, agreement for most models is worsened. Only BSI and the $\sigma_s = 0.65$ models show good agreement in this case. Yet, even with $\beta = 0.15$, the RI and $\sigma_s = 0.75$ models still compare more favorably than the $n = 1$ case with $\beta = 0$.

What do these results imply for the dwarf satellite problem? In all models, including those with truncated power, the velocity function of subhalos continues to rise below the scale of the smallest observed Milky Way satellites $V_{\text{max}} \lesssim 10$ km s$^{-1}$. Therefore, no matter how one modifies the power spectrum, some kind of feedback is required to explain the local satellite population. Different power spectra (or even different values of $\beta$) seem to indicate that different types of feedback are needed.
For example, in models with $\sigma_8 \gtrsim 0.8$ (the precise number depends on typical $\beta$ values and the degree of running/tilt), the feedback must be differential. That is, for $V_{\text{max}} \approx 8 - 30$ km s$^{-1}$, only one out of every $\sim 5 - 10$ halos in this mass range should form stars. On the other hand, in models like the $\text{dn}/\text{dln}~k = -0.03$ RI model, the BSI case, and the $\sigma_8 \approx 0.75$ model with $\beta = 0$, the discrepancy seems to set in suddenly at $V_{\text{max}} \sim 10 - 20$ km s$^{-1}$, suggesting that nearly all halos smaller than this are completely devoid of stars. In this case, the feedback mechanism must provide a sudden transition in mass scale between luminous and non-luminous galaxies.

The feedback mechanism proposed by Bullock, Kravtsov, and Weinberg (BKW) [49] accommodates the need for only $\sim 10\%$ of subsystems to actually host observable galaxies by suggesting that only those systems that formed before reionization were able to retain their gas and eventually form stars (see Figure 6.1). However, if reionization were to occur very early (e.g., $z \gtrsim 15$), many fewer than 10% of these dwarf-sized systems could have collapsed before reionization, so that almost all systems smaller than $V_{\text{max}} \sim 30$ km s$^{-1}$ would be dark. This would be more in line with what is observed for the low power models. This is an intriguing result. The best-fit running index power spectrum of the WMAP team [274] leads to similar substructure mass fractions as standard CDM, alleviates the dwarf satellite discrepancy, and forces cosmologists to consider feedback mechanisms that lead to a sharp transition between luminous and non-luminous galaxies. Additionally, the possible detection of early reionization by the WMAP team ($z_{\text{re}} \sim 17$) [169] provides a feedback mechanism that results in a sharp transition. Of course, explaining early reionization in models with low small-scale power may be problematic in itself [270]. Another feedback scenario that leads to a sharp transition in the velocity function is photo-evaporation in
which ionizing photons evaporate gas out of halos even after they have collapsed [22]. Nonetheless, the uncertainty associated with $\beta$ in determining satellite galaxy $V_{\text{max}}$ values suggests that efforts to model dwarf galaxy luminosities as well as dynamical properties will be required to resolve this issue [26, 269].

A brief word on the relevance of the lensing measurements to tests of the $\Lambda$CDM paradigm is in order. It is important to note that the lensing indications of $f > 1\%$ mass fractions may be significantly less encouraging for CDM-type models than originally expected. The results that I presented in this chapter suggest that the cumulative projected mass fraction in substructure reaches the $f \sim 1\%$ level indicated by flux-ratio anomalies only when rather massive sub-systems are included, $M_{\text{sat}} \lesssim 10^9 \, M_\odot$. At this mass scale, we expect $\sim 20$ satellite halos around the Milky Way — this is within a factor of two of the observed satellite count. While a result of this kind remains interesting, and seems to disfavor WDM and BSI models, it does not probe the mass scale associated with the hundreds of “missing” subhalos that were the focus of the dwarf satellite problem as originally posed by Klypin et al. [163] and Moore et al. [213]. At present, the substructure measurements from anomalous flux ratios in strong lenses are sensitive to even larger masses ($\sim 10^{10} M_\odot$ [80]), so that it is possible that the flux-ratio anomalies can be explained entirely by subhalos similar to those associated with the $\sim 3 - 10$ most massive, and directly observed Galactic satellites. Certainly, more detailed lens modeling will be needed to draw any firm conclusions in this regard. Nonetheless, these results highlight the need to narrow-in on the substructure masses responsible for the lensing, either by continued efforts like those of Dalal & Kochanek [81, 167, 158], modified lensing techniques [209, 215, 159],
or by probes within our own Galaxy [151, 144, 143, 203]. Modeling of the kind presented in this Chapter may play an important role in interpreting these observational results and may also find alternative applications such as that discussed in Section 6.9.
APPENDIX A

ROBERTSON-WALKER METRIC AND KINEMATICS

The standard cosmology begins with a statement of the cosmological principle. This is the assumption that the Universe is homogeneous and isotropic on large scales. The cosmic microwave background provides remarkable evidence that the Universe is very nearly isotropic (at the level of one part in $10^5$). The assumption that the Earth does not occupy a preferred point in space and therefore that the Universe is isotropic at all points in space, implies that the Universe is spatially homogeneous.

Any metric exhibiting homogeneous and isotropic spatial hypersurfaces can be written as

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

and is known as the Robertson-Walker (RW) metric. The constant $k$ is known as the curvature constant and the coordinates can be scaled such that $k$ can take on the values $k = 1$, $k = -1$, and $k = 0$. These values correspond to positively curved, negatively curved (or hyperbolic), and flat spatial hypersurfaces respectively as $k$ is related the three-dimensional Ricci curvature scalar of the three dimensional subspace by
\[ \mathcal{R}^{(3)} = \frac{6k}{R^2(t)}. \] \hfill (A.2)

A particularly convenient coordinate transformation,

\[ r = \Sigma(\chi) \] \hfill (A.3)

with

\[ \Sigma(\chi) = \sin(\chi) \text{ if } k = 1, \hfill (A.4) \]
\[ = \chi \text{ if } k = 0, \text{ and } \hfill (A.5) \]
\[ = \sinh(\chi) \text{ if } k = -1 \hfill (A.6) \]

gives the simple form

\[ ds^2 = dt^2 - R^2(t)[d\chi^2 + \Sigma^2(\chi)(d\theta^2 + \sin^2(\theta)d\phi^2)]. \] \hfill (A.7)

It is conventional to denote the scale factor today as \( R_0 \) and to refer to values of the scale factor in units of its value today \( a(t) \equiv R(t)/R_0 \).

In each case, the coordinate \( t \) simply measures the proper time measured by an observer that maintains constant spatial coordinates (\( i.e., \chi \) or \( r, \theta, \) and \( \phi \)). Such observers are known as \textit{comoving} observers because they move relative to each other only due to changes in the scale factor and not due to changes in their coordinate positions or \textit{peculiar} velocities.

The most important kinematic effect of the Robertson-Walker metric is the \textit{redshift}. Consider two wave crests of an electromagnetic wave emitted by a comoving emitter at \( \chi_e \) at times \( t_1 \) and \( t_2 \). By isotropy we may consider a wave that travels along
a path such that $\theta$ and $\phi$ remain constant. Light waves travel along null geodesics ($ds^2 = 0$) so using equation (A.7) the two crests arrive at a comoving observer at $\chi_e + \Delta \chi$ at times $t'_1$ and $t'_2$ given by

$$\Delta \chi = \int_{t_1}^{t'_1} R^{-1}(t)dt = \int_{t_2}^{t'_2} R^{-1}(t)dt.$$  \hspace{1cm} (A.8)

Therefore,

$$\int_{t_2}^{t'_2} R^{-1}dt - \int_{t_1}^{t'_1} R^{-1}dt = \int_{t'_1}^{t'_2} R^{-1}dt - \int_{t_1}^{t_2} R^{-1}dt = 0.$$  \hspace{1cm} (A.9)

In the limit that the timescale for emitting or observing the wave is small compared to the rate at which $R$ varies the gives

$$\frac{t'_2 - t'_1}{t_2 - t_1} = \frac{R(t'_1)}{R(t_1)}.$$  \hspace{1cm} (A.10)

The period of a wave is proportional to its wavelength so the observed wavelength $\lambda_o$, is related to the emitted wavelength $\lambda_e$, by

$$\frac{\lambda_o}{\lambda_e} = \frac{R(t'_1)}{R(t_1)}.$$  \hspace{1cm} (A.11)

The wavelength of the emitted photons vary as $\lambda \propto R$. The cosmological redshift is defined as

$$z \equiv \frac{\Delta \lambda}{\lambda} = \frac{R(t_{\text{obs}})}{R(t_{\text{emit}})} - 1.$$  \hspace{1cm} (A.12)

In the standard cosmological model, $R(t)$ is a monotonically increasing function of $t$ and therefore the values of $t$, $a \equiv R/R_0$, and $z \equiv 1/a - 1$ are all used to indicate different cosmologically epochs.

250
The redshift phenomenon is more general than this application to photons. Consider a massive particle with four-velocity $u^\alpha$ moving along a geodesic in space-time. The geodesic equation reads

$$\frac{du^\mu}{d\tau} + \Gamma^\mu_{\alpha\beta} u^\alpha u^\beta = 0 \quad \text{(A.13)}$$

where $\tau$ represents proper time along the geodesic. For $\mu = 0$ in the RW spacetime this reads

$$\frac{du^0}{d\tau} + \frac{d \ln R}{dt} |\vec{u}|^2 = 0, \quad \text{(A.14)}$$

where $\vec{u}$ is the spatial three-velocity of the particle. Using the fact that $u^\alpha u_\alpha = 1$, equation (A.14) reveals that

$$\frac{d \ln |\vec{u}|}{d \ln R} = -1 \quad \text{(A.15)}$$

implying that $|\vec{u}| \propto R^{-1}$ so that the three-velocity of a massive particle redshifts as well.

The redshift has a very important consequence for the phase-space distribution of particles. Consider a non-interacting particle species that has a phase space distribution described by a Fermi-Dirac or Bose-Einstein distribution:

$$f d^3 x d^3 p = [\exp(E(p)/T) \pm 1]^{-1} d^3 x d^3 p. \quad \text{(A.16)}$$

The phase space density is conserved during gravitational interactions. This implies then, that the effective temperature that can be used to describe the particles changes with redshift. In particular,
\[ T \propto R^{-1} \quad \text{(if } m \ll T) \quad \text{or} \]
\[ T \propto R^{-2} \quad \text{(if } m \gg T). \tag{A.17} \]

Another aspect of the RW spacetime can be understood through a discussion of the luminosity distance, \( d_L \). Consider a source of photons with luminosity \( L \) at the origin of coordinates (homogeneity implies that the choice of origin is arbitrary) and let the position of the earth be \( \chi_0 \). The luminosity distance is defined by demanding that the flux \( \mathcal{F} \) observed on earth is

\[ \mathcal{F} \equiv \frac{L}{4\pi d_L^2} \tag{A.19} \]

in analogy with Euclidean geometry. As I have already shown, the photons will be affected by the dynamics of the spacetime encapsulated in the scale factor. Photons emitted over a time interval \( \Delta t \) will be observed over a time interval \( \Delta t/(1+z) \) and will carry only a fraction \( 1/(1+z) \) of their initial energy. Thus the effective luminosity of the object at earth is \( L_{\text{eff}} = L/(1+z)^2 \) and this amount of luminosity must pass through the two-surface defined by \( t = t_0 \) and \( \chi = \chi_0 \). The area of this surface is

\[ A = 4\pi R_0 \Sigma^2(\chi_0). \tag{A.20} \]

The photons travel on null geodesics so \( R(t)d\chi = dt \) and

\[ \chi_0 = \int_{t_e}^{t_o} R^{-1} dt \tag{A.21} \]

where \( t_e \) and \( t_o \) are the time of emission and observation respectively. The flux seen at the observer is thus \( \mathcal{F} = L/4\pi(1+z)^2 R_0^2 \Sigma^2(\chi_0) \) and the luminosity distance is

252
\[ d_L = (1 + z)R_0 \Sigma(\chi_0). \]  

(A.22)

A practical determination of luminosity distance as a function of redshift requires information about the mapping between redshift and coordinate distance, \( \chi_0 \). This, in turn, requires input about the dynamics of the RW spacetime which is discussed in Appendix B.
APPENDIX B

THE FRIEDMANN-ROBERTSON-WALKER MODEL AND THE STANDARD COSMOLOGICAL PARAMETERS

The dynamics of the spacetime are determined by the theory of General Relativity. The Einstein field equations provide the critical link between the dynamics of the metric components $g_{\mu\nu}$, and the matter contained within the spacetime. They read

$$R_{\mu\nu} - \mathcal{R}g_{\mu\nu}/2 = 8\pi GT_{\mu\nu}$$  \hspace{1cm} (B.1)

where $R_{\mu\nu}$ is the Ricci tensor, $\mathcal{R}$ is the Ricci curvature scalar, $G$ is Newton’s gravitational constant, and $T_{\mu\nu}$ is the stress-energy tensor.

The stress-energy tensor describes the matter and energy in the Universe. Homogeneity and isotropy of space severely restrict the form of the stress-energy tensor. In particular,

$$T_{00} = \rho(t), \hspace{0.5cm} T_{i0} = 0, \hspace{0.5cm} \text{and} \hspace{0.5cm} T_{ij} = g_{ij} P(t),$$  \hspace{1cm} (B.2)

where $\rho(t)$ and $P(t)$ are functions of $t$ only. This is the form for an isotropic, homogeneous, perfect fluid with proper density $\rho(t)$ and pressure $P(t)$ as measured by observers in the comoving frame. The RW metric and conservation of stress-energy $T^{\mu\nu}_{\mu\nu}$, imply that

254
\[ \frac{d\rho}{dt} = -3 \frac{\dot{R}}{R} (\rho + P) \]  

(B.3)

where the over-dot represents differentiation with respect to \( t \).

The last ingredient that is needed to determine the spacetime dynamics is an equation of state relating \( \rho \) and \( P \). A simple and particularly useful parameterization is \( P = w \rho \). In this case the energy density of the fluid scales with \( R \) as

\[ \rho \propto R^{-3(1+w)}. \]  

(B.4)

Three particularly interesting cases are: *matter* for which the kinetic energy is much smaller than the mass energy of the fluid and so \( P \ll \rho \) and \( w \approx 0 \); radiation for which kinetic energy dominates over mass energy and \( w = 1/3 \); and vacuum energy which is constant so that \( w = -1 \). Therefore, the energy density of a Universe composed of these three components can be written down in terms of the energy densities in matter, radiation, and vacuum energy today (\( \rho_{M,0}, \rho_{R,0}, \) and \( \rho_{\Lambda} = \text{constant} \) respectively):

\[ \rho = \rho_{M,0} \left( \frac{R_0}{R} \right)^3 + \rho_{R,0} \left( \frac{R_0}{R} \right)^4 + \rho_{\Lambda}. \]  

(B.5)

Notice that the energy density in relativistic species comes to dominate at early times (i.e., small \( R \)) due to the way these different components of the energy density scale with \( R \). A vacuum energy has the same effect as a cosmological constant, so it is common practice among cosmologists to use the terms “vacuum energy” and “cosmological constant” interchangeably when discussing the dynamics of an FRW universe.

In addition to stress-energy conservation, the Einstein equations give only one other independent equation.

255
\[ H^2 \equiv \left( \frac{\dot{R}}{R} \right)^2 = \frac{8\pi G}{3} \rho - \frac{k}{R^2}, \quad (B.6) \]

known as the Friedmann equation. \( H \) is the expansion rate and its value today \( H_0 \), is known as the Hubble parameter. Since the measurements of Edwin Hubble [142] we have known that the Universe is expanding and recent measurements indicate that \( H_0 \approx 70 \text{ km/s/Mpc} \) [109]. The rate of expansion today sets a boundary condition for the Friedmann equation and allows us to extrapolate backwards in time. It is useful to cast the Friedmann equation in the form

\[ H^2 = H_0^2 (\Omega_{M,0} a^{-3} + \Omega_{R,0} a^{-4} + \Omega_{\Lambda,0} + \Omega_{k,0} a^{-2}) \quad (B.7) \]

where \( a \equiv R/R_0 \), \( \Omega_{M,0} \equiv \rho_{M,0}/\rho_{\text{crit}}, \Omega_{R,0} \equiv \rho_{R,0}/\rho_{\text{crit}}, \Omega_{\Lambda,0} = \rho_{\Lambda}/\rho_{\text{crit}}, \) and \( \Omega_{k,0} \equiv -k/R_0^2 H_0^2 \). The critical density is

\[ \rho_{\text{crit}} = \frac{3 H_0^2}{8\pi G} \quad (B.8) \]

and is significant because, from equation (B.7), if the present energy density \( \rho_0 = \rho_{M,0} + \rho_{R,0} + \rho_{\Lambda} = \rho_{\text{crit}} \) this implies that \( \Omega_{k,0} = k = 0 \) and that the Universe is spatially flat. With knowledge of \( H_0 \), the densities of the various components, and the Friedmann equation (B.6) one can calculate luminosity distances [equation (A.22)], the age of the Universe since \( R = 0 \), and other important observables. The parameters \( H_0, \Omega_{M,0}, \Omega_{\Lambda,0} \) are the fundamental parameters of cosmology and a great body of research has been devoted to determining these parameters. It is common practice to refer to the energy density of any hypothetical component of the stress-energy of the Universe in units of the critical density. Thus, \( \Omega_{i,0} \) is simply the present density of species ‘i’ in units of the critical density.

256
BIBLIOGRAPHY


257


258


261

262


263


268


