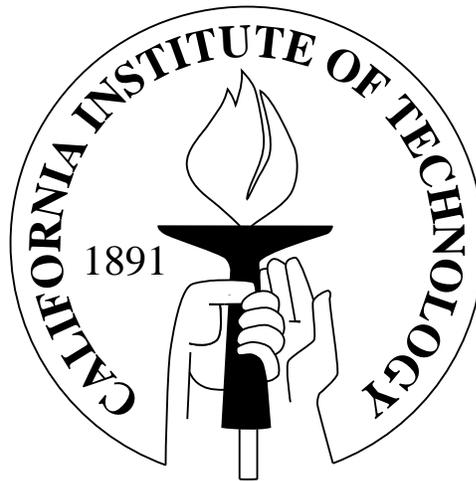


# The Consequences of Modifying Fundamental Cosmological Theories

Thesis by  
Adrienne L. Erickcek

In Partial Fulfillment of the Requirements  
for the Degree of  
Doctor of Philosophy



California Institute of Technology  
Pasadena, California

2009  
(Defended May 26, 2009)



The most incomprehensible thing about the world is that it is comprehensible.

– *Albert Einstein, 1936*

---

It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.

– *Albert Einstein, 1933*

---

An ancestor of mine maintained that if you eliminate the impossible whatever remains, however improbable, must be the truth.

– *Spock, 2293, quoting Sherlock Holmes*

## Acknowledgments

This thesis is the cumulation of five years of cosmological investigations at Caltech, and it would not exist without the interactions that I had here. First, I thank my advisor, Marc Kamionkowski, for asking extremely interesting questions and giving me the chance to find the answers. Marc's interests span the complete cosmological spectrum from field theories in the early Universe to the details of structure formation, and I have benefited from and been inspired by the breadth of his curiosity. In addition to exemplifying the type of wide-reaching cosmologist that I aspire to be, Marc has also been a generous and supportive advisor, and I am grateful for the many opportunities that he has given me to grow as a cosmologist.

I've been fortunate to collaborate with many other cosmologists in addition to Marc. When I think of the most enjoyable cosmological discussions that I've had at Caltech, my memory often takes me to Sean Carroll's office, and I thank Sean for his encouragement. Modifying Einstein's field equation is a messy business with many potential pitfalls, and I was very fortunate to have a partner in Tristan Smith. I also thank my other collaborators: Andrew Benson, Robert Caldwell, Takeshi Chiba, and Chris Hirata. In addition, I am very grateful to the broader Tapir community, especially my fellow graduate students, for creating such an exciting and stimulating scientific environment. In particular, I thank Yacine Ali-Haimoud, Nate Bode, Ben Collins, Dan Grin, Matt Johnson, Mike Kesden, Jonathan Pritchard, Annika Peter, Anthony Pullen, Hilke Schlinting, and Tristan Smith for many illuminating discussions. I also thank Shirley Hampton and Chris Mach for keeping Tapir running smoothly.

There are many people whom I thank for making the good times at Caltech splendid and the rough times bearable. First, I give my utmost gratitude and love to Nick Law. Nick, the past five years have been the best years of my life because I shared them with you, and the future is bright with you by my side. I thank you for making Pasadena "home." I also thank Jenelle Bray and Kana Takematsu, who have been dear friends and confidants; I cherished our "girls' nights" of DDR, thai food, bad TV, and long conversations. I've been very lucky to call Dan Grin a friend for ten years now, and I thank him for being an insightful and inspiring companion throughout my physics education, in addition to being a supportive and caring friend. I also thank many other friends who have made the past five years better with their presence: Neil Halelamien, Matthew Kelley, Wendy Mercer, Annika Peter, Brian Standley, Eve Stenson, Jessie Rosenberg, Christine Romano, Angelle Tanner, and Laurence Yeung.

Finally, I would like to thank a few people without whom I may not have reached Caltech. My aspirations of being a theoretical astrophysicist survived my first encounter with Newton's laws

because I took my first physics class from Mike Sinclair, and I thank him for being a fantastic teacher and mentor. And although their impact was not personal, I must thank George Lucas for convincing me at the age of four that space is the place to be, and Kip Thorne for showing me that the science is even better than the science fiction with his book, *Black Holes and Time Warps: Einstein's Outrageous Legacy*. Last, but certainly not least, I thank my parents: Jan and George Erickcek. I would have never made it this far without “the support crew in Kalamazoo,” and I thank them for always being willing to lend a helping hand and a sympathetic ear. I also thank them for nurturing my love for physics and encouraging me to pursue my dreams. My father introduced me to physics when I was very young; he told me that faster-than-light travel is difficult because time slows down at high speeds. I have been hooked ever since – Thanks, Dad.

## Abstract

In this work, we examine alternatives to three fundamental cosmological theories: extended Press-Schechter merger theory, general relativity, and single-field inflation, and derive their observational consequences. The extended Press-Schechter merger rate for dark matter haloes is mathematically inconsistent and double-valued, and yet it has been widely applied in cosmology. One such application is the merger rate of supermassive black holes, and we show that the two predictions for this rate from extended Press-Schechter merger theory are nearly equal. We then compare the supermassive-black-hole merger rate derived from the extended Press-Schechter merger formalism to the rate derived from an alternate theory, in which halo merger rates are obtained by inverting the coagulation equation.

Next, we show how two modifications to general relativity may be tested inside the Solar System. First we consider  $f(R)$  gravity, which was proposed to explain late-time cosmic acceleration. We find that several forms of  $f(R)$  gravity are inconsistent with observations, and we establish a set of criteria that determines whether or not a given form of  $f(R)$  gravity is ruled out by Solar System gravitational tests. Second, we study Chern-Simons gravity: a parity-violating theory inspired by string theory. We find that Chern-Simons gravity predicts orbital precessions that are different from those predicted by general relativity, and we use the motion of satellites to constrain the Chern-Simons coupling parameter.

Finally, we consider an alternative to single-field inflation; in the curvaton scenario, the inflaton does not generate all of the primordial perturbations. Using this theory, we propose an origin for the hemispherical power asymmetry that has been observed in the cosmic microwave background on large angular scales. While this asymmetry cannot be produced by a superhorizon fluctuation in the inflaton field, it may be generated by a superhorizon fluctuation in the curvaton field. A superhorizon fluctuation would also induce large-scale anisotropies in the cosmic microwave background; we analyze this effect and prove that our model is consistent with observations. We also show how the power asymmetry may be suppressed on smaller scales if the curvaton creates isocurvature perturbations when it decays.

# Contents

<b>Acknowledgments</b>	<b>iv</b>
<b>Abstract</b>	<b>vi</b>
<b>1 Introduction and Summary</b>	<b>1</b>
1.1 Rocking cosmology’s foundations . . . . .	1
1.2 Supermassive black hole merger rates: uncertainties from halo merger theory . . . . .	3
1.3 Solar system tests of $f(R)$ gravity . . . . .	4
1.4 The effects of Chern-Simons gravity on bodies orbiting the Earth . . . . .	5
1.5 Superhorizon perturbations and the cosmic microwave background . . . . .	6
1.6 A hemispherical power asymmetry from inflation . . . . .	7
<b>2 Supermassive Black Hole Merger Rates: Uncertainties from Halo Merger Theory</b>	<b>9</b>
2.1 Introduction . . . . .	9
2.2 Cosmological event rates . . . . .	11
2.3 The relationship between halo mass and black hole mass . . . . .	12
2.4 LISA event rates from EPS merger theory . . . . .	15
2.4.1 Review of EPS merger theory . . . . .	15
2.4.2 LISA event rates from EPS theory . . . . .	19
2.5 BKH merger theory . . . . .	26
2.5.1 Solving the coagulation equation . . . . .	26
2.5.2 BKH merger rates for power-law power spectra . . . . .	28
2.6 Comparison of LISA event rates from BKH and EPS merger theories . . . . .	33
2.7 Summary and discussion . . . . .	35
<b>3 Solar System Tests of <math>f(R)</math> Gravity</b>	<b>38</b>
3.1 Introduction . . . . .	38
3.2 A detailed example: $1/R$ gravity . . . . .	39
3.3 The weak-field solution around a spherical star in $f(R)$ gravity . . . . .	44
3.4 Case studies in $f(R)$ gravity . . . . .	50
3.5 Characteristics of viable $f(R)$ theories . . . . .	52
3.6 Summary and discussion . . . . .	56

<b>4</b>	<b>The Effects of Chern-Simons Gravity on Bodies Orbiting the Earth</b>	<b>58</b>
4.1	Introduction . . . . .	58
4.2	Chern-Simons gravity . . . . .	59
4.3	The Chern-Simons gravitomagnetic equations . . . . .	60
4.4	Gravitomagnetism due to a spinning sphere in Chern-Simons gravity . . . . .	63
4.4.1	Calculation of the vector potential . . . . .	63
4.4.2	The gravitomagnetic field . . . . .	66
4.5	Orbital and Gyroscopic precession . . . . .	69
4.5.1	Orbital precession . . . . .	69
4.5.2	Gyroscopic precession . . . . .	72
4.6	Summary and discussion . . . . .	73
<b>5</b>	<b>Superhorizon Perturbations and the Cosmic Microwave Background</b>	<b>75</b>
5.1	Introduction . . . . .	75
5.2	The Grishchuk-Zel'dovich effect: A brief review . . . . .	76
5.3	CMB anisotropies from superhorizon potential perturbations . . . . .	80
5.4	Application to curvaton perturbations . . . . .	86
5.5	Summary and discussion . . . . .	89
<b>6</b>	<b>A Hemispherical Power Asymmetry from Inflation</b>	<b>91</b>
6.1	Introduction . . . . .	91
6.2	A scale-invariant power asymmetry . . . . .	93
6.2.1	Single-field models . . . . .	93
6.2.2	The curvaton model . . . . .	96
6.3	Review of isocurvature perturbations . . . . .	100
6.3.1	Isocurvature perturbations in the curvaton scenario . . . . .	101
6.3.2	Isocurvature modes in the cosmic microwave background . . . . .	104
6.4	A power asymmetry from curvaton isocurvature . . . . .	108
6.4.1	Case 1: The curvaton creates most of the dark matter. . . . .	111
6.4.2	Case 2: The curvaton's contribution to the dark matter is negligible . . . . .	113
6.5	Summary and discussion . . . . .	117
<b>A</b>	<b>A review of <math>f(R)</math> gravity's equivalence to scalar-tensor gravity</b>	<b>121</b>
<b>B</b>	<b>The Cancellation of the CMB Temperature Dipole</b>	<b>123</b>
B.1	Dipole cancellation in a $\Lambda$ CDM Universe . . . . .	123
B.2	Dipole cancellation in a universe with an exotic fluid . . . . .	125

<b>C Attempts to Generate a Scale-Dependent Asymmetry without Isocurvature</b>	<b>128</b>
C.1 A discontinuity in $\xi(k)$ ? . . . . .	129
C.2 A smooth transition? . . . . .	131
<b>Bibliography</b>	<b>136</b>

## List of Figures

1.1	The WMAP map of the cosmic microwave background, divided into the northern and southern ecliptic hemispheres. . . . .	8
2.1	Relating halo masses to supermassive black hole masses. . . . .	13
2.2	The critical halo mass $M_*(z)$ . . . . .	16
2.3	The two EPS merger kernels. . . . .	18
2.4	The rate of SMBH mergers per comoving volume for different minimum halo masses. . . . .	20
2.5	The rate of SMBH mergers per comoving volume for different minimum SMBH masses. . . . .	21
2.6	The gravitational-wave event rate from SMBH mergers as a function of the minimum halo mass that contains a SMBH large enough to produce a detectable signal when it merges. . . . .	22
2.7	The gravitational-wave event rate from SMBH mergers as a function of the maximum redshift of a detectable merger. . . . .	23
2.8	The halo mass range that dominates the rate of SMBH mergers per comoving volume. . . . .	25
2.9	Plot of $\log \sigma(M)$ in the halo mass range that dominates the rate of SMBH mergers per comoving volume. . . . .	28
2.10	Equal-mass BKH merger kernels for $z = 0$ . . . . .	30
2.11	Equal-mass BKH merger kernels for $z = 5$ . . . . .	31
2.12	The two EPS merger kernels and the BKH merger kernel. . . . .	32
2.13	The rate of SMBH mergers per comoving volume from BKH and EPS merger rates. . . . .	34
2.14	The gravitational-wave event rate from SMBH mergers from BKH and EPS merger rates. . . . .	35
3.1	The effective potential of chameleon gravity. . . . .	54
4.1	The gravitomagnetic field generated by a rotating sphere in Chern-Simons gravity. . . . .	67
4.2	An orbital diagram showing the longitude of the ascending node. . . . .	70
4.3	The Lense-Thirring drag for the LAGEOS satellites in Chern-Simons gravity. . . . .	71
4.4	The gyroscopic precession for Gravity Probe B in Chern-Simons gravity. . . . .	73
5.1	The evolution of the gravitational potential $\Psi$ in a $\Lambda$ CDM Universe with radiation. . . . .	79
5.2	Contributions to the CMB dipole from a superhorizon perturbation. . . . .	83

5.3	The upper bounds on the fraction of the energy density in the curvaton field just prior to its decay from CMB anisotropies. . . . .	88
5.4	The parameter space for superhorizon curvaton fluctuations. . . . .	89
6.1	A power asymmetry from a superhorizon fluctuation. . . . .	92
6.2	The $R$ - $\xi$ parameter space for the curvaton model that produces a power asymmetry. . . . .	98
6.3	CMB power spectra for unit-amplitude initial perturbations. . . . .	106
6.4	A power asymmetry from a superhorizon fluctuation in the curvaton field. . . . .	108
6.5	$K_\ell$ for scenarios in which most of the dark matter comes from curvaton decay. . . . .	112
6.6	$K_\ell$ for scenarios in which the curvaton's contribution to the dark matter density is negligible and $\mathcal{T}_{SS} = \kappa R$ . . . . .	114
6.7	The ratio $\tilde{A}/\xi$ , as a function of $\xi$ , for several values of $\kappa^2\xi$ . . . . .	115
6.8	The $\xi - \kappa$ parameter space for models in which the curvaton does not contribute significantly to the dark matter density. . . . .	116
C.1	The upper bounds on $-\mathrm{d}\ln\xi/\mathrm{d}\ln k$ . . . . .	132

## Chapter 1

# Introduction and Summary

### 1.1 Rocking cosmology’s foundations

In the beginning, there was a scalar field. The scalar field’s potential energy dominated the energy density of the Universe, driving a period of nearly exponential expansion called inflation. During inflation, the Universe became homogeneous, isotropic, and spatially flat as any remnants of prior inhomogeneity or curvature were pushed beyond our cosmological horizon. Quantum fluctuations in the scalar field’s energy were also stretched outside the horizon during inflation, creating a scale-invariant spectrum of tiny Gaussian adiabatic fluctuations. Then the scalar field decayed and the Universe was filled with radiation and matter (most of it dark). Initially over-dense regions accreted more material and eventually collapsed to form stars, and then galaxies, which merged to form even larger galaxies, and then clusters. Eventually, the Universe’s expansion diluted the radiation and matter densities, revealing the presence of dark energy, which triggered a second era of accelerated expansion.

This brief history of the Universe is the foundation of what could be called “standard” cosmology: during inflation [1, 2, 3], quantum fluctuations in the inflaton field created the primordial density fluctuations [4, 5, 6, 7], which then grew to form astrophysical structures. In the canonical scenario, the expansion history of the Universe, including the current acceleration [8, 9], is attributable to the Universe’s primary components of dark matter [10] and dark energy [11], in accordance with the predictions of general relativity [12]. On top of this foundation, a more extensive theory of structure formation has been constructed: Extended Press-Schechter (EPS) theory, in which the Gaussian statistics of the density field are used to derive a number density function for dark matter haloes [13] and a halo merger rate [14].

The standard inflationary theory of cosmology has enjoyed several successes, including predictions of primordial light element production [15, 16], the observed flatness of the Universe [17], accurate descriptions of the temperature anisotropies in the cosmic microwave background [18], and observations of large-scale structure [19, 20]. Nevertheless, it is imperative that we consider alternatives to the standard cosmology. What if EPS theory does not give an accurate description of halo mergers? What if general relativity is an incomplete description of gravity? What if single-field inflation is not the origin of the primordial power spectrum? How could we recognize flaws in these fundamental cosmological theories, and how do we test their alternatives? These are the questions

that will be addressed in this thesis. We will begin here by presenting reasons to doubt the infallibility of EPS merger theory, general relativity, and single-field inflation, and in subsequent chapters, we will probe alternatives to these theories and examine their observational consequences.

The Press-Schechter halo mass function follows from the assumption that any region in which the mean density exceeds the critical threshold for spherical collapse is a halo with the mass contained in that region [13]. As the density field evolves, larger and larger regions will meet the criteria for collapse. Consequently, a given point in the density field may be in a halo of mass  $M_1$  at time  $t_1$  and then be in a halo of mass  $M_f > M_1$  at some later time  $t_2$ . Using the Gaussian properties of the density field, Lacey and Cole [14] derived an expression for the probability that a halo of mass  $M_1$  will be contained within a halo of mass  $M_f$  after a given time interval. When multiplied by the Press-Schechter halo number density for halos of mass  $M_1$ , this probability gives the EPS halo merger rate for haloes with masses  $M_1$  and  $M_2 = M_f - M_1$ . This merger rate has been applied extensively to several topics in structure formation, including galaxy formation, galactic substructure, halo density distributions, active-galactic-nuclei theory, supermassive- black-hole mergers, and the first stars (see Ref. [21] and references therein). Despite its broad application to so many aspects of cosmology, the halo merger rate provided by EPS theory is terribly flawed. First, it fails to preserve the Press-Schechter halo mass function from which it is derived. Worse, it provides two different merger rates for the same pair of halos because the expression for the halo merger rate is asymmetric in its two mass arguments [21]!

General relativity does not suffer any such unequivocal failings, but there are at least two reasons to question its veracity: dark matter and dark energy. Dark matter and dark energy are necessary ingredients of standard cosmology because general relativity cannot explain the velocity dispersions of clusters [22], the flat rotation curves of galaxies [23, 24], and the acceleration of the cosmic expansion [8, 9] given only the presence of luminous matter and radiation. Both dark matter and dark energy therefore bear a rather uncomfortable resemblance to Vulcan, the nonexistent planet that astronomers invoked in the late nineteenth century to explain an observed discrepancy between Mercury’s orbit and the predictions of Newtonian gravity [25]. Alternatives to general relativity have been proposed that would eliminate the need for dark matter [26, 27], but the observed separation of the luminous matter from the gravitational potential wells in the “bullet cluster” has challenged these theories [28]. Modifications to general relativity that hope to explain late-time cosmic acceleration without dark energy have also been proposed [29, 30, 31, 32, 33, 34, 35, 36, 37]. Furthermore, general relativity still has a serious strike against it even if the Universe does contain dark matter and dark energy: it is incompatible with quantum theory and therefore cannot be the ultimate theory of gravity.

Predating general relativity, the assumption that the Universe is homogenous and isotropic has long been a basic tenet of cosmology. With the addition of inflation to the standard cosmological

history, however, cosmic homogeneity and isotropy were promoted from tenets to predictions. Inflation effectively erases any initial inhomogeneity in the Universe, and inflationary expansion quickly eliminates any deviations from isotropy [38]. Consequently, recent indications that the Universe may not be homogeneous and isotropic [39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52] are difficult to reconcile with inflationary cosmology. Furthermore, inflation predicts that the primordial density fluctuations should be nearly perfectly Gaussian [53], and there have been reports of significant deviations from Gaussianity in the primordial power spectrum [54, 55, 56, 57, 58, 59, 60], although other analyses have yielded results that are consistent with inflation [18, 61, 62]. Finally, Kolb and Turner wrote in 1994 that “inflation remains a very attractive paradigm in search of a compelling model” [63]. Unfortunately, this description of inflation is equally apt today; the identity of the inflaton is a mystery and recent observations of the primordial power spectrum rule out some of the more attractive models for the inflaton potential [64].

In this thesis, we will examine alternatives to EPS merger theory, general relativity, and single-field inflation. The remainder of this chapter contains a summary of the thesis, which is primarily composed of six previously published articles [65, 66, 67, 68, 69, 70] and one paper that is currently in preparation [71]. We begin in Chapter 2 with an examination of the EPS predictions for the merger rate of supermassive black holes, and we show how using an alternate halo merger rate changes the predicted event rate for the Laser Interferometer Space Antenna (LISA). The next two chapters are devoted to alternative theories of gravity. We consider one class of modifications to general relativity that is capable of explaining cosmic acceleration in Chapter 3, and we show that Solar System gravitational tests severely limit these modifications. In Chapter 4, we study how Chern-Simons gravity, a modification of general relativity that is required by some classes of string theory, alters the orbits of Earth’s satellites, and we derive the first constraint on the Chern-Simons coupling parameter. Finally, Chapters 5 and 6 consider how an alternative to single-field inflation can create an inhomogeneous and anisotropic universe. The observational consequences of a density gradient across the Universe are analyzed in Chapter 5. The resulting constraints on superhorizon perturbations are applied in Chapter 6, where we show how a superhorizon fluctuation in a secondary inflationary field can explain a puzzling asymmetry that has been observed in the cosmic microwave background.

## 1.2 Supermassive black hole merger rates: uncertainties from halo merger theory

One particularly interesting application of EPS merger theory is the calculation of the merger rate of the supermassive black holes that lurk in the center of haloes. An accurate prediction of the supermassive-black-hole merger rate is essential because such events should emit gravitational waves

that are detectable by the planned LISA satellite. The expected event rate for LISA, and all hopes of using that event rate to learn about structure formation, are therefore dependent on a thorough understanding of the halo merger rate, which EPS merger theory cannot provide.

In Chapter 2, we show that the *two* EPS predictions for the LISA event rate from supermassive black hole mergers are nearly equal because mergers between haloes of similar masses dominate the event rate. An alternate merger rate may be obtained by inverting the Smoluchowski coagulation equation to find the merger rate that preserves the Press–Schechter halo abundance, but, unfortunately, these rates were initially only available for power-law power spectra [21]. However, since a limited range of halo masses dominate the supermassive-black-hole merger rate, it is possible to find a power-law power spectrum that accurately approximates the EPS prediction for the LISA event rate. In Chapter 2, we use this power-law power spectrum to compare the LISA event rates derived from EPS merger theory to those derived from the merger rates obtained by inverting the coagulation equation. We find that the LISA event rate from supermassive black hole mergers derived from EPS theory is thirty percent higher than the prediction of the alternate merger theory.

### 1.3 Solar system tests of $f(R)$ gravity

One simple way to modify general relativity is to change the action from which the gravitational field equations are derived. It is possible to explain cosmic acceleration by making the action a nonlinear function of the Ricci scalar  $R$ ; such modifications of general relativity are called  $f(R)$  gravity theories. General relativity is a sensitive creature, however; it is very difficult to modify its behavior on cosmological scales without changing gravitational effects everywhere. Shortly after  $f(R)$  gravity theories were proposed as an alternative to dark energy [33], they were shown to be equivalent to scalar-tensor gravity theories that are incompatible with Solar System gravitational tests [72]. The viability of  $f(R)$  gravity theories was subsequently questioned [73, 74, 75, 76, 77], largely because the Schwarzschild-de Sitter metric, which passes all Solar System gravitational tests [78], is a solution to the vacuum field equations in  $f(R)$  gravity.

In Chapter 3, we settle this controversy by solving the  $f(R)$  field equations around a massive body in the weak-field limit. We first consider the original  $f(R)$  gravity theory:  $1/R$  gravity, which has the gravitational action

$$S = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \left( R - \frac{\mu^4}{R} \right), \quad (1.1)$$

where  $\mu$  is a mass parameter that is chosen so that  $\mu^2 \sim H_0^2$  [33]. In this theory, the trace of the gravitational field equation is a differential equation for the Ricci scalar. We use this equation to obtain a solution for  $R$  outside a massive body, and we find that  $R$  is not constant. Therefore,

the Schwarzschild-de Sitter metric cannot be the solution to the  $1/R$  field equations in the Solar System. Instead, we find that the spacetime in the Solar System is profoundly different than the Schwarzschild solution, implying that  $1/R$  gravity is ruled out by Solar System gravitational tests.

We then apply the same methods to any  $f(R)$  gravity theory for which the function  $f(R)$  is analytic around its homogeneous background solution. We find that all  $f(R)$  gravity theories that satisfy a few conditions lead to the same spacetime metric in the Solar System as  $1/R$  gravity and are consequently excluded. We enumerate these conditions in Chapter 3, thus providing a sort of “litmus test” that  $f(R)$  theories must fail if they are to have any chance of eluding constraints from Solar System gravitational tests. Finally, we present several case studies that exemplify how this test should be applied to candidate  $f(R)$  models, and we examine how an  $f(R)$  theory proposed by Ref. [79] successfully evades Solar System tests through nonlinear effects.

*Notes on Collaboration:* The material presented in Chapter 3 was developed in close collaboration with Tristan Smith, Marc Kamionkowski, and Takeshi Chiba, and the author of this thesis is not the first author of one of the two published articles from which this chapter was derived. The mathematical results presented in this chapter were derived independently by the author and Tristan Smith and then compared to check for inconsistencies. The analysis of  $1/R$  gravity was developed from prior work by Marc Kamionkowski. The generalization of the  $1/R$  analysis to other  $f(R)$  theories was initially proposed by Takeshi Chiba, but the procedure, the conclusions, and the prose of the final publication were extensively modified by the author and Tristan Smith.

## 1.4 The effects of Chern-Simons gravity on bodies orbiting the Earth

The addition of a Chern-Simons term to the standard Einstein-Hilbert action of general relativity is a possible consequence of string theory [80, 81]. The Chern-Simons term is a contraction of the Riemann tensor with its dual, and in Chern-Simons gravity [82], the Chern-Simons term is coupled to a scalar field. The resulting gravitational theory is parity-violating. The phenomenology of Chern-Simons gravity therefore demonstrates how parity-violation in the gravitational sector may manifest itself, and studying this phenomenology provides guidance for how we should test gravitational parity.

Unfortunately, Chern-Simons gravity is difficult to distinguish from general relativity because the Chern-Simons field equations reduce to the Einstein field equations in the presence of spherical symmetry. Therefore, the standard Solar System tests of gravitational lensing and time delay do not constrain Chern-Simons gravity. In Chapter 4, we derive the first constraints to the Chern-Simons coupling parameter by considering the orbits of satellites around the Earth. The rotation of the Earth breaks spherical symmetry and generates a gravitomagnetic field. We find that the

Earth’s gravitomagnetic field in Chern-Simons gravity differs from the gravitomagnetic field implied by general relativity. The two theories therefore give different predictions for the precession of an orbit’s line of nodes and for the precession of an orbiting gyroscope’s spin axis. In Chapter 4, we use results from the LAGEOS satellites [83] to constrain the Chern-Simons coupling parameter, and we show how these constraints may be improved by Gravity Probe B [84].

*Notes on Collaboration:* Chapter 4 is adapted from a published article, and the author of this thesis is not the first author of this publication. The results presented in this article were derived in close collaboration with Tristan Smith, with additional support from Robert Caldwell and Marc Kamionkowski. The mathematical equations were often derived independently by the author and Tristan Smith and then compared to check for deviations. The first draft of the article was written by Tristan Smith, but revisions were made and additional material was added by the author of this thesis.

## 1.5 Superhorizon perturbations and the cosmic microwave background

An adiabatic fluctuation with a wavelength that is larger than the cosmological horizon manifests itself as a density gradient across the observable Universe. Since this gradient introduces a special direction in the Universe, a superhorizon fluctuation could be responsible for the deviations from statistical isotropy that have been observed in the cosmic microwave background. For example, in Chapter 6 we will show how a superhorizon fluctuation could generate a hemispherical power asymmetry in the cosmic microwave background. Before a superhorizon perturbation can be proposed as an explanation for other observations, however, one must first consider the direct observational consequences of superhorizon perturbations, which are the topic of Chapter 5.

Superhorizon perturbations induce large-scale temperature anisotropies in the cosmic microwave background (CMB) via the Grishchuk-Zel’dovich effect [85]. In Chapter 5 we analyze the CMB temperature anisotropies generated by a single-mode adiabatic superhorizon perturbation. We show that an adiabatic superhorizon perturbation in a universe with a nonzero cosmological constant does not generate a CMB temperature dipole; the intrinsic dipole created by the fluctuation is cancelled by the Doppler dipole from our motion toward the denser side of the universe. No such cancellation occurs for the higher multipole moments of the CMB temperature anisotropy, and in Chapter 5 we derive constraints to the amplitude and wavelength of a superhorizon potential perturbation from measurements of the CMB quadrupole and octupole.

In anticipation of Chapter 6, we also consider the effect a superhorizon fluctuation in a curvaton field has on the CMB in Chapter 5. The curvaton scenario [86, 87, 88, 89] is an alternative to single-field inflation in which there is a second, subdominant scalar field present during inflation. Although

its contribution to the total energy of the Universe is negligible during inflation, the relative energy of the curvaton field grows after inflation, and when the curvaton decays, the quantum fluctuations in the curvaton field generate adiabatic perturbations. If the curvaton never dominates the energy density of the Universe, a very large-amplitude fluctuation in the curvaton field corresponds to a small fluctuation in the total gravitational potential. Consequently, single-mode superhorizon curvaton fluctuations can create large-scale anisotropies in the CMB that differ from the anisotropies created by a single-mode superhorizon potential fluctuation. For a given curvaton superhorizon fluctuation amplitude, observations of the CMB quadrupole and octupole put an upper limit on the fraction of the Universe's energy that is contained in the curvaton field just prior to its decay. We derive these constraints in Chapter 5 before applying them in Chapter 6.

## 1.6 A hemispherical power asymmetry from inflation

There is a surprising anomaly in the WMAP maps of the CMB; the rms amplitude of temperature fluctuations is significantly larger on one side of the sky than on the other side [42, 43, 44, 45, 46]. Fewer than 1% of simulated realizations of an isotropic Gaussian field contain this much asymmetry between any two hemispheres, and the asymmetry has not been explained by foreground contamination or systematic errors in the WMAP data. The plane dividing the two maximally asymmetric hemispheres is not aligned with the Galactic plane and is closer to the ecliptic plane of the Solar System, with more power in the southern hemisphere. The asymmetry was found by analyzing smoothed maps of the CMB and has therefore only been detected on large angular scales. Figure 1.1 shows the WMAP 5-year map of the CMB divided into two hemispheres by the ecliptic plane. Several pronounced large-scale anisotropies are visible in the southern hemisphere, with no corresponding anisotropies in the northern hemisphere.

In Chapter 6 we explore how this hemispherical power asymmetry may have been created during inflation. We first consider a superhorizon fluctuation in the inflaton field; such a fluctuation would make the background value of the inflaton different on opposite sides of the sky, and the background value of the inflaton determines the amplitude of the primordial fluctuations in single-field inflation. The power of fluctuations from the inflaton field is only weakly dependent on the background value of the inflaton field, however, so a very large-amplitude fluctuation is required to generate the observed asymmetry. The necessary anisotropy in the inflaton field creates an adiabatic fluctuation that violates the constraints derived in Chapter 5, which implies that a superhorizon fluctuation cannot explain the power asymmetry in single-field inflation. We then turn our attention to the curvaton scenario and introduce a superhorizon fluctuation in the curvaton field. The primordial power from curvaton perturbations is more sensitive to the background value of the curvaton field, so a smaller-amplitude superhorizon fluctuation in the curvaton field is sufficient to generate the

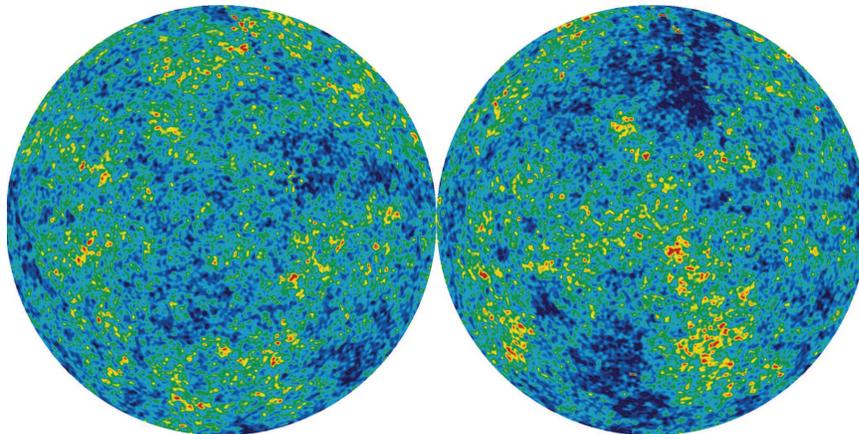


Figure 1.1: The WMAP 5-year Internal Linear Combination Map of the cosmic microwave background, divided into the northern (left) and southern (right) ecliptic hemispheres. The temperature fluctuations shown here range between  $-200$  and  $+200 \mu\text{K}$ . Note that the large-scale fluctuations appear to be more pronounced in the right hemisphere.

*Image: WMAP Science Team.*

observed asymmetry. Moreover, the curvaton contains only a fraction of the Universe's energy density, so the CMB anisotropies induced by the superhorizon curvaton fluctuation are suppressed. We apply the constraints derived in Chapter 5 and find that it is possible to generate the observed hemispherical asymmetry from a superhorizon fluctuation in the curvaton field.

The asymmetry generated by our model will be scale-invariant if the curvaton creates only adiabatic perturbations when it decays. However, a recent analysis of quasar number counts has revealed that the hemispherical power asymmetry observed in the CMB is not present on the small scales that form quasars ( $k \simeq 1.3h - 1.8h \text{ Mpc}^{-1}$ ) [90]. To break the scale-invariance of the asymmetry generated by our model, we consider curvaton models that create a mixture of isocurvature and adiabatic perturbations. In these models, the power asymmetry from a superhorizon fluctuation in the curvaton field will be partly contained in isocurvature perturbations. Since primordial isocurvature perturbations decay once they enter the horizon, the isocurvature perturbations contribute less power on small scales than the adiabatic fluctuations do. If the inflaton, which is unaffected by the superhorizon fluctuation in the curvaton field, also contributes to the adiabatic power spectrum, then the asymmetry will be diluted on small scales due to the suppression of the isocurvature modes. We find that it is possible to explain the observed power asymmetry in the CMB and the isotropy of the quasar population if the curvaton decays after dark matter freeze-out and does not significantly contribute to the dark matter density when it decays. Our model makes several predictions that will be tested by future CMB experiments, and we discuss these tests at the end of Chapter 6.

## Chapter 2

# Supermassive Black Hole Merger Rates: Uncertainties from Halo Merger Theory<sup>1</sup>

## 2.1 Introduction

Structure formation proceeds hierarchically, with small over-dense regions collapsing to form the first dark-matter haloes. These haloes then merge to form larger bound objects. The extended Press–Schechter (EPS) formalism provides a description of “bottom-up” structure formation by combining the Press–Schechter halo mass function [13] with the halo merger rates derived by Lacey and Cole [14]. Since its inception, the EPS theory has been an invaluable tool and has been applied to a wide variety of topics in structure formation (see Ref. [21] and references therein).

Unfortunately, the Lacey–Cole merger-rate formula, which is the cornerstone of EPS merger theory, is mathematically inconsistent [21]. It is possible to obtain *two* equally valid merger rates for the same pair of haloes from the EPS formalism. These two merger rates are nearly equal when the masses of the two haloes differ by less than a factor of one hundred, but they diverge rapidly for mergers between haloes with larger mass ratios. Consequently, any application of EPS merger theory gives two answers, and if the calculation involves mergers between haloes of unequal masses, the discrepancy between these two predictions may be large.

Motivated by the ambiguity in the Lacey–Cole merger rate, Benson, Kamionkowski and Hassani (hereafter BKH) [21] proposed a method to obtain self-consistent halo merger rates. Since haloes are created and destroyed through mergers, the halo merger rate determines the rate of change of the number density of haloes of a given mass. By inverting the Smoluchowski coagulation equation [91], BKH find merger rates that predict the same halo population evolution as the time derivative of the Press–Schechter mass function. In addition to eliminating the flaw that resulted in the double-valued rates in EPS theory, the BKH merger rates by definition preserve the Press–Schechter halo mass distribution when used to evolve a population of haloes. The Lacey–Cole merger rate fails this consistency test as well, and the use of EPS merger trees has been constrained by this inconsistency [92].

There are three limitations to the BKH merger rates. First, they are not uniquely determined because the Smoluchowski equation does not provide sufficient constraints on the merger rate. The

---

<sup>1</sup>This chapter includes material from *Supermassive Black Hole Merger Rates: Uncertainties from Halo Merger Theory*, Adrienne L. Erickcek, Marc Kamionkowski, and Andrew J. Benson; Mon. Not. Roy. Astron. Soc. **371**: 1992–2000 (2006). Reproduced here with permission, copyright (2006) by the Royal Astronomical Society.

BKH merger rate is the smoothest, non-negative function that satisfies the coagulation equation; it exemplifies the properties of a self-consistent merger theory, but it is not a definitive result. Second, the inversion of the Smoluchowski equation is numerically challenging and solutions have been obtained only for power-law density power spectra. Finally, the BKH merger rates are derived from the Press–Schechter halo mass function rather than the mass functions obtained from N-body simulations [93, 94].

In this paper, we explore the possible quantitative consequences of our limited understanding of merger rates for one of the astrophysical applications of merger theory: the merger rate of supermassive black holes. Since supermassive black holes (SMBHs) are believed to lie in the center of all dark-matter haloes above some critical mass, halo mergers and SMBH mergers are intimately related. By considering only halo mergers that would result in a SMBH merger, the EPS merger rates have been used to obtain SMBH merger rates [95, 92, 96, 97, 98, 99].

SMBH mergers are of great interest because they produce a gravitational-wave signal that may be detectable by the Laser Interferometry Space Antenna (LISA), which is scheduled for launch in the upcoming decade. Consequently, EPS merger theory has been used to obtain estimates for the SMBH merger event rate for LISA [95, 92, 96, 97, 98, 99]. In addition to their intrinsic interest as a probe of general relativity, there is hope that LISA’s observations of SMBH mergers will provide a new window into astrophysics at high redshifts. Wyithe and Loeb [96] used EPS merger theory to derive a redshift-dependent mass function for haloes containing supermassive black holes and then used EPS merger theory to predict the LISA event rate that arises from this SMBH population. Since SMBH formation becomes more difficult after reionization due to the limitations on cooling imposed by a hot intergalactic medium, the Wyithe–Loeb SMBH mass function and corresponding LISA event rate are highly sensitive to the redshift of reionization. Ref. [92] used EPS merger trees to demonstrate that LISA observes more SMBH merger events when SMBHs at redshift  $z = 5$  are only found in the most massive haloes as opposed to being randomly distributed among haloes. Ref. [100] also used EPS merger trees to show that higher-mass seed black holes ( $M_{\text{BH}} \sim 10^5 M_{\odot}$  as opposed to  $M_{\text{BH}} \sim 10^2 M_{\odot}$ ) at high redshifts result in significantly higher LISA SMBH-merger event rates. Unfortunately, these ambitions of using LISA SMBH-merger event rates to learn about reionization and SMBH formation rest on the shaky foundation of EPS merger theory.

We first review how the rate of mergers per comoving volume translates to an observed event rate in a  $\Lambda$ CDM universe and how the mass of the halo is related to the mass of the SMBH at its center in Sections 2.2 and 2.3. In Section 2.4, we use the EPS formalism to derive an event rate for LISA. Throughout the calculation, we present the results derived from *both* versions of the Lacey–Cole merger rate. In Section 2.5, we explore the alternative merger-rate formalism proposed by Benson, Kamionkowski, and Hassani (BKH) [21]. Since the BKH merger rates are only available for power-law density power spectra, it is not possible to use them to make a new prediction of the

SMBH merger rate and the corresponding event rate for LISA. Instead, in Section 2.6, we use the event rates for power-law power spectra derived from the EPS and BKH merger theories to gauge how the LISA event rates may be affected by switching merger formalisms. Finally, in Section 2.7, we summarize our results and discuss how these ambiguities in halo merger theory limit our ability to learn about reionization and supermassive-black-hole formation from LISA’s observations.

## 2.2 Cosmological event rates

The merger of two supermassive black holes will produce a gravitational-wave burst. The observed burst event rate depends on the number density and frequency of black-hole mergers: the number of observed gravitational-wave bursts per unit time ( $B$ ) that originate from a shell of comoving radius  $R(z)$  and width  $dR$  is

$$dB = (1+z)^{-1} \mathcal{N}(z) 4\pi R^2 dR, \quad (2.1)$$

where  $\mathcal{N}(z)$  is the SMBH merger rate per comoving volume as a function of redshift. The factor of  $(1+z)^{-1}$  in Eq. (2.1) results from cosmological time dilation. In Eq. (2.1), and throughout this article, we assume a flat  $\Lambda$ CDM universe. Given the relation between comoving distance and redshift,  $dR = [c/H(z)] dz$ , Eq. (2.1) may be converted to a differential event rate per redshift interval,

$$\frac{dB}{dz} = (1+z)^{-1} \left( \frac{4\pi [R(z)]^2 \mathcal{N}(z) c}{H_0 \sqrt{\Omega_M (1+z)^3 + \Omega_\Lambda}} \right), \quad (2.2)$$

where  $\Omega_M$  and  $\Omega_\Lambda$  are the matter and dark-energy densities today in units of the critical density. The comoving distance  $R(z)$  is obtained from

$$R(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_M (1+z')^3 + \Omega_\Lambda}}. \quad (2.3)$$

For an Einstein-de Sitter (EdS) universe, Eq. (2.2) reduces to

$$\frac{dB}{dz} = 4\pi [R(z)]^2 c \mathcal{N}(z) H_0^{-1} (1+z)^{-5/2}, \quad (2.4)$$

which is the differential event rate for an EdS universe derived by Ref. [95].

The observed gravitational-wave burst rate from SMBH mergers is obtained by integrating Eq. (2.2) over the redshifts from which the bursts are detectable. LISA will be able to detect nearly all mergers of two black holes with masses greater than  $10^4 M_\odot$  and less than  $10^8 M_\odot$  up to  $z \lesssim 9$  [95, 98, 99]. Since more massive binary-black-hole systems emit gravitational radiation at lower frequencies and the observed frequency decreases with redshift, very distant ( $z \sim 9$ ) mergers of SMBHs with masses greater than  $10^8 M_\odot$  produce signals below LISA’s frequency window [98, 99].

However, the number density of  $10^8 M_\odot$  haloes is exponentially suppressed at redshifts greater than four, so it is extremely unlikely that two black holes larger than  $10^8 M_\odot$  will merge at redshifts  $z \gtrsim 4$ . Thus, the upper bounds on the relevant redshift and SMBH mass intervals are determined by the population of supermassive black holes and not LISA’s sensitivity.

## 2.3 The relationship between halo mass and black hole mass

The transition from the rate of halo mergers to the rate of detectable SMBH mergers [ $\mathcal{N}(z)$  as defined in Eq. (2.1)] requires a relationship between the mass of a halo and the mass of the SMBH at its center. Since LISA is sensitive to SMBH mergers at high redshifts, this  $M_{\text{BH}} - M_{\text{halo}}$  relation must be applicable to high redshifts as well.

Observations of galaxies out to  $z \sim 3$  reveal a redshift-independent correlation between the mass of the central black hole and the bulge velocity dispersion [101, 102]. A recent compilation of SMBH mass measurements concludes that

$$M_{\text{BH}} = (1.66 \pm 0.32) \times 10^8 \left( \frac{\sigma_c}{200 \text{ km s}^{-1}} \right)^{4.58 \pm 0.52} M_\odot, \quad (2.5)$$

where  $\sigma_c$  is the velocity dispersion normalized to an aperture of size one-eighth the bulge effective radius [103]. The connection between  $\sigma_c$  and halo mass is mediated by the circular velocity  $v_c$ . Using a sample of thirteen spirals, Ref. [104] measured the  $v_c - \sigma_c$  relation,

$$v_c = 3.55_{-1.26}^{+1.95} (\sigma_c / \text{km s}^{-1})^{0.84 \pm 0.09} \text{ km s}^{-1}. \quad (2.6)$$

Combining Eq. (2.5) with this relation reveals that measurements are consistent with a redshift-independent  $M_{\text{BH}} \propto v_c^5$  relation.

Wyithe and Loeb [105] proposed a mechanism for black-hole-mass regulation that would result in a  $M_{\text{BH}} \propto v_c^5$  relation between central-black-hole mass and disc circular velocity for all redshifts. They postulated that a black hole ceases to accrete when the power radiated by the accretion exceeds the binding energy of the host galactic disc divided by the dynamical time of the disc. Assuming that the accretion disc shines at its Eddington luminosity, the black hole stops growing when

$$M_{\text{BH}} = 1.9 \times 10^8 \left( \frac{F_q}{0.07} \right) \left( \frac{v_c}{350 \text{ km s}^{-1}} \right)^5 M_\odot, \quad (2.7)$$

where  $F_q$  is the fraction of the radiated power which is transferred to gas in the disc. Setting  $F_q$  to 0.07 brings Eq. (2.7) into agreement with the observations presented by Ref. [104].

The final step in the determination of a halo–black-hole-mass relation is to connect the circular

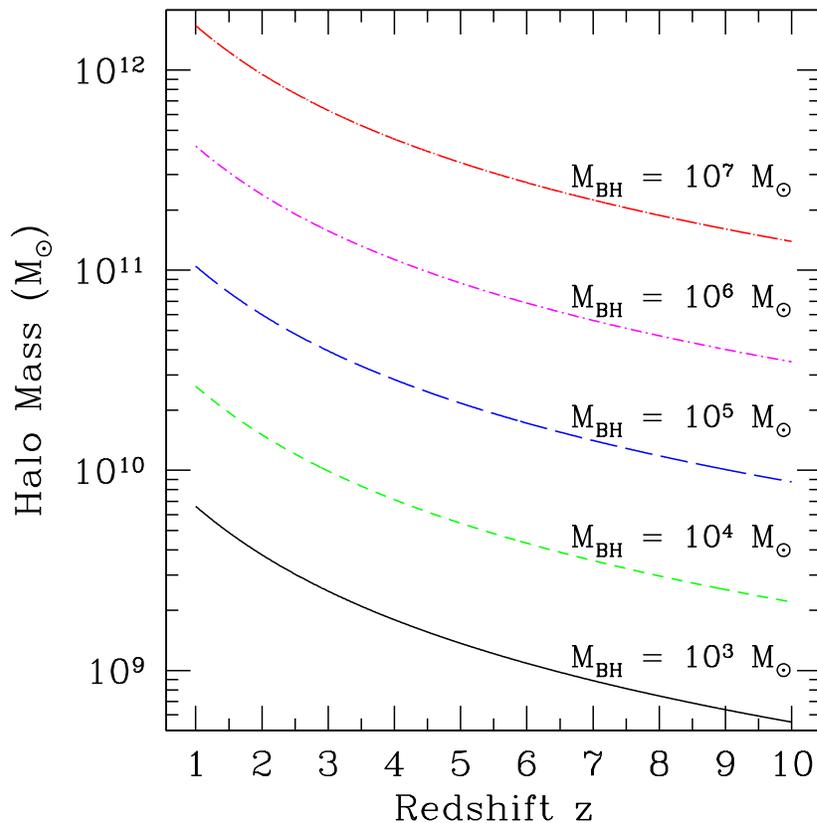


Figure 2.1: The masses of haloes that contain supermassive black holes of mass  $10^3, 10^4, 10^5, 10^6$  and  $10^7 M_\odot$ , according to the  $M_{\text{BH}}-M_{\text{halo}}$  relation proposed by Ref. [105] for a flat  $\Lambda$ CDM universe with  $\Omega_{\text{M}} = 0.27$ . This relation is normalized to fit local observations and assumes that the disc circular velocity equals the virial velocity.

velocity to the halo mass via the virial velocity [106],

$$v_{\text{vir}} = 245 \left( \frac{M_{\text{halo}}}{10^{12} M_\odot} \right)^{1/3} \left( \frac{1+z}{3} \right)^{1/2} \left( \frac{\Omega_{\text{M}}^0}{\Omega_{\text{M}}(z)} \frac{\Delta_{\text{c}}}{18\pi^2} \right)^{1/6} \text{ km s}^{-1}, \quad (2.8)$$

where  $\Omega_{\text{M}}(z)$  is the matter density divided by the critical density at redshift  $z$ ,

$$\Omega_{\text{M}}(z) \equiv \frac{\Omega_{\text{M}}(1+z)^3}{\Omega_{\text{M}}(1+z)^3 + \Omega_{\Lambda}}, \quad (2.9)$$

$\Omega_{\text{M}}^0 \equiv \Omega_{\text{M}}(z=0)$ , and  $\Delta_{\text{c}}$  is the nonlinear over-density at virialization for a spherical top-hat perturbation for a  $\Lambda$ CDM universe:

$$\Delta_{\text{c}} = 18\pi^2 + 82[\Omega_{\text{M}}(z) - 1] - 39[\Omega_{\text{M}}(z) - 1]^2. \quad (2.10)$$

The simplest possible assumption is that the circular velocity of the disc equals the virial velocity

of the halo. This assumption is made by Ref. [105], and we assume that  $v_c = v_{\text{vir}}$  throughout this chapter. However, different relations between  $v_c$  and  $v_{\text{vir}}$  have been proposed and can significantly impact the final  $M_{\text{BH}}-M_{\text{halo}}$  relation (see Ref. [104]).

Assuming that  $v_c = v_{\text{vir}}$ , the halo mass then becomes a redshift-dependent function of the mass of the central black hole:

$$\frac{M_{\text{halo}}}{10^{12} M_{\odot}} = 10.5 \left( \frac{\Omega_M^0}{\Omega_M(z)} \frac{\Delta_c}{18\pi^2} \right)^{-\frac{1}{2}} (1+z)^{-\frac{3}{2}} \left( \frac{M_{\text{BH}}}{10^8 M_{\odot}} \right)^{\frac{3}{5}}. \quad (2.11)$$

Figure 2.1 shows the masses of haloes that contain supermassive black holes of several masses. For a given black-hole mass, the corresponding halo mass decreases with increasing redshift due to the larger value for the virial velocity at earlier times. Citing the fact that the largest haloes observed at low redshifts appear to contain galaxy clusters with no central black holes, Ref. [105] argues that supermassive-black-hole growth was complete by  $z \sim 1$  and that local SMBH masses reflect the limiting values at that redshift. Consequently, when determining the mass of a halo that contains a black hole of a given mass, we use the  $z = 1$  value of Eq. (2.11) for all redshifts less than one.

Some calculations of the LISA SMBH-merger event rate impose a minimum halo virial temperature instead of a minimum black-hole mass when calculating the lower mass bound on haloes that contribute to the SMBH merger rate [96, 99]. This constraint reflects the fact that supermassive black holes only form when the gas within dark-matter haloes can cool. However, the relation between virial temperature and virial mass [106] may be used to eliminate the halo mass in Eq. (2.11) in favor of the virial temperature. The redshift-dependent terms cancel, leaving a redshift-independent relation between black-hole mass and halo virial temperature:

$$M_{\text{BH}} = (267 M_{\odot}) h^{-5/3} \left( \frac{T_{\text{vir}}}{1.98 \times 10^4 \text{ K}} \right)^{5/2}. \quad (2.12)$$

Therefore, defining  $M_{\text{min}}$  by a minimum halo virial temperature is nearly equivalent to defining  $M_{\text{min}}$  by a minimum black-hole mass via Eq. (2.11). For example, requiring that the halo's virial temperature be significantly higher than the temperature of the intergalactic medium,  $T_{\text{vir}} \gtrsim 10^5 \text{ K}$  [96], corresponds to imposing a minimum black-hole mass of  $2.6 \times 10^4 M_{\odot}$ . The only discrepancy occurs when  $z < 1$  because we assume that the  $M_{\text{BH}}-M_{\text{halo}}$  relation is fixed for redshifts less than one, while  $T_{\text{vir}}$  is still redshift dependent. However we shall see that nearly all SMBH mergers occur at redshifts greater than one, so this difference is negligible.

## 2.4 LISA event rates from EPS merger theory

### 2.4.1 Review of EPS merger theory

The first pillar of EPS merger theory is the Press–Schechter halo mass function [13], which gives the number of haloes with masses between  $M$  and  $M + dM$  per comoving volume:

$$\frac{dn_{\text{halo}}}{d \ln M} = \sqrt{\frac{2}{\pi}} \frac{\rho_0}{M} \left( \left| \frac{d \ln \sigma}{d \ln M} \right|_M \right) \frac{\delta_{\text{coll}}}{\sigma(M, z)} \exp \left[ \frac{-\delta_{\text{coll}}^2}{2\sigma^2(M, z)} \right], \quad (2.13)$$

where  $\rho_0$  is the background matter density today,  $\delta_{\text{coll}}$  is the critical over-density for collapse in the spherical-collapse model, and  $\sigma(M, z)$  is the root variance of the linear density field at redshift  $z$  in spheres containing mass  $M$  on average. In a  $\Lambda$ CDM universe,  $\delta_{\text{coll}}$  deviates slightly from its Einstein-de Sitter value of  $\sim 1.686$  when the cosmological constant begins to dominate the energy density of the Universe [107, 108]. In this work, the fitting function obtained by Ref. [107] was used to approximate  $\delta_{\text{coll}}$ :

$$\delta_{\text{coll}} \simeq \frac{3(12\pi)^{2/3}}{20} [1 + 0.0123 \log_{10} \Omega_M(z)], \quad (2.14)$$

where  $\Omega_M(z)$  is given by Eq. (2.9).

The present-day variance  $\sigma^2(M)$  is obtained by convolving the density power spectrum  $P(k)$  with a top-hat filter function of radius  $R = [3M/(4\pi\rho_0)]^{1/3}$ . The power spectrum  $P(k)$  is the product of the primordial power-law  $k^n$  and the square of the transfer function  $T(k)$ . Ref. [109] provides a smooth and simple form of the transfer function that accurately models the effect of baryon-induced suppression while neglecting the small baryon acoustic oscillations, and we use this transfer function to calculate  $\sigma(M)$ . The primordial power spectrum is assumed to be scale-invariant with  $n = 1$ . The redshift-dependent root variance  $\sigma(M, z)$  is proportional to the linear growth function  $D(z)$  for a flat  $\Lambda$ CDM universe with scale factor  $a$  [110, 111]:

$$\begin{aligned} D(z) &\propto \frac{1}{a} \frac{da}{dt} \int_0^{(1+z)^{-1}} \left( \frac{da}{dt} \right)^{-3} da, \\ D(z) &\propto H(z) \int_0^{(1+z)^{-1}} \left( \frac{\Omega_M}{a} + \Omega_\Lambda a^2 \right)^{-3/2} da. \end{aligned} \quad (2.15)$$

The function  $\sigma(M, z)$  is normalized so that it agrees with the observed value of  $\sigma_8$  when  $z = 0$  and  $M$  corresponds to a sphere of radius  $8h^{-1}$  Mpc.

A critical feature of the Press–Schechter mass function is the onset of exponential decay with increasing mass. The exponential factor in Eq. (2.13) dominates when  $\sigma(M, z) \lesssim \delta_{\text{coll}}(z)$ . We define the function  $M_*(z)$  such that  $\sigma(M_*, z) \equiv \delta_{\text{coll}}(z)$ . Since  $\sigma(M)$  is a monotonically decreasing function of mass, the Press–Schechter number density will be exponentially suppressed for all halo masses greater than  $M_*$ . The redshift dependence of  $M_*$  is primarily determined by  $\sigma(M, z)$ , which

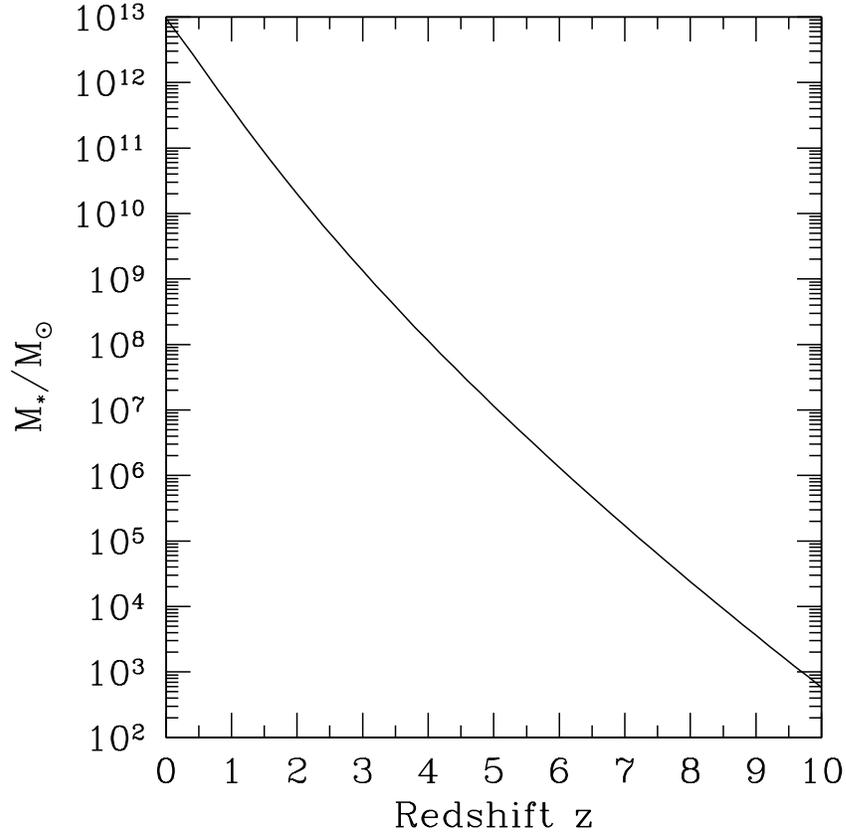


Figure 2.2: The critical mass  $M_*(z)$  for a flat  $\Lambda$ CDM universe with  $\Omega_M = 0.27$ ,  $h = 0.72$  and  $\sigma_8 = 0.9$ . The Press–Schechter number density of haloes larger than  $M_*$  is exponentially suppressed.

decreases with redshift, since  $\delta_{\text{coll}}(z)$  is nearly constant. As  $\sigma$  decreases with redshift,  $M_*$  must also decrease to keep  $\sigma(M_*, z)$  equal to  $\delta_{\text{coll}}(z)$ . Figure 2.2 shows  $M_*(z)$  for a  $\Lambda$ CDM universe.

The second pillar of EPS merger theory is the merger probability function derived by Lacey and Cole [14], which gives the probability that a halo of mass  $M_1$  will become a halo of mass  $M_f \equiv M_1 + M_2$  per unit time, per unit acquired mass:

$$\begin{aligned} \frac{d^2 p}{dt dM_2} &= \frac{1}{M_f} \sqrt{\frac{2}{\pi}} \left| \frac{\dot{\delta}_{\text{coll}}}{\delta_{\text{coll}}} - \frac{\dot{D}(z)}{D(z)} \right| \left( \left| \frac{d \ln \sigma}{d \ln M} \right|_{M_f} \right) \frac{\delta_{\text{coll}}}{\sigma(M_f, z)} \left( 1 - \frac{\sigma^2(M_f, z)}{\sigma^2(M_1, z)} \right)^{-3/2} \\ &\times \exp \left[ \frac{-\delta_{\text{coll}}^2}{2} \left( \frac{1}{\sigma^2(M_f, z)} - \frac{1}{\sigma^2(M_1, z)} \right) \right]. \end{aligned} \quad (2.16)$$

In this expression,  $D(z)$  is the linear growth function defined in Eq. (2.15). Note that for an EdS universe,  $\delta_{\text{coll}}$  is constant, and the linear growth function is simply the scale factor. In this case,

$$\begin{aligned} \left| \frac{\dot{\delta}_{\text{coll}}}{\delta_{\text{coll}}} - \frac{\dot{D}(z)}{D(z)} \right| &= \frac{\dot{a}}{a} \\ &= H_0(1+z)^{-3/2}. \end{aligned}$$

Making this substitution brings Eq. (2.16) into the form provided by Ref. [95].

Equation (2.16) is usually interpreted as the differential probability that a given halo of mass  $M_1$  will merge with a halo of mass between  $M_2$  and  $M_2 + dM_2$  per unit time, per increment mass change. Note that this quantity is defined to be asymmetric in  $M_1$  and  $M_2$  because it starts with a halo of  $M_1$  and asks if that particular halo is likely to encounter and merge with a halo of mass  $M_2$ . Thus Eq. (2.16) already includes information about the abundance of haloes of mass  $M_2$ , but not the abundance of haloes of mass  $M_1$ . Following BKH, it is revealing to examine a different quantity, which does not differentiate between the two merging haloes: the rate of mergers between haloes of masses  $M_1$  and  $M_2$  per comoving volume. This merger rate may be obtained by multiplying the Lacey–Cole probability that a specific  $M_1$  halo will merge with an  $M_2$  halo by the Press–Schechter number density of haloes of mass  $M_1$ :

$$\begin{aligned} R(M_1, M_2, t) &\equiv \frac{\text{Number of } M_1 + M_2 \text{ Mergers}}{dt \, d(\text{Comoving Volume})}, \\ &= \left( \frac{dn(M_1; t)}{dM_1} \right) \left( \frac{d^2p}{dt \, dM_2} \right) dM_1 \, dM_2. \end{aligned} \quad (2.17)$$

The EPS self-inconsistency documented by BKH manifests itself here. Although  $R(M_1, M_2, t)$  must be symmetric in its mass arguments by definition, Eq. (2.17) is not symmetric under exchange of  $M_1$  and  $M_2$ .

The mass asymmetry of EPS merger theory becomes most transparent when one defines a new function: the merger kernel. From its definition, it is apparent that  $R(M_1, M_2, t)$  should be proportional to the number densities of both haloes involved in the merger. Extracting this dependence defines the merger kernel  $Q(M_1, M_2, t)$ :

$$R(M_1, M_2, t) \equiv \left( \frac{dn(M_1; t)}{dM_1} \right) \left( \frac{dn(M_2; t)}{dM_2} \right) Q(M_1, M_2, t) \, dM_1 \, dM_2. \quad (2.18)$$

The number density of mergers expressed by  $R(M_1, M_2, t)$  is symmetric under the exchange of the two merging haloes if and only if the merger kernel  $Q(M_1, M_2, t)$  is also symmetric in its mass arguments. Comparison with the expression for  $R$  in Eq. (2.17) reveals the formula for the Lacey–Cole merger kernel:

$$\begin{aligned} Q(M_1, M_2, z) &= \frac{d^2p}{dt \, dM_2} \left( \frac{dn_{\text{halo}}}{dM_2} \right)^{-1} \\ &= \frac{1}{M_f \rho_0 \sigma_f} \left| \frac{\dot{\delta}_{\text{coll}}}{\delta_{\text{coll}}} - \frac{\dot{D}(z)}{D(z)} \right| \left( \left| \frac{d \ln \sigma}{d \ln M} \right|_{M_f} \right) \exp \left[ \frac{-1}{2} \left( \frac{\delta_{\text{coll}}^2}{\sigma_f^2} - \frac{\delta_{\text{coll}}^2}{\sigma_1^2} - \frac{\delta_{\text{coll}}^2}{\sigma_2^2} \right) \right] \\ &\quad \times \frac{M_2^2 \sigma(M_2)}{\left( 1 - \frac{\sigma^2(M_f)}{\sigma^2(M_1)} \right)^{3/2}} \left| \frac{d \ln \sigma}{d \ln M} \right|_{M_2}^{-1}, \end{aligned} \quad (2.19)$$

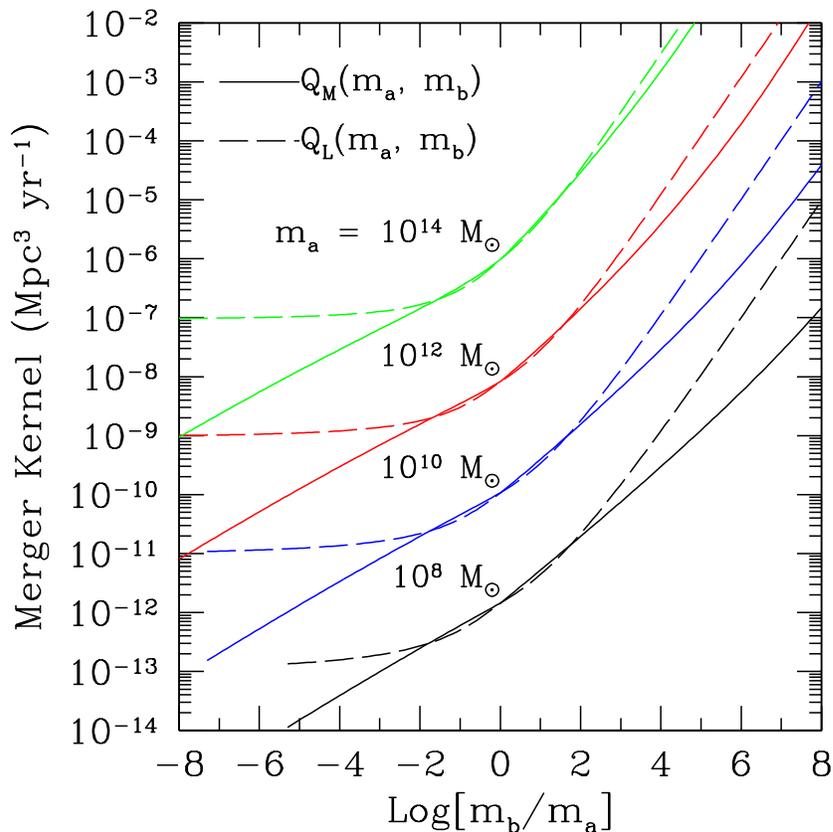


Figure 2.3: The two EPS merger kernels for  $z = 0$ . Here,  $Q_M$  is the Lacey–Cole merger kernel with the more massive halo as the first argument [as defined in Eq. (2.19)], and  $Q_L$  is the same kernel with the less massive halo as the first argument. Results are shown for a flat  $\Lambda$ CDM universe with  $\Omega_M = 0.27$ ,  $h = 0.72$  and  $\sigma_8 = 0.9$ .

where we have employed the shortened notation  $\sigma_i \equiv \sigma(M_i, z)$ . The last line of Eq. (2.19) is the source of the mass asymmetry in the Lacey–Cole merger formalism.

In effect, EPS merger theory includes two distinct merger kernels, depending on the order of the mass arguments. Thus, we define two mass-symmetric merger kernels, which are differentiated by whether the more massive halo or the less massive halo is the first mass argument:

$$Q_M(M_1, M_2) \equiv \begin{cases} Q(M_1, M_2) & \text{if } M_1 \geq M_2, \\ Q(M_2, M_1) & \text{if } M_1 < M_2, \end{cases} \quad (2.20)$$

$$Q_L(M_1, M_2) \equiv \begin{cases} Q(M_2, M_1) & \text{if } M_1 \geq M_2, \\ Q(M_1, M_2) & \text{if } M_1 < M_2, \end{cases} \quad (2.21)$$

where  $Q(M_1, M_2)$  is given by Eq. (2.19). Figure 2.3 illustrates the differences in the merger kernels  $Q_M$  and  $Q_L$ . When the masses of the two merging haloes are similar,  $Q_M$  is slightly larger than  $Q_L$ , but  $Q_L$  becomes much larger than  $Q_M$  for mergers between haloes of very different masses.

It is important to note that  $Q_M(M_1, M_2)$  and  $Q_L(M_1, M_2)$  are not smooth functions of the halo masses; the derivatives of  $Q_M$  and  $Q_L$  with respect to halo mass are discontinuous at the point  $M_1 = M_2$ . Consequently, the mass-asymmetry flaw in EPS merger theory cannot be corrected by specifying whether the first or second halo is larger. Neither  $Q_M(M_1, M_2)$  nor  $Q_L(M_1, M_2)$  are viable candidates for the true halo merger kernel. They are useful because they expose the ambiguities hidden in applications of EPS merger theory.

In order to avoid double counting mergers when calculating a merger rate, it is common to restrict one mass argument to be larger than the other. Using the standard expression for the Lacey–Cole merger probability function, as given by Eq. (2.16), in such calculations is equivalent to using  $Q_M(M_1, M_2)$  or  $Q_L(M_1, M_2)$ . Specifically, Ref. [95] effectively used  $Q_L$  to predict an event rate for LISA, while Refs. [96, 99] effectively used  $Q_M$ . Using the other version of the EPS merger kernel in either of these calculations would have yielded different results, as we show in Section 2.4.2. More generally, any application of the Lacey–Cole merger probability function uses some mixture of  $Q_M$  and  $Q_L$ , and changing the mixture will change the result of the calculation.

### 2.4.2 LISA event rates from EPS theory

The rate of halo mergers per unit volume can be obtained from EPS merger theory; this quantity is simply  $R(M_1, M_2, z)$  as defined in Eq. (2.17). The transition from the rate of halo mergers to the rate of detectable SMBH mergers [ $\mathcal{N}(z)$  as defined in Eq. (2.1)] requires a relationship between the mass of a halo and the mass of the SMBH at its center, which we derived in Section 2.3. The simplest and most general approach is to assume that all haloes above a given mass  $M_{\min}$  contain a black hole that will produce a detectable gravitational-wave signal when it merges with a black hole of equal or greater mass and to keep  $M_{\min}$  as a free parameter. With this assumption, the rate of mergers between haloes with masses larger than  $M_{\min}$  is directly related to the rate of detectable SMBH mergers.

The resulting rate of SMBH mergers per comoving volume follows from the rate of halo mergers per comoving volume given in Eq. (2.17):

$$\mathcal{N}(z) \equiv \frac{1}{2} \int_{M_{\min}}^{\infty} \int_{M_{\min}}^{\infty} \left( \frac{dn(M_1, z)}{dM} \right) \left( \frac{dn(M_2, z)}{dM} \right) Q(M_1, M_2, z) dM_1 dM_2, \quad (2.22)$$

where the factor of 1/2 accounts for the double counting of mergers. Some calculations, e.g. [99], only include mergers between haloes with mass ratios less than three and so integrate  $M_2$  from  $M_1/3$  to  $3M_1$ . This restriction is motivated by dynamical-friction calculations that indicate that when a halo merges with a halo less than a third of its size, it takes longer than a Hubble time for their central black holes to merge [112]. However, recent numerical simulations indicate that this restriction may be too strict; when gas dynamics are included, SMBHs with host-galaxy-mass ratios

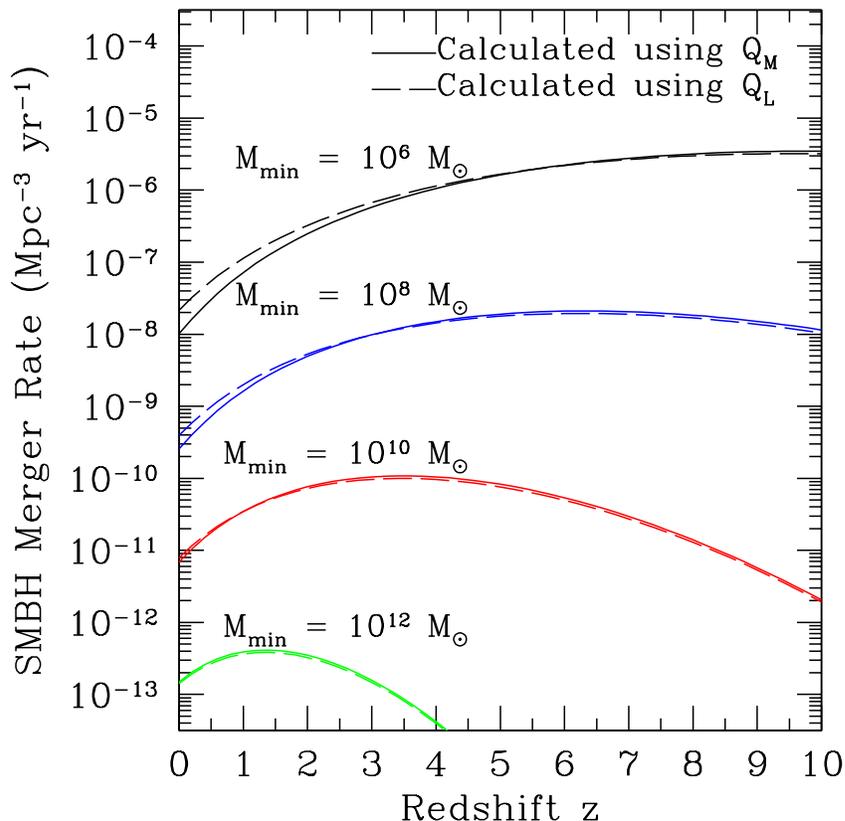


Figure 2.4: The rate  $\mathcal{N}$  of SMBH mergers per comoving volume, as defined in Eq. (2.22). The quantity  $M_{\min}$  is the minimum mass of a halo that contains a SMBH capable of producing a detectable gravitational-wave signal when it merges with a black hole of greater or equal mass. The solid (dashed) curves show the results when the first argument of the Lacey–Cole merger kernel is the more (less) massive halo. Results are shown for a flat  $\Lambda$ CDM universe with  $\Omega_M = 0.27$ ,  $h = 0.72$  and  $\sigma_8 = 0.9$ .

greater than three merge within a Hubble time [113]. We do not impose this restriction, so our event rates are upper bounds arising from the assumption that every halo merger in which both haloes contain a SMBH results in a SMBH merger.

Figure 2.4 shows the SMBH merger densities calculated from Eq. (2.22) for several values of  $M_{\min}$ . For each value  $M_{\min}$ , there are two versions of  $\mathcal{N}$  corresponding to the two versions of the EPS merger kernels defined in Eqs. (2.20) and (2.21). Clearly,  $\mathcal{N}(z)$  is strongly dependent on the choice of  $M_{\min}$ . As  $M_{\min}$  is increased, fewer halo mergers are included in the calculation of  $\mathcal{N}$ , and its value decreases accordingly. When  $M_{\min}$  is larger than  $M_*(z')$ , the paucity of larger haloes at redshifts higher than  $z'$  leads to a rapid falloff of  $\mathcal{N}(z)$  as  $z$  increases beyond  $z'$ .

As previously mentioned, estimates indicate that LISA should observe mergers between two SMBHs with masses greater than  $10^4 M_\odot$  out to redshifts of at least five [95]. Therefore, we generally use  $M_{\text{BH}} = 10^4 M_\odot$  in Eq. (2.11) to determine  $M_{\min}$ . As shown in Figure 2.1, this choice implies

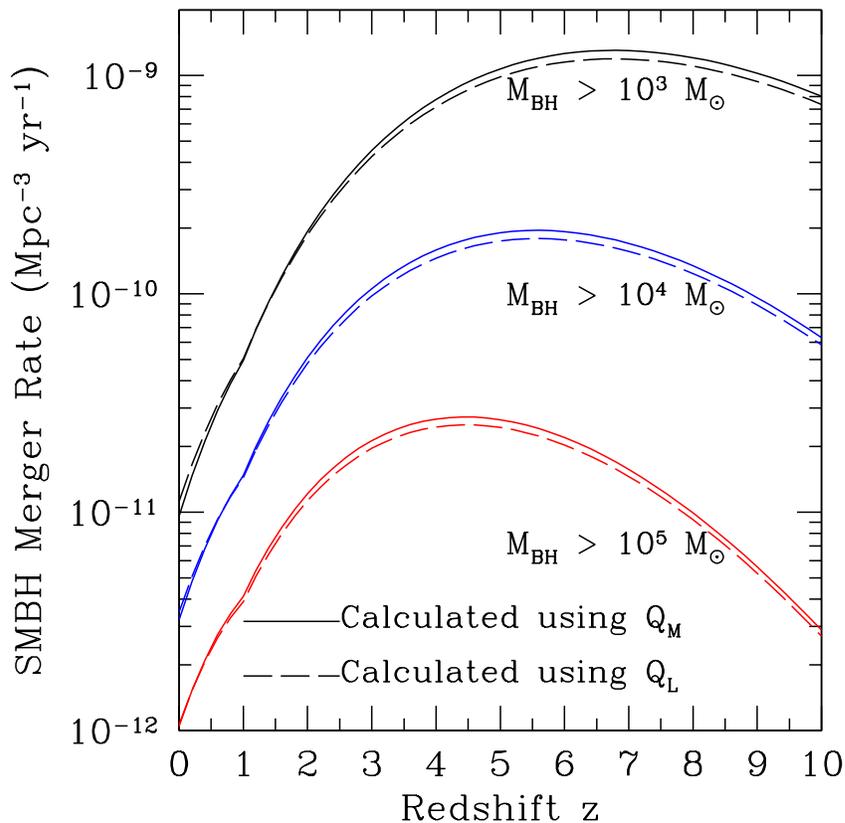


Figure 2.5: The rate of SMBH mergers per comoving volume where both merging black holes have a mass greater than  $10^3 M_\odot$ ,  $10^4 M_\odot$  and  $10^5 M_\odot$ . The solid (dashed) lines show the results when the first argument of the Lacey–Cole merger kernel is the more (less) massive halo. Results are shown for a flat  $\Lambda$ CDM universe with  $\Omega_M = 0.27$ ,  $h = 0.72$ , and  $\sigma_8 = 0.9$ .

that  $M_{\min} = 2.6 \times 10^{10} M_\odot$  for redshifts less than unity, with  $M_{\min}$  decreasing at higher redshifts. The corresponding rates of SMBH mergers per comoving volume are shown in Figure 2.5, as well as the rates which correspond to different choices for the minimum mass of a SMBH. Once again, both versions of  $\mathcal{N}$  are shown to illustrate the difference between the two Lacey–Cole merger kernels. The crimp in  $\mathcal{N}(z)$  at  $z = 1$  reflects the transition from a constant  $M_{\min}$  (evaluated at  $z = 1$ ) to the redshift-dependent form given by Eq. (2.11).

Once the rate  $\mathcal{N}(z)$  of SMBH mergers per volume is known, Eq. (2.2) may be integrated over redshift to obtain an event rate for LISA,

$$B = \int_0^{z_{\max}} (1+z)^{-1} \left( \frac{4\pi [R(z)]^2 \mathcal{N}(z) c}{H_0 \sqrt{\Omega_M (1+z)^3 + \Omega_\Lambda}} \right) dz. \quad (2.23)$$

Here,  $z_{\max}$  is the redshift of the most distant detectable merger. Figure 2.6 shows the LISA event rate for  $z_{\max}$  equal to 2, 4, 6, 8, and 10 as a function of the minimum halo mass that contains a

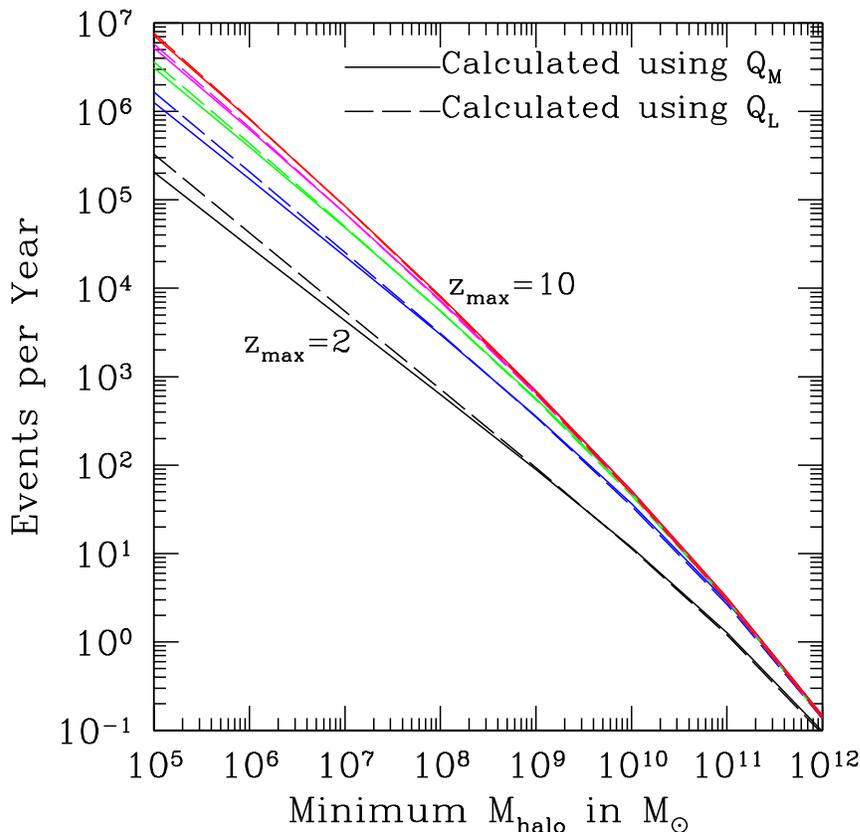


Figure 2.6: The gravitational-wave event rate from SMBH mergers as a function of the minimum halo mass that contains a SMBH large enough to produce a detectable signal when it merges. Mergers at redshifts up to  $z_{\max}$  were included in this rate, and the five pairs of lines correspond to  $z_{\max} = 2, 4, 6, 8, 10$ . The solid (dashed) lines show the results when the first argument of the Lacey–Cole merger kernel is the more (less) massive halo. Results are shown for a flat  $\Lambda$ CDM universe with  $\Omega_M = 0.27$ ,  $h = 0.72$ , and  $\sigma_8 = 0.9$ .

black hole large enough to emit an observable signal. These results were obtained using the values for  $\mathcal{N}(z)$  displayed in Figure 2.4, and they share  $\mathcal{N}$ 's strong dependence on the choice of  $M_{\min}$ .

Figure 2.7 shows the event rate as a function of  $z_{\max}$ , where  $M_{\min}$  is the mass of a halo that contains a black hole more massive than  $10^3$ ,  $10^4$ , or  $10^5 M_{\odot}$  as determined by the  $M_{\text{BH}} - M_{\text{halo}}$  relation given by Eq. (2.11). These rates correspond to the  $\mathcal{N}$  results depicted in Figure 2.5. Examination of these results reveals that increasing  $z_{\max}$  beyond  $z_{\max} = 6$  has little effect on the event rate when  $M_{\min}$  is greater than  $10^9 M_{\odot}$ , as is the case when Eq. (2.11) is used to obtain the value of  $M_{\min}$  which corresponds to a minimum black-hole mass of  $10^4 M_{\odot}$ . The leveling of the event rate for  $z_{\max} \gtrsim 6$  indicates that SMBH mergers are very rare at higher redshifts and that the event rate is dominated by mergers that occur at redshifts  $z \lesssim 6$ . Therefore, the upper bound on LISA's sensitivity to larger SMBH mergers at high redshifts will have little effect on the event rate.

The event rates shown in Figures 2.6 and 2.7 differ significantly from those calculated by Ref. [96]

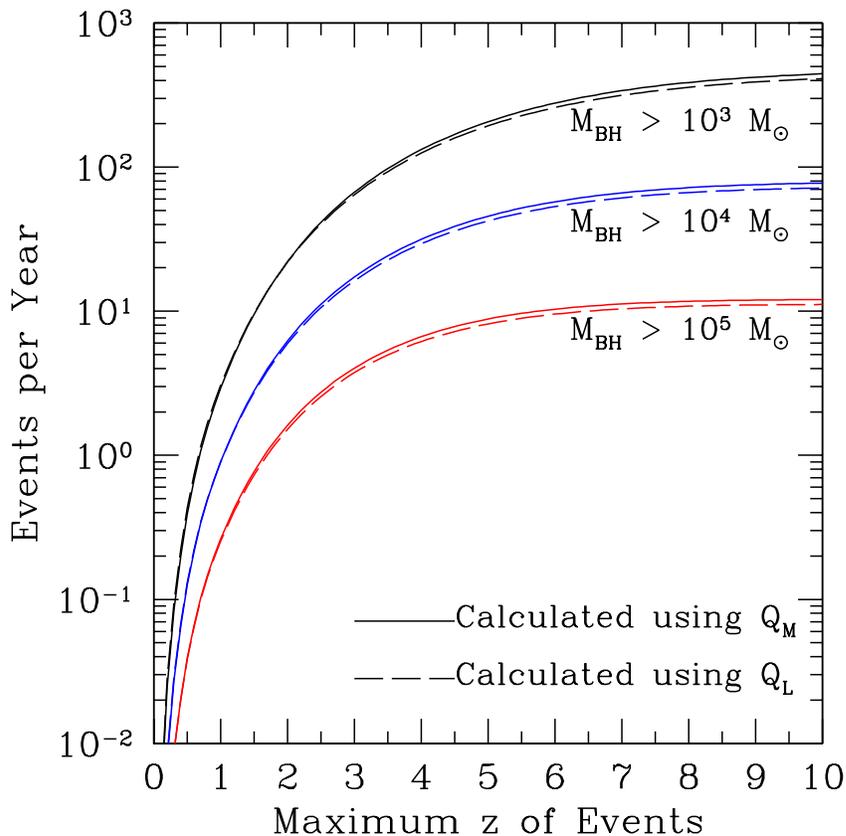


Figure 2.7: The gravitational-wave event rate from SMBH mergers as a function of the maximum redshift of a detectable merger. Only mergers in which both black holes have masses greater than the given lower bound are included. The solid (dashed) lines show the results when the first argument of the Lacey–Cole merger kernel is the more (less) massive halo. Results are shown for a flat  $\Lambda$ CDM universe with  $\Omega_M = 0.27$ ,  $h = 0.72$ , and  $\sigma_8 = 0.9$ .

and Ref. [99].<sup>1</sup> Our event rates are generally much higher than the event rates reported by Ref. [96] because we do not exclude mergers between haloes with mass ratios greater than three from our SMBH merger rate. For instance, given that LISA cannot detect SMBH mergers with  $M_{\text{BH}} \lesssim 10^3$ , Ref. [96] predicts 350 events per year (for reionization at  $z = 7$ ), compared to our 440 events per year. For haloes more massive than the minimum halo mass which corresponds to this minimum black-hole mass,  $M_{\text{halo}} \gtrsim 10^9 M_\odot$ , the mass function for black-hole-containing haloes derived by Ref. [96] is approximately equal to the Press–Schechter mass function, so the difference is primarily attributable to the exclusion of mergers with mass ratios greater than three. The event rates calculated by Ref. [99] are even lower because they do not assume that all haloes contain galaxies. The one case where our event rates are not substantially higher than those derived by Ref. [99] is when the minimum black-hole mass is taken to be very high ( $M_{\text{BH}} \gtrsim 10^5 M_\odot$ ). In that case, the minimum

<sup>1</sup>When we attempted to reproduce the differential event rates calculated by Ref. [95], we found that our rates are roughly a factor of two lower. After extensive review and two independent calculations, we were unable to find any errors in our analysis.

halo mass is so high that nearly all mergers involve haloes of similar masses ( $M_{\text{halo}} \sim 10^{11} M_{\odot}$ ), and the galaxy-occupation fraction derived by Ref. [99] indicates that nearly all haloes of this size contain galaxies for redshifts greater than three, so our event rate of 12 per year is very similar to the result of the more sophisticated treatment of Ref. [99].

Event rates obtained from both versions of the EPS merger kernel are shown in Figures 2.6 and 2.7. The differences between these results reveal the type of mergers that dominate the calculation. For smaller values of  $M_{\text{min}}$ , the event rate is slightly higher when  $Q_{\text{L}}$  is used. As shown by the comparison of  $Q_{\text{M}}$  and  $Q_{\text{L}}$  in Figure 2.3,  $Q_{\text{L}}$  is larger than  $Q_{\text{M}}$  when  $M_{\text{b}}/M_{\text{a}} \gtrsim 100$ . Therefore, a slightly larger event rate from  $Q_{\text{L}}$  indicates that mergers between haloes whose masses differ by more than a factor of a hundred dominate the event rate. However, Figure 2.3 also shows that  $Q_{\text{L}}$  and  $Q_{\text{M}}$  diverge rapidly as the mass difference increases. The difference between the event rates is always less than a factor of three, so mergers between haloes with mass ratios greater than 1000 cannot be making a significant contribution to the event rate. As  $M_{\text{min}}$  is increased, fewer and fewer of these largely unequal-mass mergers are included in the event rate, and the two merger kernels give nearly identical results. At large values of  $M_{\text{min}}$ , the event rate obtained from  $Q_{\text{M}}$  edges slightly ahead, indicating that mergers where the halo masses are within a factor of ten of each other are dominating the sum. Restricting the mass ratio to be less than three, as recommended by Ref. [112], would ensure that  $Q_{\text{M}}$  would always yield a higher event rate than  $Q_{\text{L}}$ .

The differences between the event rates obtained from the two versions of the EPS merger kernel depend on redshift as well as  $M_{\text{min}}$ . For a constant value of  $M_{\text{min}} = 10^5 M_{\odot}$ , the difference between the two versions decreases as the maximum redshift increases, as shown in Figure 2.6. This convergence indicates that the contribution from mergers between haloes of greatly unequal masses to the event rate dwindles as redshift increases. Since the lower bound on halo mass is constant with redshift, a decrease in unequal-mass mergers reflects a decrease in the population of larger haloes. Due to the exponential decline in the number density of haloes greater than  $M_{*}(z)$ , there is an effective upper bound to the integrals in Eq. (2.22), which defines  $\mathcal{N}(z)$ . This upper bound on halo mass follows  $M_{*}$  and decreases with redshift.

The evolution of this effective upper mass bound is clearest when examining the rate of SMBH mergers per comoving volume displayed in Figure 2.5 for  $M_{\text{BH}} > 10^4$ . At low redshifts,  $Q_{\text{L}}$  gives a larger value for  $\mathcal{N}$ , which indicates that mergers with mass ratios greater than 100 compose the majority of events. Recall that  $M_{\text{min}} \sim 10^{10} M_{\odot}$  when  $M_{\text{BH}} > 10^4$ , so the mass of the larger halo in mergers of this type must be greater than  $10^{12} M_{\odot}$ . As redshift increases, haloes of this size become rarer as the effective upper bound on halo mass decreases. The  $Q_{\text{L}}$  result is eventually overtaken by the  $Q_{\text{M}}$  result, indicating that the effective upper bound on halo mass has fallen below  $100M_{\text{min}}$ . Since  $\mathcal{N}$  increases with redshift (at least for  $z \lesssim 5$ ), the relative contribution of mergers between haloes of very different masses to the event rate is determined by their contribution at high redshifts.

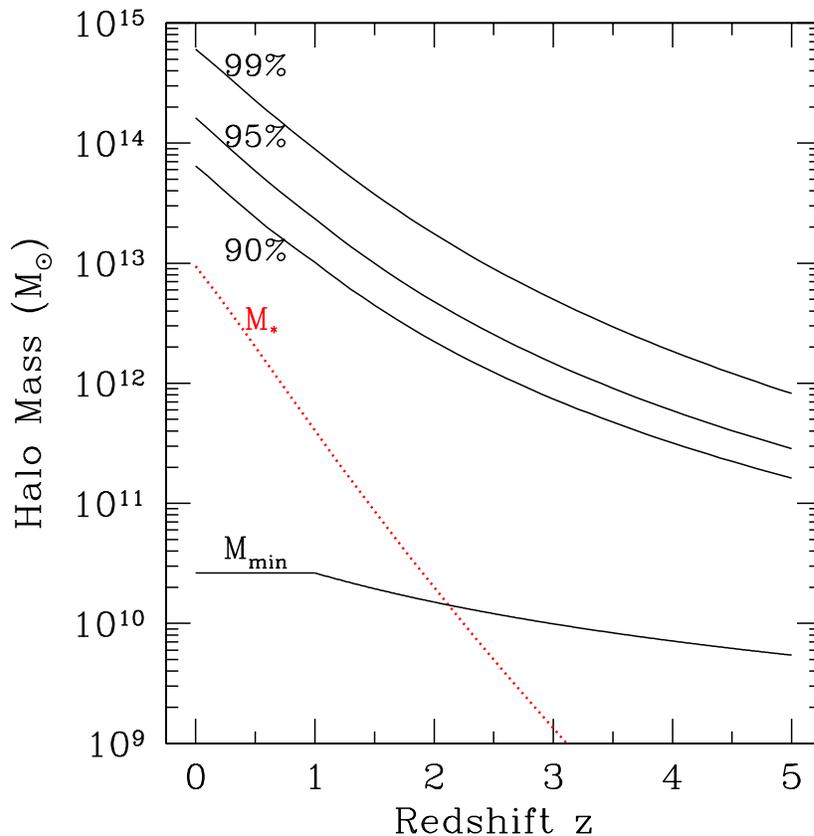


Figure 2.8: The halo mass range that dominates the rate of SMBH mergers per comoving volume. The three curves marked with percentages are the upper bounds of integration which account for 90%, 95% and 99% of  $\mathcal{N}$ . Here,  $M_{\min}$  is the mass of a halo that contains a SMBH of mass  $10^4 M_{\odot}$ . Results are shown for a flat  $\Lambda$ CDM universe with  $\Omega_M = 0.27$  and  $\sigma_8 = 0.9$ .

Consequently,  $Q_M$  predicts a higher event rate, as shown in Figure 2.7.

In summary, the slightly larger event rate given by  $Q_M$  when  $M_{\min}$  corresponds to a halo containing a  $10^4 M_{\odot}$  SMBH indicates that there is an effective upper bound on halo masses that contribute to  $\mathcal{N}$  and that this upper bound is less than  $100M_{\min} \sim 10^{12} M_{\odot}$  for  $z \gtrsim 5$ . A very limited range of halo masses dominates the integration from  $M_{\min}$  to infinity in Eq. (2.22). This mass range may be quantified by considering the ratio,

$$C(U) \equiv \frac{\frac{1}{2} \int_{M_{\min}}^U dM_1 \int_{M_{\min}}^U dM_2 \left( \frac{dn}{dM_1} \right) \left( \frac{dn}{dM_2} \right) Q(M_1, M_2)}{\frac{1}{2} \int_{M_{\min}}^{\infty} dM_1 \int_{M_{\min}}^{\infty} dM_2 \left( \frac{dn}{dM_1} \right) \left( \frac{dn}{dM_2} \right) Q(M_1, M_2)},$$

where the  $z$ -dependence of all quantities has been suppressed. Using the standard Lacey–Cole merger kernel, as given by Eq. (2.19), when evaluating  $C(U)$  is equivalent to using the arithmetic mean of  $Q_M$  and  $Q_L$ . Figure 2.8 shows the values of  $U$  for  $C = 0.9, 0.95$  and  $0.99$ . Also shown are  $M_{\min}$  and  $M_*$  as functions of redshift. Mergers between haloes with masses that lie between  $M_{\min}$  and  $U$

account for a fraction  $C$  of the SMBH merger density  $\mathcal{N}$ . The effective upper bound  $U$  exhibits the behavior deduced by the comparison of the two merger kernels. At low redshifts, when  $M_*$  is greater than  $M_{\min}$ , the upper bound traces  $M_*$ . When  $M_*$  falls below  $M_{\min}$ , the upper bound traces  $M_{\min}$ , and the dominant mass range narrows to less than two orders of magnitude.

## 2.5 BKH merger theory

The Lacey–Cole merger probability formula leads to a mass-asymmetric merger kernel, effectively giving *two* rates for the same merger. The fact that the two versions of the EPS merger kernel differ only slightly when the merging haloes have similar masses does not alleviate the problem; any mass-asymmetry in the merger kernel indicates that its derivation is flawed. In addition to the mass-asymmetry, EPS merger theory has another self-inconsistency: it fails to preserve the Press–Schechter mass function from which it is derived. When the EPS merger formalism is used to evolve a population of haloes via merger-tree algorithms, the resulting halo mass distribution does not match the Press–Schechter mass function. This divergence has limited the applications of merger trees; specifically, a merger-tree approach to SMBH mergers was limited to redshifts less than five because the discrepancy between the numerical halo population and the Press–Schechter distribution increases rapidly beyond that redshift [92].

The failure of the EPS merger theory to preserve the Press–Schechter halo mass distribution illuminates a new source of merger rates. The time evolution of the Press–Schechter halo mass distribution, given by Eq. (2.13), may be used to obtain a new merger kernel. By definition, this new merger kernel will preserve the Press–Schechter halo mass distribution, and it can be chosen to be symmetric in its mass arguments. BKH used this approach to obtain new halo merger rates for power-law power spectra. In this section, we will quickly review the derivation of the BKH merger rates and discuss how these rates may be used to estimate the SMBH merger rate.

### 2.5.1 Solving the coagulation equation

A merger kernel that preserves the Press–Schechter halo mass distribution must satisfy the Smoluchowski coagulation equation [91], which simply states that the rate of change in the number of haloes of mass  $M$  equals the rate of creation of such haloes through mergers of smaller haloes minus the rate haloes of mass  $M$  merge with other haloes. Adopting the shorthand  $n(M)$  for the Press–Schechter halo number density per interval mass and suppressing the redshift dependence of

all terms, the coagulation equation is

$$\begin{aligned} \frac{d}{dt}n(M) &= \frac{1}{2} \int_0^M n(M')n(M-M')Q(M', M-M') dM' \\ &\quad - n(M) \int_0^\infty n(M')Q(M, M') dM', \end{aligned} \quad (2.24)$$

where  $Q(M_1, M_2, z)$  is the desired merger kernel. The first term on the right-hand side is the rate of mergers per comoving volume that create a halo of mass  $M$ . The second term is the rate of mergers involving a halo of mass  $M$  per comoving volume — these mergers effectively destroy haloes of mass  $M$ .

BKH numerically invert the coagulation equation for  $Q$  for power-law density power spectra  $P(k) \propto k^n$ . When the density power spectrum is a power law, the mass variance takes a very simple form,

$$\sigma(M, z) = \sigma[M_*(z)] \left(\frac{M}{M_*}\right)^{-(3+n)/6} = \delta_{\text{coll}} \left(\frac{M}{M_*}\right)^{-(3+n)/6}.$$

Since the redshift-dependence of the Press–Schechter mass function enters via the ratio  $\delta_{\text{coll}}(z)/\sigma(M, z) = (M/M_*)^{(3+n)/6}$ , the  $z$ -dependence of the Press–Schechter mass function may be eliminated by expressing the masses in units of  $M_*(z)$ . For a judicious choice of time variables [ $\tau = -\ln \delta_{\text{coll}}/D(z)$ ], differentiating the Press–Schechter mass function introduces no  $z$ -dependence, and the coagulation equation becomes redshift-invariant. Consequently, the coagulation equation only has to be inverted once, for the resulting merger kernel  $Q(M_1/M_*, M_2/M_*)$  is applicable to all redshifts. This simplification is only possible when the power spectrum is a power law. For more complicated spectra, the coagulation equation will have to be solved at multiple redshifts.

When they numerically solve the coagulation equation on a discrete grid, BKH require that the merger kernel be symmetric in its two mass arguments. However, this restriction is not sufficient to determine  $Q$  uniquely from the coagulation equation. On an  $N \times N$  mass grid, the coagulation equation becomes  $N$  equations for the  $N$  possible values of  $M$ . Meanwhile, the symmetric  $Q$  matrix on the grid,  $Q_{ij} = Q(M_i, M_j)$ , has  $N(N+1)/2$  independent components. To break the degeneracy, BKH impose a regularization condition. By minimizing the second derivatives of  $Q$ , they find the smoothest, non-negative kernel that solves the coagulation equation.

In summary, the BKH merger kernel is mathematically self-consistent; unlike the Lacey–Cole merger kernel, it is symmetric in its mass arguments and it preserves the Press–Schechter halo distribution. It is also non-negative for all redshifts and masses. Unfortunately, these three requirements are not sufficient to uniquely determine a merger kernel. Consequently, the BKH merger kernel should be considered, for now, as an example of a self-consistent merger kernel, rather than a definitive result.

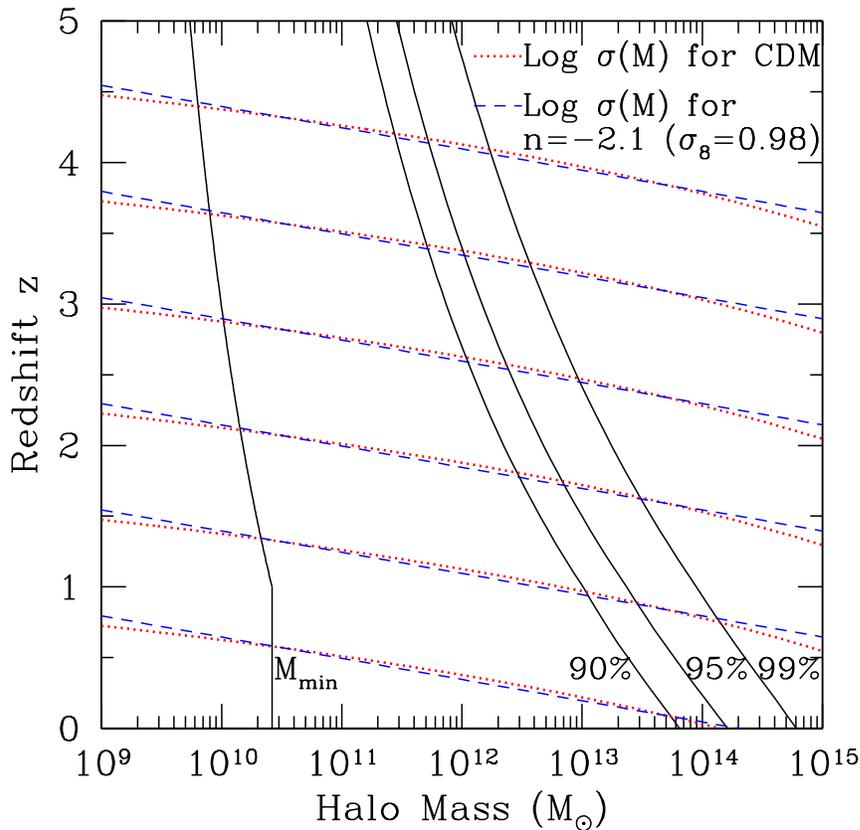


Figure 2.9: The halo mass range that dominates the rate of SMBH mergers per comoving volume. The three curves marked with percentages define the upper bounds of mass ranges that account for 90%, 95% and 99% of  $\mathcal{N}$ . Here,  $M_{\min}$  is the mass of a halo that contains a SMBH of mass  $10^4 M_{\odot}$ . The dotted curves are plots of  $\log \sigma(M)$  with arbitrary normalizations. Results are shown for a flat  $\Lambda$ CDM universe with  $\Omega_M = 0.27$  and  $\sigma_8 = 0.9$ . The dashed lines are plots of  $\log \sigma(M)$  for a power-law power spectrum with  $n = -2.1$  and  $\sigma_8 = 0.9843$ , which is the best linear fit to  $\log \sigma$  over the mass range between  $M_{\min}$  and the 99% curve for  $z \leq 5$ .

### 2.5.2 BKH merger rates for power-law power spectra

In Section 2.4.2, we demonstrated that the rate of SMBH mergers per comoving volume is dominated by mergers between haloes in a very limited mass range. Figure 2.9 shows the halo mass ranges that account for 90%, 95%, and 99% of SMBH mergers in the EPS merger theory. These are the same mass ranges depicted in Figure 2.8, but in this Figure,  $\log \sigma(M)$  is also displayed (dotted curves). If the density power spectrum were a power law with spectral index  $n$ , the  $\sigma(M)$  curves in Figure 2.9 would be straight lines with slope  $-(n+3)/6$ . Due to the transfer function [109], the power spectrum of a  $\Lambda$ CDM universe is not a power law, but within the mass ranges shown, the  $\log \sigma(M)$  curve is nearly a straight line. Therefore, it is possible to accurately approximate  $\sigma(M)$  over the relevant mass ranges as originating from a power-law power spectrum. We consider a power-law fit for  $\sigma(M)$  that extends over all masses that fall within the 99% mass range at any redshift less than

five. The fit has a lower mass bound of  $5.44 \times 10^9 M_\odot$ , which is the value of  $M_{\min}$  at  $z = 5$ , and extends to a mass of  $4.26 \times 10^{14} M_\odot$ . Over this range,  $\sigma(M)$  is best fit by spectral index  $n = -2.1$  normalized so that  $\sigma_8 = 0.9843$ , as shown by the dashed lines in Figure 2.9. This  $n = -2.1$  power-law approximation of  $\sigma(M)$  is accurate to within 16% over this mass range.

We chose to restrict these comparisons to the power law that accurately fits  $\sigma(M)$  over the mass range that dominates the EPS calculation of  $\mathcal{N}$  because the Press–Schechter mass function is employed by both merger formalisms. Like the Lacey–Cole merger rate, the Press–Schechter mass function for a  $\Lambda$ CDM universe is well-approximated by any power-law power spectrum that accurately gives  $\sigma(M)$  over the relevant halo-mass range. Using the power law that satisfies this requirement means that any major deviations from the results derived in Section 2.4 may be attributed solely to the BKH merger kernel. We chose to fit the mass range for  $z \lesssim 5$  because the SMBH merger rate peaks at redshifts less than five when the minimum black-hole mass is greater than  $10^4 M_\odot$ , so mergers at  $z \lesssim 5$  dominate the event rate. Also, when the mass range is lowered, the best-fitting spectral index decreases, and BKH merger rates have not been obtained for  $n < -2.2$ .

The density power spectrum enters the EPS merger kernel only through  $\sigma(M)$ , so any power-law approximation that accurately models  $\sigma(M)$  for  $M_1$ ,  $M_2$ , and  $M_f = M_1 + M_2$  will accurately model the Lacey–Cole merger kernel  $Q(M_1, M_2, z)$ . Unfortunately, the same is not necessarily true for the BKH merger kernels obtained by inverting the coagulation equation. Since the coagulation equation [Eq. (2.24)] involves integrals over all masses and is solved for all masses on the grid, the solution  $Q(M_1, M_2, z)$  is dependent on  $\sigma(M)$  over all masses and not just the arguments of the kernel. Therefore, while the power-law approximation accurately reflects the full  $\Lambda$ CDM result for EPS merger theory, the BKH merger rates obtained for the same power law may differ greatly from the merger rates that solve the coagulation equation for a  $\Lambda$ CDM universe. However, since the coagulation equation has not been solved for a  $\Lambda$ CDM power spectrum, we compare the EPS merger rates to the BKH merger rates for the same power law. This comparison demonstrates how the BKH merger rates differ from the EPS rates, but should not be considered a definitive description of merger rates in a  $\Lambda$ CDM universe.

BKH merger kernels for a power-law power spectrum with  $n = -2.1$  were obtained by inverting the coagulation equation on a  $91 \times 91$  grid of logarithmically-spaced  $M/M_*$  values ranging from  $10^{-12}$  to 3000. For  $M/M_*$  values greater than  $10^{-8}$ , the merger kernel values are not dependent on grid resolution, which indicates that the kernel is a numerically robust solution of the discretized coagulation equation for masses above  $10^{-8}M_*$ . The  $M_{\text{BH}}-M_{\text{halo}}$  relation [Eq. (2.11)] implies that SMBHs with masses greater than  $10^3 M_\odot$  reside in haloes with masses greater than  $10^8 M_\odot$ . Therefore, for all haloes which contain SMBHs capable of producing a gravitational-wave signal detectable by LISA,  $M/M_* \gtrsim 10^{-5}$ , so the lower mass bound on reliable kernel values is of no concern.

Unfortunately, the same is not true for the upper bound on  $M/M_*$ . The upper bound on the halo

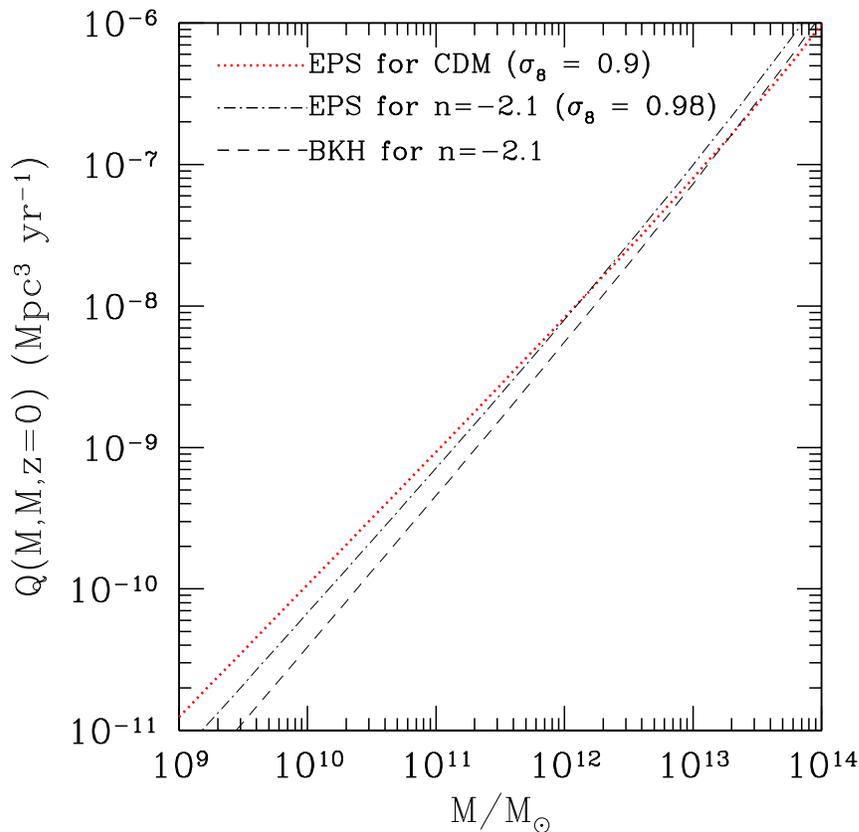


Figure 2.10: Equal-mass merger kernels for  $z = 0$ . The dotted curve shows the EPS merger kernel for a  $\Lambda$ CDM power spectrum with  $\Omega_M = 0.27$ ,  $h = 0.72$ , and  $\sigma_8 = 0.9$ . The dot-dashed curve is the EPS merger kernel for a power-law approximation. The dashed curve is the BKH merger kernel for the same power law.

masses which contribute to the SMBH merger rate  $\mathcal{N}$  in EPS theory, shown in Figure 2.8, extends to  $M/M_* \gtrsim 10^5$  for  $z \gtrsim 5$ . However, extending the mass grid to higher values of  $M/M_*$  introduces numerical noise that prevents the kernels from converging as grid resolution is increased. Therefore, we must extrapolate the BKH kernel to higher masses. We bilinearly extrapolate the logarithm of the kernel with respect to the logarithms of its mass arguments. When used to extrapolate from a grid with  $M/M_* < 100$ , this recovers the kernel to within a factor of two. Moreover, ignoring mergers of haloes with  $M/M_* > 3000$  only slightly affects the gravitational-wave event rate calculated from the BKH merger rates: the event-rate reduction is less than 3%. Therefore, the errors introduced by our extrapolation of the BKH merger kernel are negligible.

Figures 2.10 and 2.11 show the EPS and BKH equal-mass merger kernels for the  $n = -2.1$  power-law cosmology at  $z = 0$  and  $z = 5$  respectively. Over the mass range shown, the BKH equal-mass merger kernel is significantly less than the corresponding EPS merger kernel. Since this mass range corresponds to the halo masses which dominate the SMBH merger rate, the lower BKH kernel values

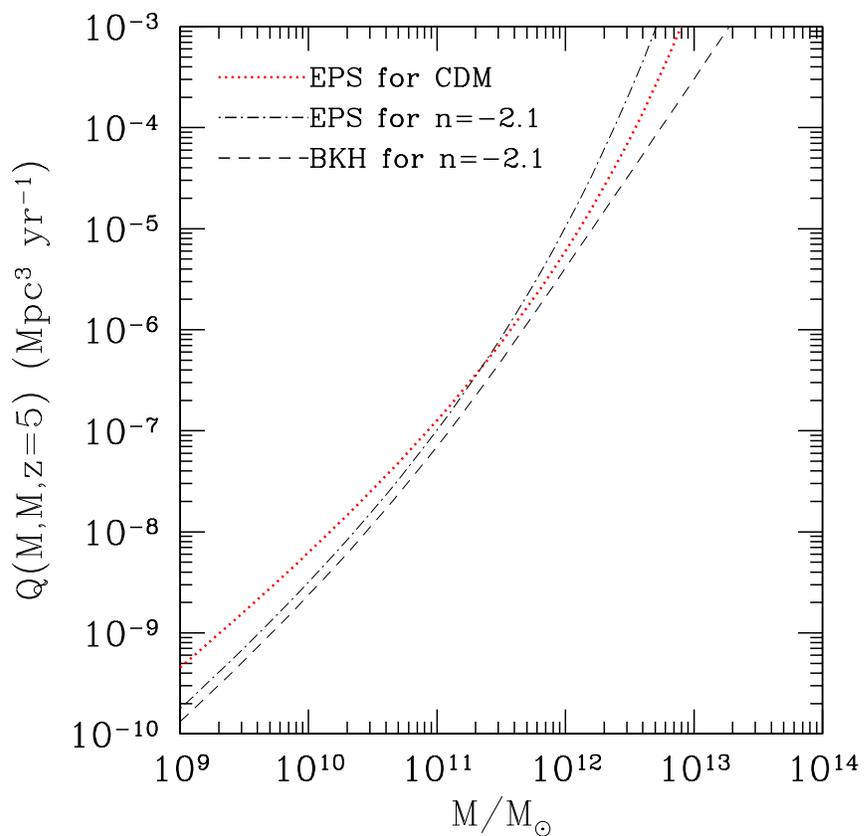


Figure 2.11: Equal-mass merger kernels for  $z = 5$ . The dotted curve shows the EPS merger kernel for a  $\Lambda$ CDM power spectrum with  $\Omega_M = 0.27$ ,  $h = 0.72$ , and  $\sigma_8 = 0.9$ . The dot-dashed curve is the EPS merger kernel for a power-law approximation with  $n = -2.1$  and  $\sigma_8 = 0.98$ . The dashed curve is the BKH merger kernel for the same power law.

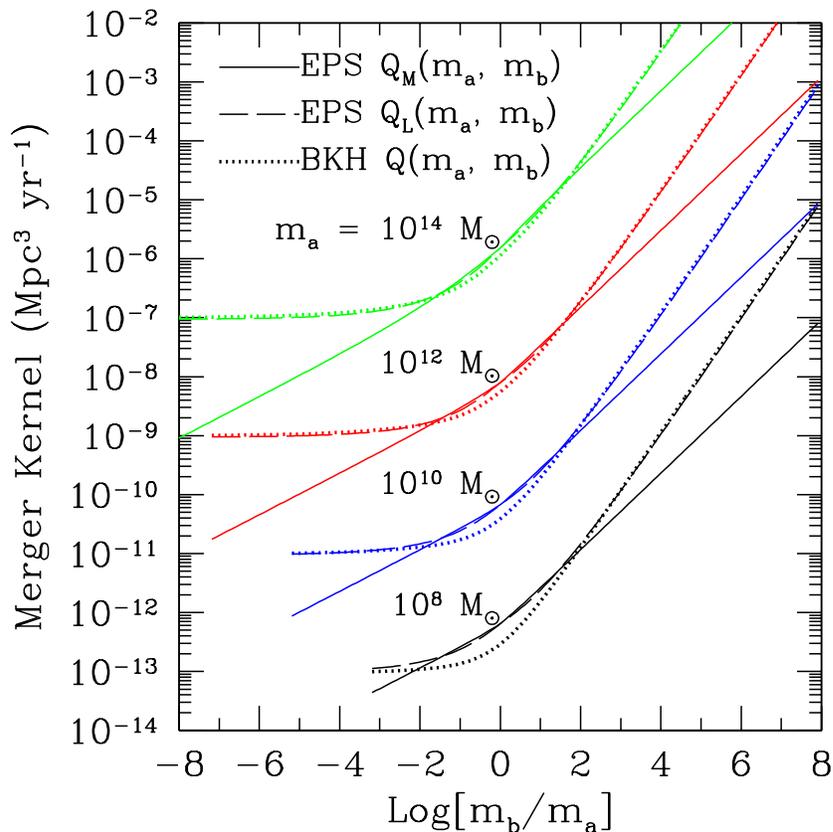


Figure 2.12: The two EPS merger kernels and the BKH merger kernel for a  $n = -2.1$  power-law power spectrum at  $z = 0$ . Here,  $Q_M$  is the Lacey–Cole merger kernel with the more massive halo as the first argument [as defined in Eq. (2.19)], and  $Q_L$  is the same kernel with the less massive halo as the first argument. Results are shown for  $\Omega_M = 0.27$ ,  $h = 0.72$  and  $\sigma_8 = 0.98$ . The low-mass cut-off of the curves arises from the  $M/M_* \gtrsim 10^{-8}$  bound on the BKH merger kernel.

indicate that using BKH merger rates instead of EPS merger rates will lower the LISA event rate from SMBH mergers. In particular, the BKH merger rate at high masses and redshifts is much lower than the corresponding EPS rates, which explains why restricting to haloes with  $M/M_* < 3000$  has less of an impact on the SMBH merger rate in BKH theory than in EPS theory.

The discrepancy between the BKH and EPS kernels also demonstrates how the Lacey–Cole merger kernel fails to solve the coagulation equation and is consequently inconsistent with the Press–Schechter halo distribution. The differences between the BKH merger kernel and both versions of the EPS merger kernel are further illustrated by Figure 2.12. The BKH merger kernel is less than both EPS kernels when the masses of the merging haloes are similar, and the difference increases as the haloes get smaller. For mergers between haloes with mass ratios greater than  $10^2$ , the BKH merger kernel is nearly equal to the EPS kernel with the least-massive halo as the first argument ( $Q_L$ ) for all masses. Therefore, for an  $n = -2.1$  power-law power-spectrum,  $Q_L$  comes closer to solving the coagulation equation than  $Q_M$ .

## 2.6 Comparison of LISA event rates from BKH and EPS merger theories

Since the BKH merger kernels for haloes of nearly equal masses are smaller than the EPS kernels for the same spectral index, applying EPS merger theory may over-estimate the LISA event rate. Figure 2.13 shows the rate  $\mathcal{N}$  of SMBH mergers per comoving volume for the power-law model discussed in the previous Section. Here,  $M_{\min}$  is the mass of a halo that contains a SMBH of mass  $10^3 M_{\odot}$ ,  $10^4 M_{\odot}$ , or  $10^5 M_{\odot}$ , as given by the  $M_{\text{halo}} - M_{\text{BH}}$  relation derived earlier [Eq. (2.11)]. For comparison, the results for a  $\Lambda$ CDM universe are also shown as a dotted curve (these are the arithmetic means of the corresponding solid and dashed curves in Figure 2.5). However, it is important to remember that although the power-law models may accurately approximate the  $\Lambda$ CDM results in the EPS theory, the same should not be assumed for the BKH merger rates. The BKH merger rates should only be compared to the EPS rates for the same power law.

At lower redshifts, the power-law EPS results closely follow the  $\Lambda$ CDM curves, but as  $M_{\min}$  decreases with increasing redshift, the mass range shifts below the region well-fit by  $n = -2.1$  when  $M_{\text{BH}} \gtrsim 10^4 M_{\odot}$ . However, Figures 2.10 and 2.11 indicate that the  $n = -2.1$  kernel is lower than the  $\Lambda$ CDM kernel for masses less than  $10^{11} M_{\odot}$ , while Figure 2.13 shows that the power law over-estimates the SMBH merger rate. The discrepancy arises from the Press–Schechter mass function: for masses below  $10^{11} M_{\odot}$ , the power-law halo number density is much greater than the  $\Lambda$ CDM halo number density, and this leads to a higher SMBH merger rate. The same mass function is used to calculate the merger rate in BKH theory, so when the power-law merger rate is higher than the  $\Lambda$ CDM rate in EPS theory, it is reasonable to assume that the same is true for the rate derived from BKH theory. Figure 2.13 also shows that the predictions for the SMBH merger rate from the BKH and EPS merger theories diverge with increasing redshift. In Section 2.4.2, we showed that as redshift increases, nearly equal-mass halo mergers dominate the event rate. The differences between the BKH merger kernel and the EPS kernel are greatest when the masses of the merging haloes are nearly equal, so as these mergers dominate the event rate at high redshifts, the BKH and EPS event rates diverge.

Figure 2.14 illustrates the potential consequences BKH merger theory has for the SMBH merger event rate observed by LISA. The difference between the BKH and EPS merger kernels for the same spectral index leads to a fairly substantial difference in the resulting event rates for LISA. For realistic values of the maximum redshift of a detectable merger ( $z_{\text{max}} \gtrsim 5$ ), the EPS prediction is about thirty percent higher than the BKH prediction for the  $n = -2.1$  power-law approximation. If the BKH merger kernel for a full  $\Lambda$ CDM power spectrum preserves the ratio of the BKS and EPS event rates for this spectral index, the LISA event rate from SMBH mergers would be reduced as well. Ref. [99] used EPS merger theory to predict that LISA will have approximately 15 SMBH-

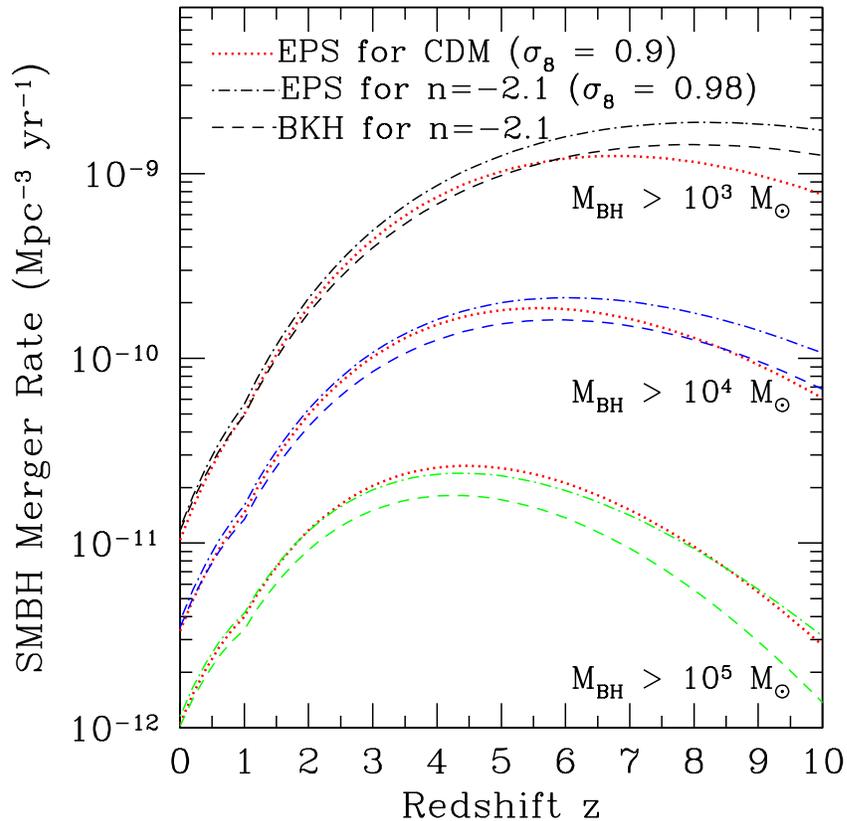


Figure 2.13: The rate of SMBH mergers per comoving volume where both merging black holes have a mass greater than  $10^3 M_{\odot}$ ,  $10^4 M_{\odot}$ , or  $10^5 M_{\odot}$ . The dotted line shows the EPS merger kernel for a  $\Lambda$ CDM power spectrum with  $\sigma_8 = 0.9$ . The dot-dashed curves are the results derived from EPS theory for a power-law approximation with  $n = -2.1$  and  $\sigma_8 = 0.98$ . The dashed curves are the BKH results for the same power law and normalization. These results all assume a flat  $\Lambda$ CDM universe with  $\Omega_M = 0.27$  and  $h = 0.72$ .

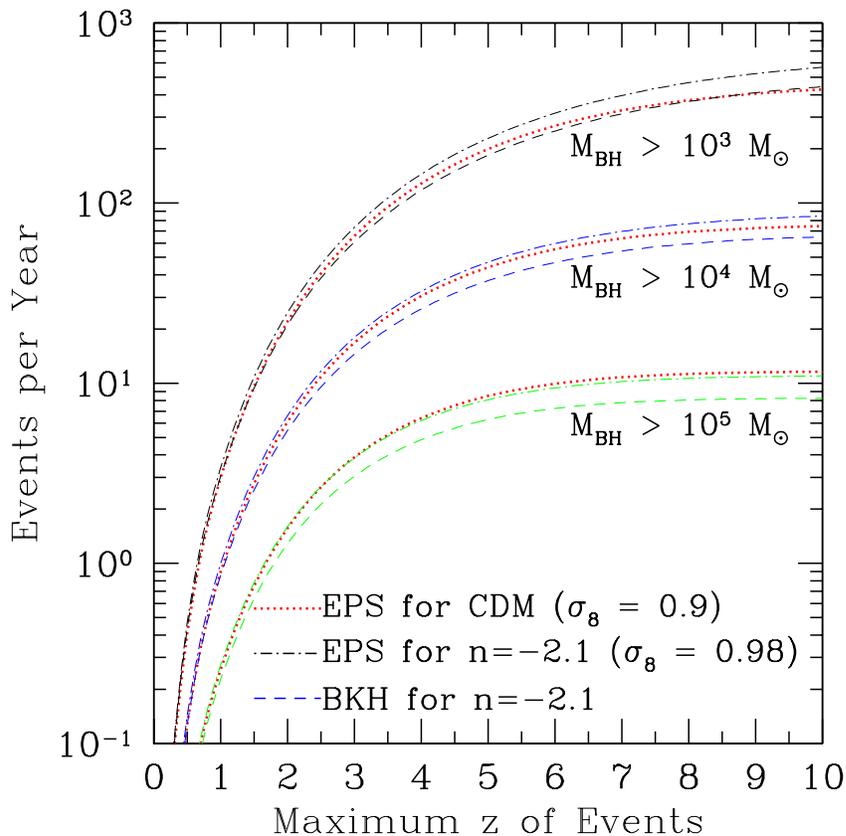


Figure 2.14: The gravitational-wave event rate from SMBH mergers as a function of the maximum redshift of a detectable merger. Only mergers in which both black holes have a mass greater than  $10^3 M_\odot$ ,  $10^4 M_\odot$ , or  $10^5 M_\odot$  are included. The dotted line shows the EPS merger kernel for a  $\Lambda$ CDM power spectrum with  $\sigma_8 = 0.9$ . The dot-dashed curves are the results derived from EPS theory for a power-law approximation:  $n = -2.1$  and  $\sigma_8 = 0.98$ . The dashed curves are the BKH results for the same power law and normalization. These results all assume a flat  $\Lambda$ CDM universe with  $\Omega_M = 0.27$  and  $h = 0.72$ .

merger detections per year at a signal to noise greater than five (they only consider mergers with  $M_{\text{BH}} \gtrsim 10^5 M_\odot$ ). These comparisons of EPS and BKH event rates indicate that LISA’s event rate may be closer to ten, with all other assumptions held fixed.

## 2.7 Summary and discussion

The EPS merger theory used to predict supermassive-black-hole merger rates is mathematically inconsistent because it contains two merger rates for the same pair of haloes. When the EPS formalism is used to derive supermassive-black-hole merger rates and the corresponding event rate for LISA, there are two potential results – the EPS predictions are ambiguous. We have found that the difference between these two predictions strongly depends on the minimum mass of a SMBH-containing halo. The relation between halo mass and black-hole mass derived by Ref. [105] and

described in Section 2.3 indicates that only haloes larger than  $10^9 M_\odot$  contain SMBHs large enough to be detectable by LISA when they merge. When this minimum halo mass is used, mergers between haloes whose masses differ by less than a factor of  $10^2$  dominate the SMBH merger rate, even when all mergers between SMBH-containing haloes are included. The difference between the EPS merger rates for mass ratios in this range is small, so the effect of the EPS mass asymmetry on the LISA event rate from SMBH mergers is minimal.

The concordance between the two EPS predictions for the SMBH merger rate is an artifact of the relative paucity of haloes with masses larger than  $10^{11} M_\odot$ . It is not an indication that the EPS merger formalism may be trusted to give realistic merger rates. In addition to its mass-asymmetry, the Lacey–Cole merger rate fails to give the same evolution of the halo population as the time derivative of the Press–Schechter mass function. Both of these flaws justify the search for a new theory of halo mergers. Benson, Kamionkowski, and Hassani (BKH) [21] inverted the coagulation equation to find merger rates that preserve the Press–Schechter halo mass function for power-law power spectra. They found that these merger rates differ significantly from the EPS rates for the same power spectrum.

The limited range of halo masses that contribute to the SMBH merger rate makes it possible to find a power-law power spectrum that accurately fits the mass variance  $\sigma(M)$  in this region. We consider such a power-law approximation with spectral index  $n = -2.1$ . Since the EPS merger formula depends only on the values of  $\sigma(M)$  for the two halo masses that are merging and the mass of the resulting halo, the power-law approximation accurately describes the result obtained from the  $\Lambda$ CDM power spectrum. The same correspondence cannot be assumed for the BKH merger rates because they are dependent on  $\sigma(M)$  at all masses.

Nevertheless, it is illuminating to compare the SMBH merger rates derived from BKH merger theory to those derived from EPS theory for the same spectral index. When  $n = -2.1$ , the BKH merger rates are lower than the corresponding EPS rates for nearly equal-mass halo mergers, which dominate the rate of SMBH mergers. This discrepancy is a clear demonstration of how the EPS rates fail to solve the coagulation equation and therefore fail to preserve the Press–Schechter halo number-density function. It also indicates how BKH theory may predict a different SMBH-merger event rate for LISA, since the difference in merger rates results in an equally large difference in event rates. Comparing the event rates derived from EPS and BKH merger theories for this spectral index indicates that the LISA event-rate predictions that employ EPS merger theory [95, 96, 99] may over-estimate the event rate by thirty percent.

Fortunately, the ambiguity carried into the SMBH-merger event-rate predictions for LISA from the uncertainty surrounding halo merger theory does not appear to immediately preclude extracting information regarding reionization or black-hole formation from LISA’s observations of SMBH mergers. Ref. [96] showed that the LISA SMBH-merger event rate with reionization occurring at  $z = 7$

is about 2.4 times higher than if reionization occurred earlier, at  $z = 12$ . This difference is larger than the uncertainties in the event rate revealed by our comparisons of BKH and EPS predictions, so it may be possible to constrain the reionization redshift from the LISA SMBH-merger event rate without a definitive theory of halo mergers. The thirty-percent uncertainty implied by these halo-merger-theory comparisons is also less than the difference in event rates for different SMBH seeding found by Ref. [92]. However, a thirty-percent uncertainty in SMBH-merger rate will significantly loosen the constraints LISA's observations of SMBH mergers could place on reionization and SMBH formation. More concerning is the fact that there is no guarantee that the merger kernel which satisfies the coagulation equation for a  $\Lambda$ CDM merger rate does not differ from the EPS merger rate by more than thirty percent.

Clearly, solving the coagulation equation for a  $\Lambda$ CDM power spectrum is imperative. Any application of EPS merger theory to astrophysical phenomena has a flawed foundation and the resulting predictions are unreliable. Specifically, we have shown that the differences between EPS merger theory and BKH merger theory for power-law power spectra indicate that switching merger theories could significantly alter the LISA SMBH-merger event rate. This theoretical uncertainty should be resolved before LISA's measurements of SMBH merger rates are used to constrain cosmological models.

### **Acknowledgments**

The author thanks Andrew Benson for providing the BKH merger rates and checking the EdS LISA event rate. The author also thanks Marc Kamionkowski and Jonathan Pritchard for useful discussions. While working on this project, the author was supported by a NSF Graduate Fellowship. This work was also supported at Caltech by DoE DE-FG03-92ER40701 and NASA NNG05GF69G.

## Chapter 3

# Solar System Tests of $f(R)$ Gravity<sup>1</sup>

### 3.1 Introduction

General relativity predicts that the expansion of the Universe should be decelerating if the Universe contains only matter and radiation. Therefore, the discovery that the expansion of the Universe is currently accelerating [8, 9] implies that either the Universe is dominated by some form of dark energy with a large negative pressure or general relativity does not provide an accurate description of gravity on cosmological scales. At the moment, the predominant hypothesis is that we live in a  $\Lambda$ CDM universe, in which a nonzero vacuum energy drives the acceleration. The  $\Lambda$ CDM model poses two serious theoretical questions: why is the vacuum energy nonzero, and why is it so miniscule? An equally plausible alternative to dark energy is a modification of general relativity that would generate cosmic acceleration [29, 30, 31, 32, 33, 34, 35, 36, 37]. (Similar modifications of general relativity have also been proposed to drive inflation [114].) Modifying general relativity in this manner eliminates the need for dark energy, but it does not explain why the vacuum energy is zero.

A possible modification to general relativity that generates an accelerated expansion is  $1/R$  gravity [33], in which a term proportional to  $1/R$ , where  $R$  is the Ricci scalar, is added to the Einstein-Hilbert action so that the  $1/R$  term dominates as the Hubble parameter decreases. Soon after the introduction of this theory, it was shown that  $1/R$  gravity is dynamically equivalent to a scalar-tensor gravity with no scalar kinetic term and a nonzero potential [72]. Moreover, the equivalence to scalar-tensor gravity applies to all modified gravity theories that replace the Einstein-Hilbert action with some function of the Ricci scalar [known as  $f(R)$  gravity], provided that  $f(R)$  has a nonzero second derivative with respect to  $R$ . When the scalar field is light, this theory makes predictions that are incompatible with Solar System tests of general relativity [115, 116, 117]. Consequently, Ref. [72] concluded that a broad class of  $f(R)$  gravity theories, including  $1/R$  gravity, are ruled out by Solar System tests.

Since then, however, the results in Ref. [72] were criticized by a number of papers [73, 74, 75, 76, 77] and some even claim that Solar System experiments do not rule out any form of  $f(R)$  gravity. The essence of the criticism is that  $f(R)$  gravity admits the Schwarzschild-de Sitter solution and hence the vacuum spacetime in the Solar System is not different from that in general relativity, although

---

<sup>1</sup>The material in this chapter, except Section 3.5, was adapted from *Solar System tests do rule out  $1/R$  gravity*, Adrienne L. Erickcek, Tristan L. Smith, and Marc Kamionkowski; Phys. Rev. D **74**, 121501 (2006) and *Solar System constraints to general  $f(R)$  gravity*, Takeshi Chiba, Tristan L. Smith, and Adrienne L. Erickcek; Phys. Rev. D **75**, 124014 (2007). Reproduced here with permission, copyright (2006) and (2007) by the American Physical Society.

there were also broader objections to the equivalence between  $f(R)$  and scalar-tensor gravity [75].

In this chapter, we will show that even though the Schwarzschild-de Sitter metric is a vacuum solution in  $f(R)$  gravity, it does not correspond to the solution around a spherically symmetric massive body. The solution for the Solar System is identical to the spacetime derived using the corresponding scalar-tensor theory. We will work directly with the field equations in the metric formalism, where the field equations are obtained by varying the action with respect to the metric and treating the Ricci scalar as a function of the metric. The Palatini formalism, which treats the Ricci scalar as a function of the connection and varies the action with respect to the connection and the metric independently, yields different field equations for  $f(R)$  gravity and has been studied extensively elsewhere (e.g. Refs. [118, 119, 120, 121, 122]).

We begin in Section 3.2 by considering  $1/R$  gravity in detail. We then generalize this analysis to a broad class of  $f(R)$  gravities, namely those theories that admit a Taylor expansion of  $f(R)$  around the background value of the Ricci scalar. In Section 3.3, we solve the linearized field equations around a spherical mass and find that the solution in the Solar System is in agreement with the solution obtained using the equivalent scalar-tensor theory. When  $f(R)$  satisfies a condition that is analogous to the scalar field being light in the equivalent scalar-tensor theory, and nonlinear effects are negligible, the resulting spacetime is incompatible with Solar System tests of general relativity. In Section 3.4, we consider how our analysis applies to several  $f(R)$  gravity theories, including general relativity. This particular example illustrates the connection between  $f(R)$  gravity and general relativity and clarifies the requirements for a general relativistic limit of an  $f(R)$  theory. Section 3.5 is devoted to a particularly interesting  $f(R)$  theory, proposed by Ref. [79], that is compatible with Solar System gravitational tests. We show how this theory uses a nonlinear effect to mask its difference from general relativity. Finally, we summarize our conclusions in Section 3.6 and list a set of conditions that, when satisfied by a given  $f(R)$  theory, imply that the theory is ruled out by Solar System tests of general relativity.

## 3.2 A detailed example: $1/R$ gravity

The gravitational action of  $1/R$  gravity,

$$S = \frac{1}{16\pi G} \int d^4x \sqrt{-g} \left( R - \frac{\mu^4}{R} \right) + \int d^4x \sqrt{-g} \mathcal{L}_M, \quad (3.1)$$

may be varied with respect to the metric  $g_{\mu\nu}$  to obtain the field equation [33]

$$8\pi G T_{\mu\nu} = \left( 1 + \frac{\mu^4}{R^2} \right) R_{\mu\nu} - \frac{1}{2} \left( 1 - \frac{\mu^4}{R^2} \right) R g_{\mu\nu} + \mu^4 (g_{\mu\nu} \nabla_\alpha \nabla^\alpha - \nabla_\mu \nabla_\nu) R^{-2}. \quad (3.2)$$

We begin by using the trace of the field equation to determine the Ricci scalar  $R$ . Contracting Eq. (3.2) with the inverse metric yields

$$\square \frac{\mu^4}{R^2} - \frac{R}{3} + \frac{\mu^4}{R} = \frac{8\pi GT}{3}, \quad (3.3)$$

where  $T \equiv g^{\mu\nu} T_{\mu\nu}$ .

The constant-curvature vacuum solution is obtained by setting  $T = 0$  and  $\nabla_\mu R = 0$ . It is  $R^2 = 3\mu^4$ , corresponding to the de Sitter spacetime with Hubble parameter  $H^2 = \mu^2/(4\sqrt{3})$ , equivalent to the general-relativistic vacuum solution with a cosmological constant  $\Lambda = 3H^2 = \sqrt{3}\mu^2/4$ . The metric for this spacetime can be written as a static spherically-symmetric spacetime:

$$ds^2 = - (1 - H^2 r^2) dt^2 + (1 - H^2 r^2)^{-1} dr^2 + r^2 d\Omega^2. \quad (3.4)$$

To match the observed acceleration of the universe, the effective cosmological constant must be set to  $\Lambda \sim \mu^2 \sim H^2 \sim 10^{-56} \text{ cm}^{-2}$ .

We now consider the spacetime in the Solar System in this theory. First of all, the distances ( $\sim 10^{13} \text{ cm}$ ) in the Solar System are tiny compared with the distance  $\mu^{-1} \sim 10^{28} \text{ cm}$ , so  $\mu r \ll 1$  everywhere in the Solar System. Moreover, the densities and velocities in the Solar System are sufficiently small that we can treat the spacetime as a small perturbation to the de Sitter spacetime. The spacetime should also be spherically symmetric and static. The most general static spherically-symmetric perturbation to the vacuum de Sitter spacetime given by Eq. (3.4) can be written

$$ds^2 = - [1 + a(r) - H^2 r^2] dt^2 + [1 + b(r) - H^2 r^2]^{-1} dr^2 + r^2 d\Omega^2, \quad (3.5)$$

where the metric-perturbation variables  $a(r), b(r) \ll 1$ . In the following, we work to linear order in  $a$  and  $b$ , and also recall that  $\mu r \ll 1$ . However,  $a$  and  $b$  are *not* necessarily small compared with  $\mu r$ .

We now return to the trace of the field equation, given by Eq. (3.3), and solve it for the Ricci scalar  $R(r)$  in the presence of the Sun. We write the trace equation in terms of a new function,

$$c(r) \equiv -\frac{1}{3} + \frac{\mu^4}{R^2(r)}, \quad (3.6)$$

and demand that  $c(r) \rightarrow 0$  as  $r \rightarrow \infty$  so that  $R$  approaches its background value of  $\sqrt{3}\mu^2$  far from the source of the perturbation. Therefore,  $c(r)$  parameterizes the departure of  $R$  from the vacuum solution, and we anticipate that  $c(r)$  will be the same order in the perturbation amplitude as the metric perturbations  $a(r)$  and  $b(r)$ . In terms of  $c(r)$ , Eq. (3.3) becomes an *exact* equation,

$$\square c(r) + \frac{\mu^2 c}{\sqrt{c + \frac{1}{3}}} = \frac{8\pi G}{3} T. \quad (3.7)$$

In the Newtonian limit appropriate for the Solar System, the pressure  $p$  is negligible compared to the energy density  $\rho$ , and so  $T = -\rho$ . Neglecting terms that are higher order in  $a(r)$ ,  $b(r)$ , and  $\mu^2 r^2$ , we are able to rewrite Eq. (3.7) as

$$\nabla^2 c + \sqrt{3}\mu^2 c = -\frac{8\pi G}{3}\rho, \quad (3.8)$$

where  $\nabla^2$  is the flat-space Laplacian operator. Note that in writing this equation, which is linear in  $c(r)$ , we have also neglected higher-order terms in  $c(r)$ . Below, we will check that the solutions we obtain have  $c(r) \ll 1$  everywhere, consistent with our assumptions. The Green's function for Eq. (3.8) is  $-\cos(3^{1/4}\mu r)/(4\pi r)$ . Convolving this with the density gives us the solution to Eq. (3.8). However, we are restricting our attention to the region where  $\mu r \ll 1$ , so the Green's function reduces to that for the Laplacian operator. Therefore the equation we need to solve is  $\nabla^2 c = -(8\pi G\rho)/3$ . Integrating the right-hand side over a spherical volume of radius  $r$  gives us  $-8\pi Gm(r)/3$ , where  $m(r)$  is the mass enclosed by a radius  $r$ . Using Gauss's law to integrate the left-hand side gives us  $4\pi r^2 c'(r)$ , where the prime denotes differentiation with respect to  $r$ . Thus, the equation for  $c(r)$  becomes

$$\frac{dc}{dr} = -\frac{2Gm(r)}{3r^2} [1 + \mathcal{O}(\mu r)]. \quad (3.9)$$

Integrating Eq. (3.9) and using the boundary condition that  $c \rightarrow 0$  as  $r \rightarrow \infty$  gives us the solution

$$c(r) = \frac{2}{3} \left( \frac{GM}{r} \right) [1 + \mathcal{O}(\mu r)], \quad \text{for } r > R_\odot. \quad (3.10)$$

We also note that integration of the equation for  $c'(r)$  to radii  $r < R_\odot$  inside the star implies that the scalar curvature  $R$  remains of order  $\mu^2$ , even inside the star. We thus see that  $c \ll 1$ , so we were justified in using the linearized equation for  $c(r)$ .

This solution for  $c(r)$  implies that

$$R = \sqrt{3}\mu^2 \left( 1 - \frac{GM}{r} \right), \quad \text{for } r > R_\odot. \quad (3.11)$$

We have thus shown that  $R$  is not constant outside the star and have already arrived at a result that is at odds with the constant-curvature Schwarzschild-de Sitter solution. Notice that had we (incorrectly) used  $\rho = 0$  in Eq. (3.8), then the equations would have admitted the solution  $c(r) = 0$ ; i.e., the constant-curvature solution. However, this would be incorrect, because even though  $\rho = 0$  at  $r > R_\odot$ , the solution to the differential equation at  $r > R_\odot$  depends on the mass distribution  $\rho(r)$  at  $r < R_\odot$ . In other words, although the Schwarzschild-de Sitter solution is a static spherically-symmetric solution to the vacuum Einstein equations, it is not the solution that correctly matches onto the solution inside the star.

The solution for  $R$  both inside and outside the star is (to linear order in  $c$ ),

$$R = \sqrt{3}\mu^2 \left[ 1 - \frac{3}{2}c(r) \right]. \quad (3.12)$$

Clearly,  $1/R$  gravity produces a spacetime inside the star that is *very* different from general relativity. This result shows that in this theory one should not assume that  $R \simeq 8\pi G\rho$ ; this has led to some confusion [123, 124, 125].

To proceed to the solutions for  $a(r)$  and  $b(r)$ , we rearrange the field equation for  $1/R$  gravity [Eq. (3.2)] to obtain equations,

$$R_{\mu\nu} = \left( 1 + \frac{\mu^4}{R^2} \right)^{-1} \left[ 8\pi GT_{\mu\nu} + \frac{1}{2} \left( 1 - \frac{\mu^4}{R^2} \right) R g_{\mu\nu} - \mu^4 (g_{\mu\nu} \nabla_\alpha \nabla^\alpha - \nabla_\mu \nabla_\nu) R^{-2} \right], \quad (3.13)$$

for the Ricci tensor in terms of the Ricci scalar. When the expression for  $R$  obtained from the trace equation is inserted into the right-hand side, we obtain equations for the nonzero components of the Ricci tensor,

$$R_t^t = 3H^2 - 6\pi G\rho - \frac{3}{4}\nabla^2 c, \quad (3.14)$$

$$R_r^r = 3H^2 - \frac{3c'(r)}{2r}, \quad (3.15)$$

$$R_\theta^\theta = R_\phi^\phi = 3H^2 - \frac{3}{4} \left( \frac{c'(r)}{r} + c''(r) \right), \quad (3.16)$$

where we have neglected terms of order  $\mu^2 c$ ,  $G\rho c$  and  $c^2$  in all three expressions.

For the perturbed metric given by Eq. (3.5), the  $tt$  component of the Ricci tensor is (to linear order in small quantities)  $R_t^t = 3H^2 - (1/2)\nabla^2 a(r)$ . Applying  $\nabla^2 c = -(8\pi G\rho)/3$  to Eq. (3.14) leaves us with an equation for  $a(r)$ ,

$$\frac{1}{2}\nabla^2 a = 4\pi G\rho, \quad (3.17)$$

plus terms that are higher order in  $GM/r$  and  $\mu r$ . The solution to this equation parallels that for  $c(r)$ ; it is

$$\frac{da}{dr} = 2G \frac{m(r)}{r^2} \quad (3.18)$$

both inside and outside the star. Outside the star, this expression may be integrated, subject to the boundary condition  $a(r) \rightarrow 0$  as  $r \rightarrow \infty$ , to obtain the metric perturbation,

$$a(r) = -\frac{2GM}{r}, \quad r > R_\odot, \quad (3.19)$$

exterior to the star. Note that this recovers the Newtonian limit for the motion of nonrelativistic bodies in the Solar System, as it should.

The  $rr$  component of the Ricci tensor is (to linear order in small quantities)  $R_r^r = 3H^2 - (b'/r) - (a''/2)$ . Given our solution for  $a'(r)$  and  $c'(r) = -(2/3)Gm(r)/r^2$ , Eq. (3.15) becomes a simple differential equation for  $b(r)$ ,

$$\begin{aligned} \frac{db}{dr} &= \frac{Gm(r)}{r^2} - \frac{Gm'(r)}{r} \\ &= \frac{d}{dr} \left[ \frac{-Gm(r)}{r} \right]. \end{aligned} \quad (3.20)$$

Integrating this equation subject to the boundary condition  $b(r) \rightarrow 0$  as  $r \rightarrow \infty$  gives an expression for  $b(r)$  that is applicable both inside and outside the star:

$$b(r) = -\frac{Gm(r)}{r}. \quad (3.21)$$

This expression for  $b(r)$  and Eq. (3.18) for  $a'(r)$  also satisfy Eq. (3.16) for the angular components of the Ricci tensor. The Ricci scalar [Eq. (3.11)] is recovered from the Ricci tensor components if terms higher order in  $\mathcal{O}(\mu r^2 GM/r)$  are included in our expressions for  $a(r)$  and  $b(r)$ .

The linearized metric outside the star thus becomes

$$ds^2 = - \left( 1 - \frac{2GM}{r} - H^2 r^2 \right) dt^2 + \left( 1 + \frac{GM}{r} + H^2 r^2 \right) dr^2 + r^2 d\Omega^2. \quad (3.22)$$

To linear order in  $GM/r$  and  $H^2 r^2$ , this metric is equivalent to the isotropic metric

$$ds^2 = - \left( 1 - \frac{2GM}{r} - H^2 r^2 \right) dt^2 + \left( 1 + \frac{GM}{r} - \frac{1}{2} H^2 r^2 \right) [dr^2 + r^2 d\Omega^2]. \quad (3.23)$$

The PPN parameter  $\gamma$  is defined by the metric,

$$ds^2 = - \left( 1 - \frac{2GM}{r} \right) dt^2 + \left( 1 + \frac{2\gamma GM}{r} \right) [dr^2 + r^2 d\Omega^2]. \quad (3.24)$$

Given that  $Hr \ll 1$  in the Solar System, we find that  $\gamma = 1/2$  for  $1/R$  gravity, in agreement with Chiba's claims [72, 126], and prior calculations in scalar-tensor gravity theories: e.g., Refs. [115, 127]. We note that recent measurements give  $\gamma = 1 + (2.1 \pm 2.3) \times 10^{-5}$  [116, 117].

It has been noted that Birkhoff's theorem [128], which states that the unique static spherically-symmetric vacuum spacetime in general relativity is the Schwarzschild spacetime, is lost in  $1/R$  gravity, and that there may be several spherically-symmetric vacuum spacetimes. This is true, and the absence of Birkhoff's theorem implies it is not sufficient to find any static, spherically-symmetric solution to the vacuum field equations  $1/R$  gravity when attempting to describe the spacetime in the Solar System. What we have shown here is that the Solar System spacetime is determined uniquely by matching the exterior vacuum solution to the solution inside the Sun. When this is

done correctly, it is found that the theory predicts a PPN parameter  $\gamma = 1/2$ , in gross violation of the measurements [116, 117], which require  $\gamma$  to be extremely close to unity.

### 3.3 The weak-field solution around a spherical star in $f(R)$ gravity

In the previous section, we solved the field equations of  $1/R$  gravity in the weak-field limit by first solving the trace of the field equations for the Ricci scalar and then using that solution to obtain the metric. We will now apply the same method to general  $f(R)$  theories. Thus, we consider gravitational theories with actions of the form

$$S = \frac{1}{2\kappa} \int d^4x \sqrt{-g} f(R) + S_m, \quad (3.25)$$

where  $f(R)$  is a function of the Ricci scalar  $R$  and  $S_m$  is the matter action. The field equation obtained by varying the action with respect to the metric is

$$f_R R_{\mu\nu} - \frac{1}{2} f g_{\mu\nu} - \nabla_\mu \nabla_\nu f_R + \square f_R g_{\mu\nu} = \kappa T_{\mu\nu}, \quad (3.26)$$

where  $f_R \equiv df/dR$ . In previous studies, predictions of Solar System dynamics in these theories were analyzed by appealing to an equivalence with scalar-tensor theories [72]. We review this equivalence in Appendix A. Since the equivalent scalar-tensor theory is incompatible with Solar System observations if the scalar field propagates on Solar System scales, Ref. [72] concluded that the corresponding  $f(R)$  theories are ruled out. We now show that this conclusion can be made *without* appealing to the equivalence between  $f(R)$  and scalar-tensor gravity. Instead, we work directly with the linearized field equations about a spherical mass distribution. Our treatment clarifies and amends a similar analysis presented in Ref. [129], and we extend it to cases where the background value of the Ricci scalar equals zero.

We now find the metric that describes the spacetime around a spherical body in  $f(R)$  gravity in the weak-field regime. To do this, we must choose a background spacetime around which to linearize the field equations. The only physically relevant choice is an isotropic and homogeneous background spacetime that solves Eq. (3.26) for some spatially uniform cosmological stress-energy tensor  $T_{\mu\nu}^{\text{cos}}$ . The evolution of the time-dependent and spatially homogeneous background scalar curvature  $R_0(t)$  is determined by the trace of Eq. (3.26),

$$f_{R0}(t)R_0(t) - 2f_0(t) + 3\square f_{R0}(t) = \kappa T^{\text{cos}}(t), \quad (3.27)$$

where  $f_{R0} \equiv df/dR|_{R=R_0}$ ,  $f_0 \equiv f(R_0)$  and  $T^{\text{cos}} \equiv g^{\mu\nu} T_{\mu\nu}^{\text{cos}}$ .

In order to investigate perturbations away from this background, we express the Ricci scalar as the sum of two components:

$$R(r, t) \equiv R_0(t) + R_1(r), \quad (3.28)$$

where  $R_0(t)$  is the spatially homogenous background curvature that solves Eq. (3.27) and  $R_1(r)$  is a time-independent perturbation to this background curvature. We assume that all derivatives of  $f(R)$  are well-defined at the present-day value of  $R_0$  so that we may use a Taylor expansion of  $f(R)$  around  $R = R_0$  to evaluate  $f(R_0 + R_1)$  and  $f_R(R_0 + R_1)$ . We will terminate the expansion by neglecting terms nonlinear in  $R_1$ . Provided that the higher-order terms of the Taylor series do not cancel in some contrived way, neglecting the higher-order terms is only justified if the sum of the zeroth-order and linear terms is greater than all other terms in the Taylor expansion. Specifically, we require that

$$f_0 + f_{R0}R_1 \gg \frac{1}{n!}f^{(n)}(R_0)R_1^n, \quad (3.29)$$

$$f_{R0} + f_{RR0}R_1 \gg \frac{1}{n!}f^{(n+1)}(R_0)R_1^n, \text{ for all } n > 1, \quad (3.30)$$

where  $f_{RR0} \equiv d^2 f/dR^2|_{R=R_0}$  and  $f^{(n)}(R_0) \equiv d^n f/dR^n|_{R=R_0}$ .

Now we consider the trace of Eq. (3.26) with both a cosmological matter source described by  $T^{\text{cos}}$  and a finite, time-independent, spherically symmetric matter source, described by  $T^{\text{s}}$ :

$$f_R R - 2f + 3\Box f_R = \kappa (T^{\text{cos}} + T^{\text{s}}). \quad (3.31)$$

Using first-order Taylor expansions to evaluate  $f_R$  and  $f$  and neglecting  $\mathcal{O}(R_1^2)$  terms, we obtain a linearized version of Eq. (3.31):

$$3f_{RR0}\Box R_1(r) - \left[ f_{R0}(t) - f_{RR0}(t)R_0(t) - 3\Box f_{RR0}(t) \right] R_1 = \kappa T^{\text{s}}. \quad (3.32)$$

To obtain this equation, we used the fact that  $R_0(t)$  solves Eq. (3.27) to eliminate terms that are independent of  $R_1$ . By dropping  $\mathcal{O}(f_{RR0}R_1^2)$  terms from Eq. (3.32) while keeping the  $f_{RR0}R_0R_1$  term, we have implicitly assumed that  $R_1 \ll R_0$  if  $R_0$  is nonzero. We will check that this condition is satisfied after the discussion following Eq. (3.55). If  $R_0$  is zero, then the  $\mathcal{O}(f_{RR0}R_1^2)$  is guaranteed to be smaller than the nonzero terms in Eq. (3.32) by virtue of Eq. (3.29). Note that if  $f_{RR0} = 0$ , as in general relativity, this equation becomes simply  $f_{R0}R_1 = -\kappa T^{\text{s}}$ . If in addition  $f_{R0}$  is nonzero then  $R_1$  must vanish outside the star and hence the Schwarzschild-de Sitter solution becomes the solution to the field equation outside the source. However, if  $f_{RR0} \neq 0$ , this is no longer necessarily the case.

Finally, we take our background metric to be a flat Friedmann-Robertson-Walker (FRW) metric.

We then consider a spherically symmetric perturbation to this background so that the linearized perturbed metric takes the form

$$ds^2 = -[1 + 2\nu(r)]dt^2 + a(t)^2\{[1 + 2\lambda(r)]dr^2 + r^2d\Omega^2\}, \quad (3.33)$$

where the present value of  $a(t)$  is one. When solving the field equations, we will keep only terms linear in the perturbations  $\nu$  and  $\lambda$ .

We will now solve Eq. (3.32) for a nonzero  $f_{RR0}$ . Since we confine our analysis to a static perturbation  $R_1(r)$ ,  $\square$  becomes the flat-space Laplacian operator  $\nabla^2$ . Restricting our analysis to a source with mass density  $\rho(r)$  and negligible pressure, we may rewrite Eq. (3.32) as

$$\nabla^2 R_1 - m^2 R_1 = -\frac{\kappa\rho}{3f_{RR0}}, \quad (3.34)$$

where we have defined a mass parameter

$$m^2 \equiv \frac{1}{3} \left( \frac{f_{R0}}{f_{RR0}} - R_0 - 3\frac{\square f_{RR0}}{f_{RR0}} \right). \quad (3.35)$$

Due to the evolution of  $R_0(t)$ , this mass parameter varies in time. However, the time-scale of variation in the cosmological background spacetime is comparable to the current Hubble time. Since this time-scale is much longer than the time-scale of Solar System dynamics, we may neglect the time variation of the background spacetime when considering the behavior of bodies within the Solar System [130]. Therefore, for the purposes of this calculation, we take  $m$  to be time-independent.

The Green's function  $G(r)$  for this differential equation depends on the sign of  $m^2$ :

$$G(r) = \begin{cases} -\cos(mr)/(4\pi r) & m^2 < 0, \\ -\exp(-mr)/(4\pi r) & m^2 > 0, \end{cases} \quad (3.36)$$

where  $m \equiv \sqrt{|m^2|}$ . If  $mr \ll 1$ , then both Green's functions are approximately  $-1/(4\pi r)$ , which is the Green's function for Laplace's equation. In this case, the term proportional to  $m^2$  in Eq. (3.34) may be neglected and the solution outside the star is given by

$$R_1 = \frac{\kappa}{12\pi f_{RR0}} \frac{M}{r}, \quad (3.37)$$

where  $M$  is the total mass of the source. We note that when applied to  $1/R$  gravity with a static de Sitter background, this result agrees with the result presented in Section 3.2.

We emphasize that in order for this solution for  $R_1$  to be valid, we must have  $mr \ll 1$ . Only when this condition is satisfied is the trace of the field equation well-approximated by Laplace's equation. This restriction was not mentioned in Ref. [129]. The physical interpretation of this constraint is

clear when one considers the equivalent scalar-tensor theory. When one switches to a frame where the scalar degree of freedom is canonical, the effective mass of the scalar field evaluated in the Jordan frame is [72]

$$m_\varphi^2 = \frac{f_{R0}}{3} \left( \frac{1}{f_{RR0}} + \frac{R_0}{f_{R0}} - \frac{4f_0}{(f_{R0})^2} - \frac{2\kappa T^{\text{cos}}}{(f_{R0})^2} \right). \quad (3.38)$$

Since  $R_0$  is the solution to Eq. (3.27), this expression may be simplified to

$$m_\varphi^2 = \frac{1}{3} \left( \frac{f_{R0}}{f_{RR0}} - R_0 - 6 \frac{\square f_{R0}}{f_{R0}} \right). \quad (3.39)$$

It is clear that both  $m_\varphi$  and  $m$  [defined by Eq. (3.35)] are of the same order. Therefore, the condition that  $mr \ll 1$  is equivalent to demanding that the scalar field be light ( $m_\varphi r \ll 1$ ). See Appendix A for more details.

In summary, Eq. (3.37) is a solution to the trace of the field equation within the Solar System only if the scalar degree of freedom propagates on Solar System scales. In terms of  $f(R)$ , the necessary condition is

$$|m^2| r^2 \equiv \left| \frac{1}{3} \left( \frac{f_{R0}}{f_{RR0}} - R_0 - 3 \frac{\square f_{RR0}}{f_{RR0}} \right) \right| r^2 \ll 1. \quad (3.40)$$

The triangle inequality tells us that the mass constraint given by Eq. (3.40) implies that

$$\left| \frac{f_{R0}}{f_{RR0}} \right| r^2 - \left| R_0 - 3 \frac{\square f_{RR0}}{f_{RR0}} \right| r^2 \ll 1. \quad (3.41)$$

Finally, since  $\square f_{RR0}/f_{RR0} \sim H^2$ , where  $H \equiv \dot{a}/a$  is the current Hubble parameter, and we know that  $R_0 r^2 \sim H^2 r^2 \ll 1$  by cosmological constraints, the mass constraint implies that

$$\left| \frac{f_{R0}}{f_{RR0}} \right| r^2 \ll 1. \quad (3.42)$$

We will now use the expression for  $R_1$  given by Eq. (3.37) to solve the field equations for the metric perturbations  $\nu$  and  $\lambda$ . As we did for the trace of the field equation, we simplify the field equations by replacing  $f(R)$  and  $f_R(R)$  with first-order Taylor expansions around the background value  $R_0$  to obtain field equations that are linear in  $R_1$ . Using Eq. (3.27) to simplify this expression, we obtain

$$f_{R0}(R_\nu^\mu - [R_0]_\nu^\mu) + f_{RR0}R_1R_\nu^\mu - \frac{1}{2}f_{R0}R_1\delta_\nu^\mu - f_{RR0}\nabla^\mu\nabla_\nu R_1 + \delta_\nu^\mu f_{RR0}\square R_1 = \kappa T_{\nu}^{\mu}, \quad (3.43)$$

where  $[R_0]_\nu^\mu$  is the unperturbed FRW Ricci tensor and  $\delta_\nu^\mu$  is the Kronecker delta. We neglected time derivatives of the background metric when deriving this equation. As previously noted, the time-scale of variations in  $R_0$  is much longer than that of Solar System dynamics, making the terms involving time derivatives of  $R_0$  irrelevant to gravitational effects within the Solar System.

We simplify Eq. (3.43) further by dropping several negligible terms. We continue to ignore terms that depend on the variation of the background spacetime by dropping terms that involve products of  $\lambda$ ,  $\nu$  and  $f_{RR0}R_1$  with  $H$  and  $dH/dt$ . Since we are working in the weak-field regime, we neglect all terms that are nonlinear functions of the metric perturbations  $\lambda$  and  $\nu$ . Keeping only terms that are linear in  $\lambda$  and  $\nu$  allows us to replace the  $\square$  with the flat-space Laplacian operator  $\nabla^2$  since the perturbation is assumed to be static. Finally, we know from Eq. (3.37) that  $f_{RR0}R_1 \sim \kappa M/r$ , and we expect  $\nu$  and  $\lambda$  to be proportional to  $\kappa M/r$  as well. Therefore,  $f_{RR0}R_1\nu$  and  $f_{RR0}R_1\lambda$  are second-order quantities, and we may neglect them. With these simplifications, the  $tt, rr, \theta\theta$  components of Eq. (3.43) are respectively

$$f_{R0}\nabla^2\nu + \frac{1}{2}f_{R0}R_1 - f_{RR0}\nabla^2R_1 = \kappa\rho, \quad (3.44)$$

$$f_{R0}\left(-\nu'' + \frac{2}{r}\lambda'\right) - \frac{1}{2}f_{R0}R_1 + \frac{2}{r}f_{RR0}R_1' = 0, \quad (3.45)$$

$$f_{R0}\left(\frac{1}{r}\lambda' - \frac{1}{r}\nu' + \frac{2}{r^2}\lambda\right) - \frac{1}{2}f_{R0}R_1 + \frac{1}{r}f_{RR0}R_1' + f_{RR0}R_1'' = 0, \quad (3.46)$$

where the prime denotes differentiation with respect to  $r$ . The  $\phi\phi$  component of Eq. (3.43) is identical to the  $\theta\theta$  component given by Eq. (3.46).

Recalling that  $R_1$  solves Eq. (3.34) with  $m^2 = 0$  so that  $\nabla^2R_1$  is proportional to the density  $\rho$ , Eq. (3.44) may be rewritten

$$f_{R0}\nabla^2\nu = \frac{2}{3}\kappa\rho - \frac{1}{2}f_{R0}R_1. \quad (3.47)$$

We express  $\nu$  as the sum of two functions:  $\nu = \nu_0 + \nu_1$ , where

$$f_{R0}\nabla^2\nu_0 = \frac{2}{3}\kappa\rho, \quad (3.48)$$

$$f_{R0}\nabla^2\nu_1 = -\frac{1}{2}f_{R0}R_1. \quad (3.49)$$

Provided that  $f_{R0} \neq 0$ , Eq. (3.48) may be integrated via Gauss's Law to give

$$\nu_0'(r) = \frac{\kappa}{6\pi f_{R0}} \frac{m(r)}{r^2}, \quad (3.50)$$

where  $m(r)$  is the mass enclosed in a sphere of radius  $r$ . If we assume that  $\nu_0$  vanishes as  $r \rightarrow \infty$ , we may integrate Eq. (3.50) to obtain

$$\nu_0 = -\frac{\kappa}{6\pi f_{R0}} \frac{M}{r}, \quad (3.51)$$

outside the star. Solving Eq. (3.49) outside the star using Eq. (3.37) for  $R_1$  yields

$$|\nu_1| = \frac{1}{48\pi f_{RR0}} \kappa M r \ll \frac{1}{f_{R0}} \frac{\kappa M}{r}, \quad (3.52)$$

where the inequality follows from Eq. (3.42). Since  $\nu_0 \sim \kappa M / (f_{R0} r)$  outside the star we have shown that  $|\nu_1| \ll |\nu_0|$ . Therefore, we may neglect  $\nu_1$  and conclude that  $\nu = \nu_0$  as given by Eq. (3.51). This expression for  $\nu$  is used to define Newton's constant:  $G \equiv \kappa / (6\pi f_{R0})$ . For  $1/R$  gravity with a static vacuum de Sitter background,  $f_{R0} = 4/3$ , so  $\kappa$  takes its standard value of  $8\pi G$  and Eq. (3.51) matches the corresponding result in Section 3.2.

We now turn our attention to Eq. (3.45), which we will solve for  $\lambda$ . First, we note that Eq. (3.37) implies that  $R'_1 = -R_1/r$ . Therefore, the ratio of the second two terms in Eq. (3.45) is

$$\left| \frac{(1/2)f_{R0}R_1}{2f_{RR0}R'_1/r} \right| \sim \left| \frac{f_{R0}}{f_{RR0}} \right| r^2 \ll 1, \quad (3.53)$$

where the inequality follows from Eq. (3.42). Consequently, the  $f_{R0}R_1$  term is negligible, and we drop it from the equation. Differentiating Eq. (3.50) to find  $\nu''$ , and using Gauss's Law to obtain  $R'_1$  from Eq. (3.34) (with  $m^2 = 0$ ), we may then rewrite Eq. (3.45) as

$$\lambda'(r) = \frac{\kappa}{12\pi f_{R0}} \frac{d}{dr} \left( \frac{m(r)}{r} \right). \quad (3.54)$$

Assuming that  $\lambda$  vanishes as  $r \rightarrow \infty$ , this equation may be integrated to obtain

$$\lambda = \frac{\kappa}{12\pi f_{R0}} \frac{M}{r}, \quad (3.55)$$

outside the star. It is easy to verify that Eqs. (3.51) and (3.55) also satisfy the third field equation, Eq. (3.46).

We may now check our assumption that  $R_1 \ll R_0$  for nonzero  $R_0$ . From the expression for  $R_1$  given by Eq. (3.37) and our definition that  $\kappa \equiv 6\pi f_{R0}G$ , we see that

$$\frac{R_1}{R_0} \lesssim \frac{1}{R_0} \left( \frac{GM}{R_s} \right) \frac{f_{R0}}{f_{RR0}}, \quad (3.56)$$

where  $R_s$  is the radius of the star. It is easy to check that this expression holds inside the star as well by integrating Eq. (3.34) into the interior of the star. Therefore, our assumption that  $R_1 \ll R_0$  places an additional condition on the ratio  $f_{R0}/f_{RR0}$ :

$$\left| \frac{f_{R0}}{f_{RR0}} \right| \ll R_0 \left( \frac{R_s}{GM} \right) \text{ for } R_0 \neq 0. \quad (3.57)$$

If  $f_{R0}/f_{RR0} \sim R_0$ , as is the case for many  $f(R)$  theories with nonzero  $R_0$ , then this condition is

always satisfied.

Thus we have shown explicitly that  $\nu = -2\lambda = -GM/r$  for all  $f(R)$  theories with nonzero  $f_{RR0}$  that satisfy the conditions given by Eqs. (3.29), (3.30), (3.40) and (3.57). Transforming the metric given by Eq. (3.33) to isotropic coordinates, taking  $a = 1$  today, and keeping only terms that are linear in  $GM/r$  gives

$$ds^2 = - \left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 + \frac{GM}{r}\right) [dr^2 + r^2 d\Omega^2]. \quad (3.58)$$

It is clear that this spacetime is equivalent to a Parameterized Post-Newtonian spacetime with PPN parameter  $\gamma = 1/2$ . This result is in gross violation of observations; as previously mentioned, Solar System tests require that  $\gamma = 1 + (2.1 \pm 2.3) \times 10^{-5}$  [116, 117]. We also note that this result is in precise agreement with the results obtained using the equivalent scalar-tensor theory [72] (see also [126]).

### 3.4 Case studies in $f(R)$ gravity

First, we show how we regain the results of general relativity if we take  $f_{RR0} = 0$  and assume that our linearized Taylor expansion is a valid approximation. We note that general relativity [ $f(R) = R$ ] satisfies both of these conditions.

Taking  $f_{RR0} = 0$ , Eq. (3.32) yields

$$f_{R0}R_1 = \kappa\rho. \quad (3.59)$$

When  $f_{RR0} = 0$ , the  $f_{R0}R_1$  terms in the field equations [Eqs. (3.45-3.46)] are no longer negligible compared to the terms proportional to  $f_{RR0}$  since these terms vanish. The field equations then become

$$f_{R0}\nabla^2\nu + \frac{1}{2}f_{R0}R_1 = \kappa\rho, \quad (3.60)$$

$$f_{R0}\left(-\nu'' + \frac{2}{r}\lambda'\right) - \frac{1}{2}f_{R0}R_1 = 0, \quad (3.61)$$

$$f_{R0}\left(\frac{1}{r}\lambda' - \frac{1}{r}\nu' + \frac{2}{r^2}\lambda\right) - \frac{1}{2}f_{R0}R_1 = 0. \quad (3.62)$$

Using Eq. (3.59), Eq. (3.60) becomes

$$f_{R0}\nabla^2\nu = \frac{\kappa}{2}\rho, \quad (3.63)$$

and the solution outside the star is

$$\nu = -\frac{\kappa}{8\pi f_{R0}} \frac{M}{r}. \quad (3.64)$$

From Eq. (3.61) and Eq. (3.62), we have

$$\frac{f_{R0}}{r^2} (r\lambda)' = \frac{\kappa}{2}\rho, \quad (3.65)$$

and the solution outside the star is

$$\lambda = \frac{\kappa}{8\pi f_{R0}} \frac{M}{r} = -\nu. \quad (3.66)$$

Since  $\nu = -\lambda = -GM/r$ , transforming to isotropic coordinates reveals that  $\gamma = 1$  as expected.

With this result it is easy to see why the  $\mu \rightarrow 0$  limit in  $1/R^n$  ( $n > 0$ ) gravity *does not* recover general relativity. In  $1/R^n$  gravity [33], we have

$$f(R) = R - \frac{\mu^{2+2n}}{R^n}, \quad n > 0. \quad (3.67)$$

The static solution to Eq. (3.27) with  $T^{\text{cos}} = 0$  is  $R_0 = (n+2)^{1/(n+1)}\mu^2$ , and  $f_{RR0} \propto \mu^{-2}$ . Therefore,  $f_{RR0}$  *diverges rather than vanishes* in the limit that  $\mu \rightarrow 0$ , and general relativity is *not* regained. The mass parameter for this theory has the dependence  $m^2 \propto \mu^2$  and hence it vanishes in the limit that  $\mu \rightarrow 0$ . Furthermore, a Taylor series of Eq. (3.67) around  $R_0$  is well-behaved and cosmological constraints tell us that  $\mu \sim H$  so that  $m^2 r^2 \ll 1$  in the Solar System. We conclude that the analysis of general  $f(R)$  gravity given in Section 3.3 applies and  $\gamma = 1/2$  for these theories in a static background.

We note however that the static solution to Eq. (3.27) may not describe the current cosmological background in  $1/R^n$  gravity. This solution is unstable, and without fine-tuning of the initial conditions, this spacetime will evolve toward a spacetime with  $R_0 \ll \mu^2$  [33]. In that case, we note that

$$\frac{(m!)^{-1} f^{(m)}(R_0) R_1^m}{f_0 + f_{R0} R_1} \lesssim \left( \frac{GM}{r} \right)^m \ll 1, \quad (3.68)$$

$$\frac{(m!)^{-1} f^{(m+1)}(R_0) R_1^m}{f_{R0} + f_{RR0} R_1} \lesssim \left( \frac{GM}{r} \right)^m \ll 1, \quad (3.69)$$

so that Eqs. (3.29) and (3.30) are still satisfied. Furthermore,  $m^2 \propto R_0$ , so, as in the static-background case, the mass is of order the Hubble parameter today. Therefore, the  $\gamma = 1/2$  result holds even during the late-time evolution of  $1/R^n$  gravity.

Next we consider Starobinsky gravity [114] which has

$$f(R) = R + \frac{R^2}{\alpha^2}. \quad (3.70)$$

The static solution to Eq. (3.27) with  $T^{\text{cos}} = 0$  is  $R_0 = 0$  for this theory. Since  $f(R)$  is a second-order

polynomial, the first-order Taylor expansion of  $f_R(R_0 + R_1)$  is exact. The  $\mathcal{O}(R_1^2)$  term in the Taylor expansion of  $f(R_0 + R_1)$  is suppressed compared to the linear term by a factor of  $GM/r$  and is therefore negligible. The mass parameter for this theory is proportional to  $\alpha^2$ , so Eq. (3.37) is a solution for  $R_1$  if  $\alpha^2 r^2 \ll 1$ . Therefore,  $\gamma = 1/2$  in this theory if  $\alpha^2 r^2 \ll 1$  inside the Solar System. If the mass parameter  $\alpha$  is made large (i.e., if  $\alpha \simeq 10^{12}$  GeV as proposed in Ref. [114]), then this condition is not satisfied and we cannot use the analysis in Section 3.3 to calculate  $\gamma$  for this theory.

Next we consider an example of a theory that uses two mass parameters: a hybrid between Starobinsky gravity and  $1/R$  gravity. In particular, consider the function

$$f(R) = R + \frac{1}{\alpha^2} R^2 - \frac{\mu^4}{R}. \quad (3.71)$$

We then find that, as in the usual  $1/R$  case, we have  $R_0 = \sqrt{3}\mu^2$  (for a static background in vacuum). However,

$$m^2 = 3\mu^2 \left( \frac{\alpha^2}{9\mu^2 - \sqrt{3}\alpha^2} \right). \quad (3.72)$$

We can make this quantity as large as we want by letting the denominator tend towards zero, which gives the condition  $\alpha \rightarrow 3^{3/4}\mu$ . Thus, in this model we can violate the conditions listed in Section 3.3 by fine-tuning the parameters.

Finally, we consider power-law gravitational actions [131]:

$$f(R) = \left( \frac{R}{\alpha} \right)^{1+\delta}. \quad (3.73)$$

Assuming that  $\delta \neq 1$ , the static vacuum solution to Eq. (3.27) is  $R_0 = 0$ . If  $\delta$  is not an integer, there will be some derivative that is not defined at  $R = 0$ , which causes the Taylor expansion to fail around that point. In particular, if it is supposed that  $\delta \ll 1$ , then at least the second derivative will be undefined so that the Taylor expansion will fail. For  $\delta = 1$  the static vacuum background value  $R_0$  is undetermined. However, if we choose  $R_0 \neq 0$  then all of the conditions listed in Section 3.3 are satisfied and we conclude that  $\gamma = 1/2$  in agreement with Ref. [132]. If  $\delta$  is an integer greater than one, then the Taylor expansion around  $f(R_0 = 0)$  is well-defined, but we cannot drop the terms that are nonlinear in  $R_1$  since the linearized function vanishes. Therefore, this analysis is incapable of determining whether  $f(R) = R^{1+\delta}$  gravity with  $\delta \neq 1$  conflicts with Solar System tests.

### 3.5 Characteristics of viable $f(R)$ theories<sup>1</sup>

In the previous section, we found that some  $f(R)$  theories do not permit a Taylor expansion around their background value and others have large mass scales. Our analysis is not applicable to these

---

<sup>1</sup>This section contains previously unpublished work by the author.

theories and consequently, they may evade Solar System tests. Their interest, however, is limited, because theories of these types have not been shown to produce late-time cosmic acceleration, which was the motivation for  $f(R)$  gravity. Now we will consider theories that fail the condition specified by Eq. (3.57), which we will refer to as the linearity condition. This condition bears a strong resemblance to the “thin-shell” condition of chameleon gravity [133, 134], so it may hold the key to designing  $f(R)$  theories that cause cosmic acceleration and evade Solar System tests. Indeed, such models do exist [135, 79, 136, 137], and in this section we consider specifically the  $f(R)$  form proposed by Hu and Sawicki [79]:

$$f(R) = R - m^2 \frac{c_1(R/m^2)^n}{c_2(R/m^2)^n + 1}. \quad (3.74)$$

In this expression,  $c_1$  and  $c_2$  are dimensionless constants,  $n$  is a positive number, and  $m^2 = 8\pi G\rho_0/3$ , where  $\rho_0$  is the present-day matter density. This theory yields cosmic expansion that mimics a  $\Lambda$ CDM universe provided that  $R \gg m^2$  [79].

Before we examine how this theory violates the linearity condition and evades Solar System tests, we will briefly review the chameleon mechanism [133, 134], which provides a conceptual picture of the underlying physics. In chameleon gravity, there is a scalar field  $\phi$  with a monotonically decreasing potential  $V(\phi)$  that couples to matter and generates a fifth force. This fifth force is suppressed because the scalar field’s dynamics are governed by an effective potential that is density-dependent:

$$V_{\text{eff}}(\phi) = V(\phi) + \rho \exp\left(\frac{\beta\phi}{m_{\text{Pl}}}\right), \quad (3.75)$$

where  $\beta$  is a coupling parameter between matter and the scalar field,  $\rho$  is the density of the environment, and  $m_{\text{Pl}}^2 = G^{-1}$ . Since  $V(\phi)$  is monotonically decreasing, this effective potential has a minimum if  $\rho$  is nonzero, as depicted in Fig. 3.1. The value of  $\phi$  at that minimum is also density-dependent, and the curvature of the effective potential around the minimum increases with the ambient density. Therefore, the scalar field is effectively more massive in a higher-density region, provided that the scalar field reaches the minimum of the effective potential in that region ( $\phi_{\text{int}}$ ). The scalar field reaches this minimum only if the “thin-shell” condition is satisfied:

$$\phi_{\text{ext}} - \phi_{\text{int}} \ll \beta m_{\text{Pl}} \Phi, \quad (3.76)$$

where  $\phi_{\text{ext}}$  is the value of  $\phi$  at the minimum of  $V_{\text{eff}}$  outside the body, and  $\Phi$  is the Newtonian gravitational potential of the massive body. In this case, the effects of the fifth force are suppressed because the scalar field is heavy inside the massive body and cannot propagate outside the body: this is the essence of the “chameleon mechanism.”

We can already see two indications that an analogous effect could occur in  $f(R)$  gravity. In

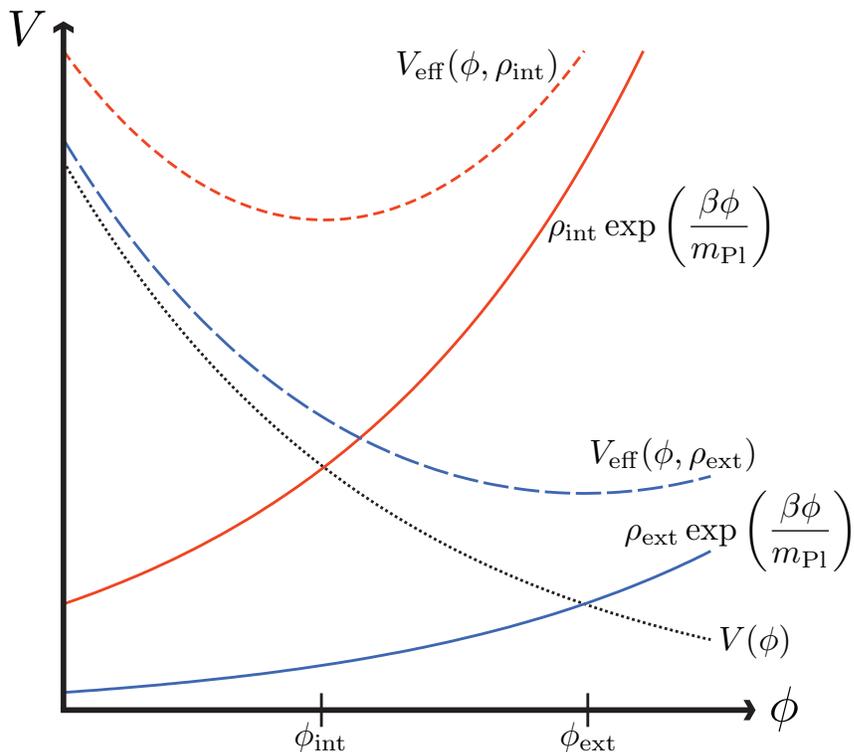


Figure 3.1: The effective potential of chameleon gravity. The effective potential is the sum of a monotonically decreasing  $V(\phi)$  (dotted curve) and  $\rho \exp[\beta\phi/m_{\text{Pl}}]$  (solid curves). The effective potential is shown for two values of the density, with  $\rho_{\text{int}} > \rho_{\text{ext}}$ . The short-dashed curve is the effective potential inside a massive body with  $\rho = \rho_{\text{int}}$ . The minimum of this potential is at  $\phi = \phi_{\text{int}}$ . The long-dashed curve is the effective potential outside the massive body, where  $\rho = \rho_{\text{ext}}$ , and its minimum is at  $\phi = \phi_{\text{ext}}$ . Note that the curvature of the effective potential inside the massive body around  $\phi_{\text{int}}$  is greater than the curvature of the effective potential outside the body around  $\phi_{\text{ext}}$ ; this implies that  $\phi$  is effectively more massive in the higher-density region, provided that  $\phi \simeq \phi_{\text{int}}$  there.

In Appendix A, we show that  $f(R)$  gravity is equivalent to a scalar-tensor theory, and we see that the effective mass of the scalar field, given by Eq. (3.38), contains a term that depends on the background density of matter. It therefore seems plausible that an  $f(R)$  theory could be designed so that the scalar field in the equivalent scalar-tensor theory behaves like the scalar field in chameleon gravity. In fact, such an  $f(R)$  theory, with an explicit connection to chameleon gravity, was first presented in Ref. [135], but this model was found to be indistinguishable from general relativity with a cosmological constant. Second, we note that the thin-shell condition is a lower bound on  $\Phi$ , while our linearity condition is an upper bound on  $\Phi = GM/R_s$ ; it is therefore possible that the two conditions are related and mutually exclusive. We will see that this is indeed the case.

It is possible for an  $f(R)$  gravity theory to give  $\gamma \simeq 1$  in the Solar System if  $R \simeq 8\pi G\rho$  inside the Sun [79]. However, in Section 3.2 we stressed that this relation is not necessarily true in  $f(R)$  gravity. On the contrary, we found that  $R \ll 8\pi G\rho$  both inside and outside the Sun in  $1/R$  gravity. The

form of  $f(R)$  given by Eq. (3.74) is special because it does permit solutions of the field equation in which  $R \simeq 8\pi G\rho$  for certain values of  $n$ ,  $c_1$  and  $c_2$  and certain values of the density inside and outside the Sun, as demonstrated numerically in Ref. [79]. The connection to the chameleon mechanism gives insight into what is happening:  $R \simeq 8\pi G\rho$  inside the Sun corresponds to  $\phi = \phi_{\text{int}}$  inside the massive body. The background value of the Ricci scalar,  $R = R_0$ , corresponds to  $\phi = \phi_{\text{ext}}$ . If  $R$  can smoothly transition from  $R_0$  far from the Sun to  $R \simeq 8\pi G\rho \gg R_0$  inside the Sun, then the chameleon mechanism hides the deviation from general relativity and  $\gamma \simeq 1$ . However, if the Sun's potential is too small to cause  $R$  to deviate significantly from  $R_0$  inside the Sun, then the linearity condition given by Eq. (3.57) is satisfied, and  $\gamma = 1/2$ .

Hu and Sawicki [79] define a “thin-shell” condition for their model:  $R \simeq 8\pi G\rho$  inside a massive body only if  $(f_R - f_{R0}) \ll \Phi$  inside the Sun. Clearly, this condition is identical in form to Eq. (3.76), which further illustrates the connection between chameleon gravity and this  $f(R)$  model. We will now show how this condition is mutually exclusive with our linearity condition. With the assumption that  $R \gg m^2$ , Eq. (3.74) implies

$$f_R \simeq 1 - n \frac{c_1}{c_2^2} \left( \frac{m^2}{R} \right)^{n+1}, \quad (3.77)$$

and since  $R_0 \ll R$  is implied by the solution  $R \simeq 8\pi G\rho$ , we see that

$$f_R - f_{R0} \simeq n \frac{c_1}{c_2^2} \left( \frac{m^2}{R_0} \right)^{n+1}. \quad (3.78)$$

Furthermore, Eq. (3.74), with  $R_0 \gg m^2$ , implies that

$$f_{RR0} \simeq n(n+1) \frac{c_1}{c_2^2} \left( \frac{m^2}{R_0} \right)^{n+1} \frac{1}{R_0}. \quad (3.79)$$

As long as  $n$  is not much greater than unity, we see that  $R_0 f_{RR0} \sim (f_R - f_{R0})$ . The thin-shell condition is therefore equivalent to  $1/f_{RR0} \gg R_0 \Phi^{-1}$ . When we note that  $f_{R0} \simeq 1$  for  $R_0 \gg m^2$ , we see that the thin-shell condition implies  $f_{R0}/f_{RR0} \gg R_0 \Phi^{-1}$ , which is the exact opposite of our linearity condition given by Eq. (3.57).

Our linearity condition provides some insight into the conditions necessary for the Hu-Sawicki  $f(R)$  theory to evade Solar System tests. A key parameter combination in this model is

$$I_0 \equiv n \frac{c_1}{c_2^2} \left( \frac{1}{41} \right)^{n+1}, \quad (3.80)$$

which was called  $|f_{R0}|$  in Ref. [79]. In terms of  $I_0$ , the linearity condition is

$$I_0 \gg \left( \frac{R_0}{41m^2} \right)^{n+1} \Phi. \quad (3.81)$$

Thus we see that the nonlinear  $R \simeq 8\pi G\rho$  solution will not exist if  $I_0$  is too large, and we expect that the maximum value will increase with increasing  $n$ ; both of these features are confirmed by the numerical analysis presented in Ref. [79].

So what value of  $I_0$  is required to violate the linearity condition, thus giving this  $f(R)$  theory a chance of evading Solar System tests? We first consider a star surrounded by a background cosmological density. To match the observed expansion of the Universe, the cosmological background value of the Ricci scalar must be  $R_0 = 41m^2$  [79]. The potential of the Sun is  $\Phi \simeq 10^{-6}$ . With these values, the linearity condition is  $I_0 \gg 10^{-6}$ . Therefore,  $I_0$  must be very small if we are to avoid  $\gamma = 1/2$ . This is not a very realistic description of our Solar System, however; the Sun is surrounded by interstellar matter, and the density of this matter is much higher than the cosmological average. If we assume that  $R_0 = 8\pi G\rho$ , with  $\rho = 10^{-24} \text{ g cm}^{-3}$ , throughout the Galaxy, then  $R_0/m^2 \simeq 10^6$ . The linearity condition is then  $I_0 \gg 10^{6n}/(41^{n+1})$ , which is much easier to violate. However, this solution is only valid if the nonlinear  $R = 8\pi G\rho$  solution is stable inside the Galaxy. Since the potential of the Galaxy is also  $\Phi \simeq 10^{-6}$ , the linearity condition for the Galaxy imbedded in a background with  $R = 41m^2$  is  $I_0 \gg 10^{-6}$ . This bound matches the numerical results presented in Ref. [79]; Hu and Sawicki find that  $R = 8\pi G\rho$  in the Galaxy only if  $I_0 \lesssim 2 \times 10^{-6}$ .

### 3.6 Summary and discussion

By analyzing the field equations around a spherically symmetric mass, we have shown that the PPN parameter  $\gamma$  of general  $f(R)$  gravity is  $\gamma = 1/2$  given the following conditions:

I. The Taylor expansions of  $f(R)$  and  $df/dR$  about the current background value  $R = R_0$ , where  $R_0$  solves Eq. (3.27), are well-defined and dominated by terms that are linear in deviations away from  $R = R_0$ . If  $R_0$  is non-zero, then the deviations from  $R_0$  are small compared to  $R_0$ . This condition may be re-expressed as Eq. (3.57) and is closely related to the third condition stated below.

II. The second derivative of  $f(R)$  with respect to  $R$  is nonzero when evaluated at the background value of  $R = R_0$ .

III. The mass parameter given by Eq. (3.35) respects the condition  $mr \ll 1$  within the Solar System.

For theories with one extra mass parameter and non-zero  $R_0$ , as in  $1/R$  gravity, it is reasonable to assume that  $f_{R0}/f_{RR0} \sim R_0$ . In that case, the latter part of the first condition is always true and the third condition is satisfied provided that  $R_0 r^2 \ll 1$  within the Solar System. However, for theories with multiple mass parameters, such as the Starobinsky- $1/R$  hybrid presented in this paper,

it is possible that this condition can be violated.

The latter part of the first condition may also be violated by multiple-parameter  $f(R)$  theories. In these theories, which do not respect Eq. (3.57), the Sun causes the Ricci scalar to deviate significantly from its background value. Consequently, terms in the field equation that are nonlinear in the perturbation to  $R$  are important. When these terms are included, the effects of the scalar degree of freedom in  $f(R)$  gravity that lead to deviations from general relativity in the Solar System may be suppressed through the chameleon mechanism [133, 134]. Since these terms were neglected in our derivation of  $\gamma = 1/2$ , it is possible that these theories could give  $\gamma \simeq 1$  in the Solar System.

Several  $f(R)$  models that explain cosmic acceleration and evade Solar System tests through these nonlinear effects have been proposed [135, 79, 136, 137]. In this chapter, we examined the Hu-Sawicki model [79] and showed how this model violates Eq. (3.57) when the parameters are chosen so that  $\gamma \simeq 1$  in the Solar System. The other viable  $f(R)$  theories exhibit similar features; they all use the chameleon mechanism to hide the local effects of the light scalar degree of freedom that is responsible for driving cosmic acceleration.

The three conditions listed above correspond to synonymous conditions in the scalar-tensor treatment. The linearity conditions established by first condition prevent the scalar potential in the equivalent scalar-tensor theory from introducing any nonlinear effects. The equivalence between  $f(R)$  and scalar-tensor gravity only holds if the second derivative of  $f(R)$  is non-zero. Finally,  $\gamma = 1/2$  only if the scalar degree of freedom is light enough to propagate through the Solar System. Therefore, we have also verified that, contrary to the claim of some authors [73, 74, 75, 76, 77], calculating the Solar System predictions of  $f(R)$  gravity using the equivalent scalar-tensor theory is a valid technique.

### **Acknowledgments**

During the final preparation of the material presented in this chapter, we learned of recent work along similar lines [138]. The author thanks Tristan Smith, Marc Kamionkowski, and Takeshi Chiba for their contributions to this research. The author also thanks Ignacy Sawicki and Wayne Hu for several illuminating discussions regarding their  $f(R)$  model. Finally, the author acknowledges useful conversations with several participants of session 86 of the Les Houches summer school. The author was supported by an NSF graduate research fellowship during the completion of this research. This research was also supported by DoE DE-FG03-92-ER40701, NASA NNG05GF69G, and the Gordon and Betty Moore Foundation.

## Chapter 4

# The Effects of Chern-Simons Gravity on Bodies Orbiting the Earth<sup>1</sup>

## 4.1 Introduction

The study of modifications of the theory of general relativity has been of interest ever since Einstein first formulated general relativity in 1915. Particularly interesting are modifications that introduce terms to the Einstein-Hilbert action that are second order in the curvature, as such modifications represent high-energy corrections to the Einstein-Hilbert action that might arise in quantum gravity. Chern-Simons gravity is an example of such a second-order modification of the Einstein-Hilbert action.

Chern-Simons modifications to gravity were first considered in 2+1 dimensions [139]. Refs. [80, 81] investigated the structure of these theories in 3+1 dimensions and showed how they could arise as a low-energy consequence of string theory. Ref. [140] considered some early-universe implications of such theories. Refs. [141, 142] investigated how Chern-Simons terms might participate in leptogenesis. Ref. [82] renewed the investigation of Chern-Simons gravity, working out the linearized equations of the theory and their implications for gravitational waves. Most recently, Refs. [143, 144] solved the linearized Chern-Simons field equations around a collection of spinning point masses. In much of the work on Chern-Simons gravity, the Chern-Simons term is coupled to a scalar field (as detailed below), and this scalar field is assumed to be spatially homogeneous but time varying. This assumption can be motivated by arguments analogous to those that have been made suggesting that the quintessence field should be coupled to the Chern-Simons term of electromagnetism [145].

Chern-Simons gravity has thus far eluded constraints from Solar System tests of weak-field gravity because it is indistinguishable from general relativity for all spacetimes that possess a maximally symmetric two-dimensional subspace and for all conformally flat spacetimes [80]. Therefore, the Schwarzschild spacetime as well as the Robertson-Walker spacetime are also solutions of the Chern-Simons gravitational field equations. Distinguishing Chern-Simons gravity from general relativity requires considerations of spacetimes that are not spherically symmetric, such as the spacetime around a spinning body. To this end, Refs. [143, 144] investigated the Chern-Simons modifications to the motion of bodies around a spinning point mass and found that the motion was indistinguishable

---

<sup>1</sup>This chapter was adapted from *The effects of Chern-Simons gravity on bodies orbiting the earth*, Tristan L. Smith, Adrienne L. Erickcek, Robert R. Caldwell and Marc Kamionkowski; Phys. Rev. D **77**, 024015 (2008). Reproduced here with permission, copyright (2008) by the American Physical Society.

from that predicted by general relativity.

In this chapter we take further steps to link Chern-Simons gravity to current and forthcoming experimental tests of weak-field gravity. We assume, as in other recent work, that the scalar field coupled to the Chern-Simons term is time varying but spatially homogeneous. We then determine the spacetime around an extended spinning mass and find that it differs from the spacetime around a spinning point mass. We determine the orbits of test particles and the precession of gyroscopes moving in this spacetime and find that the Chern-Simons modification does lead to observable deviations from the predictions of general relativity. These deviations allow us to evaluate constraints to the Chern-Simons parameter space from current satellite experiments, as well as those regions to be probed with forthcoming experiments.

We begin in Section 4.2 by defining Chern-Simons gravity and deriving the gravitational field equations. In Section 4.3, we consider the linear theory and derive the gravitomagnetic equations of motion (the Chern-Simons Ampère's law). The solution for the gravitomagnetic field around a spinning massive body is presented in Section 4.4. In Section 4.5, we consider the orbital precession of test bodies in this spacetime, as well as the orbital precession of gyroscopes, and we determine the regions of the Chern-Simons-gravity parameter space that are probed with the LAGEOS and Gravity Probe B satellites. We summarize our findings in Section 4.6.

## 4.2 Chern-Simons gravity

We consider the theory defined by the action

$$S = \int d^4x \sqrt{-g} \left[ -\frac{1}{2\kappa^2} R + \frac{\ell}{12} \theta \mathbf{R}\tilde{\mathbf{R}} - \frac{1}{2} (\partial\theta)^2 - V(\theta) + \mathcal{L}_{\text{mat}} \right], \quad (4.1)$$

where  $\mathcal{L}_{\text{mat}}$  is the Lagrangian density for matter,  $g \equiv \det g_{\mu\nu}$  is the determinant of the metric  $g_{\mu\nu}$ ,  $R$  is the Ricci scalar (with the convention  $R^\lambda{}_{\mu\nu\kappa} \equiv \Gamma^\lambda{}_{\mu\nu,\kappa} + \dots$  for the Riemann tensor), and  $\mathbf{R}\tilde{\mathbf{R}}$  is a contraction of the Riemann tensor and its dual:

$$\mathbf{R}\tilde{\mathbf{R}} \equiv R^\beta{}_\alpha{}^{\gamma\delta} \tilde{R}^\alpha{}_{\beta\gamma\delta}, \quad (4.2)$$

where the dual of the Riemann tensor is defined by

$$\tilde{R}^\mu{}_{\nu\alpha\beta} \equiv \frac{1}{2} \epsilon_{\sigma\tau\alpha\beta} R^\mu{}_{\nu}{}^{\sigma\tau}, \quad (4.3)$$

where  $\epsilon_{\sigma\tau\alpha\beta}$  is the Levi-Civita tensor, including a factor of  $\sqrt{-g}$ . Finally,  $\ell$  is a new length scale, a parameter of the theory, and  $\kappa^2 \equiv 8\pi G$ , where  $G$  is Newton's constant. Throughout this chapter we take Greek indices to range from 0 to 3. This action is different from the action considered in

Ref. [82] in that here  $\theta$  is a dynamical scalar field with a canonical kinetic term, so the  $\ell$  parameter is required to make the action dimensionless.

The equation of motion for  $\theta$  is given by

$$\square\theta = \frac{dV}{d\theta} - \frac{1}{12}\ell\mathbf{R}\tilde{\mathbf{R}}. \quad (4.4)$$

The gravitational field equations take the form

$$G_{\mu\nu} - \frac{2}{3}\ell\kappa^2 C_{\mu\nu} = -\kappa^2 T_{\mu\nu}, \quad (4.5)$$

where  $G_{\mu\nu}$  is the Einstein tensor,  $T_{\mu\nu}$  is the stress-energy tensor for the scalar field and the matter Lagrangian, and we refer to  $C_{\mu\nu}$  as the Cotton-York tensor<sup>1</sup>,

$$C^{\mu\nu} = \frac{1}{2} \left[ (\partial_\sigma\theta) \left( \epsilon^{\sigma\mu\alpha\beta} \nabla_\alpha R_\beta^\nu + \epsilon^{\sigma\nu\alpha\beta} \nabla_\alpha R_\beta^\mu \right) + \nabla_\tau(\partial_\sigma\theta) \left( \tilde{R}^{\tau\mu\sigma\nu} + \tilde{R}^{\tau\nu\sigma\mu} \right) \right]. \quad (4.6)$$

Ref. [82] notes that if  $\theta$  is a non-dynamical field (a Lagrange multiplier), the theory cannot accommodate a spacetime with a nonzero  $\mathbf{R}\tilde{\mathbf{R}}$  because the Cotton-York tensor would have a non-zero divergence. However, if  $\theta$  is a dynamical field, then the theory can indeed accommodate spacetimes with nonzero  $\mathbf{R}\tilde{\mathbf{R}}$  since we have

$$-\frac{2}{3}\ell\kappa^2 \nabla^\mu C_{\mu\nu} = \frac{\ell\kappa^2}{12} (\partial_\nu\theta) \mathbf{R}\tilde{\mathbf{R}} = -\kappa^2 \nabla^\mu T_{\mu\nu}^\theta, \quad (4.7)$$

where  $T_{\mu\nu}^\theta$  is the stress-energy tensor for  $\theta$ . We see that whereas the scalar-field stress-energy and the Cotton-York tensors are separately conserved when  $\mathbf{R}\tilde{\mathbf{R}} = 0$ , the divergence of the scalar field stress-energy tensor is precisely balanced by the divergence of the Cotton-York tensor for non-zero  $\mathbf{R}\tilde{\mathbf{R}}$  due to the novel coupling between the scalar field and gravity.

### 4.3 The Chern-Simons gravitomagnetic equations

We begin with a perturbation to the flat metric [using signature  $(-+++)$ ],

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad (4.8)$$

---

<sup>1</sup>We note that this definition differs from the usual expression for the four-dimensional Cotton-York tensor (see Ref. [82]).

and compute the linearized Einstein and Cotton-York tensors,

$$G_{\mu\nu}^{\text{linear}} = \frac{1}{2}(\square h_{\mu\nu} + \partial_\mu \partial_\nu h - \partial_\mu \partial_\alpha h_\nu^\alpha - \partial_\nu \partial_\alpha h_\mu^\alpha - \eta_{\mu\nu}[\square h - \partial_\alpha \partial_\beta h^{\alpha\beta}]), \quad (4.9)$$

$$\begin{aligned} C_{\mu\nu}^{\text{linear}} &= \frac{1}{8} \partial^\alpha \partial_\beta \theta [\eta_{\nu\gamma} \epsilon^{\gamma\beta\sigma\tau} (h_{\mu\sigma, \alpha\tau} - h_{\alpha\sigma, \mu\tau} - h_{\mu\tau, \alpha\sigma} + h_{\alpha\tau, \mu\sigma}) \\ &\quad + \eta_{\mu\gamma} \epsilon^{\gamma\beta\sigma\tau} (h_{\nu\sigma, \alpha\tau} - h_{\alpha\sigma, \nu\tau} - h_{\nu\tau, \alpha\sigma} + h_{\alpha\tau, \nu\sigma})] \\ &\quad + \frac{1}{4} \partial_\beta \theta \epsilon^{\alpha\beta\sigma\tau} [\eta_{\alpha\mu} \partial_\tau (\square h_{\nu\sigma} - \partial_\nu \partial^\lambda h_{\lambda\sigma}) + \eta_{\alpha\nu} \partial_\tau (\square h_{\mu\sigma} - \partial_\mu \partial^\lambda h_{\lambda\sigma})], \end{aligned} \quad (4.10)$$

where  $\square$  is the flat-space d'Alembertian and the comma denotes partial differentiation. Since we will require below only the gravitomagnetic fields, we will be primarily interested in the time-space components of the linearized field equations.

In this chapter, we suppose that the scalar field depends only on cosmic time,  $\theta = \theta(t)$ , the assumption being that  $\theta$  is either a quintessence field or some other field that somehow echoes the arrow of time associated with the cosmic expansion. This choice implies that the field equations are not Lorentz invariant in the Solar System since  $\partial_\sigma \theta$  points in the cosmic time direction and couples to local gravity through the Cotton-York tensor [Eq. (6)]. We note that a nonzero  $\mathbf{R}\tilde{\mathbf{R}}$  will source spatial variations in  $\theta$  through Eq. (4). By restricting  $\theta$  to be spatially homogenous, we are effectively treating  $\theta$  as a non-dynamical field, and we leave a full dynamical treatment to future work. Finally, we neglect corrections due to the motion of the Earth with respect to the rest frame of the cosmic microwave background.

We work with the trace-reversed metric perturbation,

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h, \quad (4.11)$$

and impose the Lorenz-gauge condition,  $\partial^\mu \bar{h}_{\mu\nu} = 0$ , to obtain the linearized time-space field equations,

$$G_{0i}^{\text{linear}} - \frac{2}{3} \ell \kappa^2 C_{0i}^{\text{linear}} = -\kappa^2 T_{0i}, \quad (4.12)$$

with

$$G_{0i}^{\text{linear}} = \frac{1}{2} \square \bar{h}_{0i}, \quad (4.13)$$

$$C_{0i}^{\text{linear}} = \frac{\dot{\theta}}{4} \epsilon^0{}_{ijk} \partial^j \square \bar{h}^k{}_0, \quad (4.14)$$

where the dot denotes differentiation with respect to time and Latin indices are purely spatial and range from 1 to 3. The stress-energy tensor for  $\theta(t)$  is diagonal, so it does not contribute to the time-space field equations.

Let  $t^\alpha$  be a unit vector in the coordinate time direction, and then define the 4-vector potential of this linearized theory,

$$A_\mu \equiv -\frac{1}{4}\bar{h}_{\mu\nu}t^\nu = -\frac{1}{4}\bar{h}_{\mu 0}. \quad (4.15)$$

We consider a source with mass density  $\rho$ , mass current  $\vec{J}$  and negligible pressure, so we can express the matter stress-energy tensor as

$$T_{\mu\nu} = 2t_{(\mu}J_{\nu)} - \rho t_\mu t_\nu, \quad (4.16)$$

where  $J_\mu \equiv -T_{\mu\nu}t^\nu = (-\rho, \vec{J})$ . In general relativity, the time-space components of the linearized field equations take the form

$$\partial^\mu \partial_\mu A_i = -4\pi G J_i, \quad (4.17)$$

which is (nearly) identical to Maxwell's equations for the vector potential in Lorenz gauge ( $\partial_\mu A^\mu = 0$ ). Given our definition of  $A^\mu$ , the Lorenz-gauge condition for  $A_\mu$  is implied by our earlier gauge choice for  $\bar{h}_{\mu\nu}$ .

The classically 'physical' fields (i.e., those that enter into the geodesic equation)  $\vec{E}$  and  $\vec{B}$  are given by

$$E^i = \partial_i A_0 - \partial_0 A_i, \quad (4.18)$$

$$B^i = \epsilon^{0ijk} \partial_j A_k, \quad (4.19)$$

where we have defined  $\epsilon^{0ijk} = 1$ . Two of the Maxwell equations,

$$\vec{\nabla} \cdot \vec{B} = 0, \quad (4.20)$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \quad (4.21)$$

are a direct consequence of the way in which the  $\vec{E}$  and  $\vec{B}$  fields are defined in terms of the vector potential, and so these two equations will be the same in Chern-Simons gravity. Gauss' law, which follows from the time-time component of the field equation, is now

$$\vec{\nabla} \cdot \vec{E} = 4\pi G(\rho + \rho_\theta), \quad (4.22)$$

where  $\rho_\theta$  is the energy density of the scalar field  $\theta(t)$  and is uniform throughout the Solar System. Since  $\rho_\theta$  cannot be larger than the mean cosmological energy density, it must be negligible compared to the density of the source  $\rho$ , and we do not consider it further. The only significant modification

will be to Ampère's law, which, for Chern-Simons gravity, is now given by

$$\vec{\nabla} \times \vec{B} - \frac{\partial \vec{E}}{\partial t} - \frac{1}{m_{\text{cs}}} \square \vec{B} = 4\pi G \vec{J}, \quad (4.23)$$

where we have defined  $m_{\text{cs}} \equiv -3/(\ell\kappa^2\dot{\theta})$ .

Given the metric perturbation represented by the gravitomagnetic potential and neglecting the time variation of the metric, slowly moving particles travel on geodesics such that a 'Lorentz force law' of the form,

$$\vec{a} = -\vec{E} - 4\vec{v} \times \vec{B}, \quad (4.24)$$

is obtained. Therefore, as in electrodynamics, only the physical fields, and not the potentials, have physical relevance.

We furthermore note that  $\mathbf{R}\tilde{\mathbf{R}}$  can be expressed in terms of gravito-electric and gravitomagnetic fields as

$$\mathbf{R}\tilde{\mathbf{R}} = -16(\partial_i E_j)(\partial_k B_l)(\eta^{ik}\eta^{jl} + \eta^{il}\eta^{jk}). \quad (4.25)$$

Unlike the case with Maxwell fields [146], it is not sufficient for the fields to have a non-vanishing  $\vec{E} \cdot \vec{B}$  in order to have a non-trivial coupling between gravity and the scalar field. The best example of a gravitational source which produces a non-vanishing  $\mathbf{R}\tilde{\mathbf{R}}$  is a spinning, spherical body.

## 4.4 Gravitomagnetism due to a spinning sphere in Chern-Simons gravity

### 4.4.1 Calculation of the vector potential

In Lorenz gauge ( $\partial_\mu A^\mu = 0$ ) the Chern-Simons Ampère's law, Eq. (4.23), can be written as

$$\square \left[ \vec{A} + \frac{1}{m_{\text{cs}}} \vec{B} \right] = -4\pi G \vec{J}, \quad (4.26)$$

where we have neglected the time variation in  $\dot{\theta}$  in order to place  $m_{\text{cs}}$  inside the d'Alembertian operator. We are dealing with a stationary source, and so  $\square = \nabla^2$ . We may invert Eq. (4.26) to obtain

$$\vec{A} + \frac{1}{m_{\text{cs}}} \vec{\nabla} \times \vec{A} = G \int \frac{\vec{J}}{|\vec{r} - \vec{r}'|} d^3r'. \quad (4.27)$$

We can write this as

$$\left( \mathcal{I} + \frac{1}{m_{\text{cs}}} \vec{\nabla} \times \right) \vec{A} = G \int \frac{\vec{J}}{|\vec{r} - \vec{r}'|} d^3r', \quad (4.28)$$

where  $\mathcal{I}$  is the identity matrix. Multiplying both sides of the equation by  $[\mathcal{I} - (1/m_{\text{cs}})\vec{\nabla} \times]$ , we obtain

$$\vec{A} - \frac{1}{m_{\text{cs}}^2} \vec{\nabla} \times \vec{\nabla} \times \vec{A} = G \left( \mathcal{I} - \frac{1}{m_{\text{cs}}} \vec{\nabla} \times \right) \int \frac{\vec{J}}{|\vec{r} - \vec{r}'|} d^3 r'. \quad (4.29)$$

Noting that  $\vec{\nabla} \times \vec{\nabla} \times \vec{A} = -\nabla^2 \vec{A}$  in Lorenz gauge, we have

$$\nabla^2 \vec{A} + m_{\text{cs}}^2 \vec{A} = \vec{S}, \quad (4.30)$$

where

$$\vec{S} \equiv m_{\text{cs}}^2 G \left( \mathcal{I} - \frac{1}{m_{\text{cs}}} \vec{\nabla} \times \right) \int \frac{\vec{J}}{|\vec{r} - \vec{r}'|} d^3 r'. \quad (4.31)$$

We recognize this as the inhomogeneous Helmholtz equation. We consider a homogeneous rotating sphere, and so the mass current is

$$\vec{J} = \rho [\vec{\omega} \times \vec{r}] \Theta(R - r), \quad (4.32)$$

where  $R$  is the radius of the rotating body,  $\rho$  is its density,  $\vec{\omega}$  is its angular velocity,  $r$  is the distance from the origin, and  $\Theta$  is the Heaviside step function.

The most general Green's function for the inhomogeneous Helmholtz equation is

$$G(\vec{r}, \vec{r}') = -\frac{\cos(m_{\text{cs}}|\vec{r} - \vec{r}'|) + \tilde{\gamma} \sin(m_{\text{cs}}|\vec{r} - \vec{r}'|)}{4\pi|\vec{r} - \vec{r}'|}, \quad (4.33)$$

where  $\tilde{\gamma}$  is a constant. However, the second term (that is proportional to  $\tilde{\gamma}$ ) remains constant for  $|\vec{r}' - \vec{r}| \ll m_{\text{cs}}^{-1}$ , implying that the influence of the source does not decrease with distance (for distances  $r \ll m_{\text{cs}}^{-1}$ ), which we interpret as unphysical. We therefore set  $\tilde{\gamma} = 0$ . We then use multipole expansions for the Green's function,

$$-\frac{\cos(m_{\text{cs}}|\vec{r} - \vec{r}'|)}{4\pi|\vec{r} - \vec{r}'|} = m_{\text{cs}} \sum_{\ell, m} j_{\ell}(m_{\text{cs}}r_{<}) y_{\ell}(m_{\text{cs}}r_{>}) Y_{\ell m}^*(\hat{r}') Y_{\ell m}(\hat{r}), \quad (4.34)$$

where  $j_{\ell}(x)$  and  $y_{\ell}(x)$  are, respectively, spherical Bessel function of the first and second kind,  $Y_{\ell m}(\hat{r})$  is a spherical harmonic, and the subscript  $< (>)$  means the argument is the lesser (greater) of  $r$  or  $r'$ . The solution for  $\vec{A}$  is then obtained by integrating,

$$\vec{A} = \int d^3 r' G(\vec{r}, \vec{r}') \vec{S}(\vec{r}'), \quad (4.35)$$

where all vectors are expanded in a Cartesian basis.

The resulting expression for  $\vec{A}$  may be split into a general-relativistic and a Chern-Simons term,

$\vec{A} = \vec{A}_{\text{GR}} + \vec{A}_{\text{CS}}$ , where

$$\vec{A}_{\text{GR}} = -\frac{4\pi G\rho R^3}{3}(\hat{r} \times \vec{\omega}) \times \begin{cases} \frac{r}{R} \left[ \frac{1}{2} - \frac{3}{10} \left( \frac{r}{R} \right)^2 \right], & r \leq R, \\ \frac{R^2}{5r^2}, & r \geq R, \end{cases} \quad (4.36)$$

is the gravitomagnetic vector potential in general relativity, and

$$\vec{A}_{\text{CS}} = -\frac{4\pi G\rho R^3}{m_{\text{cs}}R} [C_1(r)\vec{\omega} + C_2(r)\hat{r} \times \vec{\omega} + C_3(r)\hat{r} \times (\hat{r} \times \vec{\omega})], \quad (4.37)$$

with

$$\begin{aligned} C_1(r) &= -\frac{r^2}{5R^2} + \frac{1}{3} + \frac{2}{m_{\text{cs}}^2 R^2} + \frac{2R}{r} y_2(m_{\text{cs}}R) j_1(m_{\text{cs}}r), \\ C_2(r) &= \frac{m_{\text{cs}}r}{m_{\text{cs}}^2 R^2} + m_{\text{cs}}R y_2(m_{\text{cs}}R) j_1(m_{\text{cs}}r), \\ C_3(r) &= \frac{r^2}{5R^2} + m_{\text{cs}}R y_2(m_{\text{cs}}R) j_2(m_{\text{cs}}r), \end{aligned} \quad (4.38)$$

inside the sphere, and

$$\begin{aligned} C_1(r) &= \frac{2R^3}{15r^3} + \frac{2R}{r} j_2(m_{\text{cs}}R) y_1(m_{\text{cs}}r), \\ C_2(r) &= m_{\text{cs}}R j_2(m_{\text{cs}}R) y_1(m_{\text{cs}}r), \\ C_3(r) &= \frac{R^3}{5r^3} + m_{\text{cs}}R j_2(m_{\text{cs}}R) y_2(m_{\text{cs}}r), \end{aligned} \quad (4.39)$$

outside the sphere. We note that this solution for  $\vec{A}$  is finite at the origin and continuous across the boundary of the sphere, so it produces a finite  $\vec{B}$  at the origin and a continuous metric.

Thus far, we have not discussed any boundary conditions on the gravitomagnetic field  $\vec{B}$  at the surface of the sphere. The field equations for  $\vec{B}$  imply two such boundary conditions, and we will now prove that the continuity of  $\vec{A}$  guarantees that these two boundary conditions are satisfied. The first boundary condition follows from  $\vec{\nabla} \cdot \vec{B} = 0$ ; as in electromagnetism, this condition implies that the component of  $\vec{B}$  that is perpendicular to the surface must be continuous. The second boundary condition follows from the Chern-Simons version of Ampère's law:

$$\vec{\nabla} \times \vec{B} - \frac{1}{m_{\text{cs}}} \nabla^2 \vec{B} = 4\pi G \vec{J}. \quad (4.40)$$

Integrating this equation over a surface with vanishing area that is perpendicular to the surface of the sphere and contains the boundary implies that the components of  $[\vec{B} + (1/m_{\text{cs}})\vec{\nabla} \times \vec{B}]$  that are parallel to the sphere's surface must be continuous across the boundary.

Generally, the continuity of  $\vec{A}$  would not imply continuity of its curl. However, our  $\vec{A}$  is a solution

to Eq. (4.27), which may be rewritten as

$$\vec{A} + \frac{1}{m_{\text{cs}}} \vec{B} = \vec{A}_{\text{GR}}. \quad (4.41)$$

Since  $\vec{A}$  and  $\vec{A}_{\text{GR}}$  are both continuous across the surface of the sphere, this equation implies that  $\vec{B}$  is also continuous across the surface of the sphere. Furthermore, taking the curl of this equation shows that  $\vec{\nabla} \times \vec{B}$  is continuous provided that  $\vec{B}$  and  $\vec{\nabla} \times \vec{A}_{\text{GR}}$  are continuous. Taking the curl of Eq. (4.36) confirms that  $\vec{\nabla} \times \vec{A}_{\text{GR}}$  is continuous across the surface of the sphere. Therefore, we have shown that the continuity of  $\vec{A}$  implies that both  $\vec{B}$  and  $\vec{\nabla} \times \vec{B}$  are also continuous, which guarantees that both boundary conditions on  $\vec{B}$  are satisfied by our solution.

#### 4.4.2 The gravitomagnetic field

In the previous section, we found the vector potential by rewriting the field equation, Eq. (4.23), as an equation for  $\vec{A}$  and solving this equation with the conditions that the metric be continuous everywhere and that the gravitomagnetic field be finite and well-behaved at the origin. We note that in deriving this solution we have assumed that the time derivative of  $m_{\text{cs}}$  is negligible. The gravitomagnetic field is then obtained by taking the curl of  $\vec{A}$  and may be written as  $\vec{B} = \vec{B}_{\text{GR}} + \vec{B}_{\text{CS}}$ , where

$$\vec{B}_{\text{GR}} = \frac{4\pi G\rho R^2}{15} \times \begin{cases} \left(5 - 3\frac{r^2}{R^2}\right) \vec{\omega} + 3\frac{r^2}{R^2} \hat{r} \times (\hat{r} \times \vec{\omega}), & r \leq R, \\ \frac{R^3}{r^3} [2\vec{\omega} + 3\hat{r} \times (\hat{r} \times \vec{\omega})], & r \geq R, \end{cases} \quad (4.42)$$

is the gravitomagnetic field inside and outside a spinning sphere in general relativity, and

$$\vec{B}_{\text{CS}} = 4\pi G\rho R^2 [D_1(r) \vec{\omega} + D_2(r) \hat{r} \times \vec{\omega} + D_3(r) \hat{r} \times (\hat{r} \times \vec{\omega})], \quad (4.43)$$

is the new contribution in Chern-Simons gravity. Inside the sphere ( $r \leq R$ ),

$$\begin{aligned} D_1(r) &= \frac{2}{(m_{\text{cs}}R)^2} + \frac{2R}{r} y_2(m_{\text{cs}}R) j_1(m_{\text{cs}}r), \\ D_2(r) &= \frac{m_{\text{cs}}r}{(m_{\text{cs}}R)^2} + m_{\text{cs}}R y_2(m_{\text{cs}}R) j_1(m_{\text{cs}}r), \\ D_3(r) &= m_{\text{cs}}R y_2(m_{\text{cs}}R) j_2(m_{\text{cs}}r), \end{aligned} \quad (4.44)$$

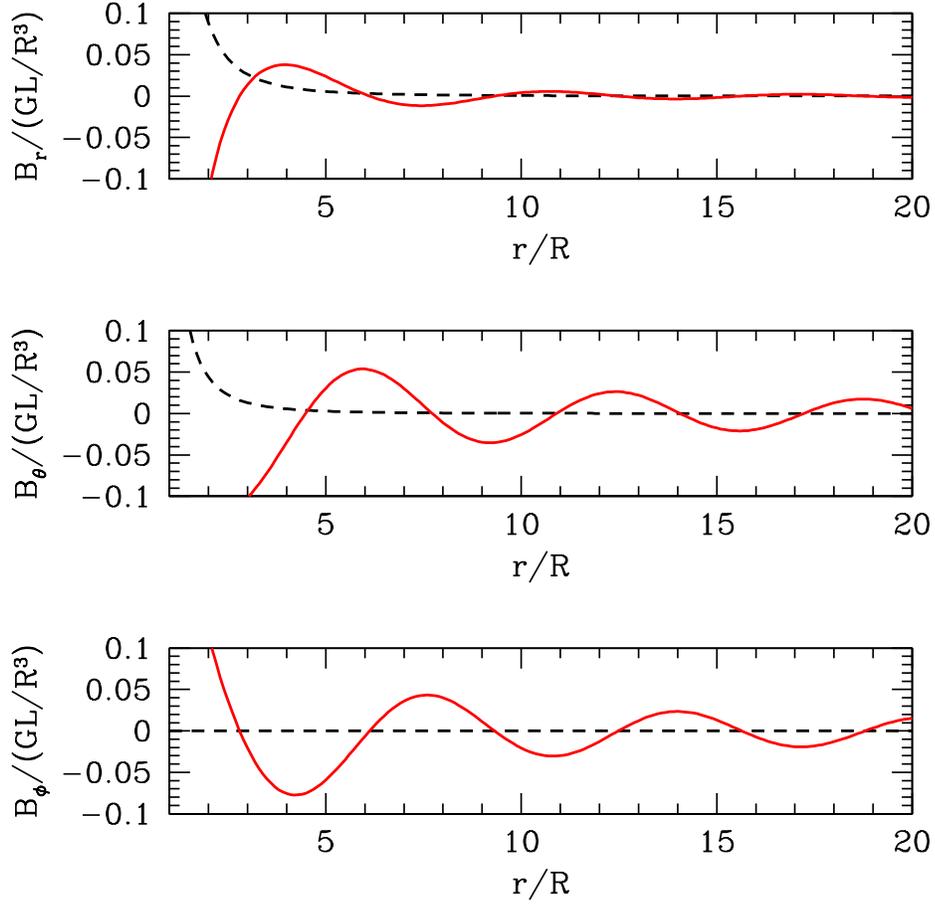


Figure 4.1: The gravitomagnetic field generated by a rotating sphere with radius  $R$  in Chern-Simons gravity. We have chosen the angular momentum of the sphere  $\vec{L}$  to point in the  $\hat{z}$  direction and evaluated the  $\vec{B}$  components at  $\theta = \pi/4$ . The dashed curves show the components of  $\vec{B}_{\text{GR}}$ , the gravitomagnetic field in general relativity. In Chern-Simons gravity, the gravitomagnetic field is the sum of  $\vec{B}_{\text{GR}}$  and  $\vec{B}_{\text{CS}}$ , and the solid curves are the components of  $\vec{B}_{\text{CS}}$ . In this plot, we set  $m_{\text{cs}}R = 1$  so that  $\vec{B}_{\text{GR}}$  and  $\vec{B}_{\text{CS}}$  would have similar amplitudes.

and outside the sphere ( $r \geq R$ )

$$\begin{aligned}
 D_1(r) &= \frac{2R}{r} j_2(m_{\text{cs}}R) y_1(m_{\text{cs}}r), \\
 D_2(r) &= m_{\text{cs}}R j_2(m_{\text{cs}}R) y_1(m_{\text{cs}}r), \\
 D_3(r) &= m_{\text{cs}}R j_2(m_{\text{cs}}R) y_2(m_{\text{cs}}r),
 \end{aligned} \tag{4.45}$$

where  $j_\ell(x)$  and  $y_\ell(x)$  are spherical Bessel functions of the first and second kind. The components of  $\vec{B}_{\text{GR}}$  and  $\vec{B}_{\text{CS}}$  in spherical coordinates for  $\vec{\omega}$  pointing in the  $\hat{z}$  direction are plotted in Fig. 4.1. In this plot, we set  $m_{\text{cs}}R = 1$ , and we note that increasing  $m_{\text{cs}}R$  will decrease the magnitude of  $\vec{B}_{\text{CS}}$  while leaving  $\vec{B}_{\text{GR}}$  unchanged.

From Eq. (4.43), we see that the Chern-Simons terms alter the components of the gravitomagnetic field along the rotation axis  $\vec{\omega}$  and  $\hat{r} \times (\hat{r} \times \vec{\omega})$ , and they also introduce a new component perpendicular to the plane defined by  $\vec{\omega}$  and  $\vec{r}$ . This new component of the gravitomagnetic field is visible in Fig. 4.1; the  $\phi$  component of  $\vec{B}_{\text{GR}}$  is zero, but the  $\phi$  component of  $\vec{B}_{\text{CS}}$  is nonzero. Thus we see that, while a toroidal mass current implies a purely poloidal gravitomagnetic field in general relativity, the parity violation introduced in Chern-Simons gravity introduces a toroidal component to the gravitomagnetic field. Something similar occurs in Chern-Simons electromagnetism [146], although the detailed fields differ since the  $\nabla^2 \vec{B}$  term in Eq. (4.23) is simply  $\vec{B}$  in the electromagnetic theory.

The Chern-Simons addition to Ampère's law, Eq. (4.23), changes that equation from a first-order differential equation for  $\vec{B}$  to a second-order differential equation. As a result, the Chern-Simons modification to the gravitomagnetic field cannot, in general, be obtained by perturbing around the general-relativistic result, as the solution in Eq. (4.43) shows. In Chern-Simons gravity, the gravitomagnetic field oscillates with distance outside the source, and we see in Fig. 4.1 that the amplitude of the oscillating field is not necessarily smaller than the general-relativistic gravitomagnetic field. Moreover, Fig. 4.1 shows that every component of  $\vec{B}_{\text{GR}}$  decreases more rapidly with distance than the corresponding component of  $\vec{B}_{\text{CS}}$ . If  $\vec{\omega}$  lies in the  $\hat{z}$  direction, Eq. (4.45) tells us that the radial component of  $\vec{B}_{\text{CS}}$  is proportional to  $1/r^2$ , while both angular components of  $\vec{B}_{\text{CS}}$  are proportional to  $1/r$ . Meanwhile,  $|\vec{B}_{\text{GR}}| \propto 1/r^3$ , so  $\vec{B}_{\text{CS}}$  will be the dominant contribution to the gravitomagnetic field sufficiently far from the source. Still, we expect from Eq. (4.23) that as  $m_{\text{CS}} \rightarrow \infty$ , the general-relativistic solution should be recovered. This occurs since the oscillatory terms vanish as  $m_{\text{CS}} \rightarrow \infty$ , and so the effects on geodesics of these new terms will vanish.

As mentioned above and shown in Fig. 4.1, the Chern-Simons gravitomagnetic field has a nonzero azimuthal component  $B_\phi$  if we take  $\vec{\omega}$  to lie in the  $\hat{z}$  direction. Since  $B_\phi \neq 0$ , one cannot find a coordinate transformation that causes both  $A_r$  and  $A_\theta$  to vanish. This is at odds with claims (see, e.g., Ref. [147]) that a metric for stationary axisymmetric spacetimes in Chern-Simons gravity can always be found with  $h_{t\theta} = h_{tr} = 0$ . In general relativity, one can always find a coordinate system in which  $A_r = A_\theta = 0$  for a stationary axisymmetric spacetime generated by a rotating perfect fluid. However, the proof of this statement assumes time-reversal invariance of the fundamental equations. This invariance implies that the metric components possess the same symmetries as the source, namely invariance under a transformation that takes  $t \rightarrow -t$  and  $\phi \rightarrow -\phi$ . In that case,  $A_r$  and  $A_\theta$  must be zero to keep the line element invariant under the same transformation. In Chern-Simons gravity, time-reversal invariance is explicitly broken by the rolling of the scalar field,  $\dot{\theta} \neq 0$ , and it is straightforward to verify that our solution for  $\vec{A}$  implies that  $A_r$  and  $A_\theta$  are both odd under time reversal. Consequently, the line element has the same symmetry as the source even though  $A_r$  and  $A_\theta$  are nonzero.

Inspection of our solution for the vector potential shows that it differs from the solution for a point-like mass-current dipole (i.e., a gravitomagnetic dipole) obtained by Alexander and Yunes (AY) [143, 144]. When applied to a single spinning source, the metric given by Refs. [143, 144] corresponds to a vector potential

$$\vec{A}_{\text{AY}} = \vec{A}_{\text{GR}} - \frac{4\pi G\rho R^3}{m_{\text{cs}}R} \left[ \frac{2R^3}{15r^3} \vec{\omega} + \frac{R^3}{5r^3} \hat{r} \times (\hat{r} \times \vec{\omega}) \right]. \quad (4.46)$$

This vector potential is an exact solution to Eq. (4.23) outside of a spinning sphere, and we can see that every term in  $\vec{A}_{\text{AY}}$  also appears in our solution for  $\vec{A}$ . The additional oscillatory terms in our solution constitute a homogeneous solution to Eq. (4.23), but without these terms,  $\vec{A}$  would not be continuous across the surface of the sphere. Furthermore, only these oscillating terms contribute to  $\vec{B}_{\text{CS}}$  because  $\vec{\nabla} \times \vec{A}_{\text{AY}} = \vec{\nabla} \times \vec{A}_{\text{GR}}$ . The inclusion of the oscillatory terms results in a Chern-Simons gravitomagnetic field that differs from general relativity, so we may use observations of the motion of test bodies in the Earth's gravitomagnetic field to constrain Chern-Simons gravity.

## 4.5 Orbital and Gyroscopic precession

### 4.5.1 Orbital precession

In order to investigate how the Chern-Simons gravitomagnetic field will affect the motion of test particles around the Earth, we will use what are known as the Gaussian perturbation equations [148, 149]. Details of how these equations are applied to gravitomagnetic forces are discussed in Ref. [150]; here we give only a brief introduction. The Gaussian perturbation equations give the time variation of the Keplerian orbital elements in the presence of a perturbing force. In our case we take the gravitomagnetic force,  $-4\vec{v} \times \vec{B}$ , as a small perturbing force and approximately solve the equations given in Ref. [150]. We will concentrate on analyzing the secular (non-periodic) time variation of the longitude of the ascending node  $\Omega$  (see Fig. 4.2), but we also note that other Keplerian elements will also vary due to the terms introduced by Chern-Simons gravity. The time variation of  $\Omega$  has been well studied since, in general relativity, it is connected with the Lense-Thirring drag [151],

$$\dot{\Omega}_{\text{GR}} = \frac{2GL}{a^3(1-e^2)^{3/2}}, \quad (4.47)$$

where  $L$  is the magnitude of the angular momentum of the central body,  $a$  is the semi-major axis of the orbit of the test body, and  $e$  is the orbit's eccentricity. Finally, in order to evaluate the secular perturbations, we approximate the orbit of the test body as circular (i.e.,  $e = 0$ , a good approximation for current measurements), and we average the perturbing force over one orbital

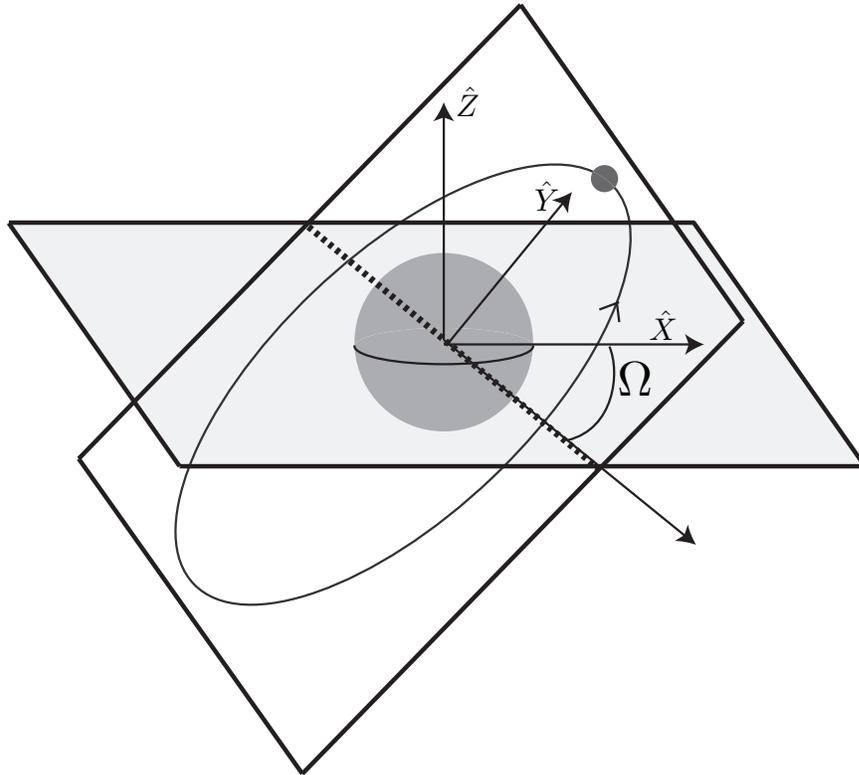


Figure 4.2: An orbital diagram showing the longitude of the ascending node ( $\Omega$ ). The stationary  $X$  and  $Y$  axes span the lightly shaded plane. The orbit of the small test body intersects this plane at two points; the line connecting these points is the line of nodes and is represented as a thick dashed line in this diagram. The longitude of the ascending node ( $\Omega$ ) is the angle between the  $X$  axis and the line of nodes segment that connects the origin of the coordinate system to the point where the orbit crosses the  $XY$  plane while traveling in the positive  $Z$  direction.

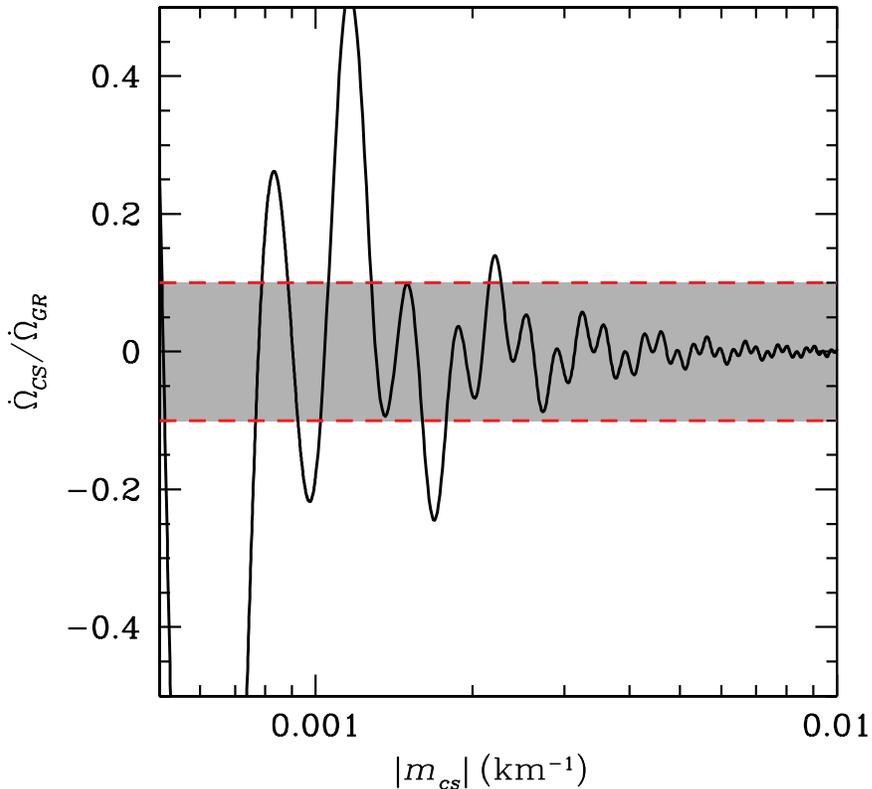


Figure 4.3: The ratio  $\dot{\Omega}_{CS}/\dot{\Omega}_{GR}$  for the LAGEOS satellites orbiting with a semimajor axis of  $a \approx 12,000$  km. A 10% verification of general relativity [83] (the shaded region) leads to a lower limit on the Chern-Simons mass of  $|m_{cs}| \gtrsim 0.001 \text{ km}^{-1}$ . A 1% verification of the Lense-Thirring drag will improve this bound on  $m_{cs}$  by a factor of roughly five.

period to obtain

$$\frac{\dot{\Omega}_{CS}}{\dot{\Omega}_{GR}} = 15 \frac{a^2}{R^2} j_2(m_{cs}R) y_1(m_{cs}a), \quad (4.48)$$

where  $\dot{\Omega}_{CS}$  is the precession due to  $\vec{B}_{CS}$ . The total precession is  $\dot{\Omega}_{GR} + \dot{\Omega}_{CS}$ . We note that  $\dot{\Omega}_{CS}$  is an even function of  $m_{cs}$ .

Recent measurements of laser ranging data to the LAGEOS I and LAGEOS II satellites have measured  $\dot{\Omega}$  to within 10% of its value in general relativity [83]. Requiring that the Chern-Simons contribution does not exceed 10% of the general relativity result, we find that we can place a lower limit to the Chern-Simons mass,  $|m_{cs}| \gtrsim 0.001 \text{ km}^{-1}$ , as shown in Fig. 4.3.

The Laser Relativity Satellite (LARES) mission [152] proposes to deploy a new laser ranging satellite and is predicted to measure  $\dot{\Omega}$  to within 1% of its value in general relativity. With this improvement the bound on  $m_{cs}$  is increased by a factor of roughly five.

### 4.5.2 Gyroscopic precession

The Earth's gravitomagnetic field will also cause a precession of gyroscopes moving in the spacetime. A gyroscope will undergo precession due to two torques. One is known as the geodetic precession and is independent of the Earth's gravitomagnetic field. The other torque is due to a coupling to the gravitomagnetic field and results in a rate of change of the spin of a gyroscope given by [153, 154, 155]

$$\dot{\vec{S}} = 2\vec{B} \times \vec{S}, \quad (4.49)$$

where  $\vec{S}$  is the angular momentum of the gyroscope.

NASA's Gravity Probe B (GPB) mission is currently attempting to measure this gyroscopic precession [84]. GPB consists of a satellite, in a polar orbit at an altitude of about 640 km, that contains four drag-free gyroscopes and a telescope. The gyroscopes are initially oriented such that their spins are aligned parallel to the optical axis of the telescope, which is pointing within the plane of the orbit. The telescope points towards a guide star, allowing a measurement of the precession of the direction of the spins of the gyroscopes. Geodetic precession results in an annual precession in the North-South direction of about 6600 milliarcseconds (mas) whereas the general relativistic gravitomagnetic field causes an annual East-West precession of around 42 mas [84].

With the Chern-Simons expression for the gravitomagnetic field, given in Eq. (4.43), it is straightforward to calculate the resulting gyroscopic precession for a polar orbit (applicable to GPB). Relative to the general relativity result, we find

$$\frac{\dot{\Phi}_{\text{CS}}}{\dot{\Phi}_{\text{GR}}} = 15 \frac{a^2}{R^2} j_2(m_{\text{cs}} R) [y_1(m_{\text{cs}} a) + m_{\text{cs}} a y_0(m_{\text{cs}} a)], \quad (4.50)$$

where  $\dot{\Phi} \equiv |\dot{\vec{S}}|/|\vec{S}| = \dot{\Phi}_{\text{GR}} + \dot{\Phi}_{\text{CS}}$  is the rate at which the angle of axis  $\Phi$  changes in time due to the gravitomagnetic field. We note that  $\dot{\Phi}_{\text{CS}}$  is an even function of  $m_{\text{cs}}$ .

It was initially projected that GPB would achieve a percent-level measurement of the gravitomagnetic contribution to  $\dot{\Phi}_{\text{GR}}$ . However, since its launch in 2004, it has encountered several unexpected complications that will degrade the precision of the tests of gravity [156], although the extent of the degradation has yet to be reported. In Fig. 4.4, we plot Eq. (4.50) for a GPB detection of the gravitomagnetic precession to within 10% of its value in general relativity.

We have idealized the Earth to be a sphere of constant density throughout this work, when in reality, it is an oblate spheroid with layers that have different mean densities. However, we expect that the non-spherical corrections would affect both the general relativity and Chern-Simons calculations similarly and, to the accuracy we require, are negligible when we consider the ratio between general relativity and Chern-Simons results. Furthermore, it is easy to generalize our results to spheres with layered density profiles because  $\vec{B}$  depends linearly on  $\rho$ . We replaced our model

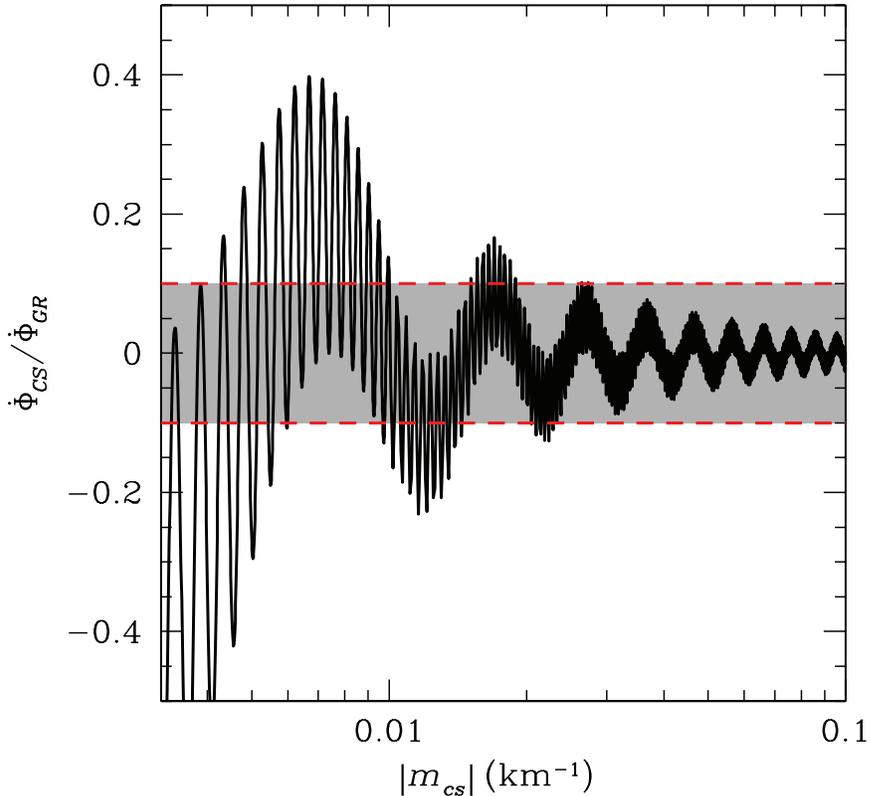


Figure 4.4: The ratio  $\dot{\Phi}_{CS}/\dot{\Phi}_{GR}$  for Gravity Probe B in a polar orbit at an altitude of approximately 640 km. A 10% verification of general relativity (the shaded region) leads to a lower limit on the Chern-Simons mass of  $|m_{cs}| \gtrsim 0.01 \text{ km}^{-1}$ , an order of magnitude improvement over the LAGEOS result.

of a homogeneous Earth with a model of the core and mantle and we found that the amplitudes of the oscillations in  $\dot{\Omega}_{CS}$  and  $\dot{\Phi}_{CS}$  were not affected. We conclude that our constraints on  $m_{cs}$  are not sensitive to the details of the density profile of the Earth.

## 4.6 Summary and discussion

The addition of a Chern-Simons term to the action for gravity is of interest as it may arise as a low-energy limit of string theory. The theory and formalism of this modification of gravity have been worked out in a number of previous papers, and some of the early-Universe consequences of such a term have been investigated. However, there has been little work on tests of such modifications in the present Universe.

In this chapter, we have calculated the linear-theory spacetime around a spinning massive body, finding new corrections that were overlooked in previous work. The gravitomagnetic field in Chern-Simons gravity differs from that in general relativity in two ways: (1) there is an oscillating component, and (2) there is a toroidal component to the gravitomagnetic field that arises as a consequence

of the parity-breaking nature of the theory and that has no counterpart in ordinary general relativity.

We then determined the precession of orbits of test particles in this spacetime and also of gyroscopes moving in this spacetime. We showed that current constraints from the LAGEOS satellites restrict the inverse Chern-Simons mass parameter  $m_{cs}^{-1}$  to be less than roughly 1000 km, corresponding to a mass constraint  $m_{cs} \gtrsim 2 \times 10^{-22}$  GeV. This bound may be improved by a factor of 5-10 by future observations.

The mass parameter  $m_{cs}$  is related to the more fundamental parameters  $\ell$  and  $\dot{\theta}$  of the theory through  $m_{cs} = -3/(8\pi G\ell\dot{\theta})$ , where  $\ell$  is a length parameter that enters into the Chern-Simons Lagrangian, and  $\dot{\theta}$  is presumably related to the time variation of the quintessence field. In principle, a precise constraint to  $\ell$  can be derived once the precise nature of the field (a quintessence field?)  $\theta$  and its time evolution are specified. We leave such model building for future work.

### **Acknowledgments**

The author thanks Tristan Smith, Robert Caldwell and Marc Kamionkowski for their contributions to the research described in this chapter. During the preparation of this chapter, the author was supported by an NSF graduate fellowship. This work also was supported by DoE DE-FG03-92-ER40701, NASA NNG05GF69G, the Gordon and Betty Moore Foundation, and NSF AST-0349213.

## Chapter 5

# Superhorizon Perturbations and the Cosmic Microwave Background<sup>1</sup>

## 5.1 Introduction

The finite age of the Universe implies the existence of a cosmological particle horizon beyond which we cannot observe. Inhomogeneities with wavelengths longer than the horizon are not completely invisible, however. The generation of large-scale temperature fluctuations in the cosmic microwave background (CMB) by superhorizon perturbations is known as the Grishchuk-Zel'dovich effect [85]. Through this effect, measurements of the low-multipole moments of the CMB [157, 158] place constraints on the amplitudes and wavelengths of superhorizon perturbations.

A well-known application of the Grishchuk-Zel'dovich effect uses CMB observations to place a lower bound on the size of the nearly homogeneous patch that contains the observable Universe. This bound was first derived for an Einstein-de Sitter universe [85, 159], and then for an open universe [160, 161]. Most recently, an analysis of the WMAP first-year data [158] found that our nearly homogeneous patch of the Universe extends to 3900 times the cosmological horizon [162]. All of these analyses considered a statistically isotropic distribution of power in superhorizon perturbations and then asked how large the wavelength of order-unity perturbations needed to be in order to be consistent with the observed CMB anisotropies.

In this paper, we analyze the CMB anisotropies induced by a single superhorizon adiabatic perturbation mode rather than an isotropic distribution of superhorizon inhomogeneities. A single-mode superhorizon perturbation to the gravitational potential would naively be expected to generate a dipolar CMB anisotropy with an amplitude comparable to the perturbation amplitude across the observable Universe. This is not the case in an Einstein-de Sitter universe, however, because the intrinsic dipole in the CMB produced by the perturbation is exactly cancelled by the Doppler dipole induced by our peculiar motion [85, 159, 163]. We show that the same cancellation occurs for an adiabatic superhorizon perturbation in a flat universe with a cosmological constant ( $\Lambda$ ), cold dark matter (CDM), and radiation. The strongest constraints to the amplitude and wavelength of a single superhorizon mode therefore arise from measurements of the CMB quadrupole and octupole. These constraints are less stringent than those derived for modes in a realization of a random-phase

---

<sup>1</sup>This chapter was adapted from *Superhorizon perturbations and the cosmic microwave background*, Adrienne L. Erickcek, Sean M. Carroll, and Marc Kamionkowski; Phys. Rev. D **78**, 083012 (2008). Reproduced here with permission, copyright (2008) by the American Physical Society.

random field because it is possible to choose the phase of a single sinusoidal perturbation in such a way that there is no resulting quadrupole anisotropy.

Single-mode superhorizon perturbations have received attention recently [164, 165, 166, 167, 168] because they introduce a special direction in our Universe and could be responsible for observed deviations from statistical isotropy in the CMB [39, 40, 41, 42, 43, 44, 45, 46, 52] and for the anomalous bulk velocity detected through the kinetic Sunyaev-Zel’dovich effect [169] and in peculiar velocity surveys [170]. In the next chapter, we will investigate how a superhorizon perturbation during slow-roll inflation can generate an anomalous feature of the CMB: the fluctuation amplitude on large scales ( $\ell \lesssim 64$ ) is 7% larger on one side of the sky than on the other side [46]. We will first consider a perturbation to the inflaton field, but we will find that the perturbation required to generate the observed power asymmetry induces large-scale anisotropies in the CMB that are too large to be consistent with measurements of the CMB octupole. We will then consider a multi-field model of inflation in which a subdominant field, called the curvaton, is responsible for generating primordial perturbations [86, 87, 88, 89]. We will find that a superhorizon perturbation in the curvaton field can generate the observed power asymmetry without inducing prohibitively large CMB anisotropies. In this chapter, we will use these findings as an example of how one may apply the CMB constraints to single-mode superhorizon perturbations derived here.

We begin in Section 5.2 by reviewing the Grishchuk-Zel’dovich effect for adiabatic perturbations. In Section 5.3, we derive the CMB anisotropy induced by a sinusoidal superhorizon perturbation in the gravitational potential, as would arise from a sinusoidal inflaton fluctuation. We also show in Section 5.3 that a superhorizon adiabatic perturbation does not generate a large dipolar anisotropy in a  $\Lambda$ CDM universe because the leading-order intrinsic dipole anisotropy is cancelled by the anisotropy induced by the Doppler effect. A sinusoidal curvaton fluctuation generates a potential perturbation that is not sinusoidal, and we derive the constraints to single-mode perturbations to the curvaton field in Section 5.4. We summarize our results in Section 5.5. Finally, an analytic demonstration of the dipole cancellation in a  $\Lambda$ CDM universe is presented in Appendix B.1, and the cancellation is shown to occur in flat universes containing a single fluid with an arbitrary constant equation of state in Appendix B.2.

## 5.2 The Grishchuk-Zel’dovich effect: A brief review

Working in conformal Newtonian gauge, we take the perturbed Friedmann-Robertson-Walker (FRW) metric to be

$$ds^2 = -(1 + 2\Psi)dt^2 + a^2(t)\delta_{ij}(1 - 2\Phi)dx^i dx^j, \quad (5.1)$$

where  $a$  is normalized to equal one today. In the absence of anisotropic stress,  $\Psi = \Phi$ . The primary sources of anisotropic stress are the quadrupole moments of the photon and neutrino distributions.

Since the perturbations we consider are superhorizon, the distance travelled by photons and neutrinos arriving at a point is far smaller than the wavelength of the perturbation. Therefore, the quadrupole moments of the photon and neutrino distributions are much smaller than the monopole moments, and the anisotropic stress is negligible [171, 172]. We will assume that  $\Psi = \Phi$  throughout this paper.

On large scales, intrinsic fluctuations in the CMB temperature are generated by metric perturbations through the Sachs-Wolfe effect [173]. The current temperature fluctuation at a particular point in the sky (specified by  $\hat{n}$ ) is given by

$$\left[ \frac{\Delta T}{T}(\hat{n}) \right]_{\text{SW+ISW}} = \frac{\Delta T}{T}(\tau_{\text{dec}}, \hat{n}x_{\text{dec}}) + \Psi(\tau_{\text{dec}}, \hat{n}x_{\text{dec}}) + 2 \int_{\tau_{\text{dec}}}^{\tau_0} \frac{d\Psi}{d\tau}[\tau, \hat{n}(\tau_0 - \tau)]d\tau, \quad (5.2)$$

where  $x_{\text{dec}}$  is the comoving distance to the surface of last scattering, and  $\tau = \int dt/a$  is the conformal time:  $\tau_{\text{dec}}$  is the conformal time at decoupling and  $\tau_0$  is the current conformal time. Given that the early universe was radiation-dominated, the Boltzmann and Einstein equations imply

$$\frac{\Delta T}{T}(\tau_{\text{dec}}) + \Psi(\tau_{\text{dec}}) = \Psi(\tau_{\text{dec}}) \left[ 2 - \frac{5}{3} \left\{ \frac{9}{10} \frac{\Psi_{\text{p}}}{\Psi(\tau_{\text{dec}})} \right\} \right], \quad (5.3)$$

where  $\Psi_{\text{p}}$  is the primordial value of  $\Psi$  at  $a = \tau = 0$ . This expression simplifies to the familiar  $\Psi(\tau_{\text{dec}})/3$  in the limit that the Universe was matter-dominated at the time of decoupling. We will refer to this as the Sachs-Wolfe (SW) effect, and the last term in Eq. (5.2) will be referred to as the integrated Sachs-Wolfe (ISW) effect.

We also observe a temperature fluctuation due to our peculiar motion [174, 175], which we will refer to as the Doppler effect:

$$\left[ \frac{\Delta T}{T}(\hat{n}) \right]_{\text{D}} = \hat{n} \cdot \vec{v}_{\text{net}}(\hat{n}) + v_{\text{net}}^2 \left[ (\hat{n} \cdot \hat{v}_{\text{net}})^2 - \frac{1}{2} \right] + \mathcal{O}(v_{\text{net}}^3), \quad (5.4)$$

where  $\vec{v}_{\text{net}}(\hat{n})$  is our current velocity relative to the fluid at the surface of last scattering in a given direction. If  $\vec{v}(\tau, \vec{x})$  is the proper peculiar velocity of an observer at conformal time  $\tau$  and position  $\vec{x}$  in the frame defined by Eq. (5.1), then

$$\vec{v}_{\text{net}}(\hat{n}) = \vec{v}(\tau_0, \vec{0}) - \vec{v}(\tau_{\text{dec}}, \hat{n}x_{\text{dec}}). \quad (5.5)$$

For superhorizon perturbations, there is a direct relationship between the potential perturbation  $\Psi$  and the proper peculiar velocity of an observer falling into the potential well:

$$\vec{v}(\tau, \vec{x}) = -\frac{2a^2}{H_0\Omega_{\text{M}}} \frac{H(a)}{H_0} \left( \frac{y}{4+3y} \right) \left[ \vec{\nabla}\Psi + \frac{d}{d \ln a} \vec{\nabla}\Psi \right]. \quad (5.6)$$

Throughout this paper,  $y \equiv a(1+z_{\text{eq}})$ , where  $z_{\text{eq}}$  is the redshift of matter-radiation equality,  $H(a) \equiv$

$(1/a)da/dt$ , and  $\Omega_M$  is the present-day ratio of the matter density to the critical density.

Thus we see that the SW, ISW, and Doppler effects on the CMB are all determined completely by the evolution of the gravitational potential  $\Psi$ . The rest of this Section is devoted to the derivation of  $\Psi_{\vec{k}}(a)$  for a single superhorizon adiabatic perturbation mode ( $k \ll H_0$ ) in a flat  $\Lambda$ CDM Universe that includes radiation. In this case, the Hubble parameter is given by

$$H^2(a) = H_0^2 \left[ \frac{\Omega_M}{a^4} \left( \frac{1}{1+z_{\text{eq}}} + a \right) + \Omega_\Lambda \right] \quad (5.7)$$

with  $\Omega_\Lambda = 1 - \Omega_M - \Omega_M(1+z_{\text{eq}})^{-1}$ .

Since no causal processes can separate the components of the density perturbation, the superhorizon perturbation may be treated as a perturbation to a single fluid with over-density in the fluid's rest frame  $\Delta_{\vec{k}}$  and peculiar velocity  $\vec{v}_{\vec{k}}$ . These two quantities are related through two coupled equations [171]: in a flat Universe with no entropy perturbations these equations are

$$\dot{\Delta}_{\vec{k}} - 3waH\Delta_{\vec{k}} = -(1+w)kv_{\vec{k}}, \quad (5.8)$$

$$\dot{v}_{\vec{k}} + aHv_{\vec{k}} = \frac{4}{3} \frac{w}{(1+w)^2} k\Delta_{\vec{k}} + k\Psi_{\vec{k}}, \quad (5.9)$$

where an overdot denotes differentiation with respect to  $\tau$ , and  $w \equiv p/\rho$  is the equation of state parameter for the perturbed fluid. Since matter and radiation are the only perturbed density components,  $w = 1/[3(1+y)]$ .

Differentiating Eq. (5.8) with respect to  $\tau$  gives an expression for  $\dot{v}_{\vec{k}}$  that may be used, along with Eq. (5.8) itself, to eliminate  $v_{\vec{k}}$  from Eq. (5.9). If matter and radiation are the only perturbed density components, the potential  $\Psi$  is related to  $\Delta_{\vec{k}}$  through [171]

$$\Psi_{\vec{k}} = -\frac{3}{4} \left( \frac{k_{\text{eq}}}{k} \right)^2 \frac{1+y}{y^2} \Delta_{\vec{k}}, \quad (5.10)$$

where  $k_{\text{eq}} = (1+z_{\text{eq}})^{-1}H_{\text{eq}}$ . This expression may be inserted into Eq. (5.8) to obtain Eq. (5.6) for  $\vec{v}$ , and it may be used to eliminate  $\Psi_{\vec{k}}$  from Eq. (5.9). The resulting differential equation for  $\Delta_{\vec{k}}$  is

$$\begin{aligned} \ddot{\Delta}_{\vec{k}} + \frac{y(5+3y)}{(4+3y)(1+y)} aH\dot{\Delta}_{\vec{k}} + \left[ \frac{-\dot{H}}{H^2 a(1+y)} - \frac{8+3y}{(4+3y)(1+y)} \right] a^2 H^2 \Delta_{\vec{k}} \\ + \frac{H_0^2}{4+3y} \left[ \frac{4}{3} \frac{k^2}{H_0^2} - \frac{(4+3y)^2}{4y^2} \frac{H_{\text{eq}}^2}{(1+z_{\text{eq}})^2 H_0^2} \right] \Delta_{\vec{k}} = 0, \quad (5.11) \end{aligned}$$

where  $H_{\text{eq}}^2 \simeq 2\Omega_M(1+z_{\text{eq}})^3 H_0^2$  is the Hubble parameter at the time of matter-radiation equality. Thus we see that the last term on the second line is always much larger than one, while  $k/H_0 \ll 1$  for a superhorizon mode. Therefore, we will neglect the  $k^2$  term in this equation.

Inserting Eq. (5.10) into Eq. (5.11) yields a differential equation for  $\Psi$ . It will be more convenient

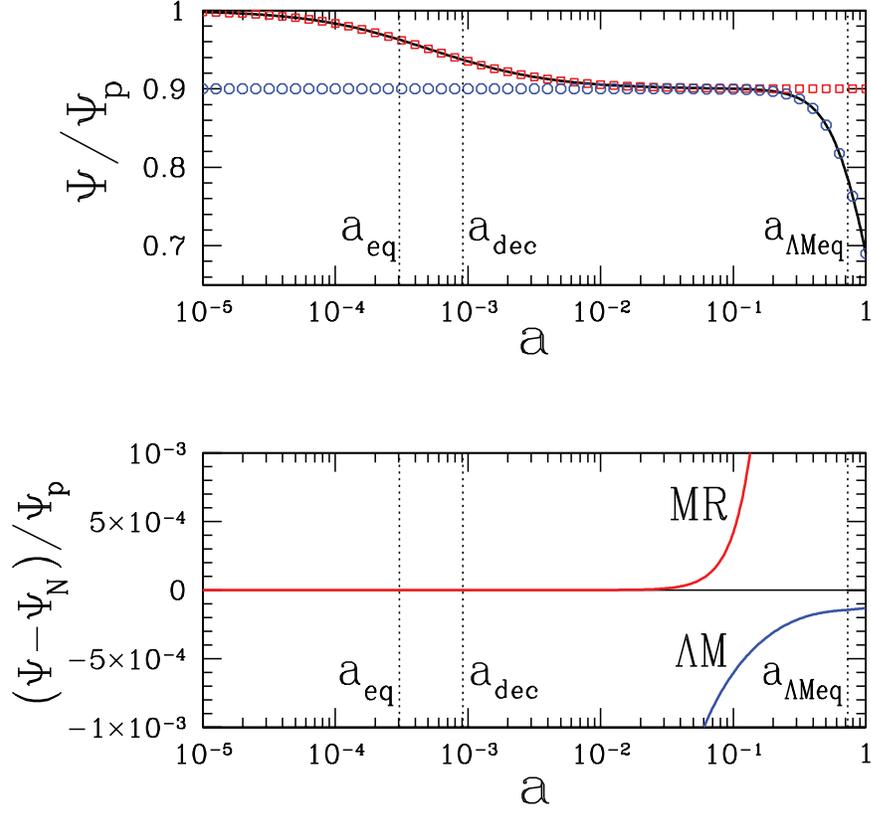


Figure 5.1: The evolution of  $\Psi$  in a  $\Lambda$ CDM Universe with radiation. In the top panel, the solid curve is the numerical solution to Eq. (5.12), the squares are the analytic solution for matter and radiation (MR) only, and the circles are the solution for  $\Lambda$  and matter ( $\Lambda M$ ) only. The bottom panel shows the difference between these two approximate solutions and the numerical solution.

to work with  $\Psi(a)$  than  $\Psi(\tau)$ :

$$\Psi_{\vec{k}}''(a) + \frac{1}{a} \left( 2 + \frac{16 + 9y}{4 + 3y} + \frac{d \ln H}{d \ln a} \right) \Psi_{\vec{k}}'(a) + \frac{1}{a^2} \left( 3 + \frac{4}{4 + 3y} - \frac{4 + 3y}{4y^4} \frac{H_{\text{eq}}^2}{H^2} + \frac{d \ln H}{d \ln a} \right) \Psi_{\vec{k}} = 0, \quad (5.12)$$

where a prime denotes differentiation with respect to  $a$ . Long before  $\Lambda$ -domination ( $\Omega_M/a^3 \gg \Omega_\Lambda$ ), this equation yields the usual expression for  $\Psi$  in a Universe with only matter and radiation [176]:

$$\Psi_{\text{MR}}(y) = \frac{\Psi_p}{10y^3} \left( 16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16 \right). \quad (5.13)$$

Long after matter-radiation equality ( $y \gg 1$ ), Eq. (5.12) is solved by

$$\Psi_{\Lambda M}(a) = \left[ \frac{9}{10} \frac{\Psi_p}{a} \right] \frac{5}{2} \Omega_M H_0^2 \tilde{H}(a) \int_0^a \frac{da'}{[\tilde{H}(a')a']^3}, \quad (5.14)$$

where  $\tilde{H}(a)$  is the radiation-free ( $z_{\text{eq}} \rightarrow \infty$ ) limit of  $H(a)$ .

We numerically solve Eq. (5.12) to obtain  $\Psi(a)$ , which is shown in Fig. 5.1. We use WMAP5+BAO+SN [18] best-fit values  $\Omega_M = 0.28$ ,  $z_{\text{eq}} = 3280$  and  $z_{\text{dec}} = 1090$ . We begin the numerical integration at  $a_i = 10^{-10}$ ; Taylor expansions of Eq. (5.13) and its derivative with respect to  $a$  around  $a = 0$  were used to set initial conditions for  $\Psi(a_i)$  and  $\Psi'(a_i)$  in terms of the primordial  $\Psi_p$ . Figure 5.1 also shows the solutions given by Eqs. (5.13) and (5.14) for comparison. A key feature of the numerical solution is the value of  $\Psi$  at decoupling:  $\Psi(\tau_{\text{dec}}) = 0.937\Psi_p$ , which is larger than the value  $\Psi$  attains during matter-domination [ $0.9\Psi_p$ ].

### 5.3 CMB anisotropies from superhorizon potential perturbations

Since  $kH_0^{-1} \ll 1$  for a superhorizon perturbation, it is desirable to expand  $\Psi(\tau, \vec{x})$  in powers of  $(\vec{k} \cdot \vec{x})$ . We generalize the expansion of a sine wave perturbation  $\Psi = \Psi_{\vec{k}} \sin(\vec{k} \cdot \vec{x} + \varpi)$  by considering a superhorizon potential perturbation of the form

$$\Psi(\tau, \vec{x}) = \Psi_{\vec{k}}(\tau) \left[ \sin \varpi_0 + \cos \varpi_1 (\vec{k} \cdot \vec{x}) - \frac{\sin \varpi_2}{2} (\vec{k} \cdot \vec{x})^2 - \frac{\cos \varpi_3}{6} (\vec{k} \cdot \vec{x})^3 + \mathcal{O}(k^4 x^4) \right]. \quad (5.15)$$

If the potential perturbation is a single sine wave, as would result from a sinusoidal fluctuation in the inflaton field, then all the  $\varpi_i$  phases are equal and correspond to the phase of the wave  $\varpi$ . We use a more general expression here because it will be useful when considering curvaton perturbations in the next section.

The expansion in powers of  $(\vec{k} \cdot \vec{x})$  of the CMB temperature anisotropy due to the SW effect follows directly from Eqs. (5.3) and (5.15). The corresponding expansion of the ISW effect is a little more involved. We start by rewriting the ISW term in Eq. (5.2) as

$$\left[ \frac{\Delta T}{T}(\hat{n}) \right]_{\text{ISW}} = 2 \int_{a_{\text{dec}}}^1 \frac{d\Psi}{da} [a, H_0^{-1} \{\chi_0 - \chi(a)\} \hat{n}] da, \quad (5.16)$$

where

$$\chi(a) \equiv H_0[\tau(a) - \tau_{\text{dec}}] = \int_{a_{\text{dec}}}^a \frac{da'}{(a')^2 H(a')/H_0}, \quad (5.17)$$

and  $\chi_0 \equiv \chi(a=1) = H_0 x_{\text{dec}}$ . We then use Eq. (5.15) to expand the integrand in powers of  $(\vec{k} \cdot \vec{x})$ . The resulting expression for the ISW effect is

$$\begin{aligned} \frac{[\Delta T/T]_{\text{ISW}}}{\Psi_{\vec{k}}(\tau_{\text{dec}})} &= \mathcal{I}_0 \sin \varpi_0 + (\mathcal{I}_0 - \mathcal{I}_1) \cos \varpi_1 (\vec{k} \cdot \vec{x}_{\text{dec}}) \\ &\quad - (\mathcal{I}_0 - 2\mathcal{I}_1 + \mathcal{I}_2) \frac{\sin \varpi_2}{2} (\vec{k} \cdot \vec{x}_{\text{dec}})^2 \\ &\quad - (\mathcal{I}_0 - 3\mathcal{I}_1 + 3\mathcal{I}_2 - \mathcal{I}_3) \frac{\cos \varpi_3}{6} (\vec{k} \cdot \vec{x}_{\text{dec}})^3, \end{aligned} \quad (5.18)$$

where  $\vec{x}_{\text{dec}} \equiv \hat{n}x_{\text{dec}}$  and we have defined

$$\mathcal{I}_n \equiv \frac{2}{(H_0 x_{\text{dec}})^n} \int_{a_{\text{dec}}}^1 \frac{\Psi'_{\vec{k}}(a)}{\Psi_{\vec{k}}(\tau_{\text{dec}})} \chi^n(a) da. \quad (5.19)$$

Finally, we need to expand the Doppler effect in powers of  $(\vec{k} \cdot \vec{x})$ . Recall from Eq. (5.4) that, to first order in  $v$ ,  $[\Delta T/T]_{\text{D}} = [\vec{v}(\tau_0, \vec{0}) - \vec{v}(\tau_{\text{dec}}, \vec{x}_{\text{dec}})] \cdot \hat{n}$ . Since the time and spatial dependence of  $\Psi$  are separable, as shown in Eq. (5.15), we can isolate the time dependence of  $\vec{v}(\tau, \vec{x})$  by defining a dimensionless quantity

$$\mathcal{V}(\tau) \equiv \frac{-2a^2}{x_{\text{dec}} H_0 \Omega_{\text{M}}} \frac{H(a)}{H_0} \left( \frac{y}{4+3y} \right) \left[ \frac{\Psi_{\vec{k}}(a)}{\Psi_{\vec{k}}(\tau_{\text{dec}})} + \frac{d}{d \ln a} \frac{\Psi_{\vec{k}}(a)}{\Psi_{\vec{k}}(\tau_{\text{dec}})} \right]. \quad (5.20)$$

It follows from Eq. (5.6) for  $\vec{v}(\tau, \vec{x})$  and Eq. (5.15) for  $\Psi$  that

$$\vec{v}(\tau, \vec{x}) = x_{\text{dec}} \mathcal{V}(\tau) \vec{\nabla} \Psi(\tau_{\text{dec}}, \vec{x}). \quad (5.21)$$

Since  $\vec{\nabla} \Psi(\tau_{\text{dec}}, \vec{0}) = \vec{k} \Psi_{\vec{k}}(\tau_{\text{dec}}) \cos \varpi_1$ , our current velocity only contributes to the  $(\vec{k} \cdot \vec{x})$  term of the expansion. Meanwhile,

$$\vec{\nabla} \Psi(\tau_{\text{dec}}, \vec{x}_{\text{dec}}) = \vec{k} \Psi_{\vec{k}}(\tau_{\text{dec}}) \left[ \cos \varpi_1 - \sin \varpi_2 (\vec{k} \cdot \vec{x}_{\text{dec}}) - \frac{\cos \varpi_3}{2} (\vec{k} \cdot \vec{x}_{\text{dec}})^2 \right], \quad (5.22)$$

so the velocity of the fluid at the surface of last scattering contributes to all terms in the  $(\vec{k} \cdot \vec{x})$  expansion. The temperature anisotropy due to the Doppler effect is therefore

$$\begin{aligned} \frac{1}{\Psi_{\vec{k}}(\tau_{\text{dec}})} \left[ \frac{\Delta T}{T}(\hat{n}) \right]_{\text{D}} &= [\mathcal{V}(\tau_0) - \mathcal{V}(\tau_{\text{dec}})] \cos \varpi_1 (\vec{k} \cdot \vec{x}_{\text{dec}}) \\ &+ 2\mathcal{V}(\tau_{\text{dec}}) \frac{\sin \varpi_2}{2} (\vec{k} \cdot \vec{x}_{\text{dec}})^2 \\ &+ 3\mathcal{V}(\tau_{\text{dec}}) \frac{\cos \varpi_3}{6} (\vec{k} \cdot \vec{x}_{\text{dec}})^3. \end{aligned} \quad (5.23)$$

Combining the SW effect, the ISW effect, and the Doppler effect gives the total CMB temperature anisotropy produced by a potential perturbation of the form given in Eq. (5.15):

$$\frac{\Delta T}{T}(\hat{n}) = \Psi_{\vec{k}}(\tau_{\text{dec}}) \left[ \mu(kx_{\text{dec}}) \delta_1 \cos \varpi_1 - \mu^2(kx_{\text{dec}})^2 \delta_2 \frac{\sin \varpi_2}{2} - \mu^3(kx_{\text{dec}})^3 \delta_3 \frac{\cos \varpi_3}{6} \right], \quad (5.24)$$

where  $\mu \equiv \hat{k} \cdot \hat{n}$ . We have discarded the monopolar components of the SW and ISW effects since

they only shift the mean CMB temperature and are therefore unobservable. The  $\delta_i$  are

$$\delta_1 = \mathcal{S} + (\mathcal{I}_0 - \mathcal{I}_1) + [\mathcal{V}(\tau_0) - \mathcal{V}(\tau_{\text{dec}})] \quad (5.25)$$

$$\delta_2 = \mathcal{S} + (\mathcal{I}_0 - 2\mathcal{I}_1 + \mathcal{I}_2) - 2\mathcal{V}(\tau_{\text{dec}}) \quad (5.26)$$

$$\delta_3 = \mathcal{S} + (\mathcal{I}_0 - 3\mathcal{I}_1 + 3\mathcal{I}_2 - \mathcal{I}_3) - 3\mathcal{V}(\tau_{\text{dec}}) \quad (5.27)$$

where, from Eq. (5.3), we have defined

$$\mathcal{S} \equiv 2 - \frac{5}{3} \left[ \frac{\frac{9}{10}\Psi_{\text{p}}}{\Psi(\tau_{\text{dec}})} \right] \quad (5.28)$$

to be the SW effect's contribution to the anisotropy. The contribution of the ISW effect to the  $\delta_i$  coefficients is contained in the  $\mathcal{I}_n$  terms and follows from Eq. (5.18). Finally, the  $\mathcal{V}$  terms are the contribution from the Doppler effect and follow from Eq. (5.23).

In an Einstein-de Sitter universe, the dipole induced by a superhorizon perturbation through the SW effect is exactly cancelled by the Doppler dipole [85, 159, 163]; since there is no ISW effect, we have  $\delta_1 = 0$  in this case. We find that this cancellation extends to flat universes that contain radiation and a cosmological constant in addition to matter, provided that the superhorizon perturbation is adiabatic. Figure 5.2 shows the SW+Doppler and ISW terms in  $\delta_1$ ; in the presence of radiation and  $\Lambda$ , the SW+Doppler term is no longer zero, but it is equal and opposite to the ISW term for all values of  $\Omega_{\text{M}}$ ,  $z_{\text{eq}}$ , and  $z_{\text{dec}}$ . Figure 5.2 also shows the contributions to  $\delta_1$  for a  $\Lambda$ CDM universe with no radiation; an analytic derivation of  $\delta_1 = 0$  in the absence of radiation is given in Appendix B.1.

The cancellation of the  $\mathcal{O}(kx_{\text{dec}})$  temperature anisotropies also occurs if the universe contains a fluid other than matter or radiation, as shown in Appendix B.2, which leads us to suspect that  $\delta_1 = 0$  is a ubiquitous feature of adiabatic perturbations in flat cosmologies. This conclusion is also supported by going to synchronous gauge, in which galaxies have zero peculiar velocity by construction. As shown in Ref. [177], any individual observer in a flat spacetime with a superhorizon adiabatic perturbation can choose Riemann normal coordinates consistent with synchronous gauge such that the metric in the neighborhood of that observer is described by an unperturbed FRW metric plus corrections that are  $\mathcal{O}(k^2H_0^{-2})$ . Since there are no  $\mathcal{O}(k)$  perturbations and the observer has no extragalactic peculiar velocity in this gauge and therefore sees no Doppler effects due to the superhorizon perturbation, we would not expect the observer to see a temperature anisotropy that is proportional to  $\mathcal{O}(kx_{\text{dec}})$ . It follows that the  $\mathcal{O}(kx_{\text{dec}})$  temperature anisotropies in any other gauge must sum to zero.

Since  $\delta_1 = 0$ , the leading order anisotropy will be the  $\delta_2$  term, unless  $\sin \varpi_2 = 0$ , in which case the  $\delta_3$  term will be the leading order contribution. For a flat  $\Lambda$ CDM universe with  $\Omega_{\text{M}} = 0.28$ ,  $z_{\text{eq}} = 3280$ , and  $z_{\text{dec}} = 1090$ ,  $\delta_2 = 0.33$  and  $\delta_3 = 0.35$ .

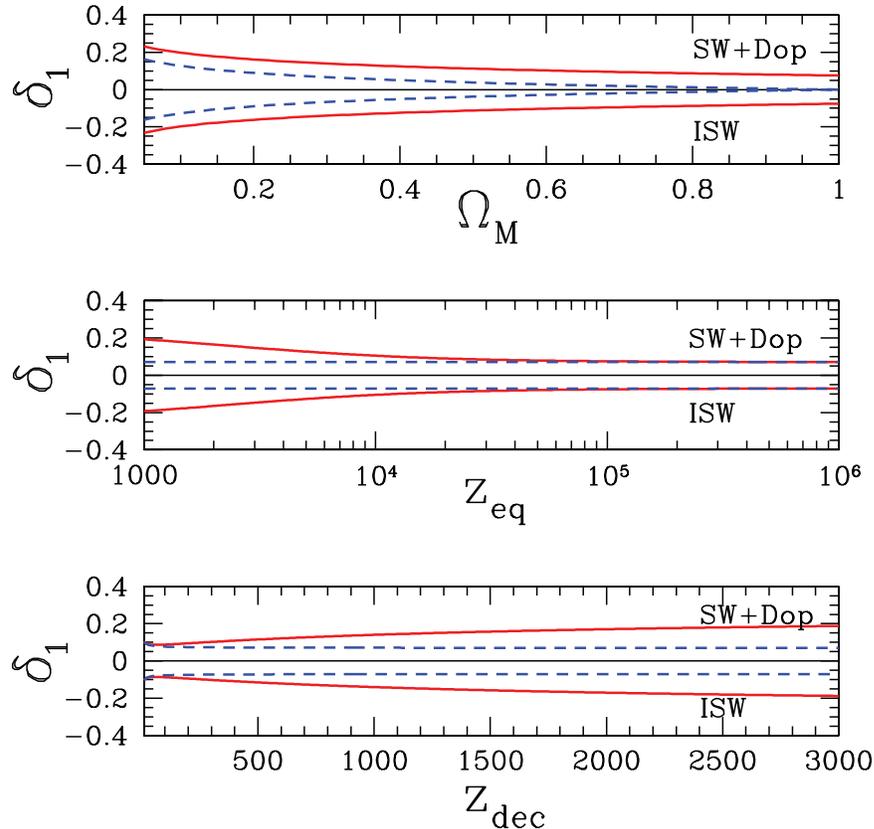


Figure 5.2: The contributions to  $\delta_1$ , as defined in Eqs. (5.24) and (5.25), from the Sachs-Wolfe (SW) effect, the Doppler (Dop) effect, and the integrated Sachs-Wolfe (ISW) effect. The solid curves correspond to a  $\Lambda$ CDM universe with radiation (fiducial parameters  $\Omega_M = 0.28$ ,  $z_{\text{eq}} = 3280$ , and  $z_{\text{dec}} = 1090$ ). The dashed curves correspond to a  $\Lambda$ CDM universe with no radiation. In all cases, the SW+Dop effect is exactly cancelled by the ISW effect so that  $\delta_1 = 0$ . An approximate solution for  $\Psi(a)$ , accurate to within 0.05%, was used to generate this figure.

As a brief aside, let us consider the  $v_{\text{net}}^2$  term in Eq. (5.4) for the Doppler effect. Like the  $\delta_2$  term in Eq. (5.24), this term is proportional to  $(\vec{k} \cdot \vec{x}_{\text{dec}})^2$ . Since  $v_{\text{net}}$  could be near unity if the gradient of the superhorizon perturbation is large, we may be concerned that this term will produce a quadrupolar anisotropy that is comparable to, or even larger than, the quadrupole given by the  $\delta_2$  term in Eq. (5.24). Fortunately, this concern is unfounded because the Doppler shift of the intrinsic dipole also produces a quadrupole: if there is an intrinsic temperature pattern  $T_{\text{int}}(\hat{n})$  in the CMB, then an observer moving with velocity  $\vec{v}_{\text{net}}$  with respect to the CMB will measure a temperature pattern  $T_{\text{int}}(\hat{n})/[\gamma(1 - \vec{v}_{\text{net}} \cdot \hat{n})]$ . Taylor expanding in  $v_{\text{net}}$  and dropping terms that are isotropic gives

the observed temperature fluctuation if the CMB has an intrinsic anisotropy:

$$\begin{aligned} \frac{T_{\text{obs}}}{T}(\hat{n}) &= \left[ 1 + \frac{\Delta T_{\text{int}}(\hat{n})}{T} \right] [1 + \vec{v}_{\text{net}} \cdot \hat{n} + v_{\text{net}}^2 (\hat{v}_{\text{net}} \cdot \hat{n})^2], \\ &= 1 + (\vec{k} \cdot \vec{x}_{\text{dec}}) \delta_1 \Psi_{\vec{k}}(\tau_{\text{dec}}) \cos \varpi_1 + \dots + \\ &\quad (\vec{k} \cdot \vec{x}_{\text{dec}})^2 \delta_1 [\mathcal{V}(\tau_0) - \mathcal{V}(\tau_{\text{dec}})] \Psi_{\vec{k}}^2(\tau_{\text{dec}}) \cos^2 \varpi_1, \end{aligned}$$

where the second line follows from  $\Delta T_{\text{int}}(\hat{n})/T \simeq (\vec{k} \cdot \vec{x}_{\text{dec}}) \Psi_{\vec{k}}(\tau_{\text{dec}}) \cos \varpi_1 [\mathcal{S} + \mathcal{I}_0 - \mathcal{I}_1]$  and the ellipses contain the linear quadrupolar and octupolar anisotropies given in Eq. (5.24). Given that  $\delta_1 = 0$ , we see that the Doppler quadrupole is exactly cancelled by the Doppler-shifted intrinsic dipole. Of course, this Doppler quadrupole is nonlinear in  $\Psi$ , so it is questionable to analyze it using linear theory. Nevertheless, this argument shows that this term is no more alarming than any other higher-order term in  $\Psi$ .

To compare the CMB anisotropy given by Eq. (5.24) to observations, we must decompose this anisotropy into multipole moments:

$$\frac{\Delta T}{T}(\hat{n}) = \sum_{\ell, m} a_{\ell m} Y_{\ell m}(\hat{n}). \quad (5.29)$$

Given the addition theorem of spherical harmonics,

$$P_{\ell}(\mu) = \frac{4\pi}{2\ell + 1} \sum_{m=-\ell}^{\ell} Y_{\ell m}^*(\hat{k}) Y_{\ell m}(\hat{n}), \quad (5.30)$$

the values of  $a_{1m}$ ,  $a_{2m}$  and  $a_{3m}$  are easily obtained from Eq. (5.24). It is also clear that each  $a_{\ell m}$  is proportional to  $Y_{\ell m}^*(\hat{k})$ . Consequently, if  $\hat{k}$  is chosen to lie on the  $z$  axis, then the only nonzero moments are  $a_{10}$ ,  $a_{20}$  and  $a_{30}$ . In this case, with  $\delta_1 = 0$ ,

$$a_{10} = -\sqrt{\frac{4\pi}{3}} (kx_{\text{dec}})^3 \delta_3 \frac{\cos \varpi_3}{10} \Psi_{\vec{k}}(\tau_{\text{dec}}), \quad (5.31)$$

$$a_{20} = -\sqrt{\frac{4\pi}{5}} (kx_{\text{dec}})^2 \delta_2 \frac{\sin \varpi_2}{3} \Psi_{\vec{k}}(\tau_{\text{dec}}), \quad (5.32)$$

$$a_{30} = -\sqrt{\frac{4\pi}{7}} (kx_{\text{dec}})^3 \delta_3 \frac{\cos \varpi_3}{15} \Psi_{\vec{k}}(\tau_{\text{dec}}). \quad (5.33)$$

Thus we see that even though  $\delta_1 = 0$ , a superhorizon potential perturbation still induces a dipolar anisotropy in the CMB. However, this anisotropy is suppressed by a factor of  $(kx_{\text{dec}})^3$ . Moreover, it is comparable in magnitude to the induced octupolar anisotropy. Since measurements of  $|a_{10}|$  are contaminated by our peculiar velocity, the upper bound on  $|a_{10}|$  ( $|a_{10}| \lesssim 10^{-3}$ ) is much higher than the upper bound on  $|a_{30}|$ . Therefore, the most restrictive constraints on  $\Psi_{\vec{k}}(\tau_{\text{dec}})$  come from

Eqs. (5.32) and (5.33):

$$(kx_{\text{dec}})^2 |\Psi_k(\tau_{\text{dec}}) \sin \varpi_2| \lesssim 5.8 Q \quad (5.34)$$

$$(kx_{\text{dec}})^3 |\Psi_k(\tau_{\text{dec}}) \cos \varpi_3| \lesssim 32 \mathcal{O} \quad (5.35)$$

where  $Q$  and  $\mathcal{O}$  are upper bounds on  $|a_{20}|$  and  $|a_{30}|$ , respectively, in a coordinate system aligned with the superhorizon perturbation ( $\hat{k} = \hat{z}$ ).

Since other primordial perturbations, including smaller-scale modes, may also contribute to the measured values of  $|a_{20}|$  and  $|a_{30}|$  in a way that suppresses the perturbation from the single superhorizon mode we have been considering, the  $|a_{20}|$  and  $|a_{30}|$  values from the superhorizon mode may be as large as the largest values of  $|a_{20}|$  and  $|a_{30}|$  that are consistent with the measured variance in these moments. We take  $Q = 3\sqrt{C_2} \lesssim 1.8 \times 10^{-5}$  and  $\mathcal{O} = 3\sqrt{C_3} \lesssim 2.7 \times 10^{-5}$ , three times the measured rms values of the quadrupole and octupole [178], as  $3\sigma$  upper limits.

When a superhorizon potential perturbation is invoked to generate a power asymmetry in the CMB, as in Chapter 6, it is the variation of  $\Psi$  across the surface of last scattering,

$$\Delta\Psi(\tau_{\text{dec}}) \simeq |\Psi_k(\tau_{\text{dec}})(kx_{\text{dec}}) \cos \varpi_1|, \quad (5.36)$$

that is the relevant quantity. For a given  $\Delta\Psi(\tau_{\text{dec}})$ , the induced CMB quadrupole and octupole can be made arbitrarily small by decreasing  $(kx_{\text{dec}})$ . However, if the superhorizon perturbation is a single mode of the form  $\Psi = \Psi_{\vec{k}} \sin(\vec{k} \cdot \vec{x} + \varpi)$ , then demanding that  $\Psi \lesssim 1$  everywhere, even outside our Hubble volume, leads to an additional constraint:  $\Psi_{\vec{k}}(\tau_{\text{dec}}) \lesssim 1$  implies that  $(kx_{\text{dec}}) \gtrsim \Delta\Psi / \cos \varpi$ . Moreover, since  $\varpi_2 = \varpi_3 = \varpi$  in this case, the bounds given by Eqs. (5.34) and (5.35) imply that

$$\Delta\Psi(\tau_{\text{dec}})(kx_{\text{dec}}) |\tan \varpi| \lesssim 5.8 Q \quad (5.37)$$

$$\Delta\Psi(\tau_{\text{dec}})(kx_{\text{dec}})^2 \lesssim 32 \mathcal{O}. \quad (5.38)$$

Since the octupole constraint on  $\Delta\Psi(\tau_{\text{dec}})$  is independent of  $\varpi$  while the quadrupole constraint vanishes if  $\varpi = 0$ , the maximum allowed value for  $\Delta\Psi(\tau_{\text{dec}})$  is obtained when  $\varpi = 0$ . In this case, we have

$$\Delta\Psi(\tau_{\text{dec}}) \lesssim \text{Min} \left[ (kx_{\text{dec}}), \frac{32 \mathcal{O}}{(kx_{\text{dec}})^2} \right], \quad (5.39)$$

where the first bound follows from  $\Psi_{\vec{k}}(\tau_{\text{dec}}) \lesssim 1$  and the second follows from Eq. (5.38). Consequently, the maximum value for  $\Delta\Psi(\tau_{\text{dec}})$  is obtained when  $\Delta\Psi(\tau_{\text{dec}}) \simeq [32 \mathcal{O}]^{1/3} \simeq 0.095$ . In this case,  $\Delta\Psi(\tau_{\text{dec}}) \simeq (kx_{\text{dec}})$ , which implies that the wavelength of the superhorizon mode is 65 times larger than the particle horizon. This is much smaller than the lower bound on the wavelength of order-unity perturbations found in Ref. [162] because we have placed ourselves at the node ( $\varpi = 0$ )

of a single-mode perturbation, thus eliminating the induced quadrupolar anisotropy that would have otherwise provided a much more stringent constraint. Nevertheless, the octupole constraint alone is sufficient to rule out an inflaton perturbation large enough to generate the observed power asymmetry in the CMB, as we will see in Chapter 6.

## 5.4 Application to curvaton perturbations

The constraints given by Eqs. (5.34) and (5.35) may also be applied to superhorizon perturbations that are not describable by a single sine wave fluctuation in  $\Psi$ . In this section, we will show how these constraints limit the amplitude of a superhorizon fluctuation in a curvaton field. Superhorizon curvaton fluctuations may be a generic feature of the curvaton model of inflation [179], and in Chapter 6, we will see that a superhorizon curvaton fluctuation is capable of generating the observed power asymmetry in the CMB.

In the curvaton model [86, 87, 88, 89], there are two fields present during inflation: the inflaton dominates the energy density of the Universe and drives inflation while the curvaton ( $\sigma$ ) generates some or all of the primordial perturbations. The curvaton potential is assumed to be  $V(\sigma) = (1/2)m_\sigma^2\sigma^2$ , with  $m_\sigma \ll H_{\text{inf}}$ , where  $H_{\text{inf}}$  is the value of the Hubble parameter during inflation. Consequently, the curvaton is effectively massless during inflation and remains frozen at its initial value. After inflation ends and  $m_\sigma \simeq H$ , the curvaton will oscillate about its minimum, behaving like a cold gas of  $\sigma$  particles. The curvaton is then assumed to decay into radiation prior to neutrino decoupling, generating a gauge-invariant curvature perturbation [180, 181]  $\zeta \simeq (R/3)\delta\rho_\sigma/\rho_\sigma$ , where  $R \equiv 0.75(\rho_\sigma/\rho_{\text{tot}})$  is the fraction of the total energy density in the curvaton field just prior to its decay.

During radiation domination, a curvature perturbation  $\zeta$  corresponds to a potential perturbation  $\Psi = -(2/3)\zeta$ . We assume that the curvaton decay occurred early enough that  $\Psi_p \simeq -(2R/9)\delta\rho_\sigma/\rho_\sigma$  (see Chapter 6 for more details). In this case, a perturbation in the curvaton field induces a potential perturbation at decoupling given by

$$\Psi(\tau_{\text{dec}}) \simeq -\frac{R}{5} \left[ \frac{\Psi(\tau_{\text{dec}})}{\frac{9}{10}\Psi_p} \right] \left[ 2 \left( \frac{\delta\sigma}{\bar{\sigma}} \right) + \left( \frac{\delta\sigma}{\bar{\sigma}} \right)^2 \right], \quad (5.40)$$

where  $\bar{\sigma}$  is the spatially homogeneous background value of the curvaton field and  $\sigma(\vec{x}) = \bar{\sigma} + \delta\sigma(\vec{x})$ . We consider a superhorizon sinusoidal perturbation to the curvaton field  $\delta\sigma = \sigma_{\vec{k}} \sin(\vec{k} \cdot \vec{x} + \varpi_\sigma)$ . Since adding  $\pi$  to  $\varpi_\sigma$  changes the sign of  $\delta\sigma$ , we may assume that  $\sigma_{\vec{k}}/\bar{\sigma} > 0$  without loss of generality. The potential perturbation induced by  $\delta\sigma$  may be expanded in powers of  $(\vec{k} \cdot \vec{x})$  and the resulting

expression is Eq. (5.15) with

$$\sin \varpi_0 = \sin \varpi_\sigma + \left(\frac{\sigma_{\vec{k}}}{2\bar{\sigma}}\right) \sin^2 \varpi_\sigma, \quad (5.41)$$

$$\cos \varpi_1 = \cos \varpi_\sigma + \left(\frac{\sigma_{\vec{k}}}{2\bar{\sigma}}\right) \sin 2\varpi_\sigma, \quad (5.42)$$

$$\sin \varpi_2 = \sin \varpi_\sigma - \left(\frac{\sigma_{\vec{k}}}{\bar{\sigma}}\right) \cos 2\varpi_\sigma, \quad (5.43)$$

$$\cos \varpi_3 = \cos \varpi_\sigma + 2 \left(\frac{\sigma_{\vec{k}}}{\bar{\sigma}}\right) \sin 2\varpi_\sigma. \quad (5.44)$$

Since  $a_{20} \propto \sin \varpi_2$ , we can see that it will not be possible to choose a phase  $\varpi_\sigma$  such that the CMB quadrupole vanishes for all values of  $\sigma_{\vec{k}}$ . For specific values of  $\sigma_{\vec{k}}$ , however, there will be values of  $\varpi_\sigma$  for which  $\sin \varpi_2 = 0$ .

The constraint from the CMB quadrupole [Eq. (5.34)] implies

$$R \left| \frac{\Delta\sigma}{\bar{\sigma}} \right| \left| \tan \varpi_\sigma - \frac{\Delta\sigma}{\bar{\sigma}(kx_{\text{dec}})} \frac{\cos 2\varpi_\sigma}{\cos^2 \varpi_\sigma} \right| \lesssim \frac{5}{2} \left( \frac{5.8 Q}{kx_{\text{dec}}} \right) \left[ \frac{\frac{9}{10} \Psi_{\text{p}}}{\Psi(\tau_{\text{dec}})} \right], \quad (5.45)$$

where  $\Delta\sigma = \sigma_{\vec{k}}(kx_{\text{dec}}) \cos \varpi_\sigma$  is the variation in the curvaton field across the surface of last scattering. If  $\varpi_\sigma = 0$ , then the CMB quadrupole anisotropy is sourced exclusively by the term in Eq. (5.40) that is proportional to  $(\delta\sigma)^2$ , and the resulting constraint is independent of  $kx_{\text{dec}}$ . In this case, Eq. (5.45) reduces to

$$R \left( \frac{\Delta\sigma}{\bar{\sigma}} \right)^2 \lesssim \frac{5}{2} (5.6 Q) \text{ for } \varpi_\sigma = 0, \quad (5.46)$$

where we have used  $\Psi(\tau_{\text{dec}}) = 0.937\Psi_{\text{p}}$ , as derived in Section 5.2.

If  $\varpi_\sigma$  is nonzero, then the quadrupole constraint given by Eq. (5.45) will depend on  $kx_{\text{dec}}$ . For instance, if  $\varpi_\sigma = \pi/4$ , Eq. (5.45) reduces to

$$R \left| \frac{\Delta\sigma}{\bar{\sigma}} \right| \lesssim \frac{5}{2} \left( \frac{5.6 Q}{kx_{\text{dec}}} \right) \text{ for } \varpi_\sigma = \frac{\pi}{4}. \quad (5.47)$$

This upper bound may be made arbitrarily large by decreasing  $kx_{\text{dec}}$ . However, the condition that  $\delta\sigma \lesssim \bar{\sigma}$  leads to a lower bound on  $kx_{\text{dec}}$ :  $kx_{\text{dec}} \gtrsim \Delta\sigma/(\bar{\sigma} \cos \varpi_\sigma)$ . We now set  $kx_{\text{dec}} = \Delta\sigma/(\bar{\sigma} \cos \varpi_\sigma)$  (or equivalently  $\sigma_{\vec{k}} = \bar{\sigma}$ ) and consider how the quadrupole constraint depends on the phase of the curvaton wave. Figure 5.3 shows the quadrupole bound on  $R$  as a function of  $\varpi_\sigma$  for  $|\Delta\sigma/\bar{\sigma}| = 0.2$ . As mentioned earlier, there are some values of  $\varpi_\sigma$  for which there is no induced CMB quadrupole; in Fig. 5.3 we see that setting  $\sigma_{\vec{k}} = \bar{\sigma}$  implies that the quadrupole constraint is lifted if  $\varpi_\sigma = \pi/6$  or  $5\pi/6$ . We also note that setting  $\sigma_{\vec{k}} = \bar{\sigma}$  implies an additional constraint on  $\Delta\sigma/\bar{\sigma}$ : the curvaton perturbation is superhorizon only if  $|\Delta\sigma/\bar{\sigma}| < |\cos \varpi_\sigma|$ .

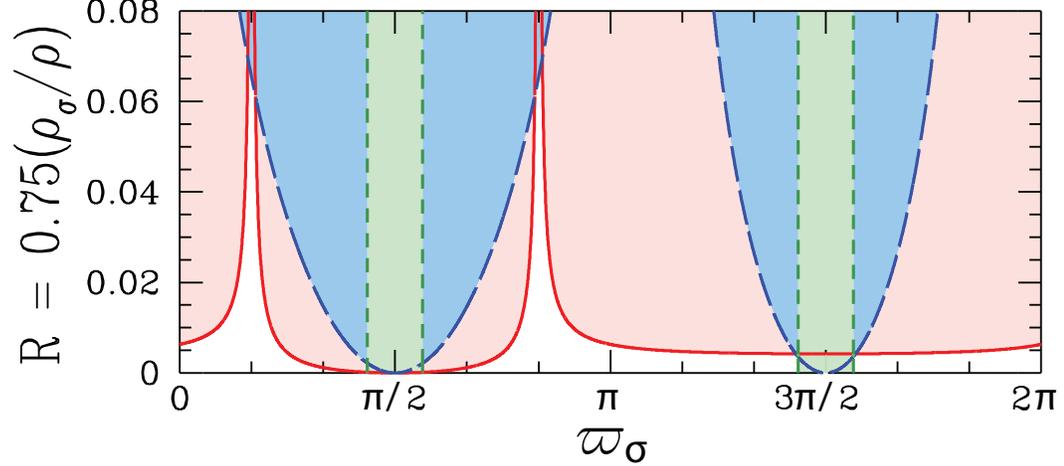


Figure 5.3: The upper bounds on  $R$  (the fraction of the energy density in the curvaton field just prior to its decay) for a superhorizon curvaton fluctuation with  $\delta\sigma = \bar{\sigma} \sin(\vec{k} \cdot \vec{x} + \varpi_\sigma)$ . In this plot,  $|\Delta\sigma/\bar{\sigma}| = (kx_{\text{dec}}) |\cos \varpi_\sigma| = 0.2$  and the shaded regions are excluded. The solid curve is the bound placed by the CMB quadrupole, and it vanishes for  $\varpi_\sigma = \pi/6$  and  $\varpi_\sigma = 5\pi/6$ . The long-dashed curve is the bound from the CMB octupole. The shaded areas enclosed by the short-dashed lines around  $\varpi_\sigma = \pi/2$  and  $\varpi_\sigma = 3\pi/2$  are not allowed because  $(kx_{\text{dec}}) > 1$  in these regions.

The CMB octupole constraint implied by Eq. (5.35) for a curvaton perturbation is

$$R \left| \frac{\Delta\sigma}{\bar{\sigma}} \right| \left| 1 + 2 \frac{\Delta\sigma}{\bar{\sigma}(kx_{\text{dec}})} \frac{\sin 2\varpi_\sigma}{\cos^2 \varpi_\sigma} \right| \lesssim \frac{5}{2} \left[ \frac{32 \mathcal{O}}{(kx_{\text{dec}})^2} \right] \left[ \frac{\frac{9}{10} \Psi_{\text{p}}}{\Psi(\tau_{\text{dec}})} \right], \quad (5.48)$$

and the resulting upper bound on  $R$  for  $\sigma_{\vec{k}} = \bar{\sigma}$  and  $|\Delta\sigma/\bar{\sigma}| = 0.2$  is shown in Fig. 5.3. Since the CMB octupole generated by a superhorizon perturbation is suppressed by a factor of  $kx_{\text{dec}}$  relative to the CMB quadrupole, the octupole constraint is weaker than quadrupole constraint for most values of  $\varpi_\sigma$ . The only exceptions are  $\varpi_\sigma = \pi/6$  or  $5\pi/6$  since the quadrupole constraint is lifted for these phases if  $\sigma_{\vec{k}} = \bar{\sigma}$ . In this case, Eq. (5.48) implies

$$R \left| \frac{\Delta\sigma}{\bar{\sigma}} \right|^3 \lesssim \frac{5}{8} (31 \mathcal{O}) \text{ for } \varpi_\sigma = \left\{ \frac{\pi}{6} \text{ or } \frac{5\pi}{6} \right\} \text{ and } \sigma_{\vec{k}} = \bar{\sigma}, \quad (5.49)$$

and we see in Fig. 5.3 that this bound on  $R$  is far less restrictive than the quadrupole bound for other values of  $\varpi_\sigma$ .

In the next chapter, we will see that the CMB power asymmetry induced by a superhorizon curvaton fluctuation is proportional to  $(\Delta\sigma/\bar{\sigma})$ . Consequently, the quadrupole and octupole constraints establish upper bounds on  $R$  for a given power asymmetry, as shown in Fig. 5.4. This upper bound increases as the required power asymmetry decreases. There is also a lower bound on  $R$  that arises from limits to non-Gaussianity in the CMB, but there is a small range of  $R$  values for which it is possible to generate the observed CMB power asymmetry with a superhorizon curvaton fluctuation even if the induced CMB quadrupole does not vanish, as we will see in Chapter 6.

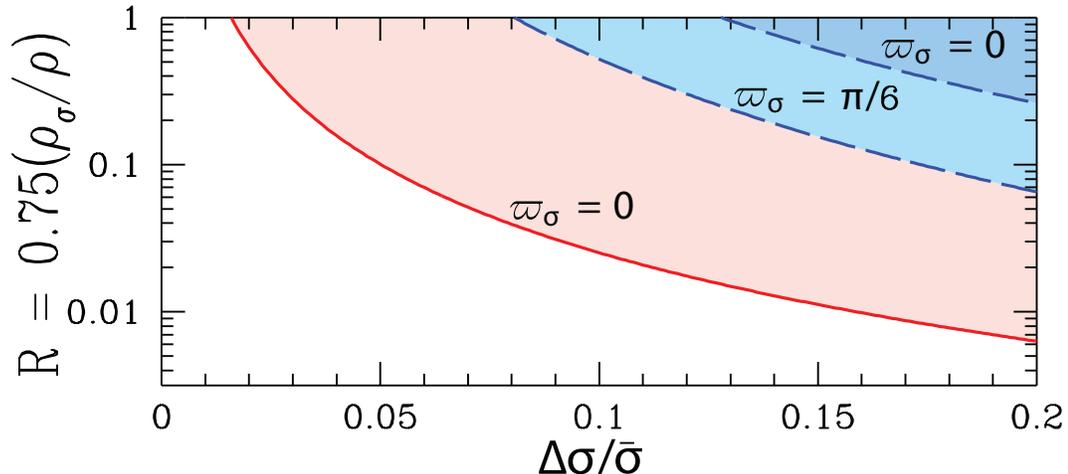


Figure 5.4: The  $R$ - $\Delta\sigma/\bar{\sigma}$  parameter space for a superhorizon curvaton fluctuation with  $\delta\sigma = \bar{\sigma} \sin(\vec{k} \cdot \vec{x} + \varpi_\sigma)$ ; the shaded regions are excluded.  $\Delta\sigma/\bar{\sigma}$  is the variation of the curvaton field across the surface of last scattering divided by the background value of the curvaton and  $R$  is the same as in Fig. 5.3. The solid curve is the bound from the CMB quadrupole, and the long-dashed curves are bounds from the CMB octupole. The bounds are plotted for two values of  $\varpi_\sigma$ ; there is no quadrupole bound if  $\varpi_\sigma = \pi/6$ .

## 5.5 Summary and discussion

Superhorizon perturbations generate large-scale anisotropies in the CMB through the Grishchuk-Zel'dovich effect [85]. In this chapter, we have derived the constraints to single-mode adiabatic superhorizon perturbations that arise from measurements of the CMB quadrupole and octupole. These constraints differ from those previously derived for an isotropic distribution of superhorizon perturbations [85, 159, 160, 161, 162] because the CMB anisotropies generated by a single-mode perturbation depend on the perturbation's phase.

We started by considering a sinusoidal superhorizon gravitational potential perturbation with wavenumber  $k \ll H_0$ . Since the leading-order term in the potential perturbation is proportional to  $(\vec{k} \cdot \vec{x})$ , it would be expected to generate a dipolar anisotropy of comparable amplitude in the CMB through the Sachs-Wolfe effect. However, the superhorizon perturbation also gives us a velocity with respect to the CMB, and the resulting Doppler dipole exactly cancels the leading-order intrinsic anisotropy generated by the SW and ISW effects, provided that the perturbation is adiabatic. This cancellation was known to occur in an Einstein-de Sitter universe [85, 159, 163], but we found that it also applies to flat  $\Lambda$ CDM universes with and without radiation, as well as in more exotic flat cosmological models.

Due to this cancellation of the CMB dipole, the leading-order constraints on adiabatic superhorizon fluctuations arise from measurements of the CMB quadrupole and octupole. If the potential perturbation is sinusoidal, as would be created by a sinusoidal fluctuation in the inflaton, then

putting ourselves at the node of the sine wave maximizes the difference in potential across the Universe while also eliminating the induced quadrupole anisotropy in the CMB. In this case, the CMB octupole provides the strongest constraint on the amplitude of the superhorizon perturbation:  $\Delta\Psi \lesssim 0.095$ , where  $\Delta\Psi$  is the variation of the potential  $\Psi$  across the surface of last scattering.

A fluctuation in a field that contains only a small fraction of the energy density of the Universe generates a smaller potential perturbation and, consequently, smaller CMB anisotropies. We consider a multi-field model of inflation in which a subdominant curvaton field generates primordial perturbations [86, 87, 88, 89]. For a given superhorizon fluctuation in the curvaton field, the measured values of the CMB quadrupole and octupole place upper bounds on the fraction of the total energy density contained in the curvaton field prior to its decay. Since a sinusoidal perturbation in the curvaton field generates a potential perturbation that is not sinusoidal, there is no value for the phase of the curvaton fluctuation that eliminates the induced CMB quadrupole for any superhorizon curvaton fluctuation. However, once the amplitude of the curvaton perturbation is specified, it is possible to choose a phase for which the induced CMB quadrupole vanishes. In this case, measurements of the CMB octupole still place an upper bound on the curvaton energy density, but this bound is significantly weaker than the bound from the CMB quadrupole that applies to curvaton fluctuations with different phases.

Superhorizon perturbations have generated interest recently because they are a simple way to introduce a preferred direction in the Universe and may generate the deviations from statistical isotropy that have been observed in the CMB. In particular, in Chapter 6, we will show that a superhorizon perturbation to an inflationary field can generate the hemispherical power asymmetry found in the WMAP data [42, 43, 44, 45, 46]. In this chapter, we have demonstrated how the CMB constrains such superhorizon perturbations: the octupole constraint on  $\Delta\Psi$  is sufficient to rule out an inflaton perturbation as the source of the observed power asymmetry, but it is possible to generate the observed power asymmetry with a superhorizon curvaton perturbation. These constraints may also be applied to other scenarios that invoke superhorizon perturbations. For instance, order-unity superhorizon fluctuations in the mean value of the curvaton may be a generic feature of the curvaton model [179].

### **Acknowledgments**

The author thanks Sean Carroll, Marc Kamionkowski, Chris Hirata, and Mike Kesden for useful discussions. This work was supported by DoE DE-FG03-92-ER40701 and the Gordon and Betty Moore Foundation.

## Chapter 6

# A Hemispherical Power Asymmetry from Inflation<sup>1</sup>

## 6.1 Introduction

The cosmic microwave background (CMB) [17, 18] and the distribution of galaxies [19, 20] tell us that the early Universe was homogeneous on superhorizon scales, spatially flat, and contained a nearly scale-invariant spectrum of adiabatic fluctuations. These features of the early Universe provide compelling evidence for inflation [1, 2, 3]. In the standard inflation scenario, the Universe underwent a very long inflationary expansion before the comoving observable Universe exited the horizon during inflation. Thus, any remnants of the pre-inflationary Universe were inflated away before there could be observable consequences. This accounts for the smoothness of the primordial Universe as well as its flatness. It also suggests that the primordial density perturbations should show no preferred direction.

There are indications, however, that the Universe is not perfectly isotropic [39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52]. In this chapter, we will focus our attention on one of these anomalies: the rms temperature fluctuation in the CMB on one side of the sky is larger than on the other side [42, 43, 44, 45, 46]. This hemispherical power asymmetry can be parametrized as a dipolar modulation of the temperature anisotropy field [44, 46]; the temperature fluctuation in the  $\hat{n}$  direction is

$$\frac{\delta T}{T}(\hat{n}) = s(\hat{n})[1 + A(\hat{n} \cdot \hat{p})], \quad (6.1)$$

where  $s(\hat{n})$  is an isotropic Gaussian random field. The magnitude of the asymmetry is given by  $A$  and its direction is specified by  $\hat{p}$ ; the most recent analysis, using WMAP5 data, found  $A = 0.072 \pm 0.022$  for  $\ell \lesssim 64$  with  $\hat{p}$  pointing at  $(\ell, b) = (224^\circ, -22^\circ) \pm 24^\circ$  [46]. The hemispherical power asymmetry is unrelated to the “axis of evil” [39, 40, 41], an apparent alignment of the lowest multipole moments, and has received considerably less theoretical attention (although see [165, 166]).

A superhorizon perturbation would introduce a preferred direction in the Universe and has been considered as a possible origin of the “axis of evil” [164]. In this chapter, we investigate how the hemispherical power asymmetry could result from a superhorizon perturbation during inflation, as

---

<sup>1</sup>The first two sections of this chapter were adapted from *A hemispherical power asymmetry from inflation*, Adrienne L. Erickcek, Marc Kamionkowski, and Sean Carroll; Phys. Rev. D **78**, 123520 (2008). Reproduced here with permission, copyright (2008) by the American Physical Society. The subsequent sections were adapted from an article that is currently in preparation by Adrienne L. Erickcek, Christopher M. Hirata, and Marc Kamionkowski.

depicted in Fig. 6.1. Since the amplitude of the primordial fluctuations depends on the background value of the fluctuating inflationary field, a large-amplitude superhorizon fluctuation would generate a power asymmetry by varying the background value of the field across the observable Universe. Of course, the superhorizon fluctuation would make the Universe inhomogeneous, and the near-uniformity of the CMB constrains such departures from homogeneity, as described in Chapter 5.

We first consider how a superhorizon fluctuation could generate a scale-invariant power asymmetry. We begin by showing in Section 6.2.1 that the power asymmetry cannot be reconciled with single-field slow-roll inflation without violating the constraints derived in Chapter 5. We then consider an alternative inflationary theory, the curvaton model [86, 87, 88, 89], which has been suggested as a possible source of a power asymmetry [165]. In the curvaton model, the inflaton field dominates the Universe’s energy density during inflation and drives the inflationary expansion, but the primordial fluctuations arise from quantum fluctuations in a subdominant scalar field called the curvaton. In Section 6.2.2, we demonstrate that a superhorizon fluctuation in the curvaton field can generate the observed asymmetry without violating the homogeneity constraints imposed by the CMB. The required superhorizon fluctuation in the curvaton field may occur, for example, as a remnant of the pre-inflationary epoch or as a signature of superhorizon curvaton-web structures [179]. The proposed model predicts several signatures, which may soon be tested, in the CMB, and we discuss these signatures at the end of Section 6.2.2.

Of course, a key prediction of the model discussed in Section 6.2.2 is that the magnitude and direction of the power asymmetry are scale-invariant. There are indications, however, that the asymmetry in the CMB temperature fluctuations does not extend to  $\ell \gtrsim 600$  [182, 45]. Furthermore, an analysis of quasar number counts reveals that any asymmetry in the direction  $(\ell, b) = (225^\circ, -27^\circ)$  in the rms amplitude of primordial density fluctuations on scales that form quasars ( $k \simeq 1.3h - 1.8h \text{ Mpc}^{-1}$ ) must have  $A \lesssim 0.012$  at the 95% confidence level [90]. In light of these findings, we

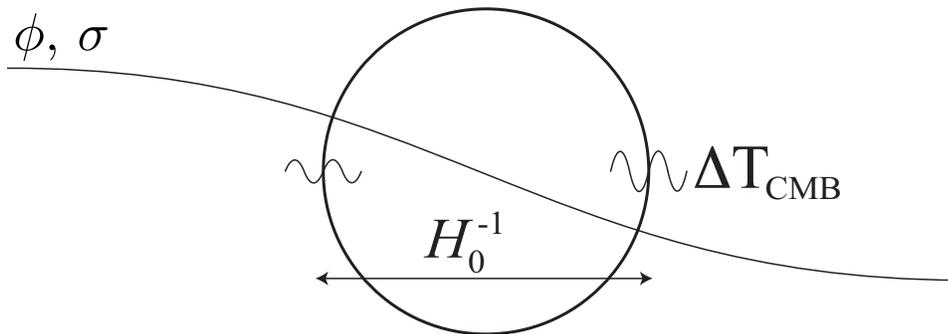


Figure 6.1: Measurements of temperature fluctuations in the cosmic microwave background (CMB) show that the rms temperature-fluctuation amplitude is larger in one side of the sky than in the other. We investigate here whether this may arise as a consequence of a large-scale fluctuation in the inflaton field  $\phi$  or the curvaton field  $\sigma$ .

consider how a superhorizon fluctuation in the curvaton field could produce a scale-dependent power asymmetry that is more pronounced on large scales than on small scales.

It is possible to dilute the power asymmetry on smaller scales by introducing discontinuities in the inflaton potential and its derivative that change relative contributions of the curvaton and inflaton field to the primordial perturbations [165]. We examine this proposal in Appendix C and find that the discontinuity in the inflaton potential required to satisfy the quasar constraint on the asymmetry violates constraints from ringing in the power spectrum [183, 184]. In Appendix C we also find that it is not possible to sufficiently dilute the asymmetry on small scales by smoothly changing the relative contributions of the curvaton and inflaton fluctuations to the primordial power spectrum.

We then consider the dark-matter isocurvature perturbations generated by some curvaton scenarios [89, 185, 186, 187, 188, 189, 190]. In the presence of a superhorizon fluctuation in the curvaton field, the power in these isocurvature perturbations will be asymmetric. Since isocurvature perturbations decay once they enter the horizon, they will contribute more to the large-scale ( $\ell \lesssim 100$ ) CMB anisotropies than to the smaller scales probed by quasars. Consequently, the desired scale-dependence of the asymmetry is a natural feature of isocurvature perturbations.

We review how isocurvature perturbations are generated in the curvaton scenario in Section 6.3.1, and we review the CMB signatures of isocurvature perturbations in Section 6.3.2. In Section 6.4, we examine how a hemispherical power asymmetry could be created by a superhorizon fluctuation in the curvaton field in two limiting cases of the curvaton scenario. We find in Section 6.4.1 that it is not possible to generate the observed hemispherical power asymmetry if the curvaton decay created the dark matter because the necessary superhorizon isocurvature fluctuation induces an unacceptably large dipolar anisotropy in the CMB. In Section 6.4.2, we show that the observed asymmetry can be generated by a superhorizon curvaton fluctuation if the curvaton's contribution to the dark matter is negligible. Our model predicts that the asymmetry will have a specific spectrum and that the current bounds on the contribution of isocurvature perturbations to the CMB power spectrum are nearly saturated. We summarize our findings and discuss these future tests of our model in Section 6.5.

## 6.2 A scale-invariant power asymmetry

### 6.2.1 Single-field models

Inflation postulates that the energy density in the early Universe was dominated by a scalar field  $\phi$ , the inflaton. The energy density of the inflaton field contains kinetic energy,  $(1/2)\dot{\phi}^2$ , plus some potential energy  $V(\phi)$ . If the slow-roll parameters,  $\epsilon_V \equiv (m_{\text{Pl}}^2/16\pi)(V'/V)^2$  and  $\eta_V \equiv (m_{\text{Pl}}^2/8\pi)(V''/V)$ , are small, then the field rolls slowly. The energy density is then dominated

by the potential energy, the pressure is negative, and the expansion of the Universe is inflationary.

Quantum fluctuations in the inflaton give rise to primordial density perturbations characterized by a gravitational-potential power spectrum  $P_\Psi(k) \propto V/\epsilon_V$ , where  $V$  and  $\epsilon_V$  are evaluated at the value the inflaton took when the comoving wavenumber  $k$  exited the horizon during inflation. Differentiation of the expression for  $P_\Psi(k)$  suggests that the power spectrum can be approximated as  $P_\Psi(k) \propto k^{n_s-1}$ , where the scalar spectral index  $n_s = 1 - 6\epsilon_V + 2\eta_V$  is close to unity, consistent with current measurements [18, 191].

The power spectrum  $P_\Psi(k)$  may vary with  $k$  because different values of  $k$  sample the quantity  $V/\epsilon$  at different values of the inflaton  $\phi$ . This suggests that the power asymmetry might be explained by a large-amplitude mode of  $\phi$  with a comoving wavelength that is long compared with the current Hubble distance ( $k \ll H_0$ ). Then one side of the CMB sky would reflect the imprint of a different value of  $\phi$  than the other side. From  $P_\Psi(k) \propto V/\epsilon_V$ , we infer a fractional power asymmetry,

$$\frac{\Delta P_\Psi}{P_\Psi} = -2\sqrt{\frac{\pi}{\epsilon}}(1 - n_s)\frac{\Delta\phi}{m_{\text{Pl}}}, \quad (6.2)$$

where  $\Delta\phi$  is the change in the inflaton field across the observable Universe and  $m_{\text{Pl}}^2 = G^{-1}$  in natural units. The observed 7.2% variation in the amplitude of the CMB temperature fluctuations [46] corresponds to a power asymmetry  $\Delta P_\Psi/P_\Psi \simeq 2A = 0.144$ , where  $A$  is defined by Eq. (6.1).

The gravitational-potential perturbation  $\Psi$  during matter domination is related to the inflaton perturbation  $\delta\phi$  through  $\Psi = (6/5)\sqrt{\pi/\epsilon}(\delta\phi/m_{\text{Pl}})$ . Thus, a long-wavelength perturbation  $\delta\phi \propto \sin[\vec{k} \cdot \vec{x} + \varpi]$ , with  $kx_{\text{dec}} \ll 1$  (where  $x_{\text{dec}}$  is the distance to the surface of last scatter), introduces a gravitational-potential perturbation with the same spatial dependence. It follows from Eq. (6.2) that  $\Delta\Psi = 6A/[5(n_s - 1)]$ . An immediate concern, therefore, is whether this large-amplitude perturbation is consistent with the isotropy of the CMB.

Gravitational-potential perturbations give rise to temperature fluctuations in the CMB through the Sachs-Wolfe effect [173] ( $\delta T/T \simeq \Psi/3$ ). A large-scale potential perturbation might thus be expected to produce a CMB temperature dipole of similar magnitude. However, for the Einstein-de Sitter universe, the potential perturbation induces a peculiar velocity whose Doppler shift cancels the intrinsic temperature dipole [85, 159]. The same is true for a flat universe with a cosmological constant, as we saw in Chapter 5.

Although the dipole vanishes, measurements of the CMB temperature quadrupole and octupole constrain the cosmological potential gradient [85, 159, 162]. Here we will briefly review how these constraints are derived before applying them to the superhorizon mode necessary to create the observed power asymmetry; the full calculation is presented in Chapter 5. Since  $kx_{\text{dec}} \ll 1$ , we first expand the sinusoidal dependence  $\Psi(\vec{x}) = \Psi_{\vec{k}} \sin(\vec{k} \cdot \vec{x} + \varpi)$  in powers of  $\vec{k} \cdot \vec{x}$  as in Eq. 5.15. The

terms that contribute to the CMB quadrupole and octupole are

$$\Psi(\vec{x}) = -\Psi_{\vec{k}} \left\{ \frac{(\vec{k} \cdot \vec{x})^2}{2} \sin \varpi + \frac{(\vec{k} \cdot \vec{x})^3}{6} \cos \varpi \right\}. \quad (6.3)$$

The CMB temperature anisotropy produced by the potential in Eq. (6.3) is

$$\frac{\Delta T}{T}(\hat{n}) = -\Psi_{\vec{k}} \left[ \frac{\mu^2}{2} (kx_{\text{dec}})^2 \delta_2 \sin \varpi + \frac{\mu^3}{6} (kx_{\text{dec}})^3 \delta_3 \cos \varpi \right], \quad (6.4)$$

where  $\mu \equiv \hat{k} \cdot \hat{n}$  and  $\Psi_{\vec{k}}$  is evaluated at the time of decoupling ( $\tau_{\text{dec}}$ ). As described in Section 5.3, the  $\delta_2$  and  $\delta_3$  coefficients account for the Sachs-Wolfe (including integrated) effect and the Doppler effect induced by  $\Psi_{\vec{k}}$ ; for a  $\Lambda$ CDM Universe with  $\Omega_M = 0.28$ , matter-radiation equality redshift  $z_{\text{eq}} = 3280$ , and decoupling redshift  $z_{\text{dec}} = 1090$ , we find that  $\delta_2 = 0.33$  and  $\delta_3 = 0.35$ . Choosing  $\hat{k} = \hat{z}$ , Eq. (6.4) gives nonzero values for the spherical-harmonic coefficients  $a_{20}$  and  $a_{30}$ . The relevant observational constraints are therefore given by Eqs. (5.34) and (5.35):

$$(kx_{\text{dec}})^2 |\Psi_{\vec{k}}(\tau_{\text{dec}}) \sin \varpi| \lesssim 5.8 Q \quad (6.5)$$

$$(kx_{\text{dec}})^3 |\Psi_{\vec{k}}(\tau_{\text{dec}}) \cos \varpi| \lesssim 32 \mathcal{O} \quad (6.6)$$

where  $Q$  and  $\mathcal{O}$  are upper bounds on  $|a_{20}|$  and  $|a_{30}|$ , respectively, in a coordinate system aligned with the power asymmetry.

As in Section 5.3, we take  $Q = 3\sqrt{C_2} \lesssim 1.8 \times 10^{-5}$  and  $\mathcal{O} = 3\sqrt{C_3} \lesssim 2.7 \times 10^{-5}$ , three times the measured rms values of the quadrupole and octupole [178], as  $3\sigma$  upper limits; this accounts for cosmic variance in the quadrupole and octupole due to smaller-scale modes. The temperature quadrupole and octupole induced by the superhorizon mode can be made arbitrarily small for fixed  $\Delta\Psi \simeq \Psi_{\vec{k}}(kx_{\text{dec}}) \cos \varpi$  by choosing  $k$  to be sufficiently small. However, we also demand that  $\Psi_{\vec{k}} \lesssim 1$  everywhere, and this sets a lower bound on  $(kx_{\text{dec}})$ .

We now return to the power asymmetry generated by an inflaton perturbation. The largest value of  $\Delta\Psi$  is obtained if  $\varpi = 0$ , in which case the perturbation produces no quadrupole. The octupole constraint [Eq. (6.6)] combined with  $(kx_{\text{dec}}) \gtrsim |\Delta\Psi|$  [i.e., the requirement  $\Psi_{\vec{k}} \lesssim 1$ ] implies that  $|\Delta\Psi| \lesssim (32\mathcal{O})^{1/3}$ . Given that  $(1 - n_s) \lesssim 0.06$ , we see that the maximum possible power asymmetry obtainable with a single superhorizon mode is  $A_{\text{max}} \simeq 0.05(32\mathcal{O})^{1/3} \simeq 0.0048$ . This is too small, by more than an order of magnitude, to account for the observed asymmetry ( $A = 0.072 \pm 0.022$ ). The limit can be circumvented if a number of Fourier modes conspire to make the density gradient across the observable Universe smoother. This would require, however, that we live in a very special place in a very unusual density distribution.

### 6.2.2 The curvaton model

We thus turn our attention to the curvaton model of inflation. In the curvaton scenario [86, 87, 88, 89], there are two light scalar fields during inflation. The inflaton dominates the energy density of the Universe and drives the inflationary expansion. The curvaton ( $\sigma$ ) is a spectator field during inflation; it remains fixed at its initial value  $\sigma_*$  and its energy is given by its potential  $V(\sigma) = (1/2)m_\sigma^2\sigma^2$ , with  $m_\sigma \ll H_{\text{inf}}$ , where  $H_{\text{inf}}$  is the Hubble parameter during inflation. When  $H \simeq m_\sigma$  after inflation, the curvaton field begins to oscillate in its potential well, and it behaves like a pressureless fluid until it decays. We will assume that the curvaton field is non-interacting prior to its decay.

During inflation, quantum fluctuations in the curvaton field  $[(\delta\sigma)_{\text{rms}} = H_{\text{inf}}/(2\pi)]$  generate a nearly scale-invariant spectrum of isocurvature fluctuations. After the inflaton decays into radiation, the growth of the curvaton energy density relative to the radiation density creates adiabatic perturbations from these isocurvature fluctuations. If the curvaton decays before any particle species decouples from radiation, then the isocurvature fluctuation is erased after the curvaton decays because isocurvature fluctuations between interacting fluids in thermal equilibrium decay quickly [192, 190]. If the curvaton decays after a particle species decouples from the radiation, however, there is a lasting isocurvature fluctuation between that species and the radiation in addition to the adiabatic perturbation generated by the growth of  $\Omega_\sigma$  after inflation [89, 185, 186, 187, 188, 189, 190].

In this section, we will assume that the curvaton decayed prior to the decoupling of any particles from the radiation so that the fluctuations in the curvaton field produce only adiabatic fluctuations. The resulting power spectrum of gravitational-potential perturbations during matter domination is

$$P_{\Psi,\sigma} \propto R^2 \left\langle \left[ \frac{\delta V}{V(\bar{\sigma}_*)} \right]^2 \right\rangle \sim R^2 \left( \frac{H_{\text{inf}}}{\pi \bar{\sigma}_*} \right)^2, \quad (6.7)$$

provided that  $\bar{\sigma}_* \gg H_{\text{inf}}$  [193]. Here  $R \simeq (3/4)(\rho_\sigma/\rho_{\text{tot}})$  is 3/4 times the energy density of the curvaton field just prior to its decay divided by the total energy density of the Universe at that time.

We hypothesize that the density due to curvaton decay is small compared with the density due to inflaton decay; i.e.,  $R \ll 1$ . In this case, the perturbation in the total energy density, and thus the potential perturbation  $\Psi$ , due to a fluctuation in  $\rho_\sigma$  will be suppressed, making it possible to satisfy the homogeneity conditions set by the CMB [Eqs. (6.5) and (6.6)], even if  $\rho_\sigma$  has order-unity variations. We then hypothesize that the power asymmetry comes from a variation  $\Delta\bar{\sigma}_*$  in the initial value of the mean curvaton field across the observable Universe.

For  $R \ll 1$ , the Universe is radiation-dominated between the end of inflation and the decay of the curvaton, and

$$R = \pi \left( \frac{\bar{\sigma}_*}{m_{\text{Pl}}} \right)^2 \sqrt{\frac{1.4 m_\sigma}{\Gamma_\sigma}}, \quad (6.8)$$

where  $\Gamma_\sigma$  is the curvaton decay rate [88]. Since  $R \propto \bar{\sigma}_*^2$ , the power spectrum for gravitational-potential perturbations produced by the curvaton is proportional to  $\bar{\sigma}_*^2$ . A variation  $\Delta\bar{\sigma}$  in the value of the mean curvaton field across the observable Universe therefore induces a fractional power asymmetry  $\Delta P_{\Psi,\sigma}/P_{\Psi,\sigma} \simeq 2(\Delta\bar{\sigma}_*/\bar{\sigma}_*)$ . For superhorizon fluctuations,  $\delta\sigma$  and  $\bar{\sigma}$  obey the same evolution equation, so the ratio  $\delta\sigma/\bar{\sigma}$  is conserved [187]. We will therefore omit the “\*” subscript from future expressions in this section.

First we must ensure that this inhomogeneity does not violate Eqs. (6.5) and (6.6). The potential fluctuation at the time of decoupling produced by a fluctuation  $\delta\sigma$  in the curvaton field is given by Eq. (5.40):

$$\Psi(\tau_{\text{dec}}) \simeq -\frac{R}{5} \left[ \frac{\Psi(\tau_{\text{dec}})}{\frac{9}{10}\Psi_{\text{p}}} \right] \left[ 2 \left( \frac{\delta\sigma}{\bar{\sigma}} \right) + \left( \frac{\delta\sigma}{\bar{\sigma}} \right)^2 \right], \quad (6.9)$$

where  $\Psi_{\text{p}}$  is the potential fluctuation in the radiation-dominated era [ $\Psi(\tau_{\text{dec}}) = 0.937\Psi_{\text{p}}$  from Section 5.2]. In Section 5.4, we examined how this potential fluctuation induces large-scale anisotropies in the CMB, and we now briefly review how  $\delta\sigma$  is constrained by the CMB quadrupole. Consider a superhorizon sinusoidal perturbation to the curvaton field  $\delta\sigma = \sigma_k \sin(\vec{k} \cdot \vec{x} + \varpi)$ . If we ignored the term in Eq. (6.9) quadratic in  $\delta\sigma$ , then the upper bound to  $\delta\bar{\sigma}$  would be obtained by setting  $\varpi = 0$ . As with the inflaton, the constraint would then arise from the CMB octupole. However, the term in Eq. (6.9) that is quadratic in  $\delta\sigma$  gives rise to a term in  $\Psi$  that is quadratic in  $(\vec{k} \cdot \vec{x})$ —i.e.,  $\Psi_{\text{quad}} = -(R/5)[\Psi(\tau_{\text{dec}})/(0.9\Psi_{\text{p}})](\sigma_k/\bar{\sigma})^2(\vec{k} \cdot \vec{x})^2$  for  $\varpi = 0$ . Noting that  $(\Delta\bar{\sigma}/\bar{\sigma}) = (\sigma_k/\bar{\sigma})(\vec{k} \cdot \vec{x}_{\text{dec}})$ , the quadrupole bound in Eq. (6.5) yields an upper limit,

$$R \left( \frac{\Delta\bar{\sigma}}{\bar{\sigma}} \right)^2 \lesssim \frac{5}{2}(5.6 Q), \quad (6.10)$$

just as in Eq. (5.46). While this bound was derived for  $\varpi = 0$ , most other values for  $\varpi$  yield similar constraints, as shown in Section 5.4.

Most generally, the primordial power will be some combination of that due to the inflaton and curvaton [194],  $P_\Psi = P_{\Psi,\phi} + P_{\Psi,\sigma} \simeq 10^{-9}$ , with a fraction  $\xi \equiv P_{\Psi,\sigma}/P_\Psi$  due to the curvaton. Since the fluctuations in  $\Psi$  from the inflaton field are not affected by the superhorizon fluctuation in the curvaton field, the power asymmetry will be diluted by the inflaton’s contribution to the power spectrum. The total power asymmetry is therefore  $\Delta P_{\Psi,\sigma}/P_\Psi \simeq 2\xi(\Delta\bar{\sigma}/\bar{\sigma}) \simeq 2A$ . This asymmetry can be obtained without violating Eq. (6.10) by choosing  $R \lesssim 14 Q\xi^2/A^2$ , as shown in Fig. 6.2.

The only remaining issue is the Gaussianity of primordial perturbations. The curvaton fluctuation  $\delta\sigma$  is a Gaussian random variable. Since the curvaton-induced density perturbation has a contribution that is quadratic in  $\delta\sigma$ , it implies a non-Gaussian contribution to the density fluctuation. The departure from Gaussianity can be estimated from the parameter  $f_{\text{NL}}$  [195, 196], which for the curvaton model is  $f_{\text{NL}} \simeq 5\xi^2/(4R)$  [186, 197, 198]. The current upper limit,  $f_{\text{NL}} \lesssim 100$

[199, 59, 18, 61, 62], leads to the lower limit to  $R$  shown in Fig. 6.2.

Figure 6.2 shows that there are values of  $R$  and  $\xi$  that lead to a power asymmetry  $A = 0.072$  and are consistent with measurements of the CMB quadrupole and  $f_{\text{NL}}$ . For any value of  $A$ , the allowed region of  $R$ - $\xi$  parameter space is

$$\frac{5}{4 f_{\text{NL,max}}} \lesssim \frac{R}{\xi^2} \lesssim 14 \frac{Q}{A^2}, \quad (6.11)$$

where  $f_{\text{NL,max}}$  is the largest allowed value for  $f_{\text{NL}}$ . Thus we see that measurements of the CMB

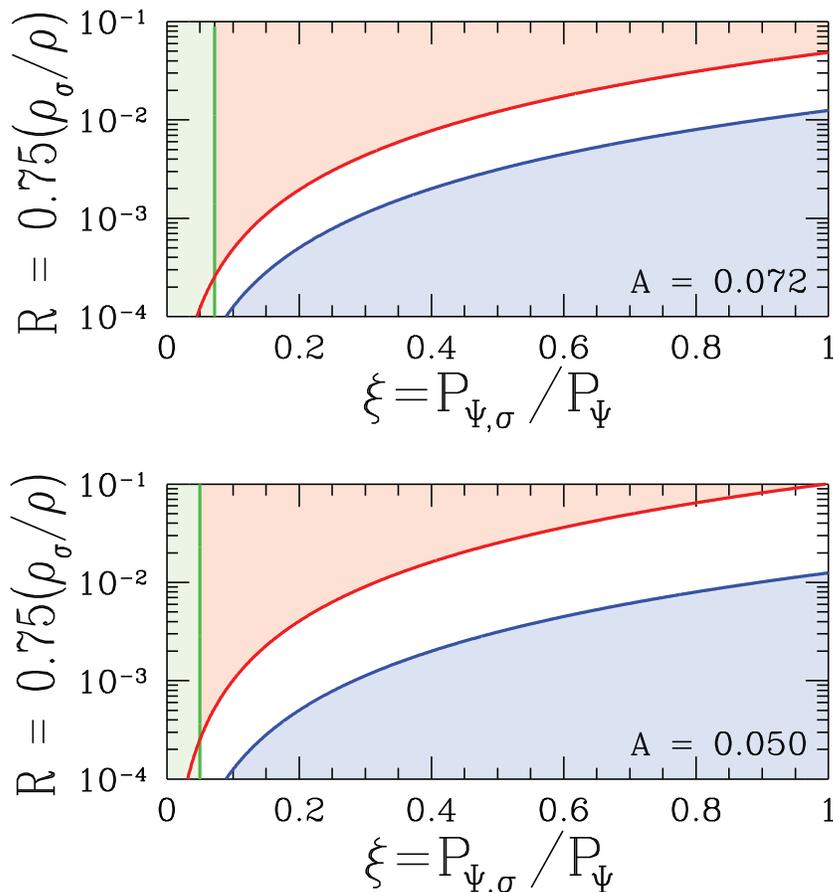


Figure 6.2: The  $R$ - $\xi$  parameter space for the curvaton model that produces a power asymmetry  $A = 0.072$  (top) and  $A_p = 0.050$  (bottom). The observed asymmetry is  $A = 0.072 \pm 0.022$  [46]. Here  $R$  is  $3/4$  times the fraction of the cosmological density due to curvaton decay, and  $\xi$  is the fraction of the power due to the curvaton. The shaded regions in this plot are excluded. The upper limit to  $R$  comes from the CMB-quadrupole constraint. The lower bound comes from  $f_{\text{NL}} \leq 100$ . The lower limit to  $\xi$  comes from the requirement that the fractional change in the curvaton field across the observable Universe be less than one. If  $A$  is lowered, the lower bound to  $R$  remains unchanged, but the upper bound increases, proportional to  $A^{-2}$ . The lower limit to  $\xi$  also decreases as  $A$  decreases, proportional to  $A$ .

quadrupole and  $f_{\text{NL}}$  place an upper bound,

$$A \lesssim \sqrt{(14Q) \left( \frac{4f_{\text{NL,max}}}{5} \right)}, \quad (6.12)$$

on the power asymmetry that may be generated by a superhorizon curvaton fluctuation. Given that  $f_{\text{NL}} \lesssim 100$  and  $Q = 1.8 \times 10^{-5}$ , we have  $A \lesssim 0.14$ , which is about twice the observed value.

The allowed region of  $R$ - $\xi$  parameter space disappears if  $f_{\text{NL,max}}$  is too small, so Eq. (6.11) also implies a lower bound on  $f_{\text{NL}}$ :

$$f_{\text{NL}} \gtrsim \frac{5}{4} \left( \frac{A^2}{14Q} \right). \quad (6.13)$$

For  $Q = 1.8 \times 10^{-5}$ , we predict (for  $A = 0.072$ )  $f_{\text{NL}} \gtrsim 26$ , much larger than the  $f_{\text{NL}}$  predicted by standard slow-roll inflation ( $f_{\text{NL}} \ll 1$ ) [53]. Values as small as  $f_{\text{NL}} \simeq 5$  should be accessible to the forthcoming Planck satellite [196, 200, 201], and so there should be a clear non-Gaussian signature in Planck if the power asymmetry was generated by a curvaton perturbation and  $A = 0.072$ .

If  $(\delta\sigma/\bar{\sigma}) \ll 1$ , the power due to the curvaton is  $P_{\Psi,\sigma} \simeq (2R/5)^2 \langle (\delta\sigma/\bar{\sigma})^2 \rangle$ , and the power required from the curvaton fixes  $R(\delta\sigma/\bar{\sigma})_{\text{rms}} \simeq 8 \times 10^{-5} \xi^{1/2}$ . It follows that  $(\delta\sigma/\bar{\sigma})_{\text{rms}} \lesssim 0.33$  for the allowed parameter space in Fig. 6.2 for  $A = 0.072$ , thus verifying that this parameter is small. We find from  $(\Delta\bar{\sigma}/\bar{\sigma}) = A/\xi \lesssim 1$  that the required cross-horizon variation  $\Delta\bar{\sigma}/\bar{\sigma}$  in the curvaton is large compared with the characteristic quantum-mechanical curvaton fluctuation  $(\delta\sigma/\bar{\sigma})_{\text{rms}}$ ; the required  $\Delta\bar{\sigma}$  is at least a  $\sim 3\sigma$  fluctuation. While such a large quantum fluctuation is unlikely, we note that the observed asymmetry is a  $3.3\sigma$  deviation from an isotropic primordial power spectrum; it is therefore equally unlikely that the asymmetry is a statistical fluke. The superhorizon fluctuation in the curvaton field could also be a superhorizon inhomogeneity not completely erased by inflation. Another possibility is that positive- and negative-value cells of  $\bar{\sigma}$  created during inflation may be large enough to encompass the observable Universe; if so, we would observe an order-unity fluctuation in  $\bar{\sigma}$  near the  $\bar{\sigma} = 0$  wall that divides two cells [179].

We have considered the specific asymmetry  $A \simeq 0.072$  reported by Ref. [46], but our results can be scaled for different values of  $A$ , should the measured value for the asymmetry change in the future. In particular, the  $f_{\text{NL}}$  constraint (the lower bound to  $R$ ) in Fig. 6.2 remains the same, but the upper bound (from the quadrupole) increases as  $A$  is decreased. The lower limit to  $\xi$  also decreases as  $A$  is decreased. Here we have also considered a general model in which primordial perturbations come from some combination of the inflaton and curvaton. Although it may seem unnatural to expect the two field decays to produce comparable fluctuation amplitudes, our mechanism works even if  $\xi = 1$  (the fluctuations are due entirely to the curvaton). Thus, the coincidence is not a requirement of the model.

If the power asymmetry can indeed be attributed to a superhorizon curvaton mode, then the

workings of inflation are more subtle than the simplest models would suggest. Fortunately, the theory makes a number of predictions that can be pursued with future experiments. To begin, the modulated power should produce signatures in the CMB polarization and temperature-polarization correlations [202, 203]. The curvaton model predicts non-Gaussianity, of amplitude  $f_{\text{NL}} \gtrsim 26$  for  $A \simeq 0.072$ , which will soon be experimentally accessible. However, the theory also predicts that the small-scale non-Gaussianity will be modulated across the sky by the variation in  $\bar{\sigma}$  (and thus in  $\xi$  and  $R$ ). The presence of curvaton fluctuations also changes other features of the CMB [198]. The ratio of tensor and scalar perturbations ( $r$ ) is reduced by a factor of  $(1 - \xi)$  and the scalar spectral index is  $n_s = 1 - 2\epsilon_V - (1 - \xi)(4\epsilon_V - 2\eta_V)$ . The tensor spectral index ( $n_T$ ), however, is unaltered by the presence of the curvaton, and so this model alters the inflationary consistency relation between  $n_T$  and  $r$  and possibly the prospects for testing it [204].

### 6.3 Review of isocurvature perturbations

Shortly after we proposed that the observed hemispherical power asymmetry could be a signature of a superhorizon fluctuation in the curvaton field, it was discovered that the asymmetry does not extend to smaller scales, as discussed in Section 6.1. If the curvaton creates only adiabatic perturbations, as assumed in the previous section, then  $A = \xi(\Delta\bar{\sigma}/\bar{\sigma})$ . Clearly, the only way to break the scale invariance of the asymmetry to make  $\xi$  depend on scale; since the inflaton field is unaffected by the superhorizon fluctuation in the curvaton field, the fluctuations in the inflaton field are symmetric, and the magnitude of the asymmetry will be diluted on small scales if the inflaton's contribution to the total power is larger on these scales. This approach is studied in Appendix C and is found to be incompatible with current constraints on the running of the scalar spectral index and the smoothness of the primordial power spectrum.

In Section 6.1, we introduced an alternate explanation for the scale-dependence of the power asymmetry: the asymmetry could be partly attributable to isocurvature perturbations. We devote the remainder of this chapter to an analysis of this proposal. We will restrict our attention to scenarios in which the curvaton decays after dark matter freeze-out, but prior to the decoupling of any other particle species. In this case, an isocurvature fluctuation between dark matter and radiation is created [89, 185, 186, 187, 188, 189, 190]. (We will neglect baryon isocurvature modes, which may arise due to the annihilations of baryons and antibaryons created during curvaton decay [205]). We will now review how the dark-matter isocurvature fluctuation relates to the initial curvaton perturbation and how isocurvature fluctuations impact the CMB before considering power asymmetry production in the next section.

### 6.3.1 Isocurvature perturbations in the curvaton scenario

We continue to work in conformal Newtonian gauge; as in Chapter 5, we take the perturbed Friedmann-Robertson-Walker (FRW) metric to be

$$ds^2 = -(1 + 2\Psi)dt^2 + a^2(t)\delta_{ij}(1 - 2\Phi)dx^i dx^j, \quad (6.14)$$

where  $a$  is normalized to equal one today. We define

$$\zeta_i \equiv -\Psi - H \frac{\delta\rho_i}{\dot{\rho}_i} \quad (6.15)$$

to be the curvature perturbation on surfaces of uniform  $i$ -fluid density, and

$$\zeta \equiv -\Psi - H \frac{\delta\rho}{\dot{\rho}} = \sum_i \frac{\dot{\rho}_i}{\dot{\rho}} \zeta_i \quad (6.16)$$

is the curvature perturbation on surfaces of uniform total density. Throughout this paper, a dot refers to differentiation with respect to proper time  $t$ . We use the notation  $S_{i\gamma} \equiv 3(\zeta_i - \zeta_\gamma)$ , where a subscript  $\gamma$  refers to radiation, to describe isocurvature fluctuations. For any non-interacting fluid,  $\zeta_i$  is conserved on superhorizon scales. In the absence of isocurvature perturbations,  $\zeta$  is constant on superhorizon scales, but if there is an isocurvature perturbation, then  $\zeta$  evolves due to the changing value of  $\dot{\rho}_i/\dot{\rho}$ .

Immediately after inflation, there are superhorizon adiabatic fluctuations from inhomogeneities in the inflaton field  $\zeta^{(i)} \simeq \zeta_\gamma^{(i)}$  and superhorizon isocurvature fluctuations in the curvaton field given by  $S_{\sigma\gamma}$ . After curvaton decay, there are superhorizon adiabatic perturbations  $\zeta^{(f)}$  and superhorizon dark-matter isocurvature perturbations  $S_{m\gamma}$ . These perturbations are related through a transfer matrix:

$$\begin{pmatrix} \zeta^{(f)} \\ S_{m\gamma} \end{pmatrix} = \begin{pmatrix} 1 & \mathcal{T}_{\zeta S} \\ 0 & \mathcal{T}_{SS} \end{pmatrix} \begin{pmatrix} \zeta^{(i)} \\ S_{\sigma\gamma} \end{pmatrix}. \quad (6.17)$$

This transfer matrix is completely general and applicable to the evolution of any mixture of isocurvature and adiabatic perturbations. The left column indicates that superhorizon adiabatic perturbations do not evolve in the absence of isocurvature fluctuations and that they are incapable of generating isocurvature fluctuations. The expressions for  $\mathcal{T}_{\zeta S}$  and  $\mathcal{T}_{SS}$  are model dependent.

In the limit of instantaneous curvaton decay, the total curvature perturbation cannot change during the decay of the curvaton. We can therefore obtain  $\mathcal{T}_{\zeta S}$  by equating  $\zeta^{(f)}$ , which is evaluated just after curvaton decay, to  $\zeta^{(\text{bd})}$ , which is evaluated just prior to curvaton decay. From Eq. (6.16) we have

$$\zeta^{(f)} = \zeta^{(i)} + \left[ \frac{R}{3} + \left( \frac{\Omega_{\text{cdm}}}{4\Omega_\gamma + 3\Omega_\sigma + 3\Omega_{\text{cdm}}} \right)^{(\text{bd})} \mathcal{T}_{\text{fr}} \right] S_{\sigma\gamma}, \quad (6.18)$$

where

$$R \equiv \left[ \frac{3\Omega_\sigma}{4\Omega_\gamma + 3\Omega_\sigma + 3\Omega_{\text{cdm}}} \right]^{(\text{bd})}. \quad (6.19)$$

All quantities with a “bd” superscript are evaluated just prior to curvaton decay. We will see below that  $\zeta_{\text{cdm}}^{(\text{bd})}$  is not equal to its initial value  $\zeta_{\text{cdm}}^{(i)} = \zeta^{(i)}$ , and we have defined  $\mathcal{T}_{\text{fr}}$  through

$$\zeta_{\text{cdm}}^{(\text{bd})} = \zeta_{\text{cdm}}^{(i)} + \frac{\mathcal{T}_{\text{fr}}}{3} S_{\sigma\gamma}. \quad (6.20)$$

As in the previous section, we will assume that  $R \ll 1$ , so that the curvaton never dominates the energy density of the Universe. Since the Universe must be radiation-dominated after curvaton decay, the assumption that  $R \ll 1$  also implies that  $\Omega_{\text{cdm}}^{(\text{bd})} \ll 1$ . Furthermore, we will see that  $\mathcal{T}_{\text{fr}} \ll R$  if  $R \ll 1$ . With these assumptions,

$$\mathcal{T}_{\zeta S} \simeq \frac{R}{3} \simeq \frac{1}{4} \Omega_\sigma^{(\text{bd})}, \quad (6.21)$$

and we see that  $R \simeq (3/4)\rho_\sigma/\rho$  just prior to curvaton decay, as defined earlier. Numerical studies of curvaton decay in the absence of dark matter and perturbations from the inflaton [ $\zeta^{(i)} = 0$ ] indicate that this instant-decay expression for  $\mathcal{T}_{\zeta S}$  is accurate to within 10% provided that  $R$  is evaluated when  $H = \Gamma_\sigma/1.4$ , where  $\Gamma_\sigma$  is the rate of curvaton decay [193].

If the dark matter freezes out prior to curvaton decay, then a dark-matter isocurvature perturbation is created in two stages. The first stage occurs at dark-matter freeze-out [187]. The abundance of dark matter after freeze-out is determined by the expansion rate at freeze-out, so immediately after freeze-out, the hypersurface of constant dark-matter density coincides with the hypersurface of constant total density. In the presence of the curvaton field, this hypersurface is not a hypersurface of constant radiation density and a dark-matter isocurvature perturbation is created. The change in  $\zeta_{\text{cdm}}$  during freeze-out is given by  $\Delta\zeta_{\text{cdm}} = (\mathcal{T}_{\text{fr}}/3)S_{\sigma\gamma}$  where

$$\mathcal{T}_{\text{fr}} = \frac{(\alpha - 3)\Omega_\sigma^{(\text{fr})}}{2(\alpha - 2) + \Omega_\sigma^{(\text{fr})}}. \quad (6.22)$$

In this expression,  $\alpha \equiv \left. \frac{d \ln \Gamma_{\text{cdm}}}{d \ln T} \right|_{(\text{fr})}$  gives the dependence of the rate for dark matter annihilations  $\Gamma_{\text{cdm}}$  on temperature  $T$  for s-wave annihilations ( $\alpha \simeq 21$  for neutralino dark matter). All quantities with an “(fr)” superscript are to be evaluated when the dark matter freezes out. Since we assume that the Universe does not cease to be radiation dominated prior to curvaton decay,

$$\Omega_\sigma^{(\text{fr})} = \Omega_\sigma^{(\text{bd})} \sqrt{\frac{H^{(\text{bd})}}{H^{(\text{fr})}}} \simeq \frac{4}{3} R \sqrt{\frac{\Gamma_\sigma}{1.4\Gamma_{\text{cdm}}^{(\text{fr})}}}. \quad (6.23)$$

Between freeze-out and curvaton decay, the dark matter is non-interacting and  $\zeta_{\text{cdm}}$  is conserved;

Eq. (6.20) therefore relates  $\zeta_{\text{cdm}}$  just prior to curvaton decay to its initial value. In the case that  $R \ll 1$ ,  $\mathcal{T}_{\text{fr}} \ll (2/3)R$  since  $\Gamma_\sigma \ll \Gamma_{\text{cdm}}^{(\text{fr})}$  (dark matter freezes out prior to curvaton decay).

The second stage of dark-matter isocurvature creation occurs at curvaton decay. Let  $B_m \equiv \Gamma_{\sigma \rightarrow \text{m}}/\Gamma_\sigma$  be the fraction of the curvaton energy that is turned into dark matter when the curvaton decays. In that case, the change in  $\zeta_{\text{cdm}}$  during curvaton decay is given by [190]

$$\zeta_{\text{cdm}}^{(\text{f})} - \zeta_{\text{cdm}}^{(\text{bd})} = \frac{B_m \Omega_\sigma^{(\text{bd})}}{\Omega_{\text{cdm}}^{(\text{bd})} + B_m \Omega_\sigma^{(\text{bd})}} \left[ \zeta_\sigma^{(\text{bd})} - \zeta_{\text{cdm}}^{(\text{bd})} \right]. \quad (6.24)$$

This expression differs slightly from the expression in Ref. [190], but the two expressions are equivalent because  $\Omega_\sigma^{(\text{fr})}/\Omega_\sigma^{(\text{bd})} = \Omega_{\text{cdm}}^{(\text{fr})}/\Omega_{\text{cdm}}^{(\text{bd})}$ . We can relate  $S_{\text{m}\gamma}$  to  $S_{\sigma\gamma}$  by combining Eq. (6.18) for  $\zeta^{(\text{f})} \simeq \zeta_\gamma^{(\text{f})}$  and Eqs. (6.20) and (6.24) for  $\zeta_{\text{cdm}}^{(\text{f})}$ . We find that

$$\mathcal{T}_{SS} = \frac{\mathcal{T}_{\text{fr}} \Omega_{\text{cdm}}^{(\text{bd})}}{\Omega_{\text{cdm}}^{(\text{bd})} + B_m \Omega_\sigma^{(\text{bd})}} + \frac{B_m \Omega_\sigma^{(\text{bd})}}{\Omega_{\text{cdm}}^{(\text{bd})} + B_m \Omega_\sigma^{(\text{bd})}} - R. \quad (6.25)$$

The derivation of this expression for  $\mathcal{T}_{SS}$  neglected the possibility that the injection of dark matter particles at curvaton decay could raise the dark matter particle number density to the point that dark matter particles begin to self-annihilate again. If dark matter annihilations resume after curvaton decay, then the final value of  $S_{\text{m}\gamma}$  is suppressed. It is possible to approximate the effect of a second stage of dark matter self-annihilation by multiplying  $\mathcal{T}_{SS}$  by a factor of  $1/(1+\Upsilon)$ , where  $\Upsilon \equiv \Gamma_{\text{cdm}}/H$  evaluated just after curvaton decay [190]. Since we are most interested in curvaton models in which the curvaton produces very little dark matter, we will assume that  $\Upsilon \ll 1$ . We note, however, that our results can be easily adapted to cases where the self-annihilation of the dark matter after curvaton decay is significant.

It will be useful to make the  $R$  dependence of  $\mathcal{T}_{SS}$  explicit by defining  $\tilde{B}R \equiv B_m \Omega_\sigma^{(\text{bd})}/\Omega_{\text{cdm}}^{(\text{bd})}$  to be the dark matter density from curvaton decay divided by the dark matter density prior to curvaton decay. We will also define  $\tilde{\lambda} \equiv (4/3)\sqrt{H^{(\text{bd})}/H^{(\text{fr})}}$  so that  $\Omega_\sigma^{(\text{fr})} = \tilde{\lambda}R$ . In this notation,

$$\mathcal{T}_{SS} = \left[ \frac{(\alpha - 3)\tilde{\lambda}R}{2(\alpha - 2) + \tilde{\lambda}R} \right] \left( \frac{1}{1 + \tilde{B}R} \right) + \left( \frac{\tilde{B}R}{1 + \tilde{B}R} \right) - R. \quad (6.26)$$

In our analysis, we will consider two limiting cases:  $\tilde{B}R \gg 1$  (i.e., the curvaton creates nearly all the dark matter), and  $\tilde{B}R \ll 1$  (i.e., the curvaton creates an insignificant fraction of the dark matter). In both cases, we will still assume that  $R \ll 1$ .

If the curvaton creates nearly all the dark matter so that  $\tilde{B}R \gg 1$ , then

$$\lim_{\tilde{B}R \gg 1} \mathcal{T}_{SS} = \left[ \frac{(\alpha - 3)\tilde{\lambda}R}{2(\alpha - 2) + \tilde{\lambda}R} \right] \left( \frac{1}{\tilde{B}R} \right) + 1 - R. \quad (6.27)$$

When we recall that  $\tilde{\lambda} < 1$  is required to make the dark matter freeze-out prior to curvaton decay, we see that the first term in Eq. (6.27) is proportional to  $\tilde{\lambda}/\tilde{B}$ , which is much smaller than  $R$  if  $\tilde{B}R \gg 1$ . The first term is therefore negligible, and we are left with

$$\lim_{\tilde{B}R \gg 1} \mathcal{T}_{SS} = 1 - R. \quad (6.28)$$

In the opposite limit, in which the curvaton's contribution to the dark matter density is negligible, we have

$$\lim_{\tilde{B}R \ll 1} \mathcal{T}_{SS} = \left[ \frac{(\alpha - 3)\tilde{\lambda}}{2(\alpha - 2)} + \tilde{B} - 1 \right] R \equiv \kappa R. \quad (6.29)$$

The first two terms in Eq. (6.29) are positive by definition, so  $\kappa \gtrsim -1$ . The first term is always less than 0.5 since  $\tilde{\lambda} < 1$ , but  $\tilde{B} = (4/3)B_m/\Omega_{\text{cdm}}^{(\text{bd})}$  could be much larger than unity since  $\Omega_{\text{cdm}}^{(\text{bd})} \ll 1$ . The only upper limit on  $\kappa$  is given by  $\tilde{B}R \ll 1$  which implies that  $\kappa \ll 1/R$ .

### 6.3.2 Isocurvature modes in the cosmic microwave background

Now that we have defined  $\mathcal{T}_{\zeta S}$  and  $\mathcal{T}_{SS}$  in Eq. (6.17), we can relate the early-time perturbations in the matter-radiation fluid to the inflaton and curvaton perturbations created during inflation. The power spectra of the early-time perturbations in the matter-radiation fluid ( $\zeta^{(\text{f})}$  and  $S_{\text{m}\gamma}$ ) are the spectra that we will use as initial conditions to calculate the CMB power spectrum.

Following Ref. [206], we define

$$\mathcal{P}_{\zeta}(k) \equiv \frac{k^3}{2\pi^2} \langle \zeta^{(\text{f})}(\vec{k}) \zeta^{(\text{f})*}(\vec{k}) \rangle, \quad (6.30)$$

$$\mathcal{P}_S(k) \equiv \frac{k^3}{2\pi^2} \langle S_{\text{m}\gamma}(\vec{k}) S_{\text{m}\gamma}^*(\vec{k}) \rangle, \quad (6.31)$$

$$\mathcal{C}_{\zeta S}(k) \equiv \frac{k^3}{2\pi^2} \langle \zeta^{(\text{f})}(\vec{k}) S_{\text{m}\gamma}^*(\vec{k}) \rangle. \quad (6.32)$$

We will use a similar convention for the perturbations from inflation:

$$\mathcal{A}^2 \left( \frac{k}{k_0} \right)^{n_\phi - 1} \equiv \frac{k^3}{2\pi^2} \langle \zeta^{(\text{i})}(\vec{k}) \zeta^{(\text{i})*}(\vec{k}) \rangle, \quad (6.33)$$

$$\mathcal{B}^2 \left( \frac{k}{k_0} \right)^{n_\sigma - 1} \equiv \frac{k^3}{2\pi^2} \langle S_{\sigma\gamma}(\vec{k}) S_{\sigma\gamma}^*(\vec{k}) \rangle. \quad (6.34)$$

Both spectra produced during inflation are nearly flat (e.g., [88]), and we will assume that  $n_\phi \simeq n_\sigma \simeq 1$ . The initial curvature fluctuations are created by the inflaton; the standard slow-roll power spectrum is

$$\mathcal{A}^2 = \frac{GH_{\text{inf}}^2}{\pi\epsilon_H}, \quad (6.35)$$

where  $\epsilon_H \equiv -\dot{H}_{\text{inf}}/H_{\text{inf}}^2 \simeq \epsilon_V$  is a slow-roll parameter. When both the radiation from curvaton decay

and the curvaton field are perturbed,  $S_{\sigma\gamma} \simeq 2\delta\sigma_*/\bar{\sigma}_*$ , where  $\delta\sigma_*$  and  $\bar{\sigma}_*$  are evaluated at horizon exit [194]. However, as mentioned in Section 6.2.2, the ratio  $\delta\sigma/\bar{\sigma}$  is conserved for superhorizon perturbations [187]. Given that  $\mathcal{P}_{\delta\sigma} = [H_{\text{inf}}/(2\pi)]^2$ , we have

$$\mathcal{B}^2 = \frac{H_{\text{inf}}^2}{\pi^2 \bar{\sigma}_*^2}. \quad (6.36)$$

Since  $\zeta^{(i)}$  is determined by the inflaton fluctuation and  $S_{\sigma\gamma}$  is determined by the curvaton fluctuation,  $S_{\sigma\gamma}$  and  $\zeta^{(i)}$  are uncorrelated. From Eq. (6.17), we see that

$$\mathcal{P}_\zeta(k) = \mathcal{A}^2 + \mathcal{T}_{\zeta S}^2 \mathcal{B}^2, \quad (6.37)$$

$$\mathcal{P}_S(k) = \mathcal{T}_{SS}^2 \mathcal{B}^2, \quad (6.38)$$

$$\mathcal{C}_{\zeta S}(k) = \mathcal{T}_{\zeta S} \mathcal{T}_{SS} \mathcal{B}^2. \quad (6.39)$$

The CMB power spectrum may be divided into contributions from adiabatic and isocurvature perturbations [206]:

$$C_\ell = (\mathcal{A}^2 + \mathcal{T}_{\zeta S}^2 \mathcal{B}^2) \hat{C}_\ell^{\text{ad}} + \mathcal{T}_{SS}^2 \mathcal{B}^2 \hat{C}_\ell^{\text{iso}} + \mathcal{T}_{\zeta S} \mathcal{T}_{SS} \mathcal{B}^2 \hat{C}_\ell^{\text{cor}}. \quad (6.40)$$

In this decomposition,  $\hat{C}_\ell^{\text{ad}}$  is the CMB power spectrum derived from a flat spectrum of adiabatic fluctuations with  $\mathcal{P}_\zeta(k) = 1$ , and  $\hat{C}_\ell^{\text{iso}}$  is the CMB power spectrum derived from a flat spectrum of dark-matter isocurvature perturbations with  $\mathcal{P}_S(k) = 1$ . If both isocurvature and adiabatic perturbations are present, with  $\mathcal{P}_\zeta(k) = \mathcal{P}_S(k) = \mathcal{C}_{\zeta S}(k) = 1$ , then the CMB power spectrum is  $\hat{C}_\ell^{\text{ad}} + \hat{C}_\ell^{\text{iso}} + \hat{C}_\ell^{\text{cor}}$ . Figure 6.3 shows these three component spectra, as calculated by CMBFast [207] with WMAP5 best-fit cosmological parameters [18]:  $\Omega_b = 0.0462$ ,  $\Omega_{\text{cdm}} = 0.233$ ,  $\Omega_\Lambda = 0.721$  and  $H_0 = 70.1$  km/s/Mpc.

Figure 6.3 clearly shows that isocurvature perturbations leave a distinctive imprint on the CMB power spectrum. It is therefore possible to constrain the properties of  $\mathcal{P}_\zeta(k)$  and  $\mathcal{P}_S(k)$  using CMB data. These constraints are often reported as bounds on the isocurvature fraction  $\alpha$  and the correlation parameter  $\gamma$ :

$$\alpha \equiv \frac{\mathcal{T}_{SS}^2 \mathcal{B}^2}{\mathcal{A}^2 + \mathcal{T}_{\zeta S}^2 \mathcal{B}^2 + \mathcal{T}_{SS}^2 \mathcal{B}^2}, \quad (6.41)$$

$$\gamma \equiv \text{sign}(\mathcal{T}_{\zeta S} \mathcal{T}_{SS}) \frac{\mathcal{T}_{\zeta S} \mathcal{B}^2}{\mathcal{A}^2 + \mathcal{T}_{\zeta S}^2 \mathcal{B}^2}. \quad (6.42)$$

We will find it useful to continue to define  $\xi$  as the fraction of adiabatic power from the curvaton:

$$\xi \equiv \frac{\mathcal{T}_{\zeta S}^2 \mathcal{B}^2}{\mathcal{A}^2 + \mathcal{T}_{\zeta S}^2 \mathcal{B}^2}, \quad (6.43)$$

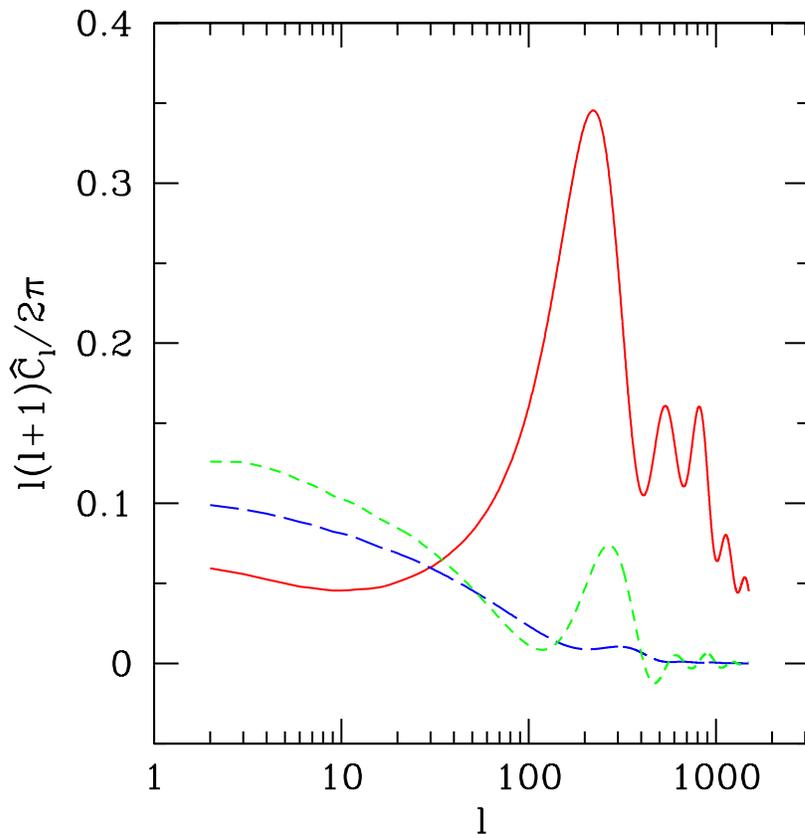


Figure 6.3: CMB power spectra for unit-amplitude initial perturbations. The solid curve is  $\hat{C}_\ell^{\text{ad}}$ : the power spectrum derived from  $\mathcal{P}_\zeta(k) = 1$ . The long-dashed curve is  $\hat{C}_\ell^{\text{iso}}$ : the power spectrum derived from  $\mathcal{P}_S(k) = 1$ . The short-dashed curve is  $\hat{C}_\ell^{\text{cor}}$ : the difference between the power spectrum derived from  $\mathcal{P}_\zeta(k) = \mathcal{P}_S(k) = \mathcal{C}_{\zeta S}(k) = 1$  and  $\hat{C}_\ell^{\text{ad}} + \hat{C}_\ell^{\text{iso}}$ .

with  $\mathcal{T}_{\zeta S} = R/3$ . We then see that

$$\alpha = \frac{9(\xi/R^2)\mathcal{T}_{SS}^2}{1 + 9(\xi/R^2)\mathcal{T}_{SS}^2} \quad (6.44)$$

$$\gamma = \text{sign}(\mathcal{T}_{SS})\xi. \quad (6.45)$$

Ideally, we would like to use constraints for  $\alpha$  and  $\gamma$  that were derived assuming only that  $n_{\text{ad}} \simeq n_{\text{iso}} \simeq 1$ , where  $n_{\text{ad}}$  and  $n_{\text{iso}}$  are the spectral indices for  $\mathcal{P}_\zeta(k)$  and  $\mathcal{P}_S(k)$  respectively. Unfortunately, such an analysis does not exist. The most general analyses [206, 208, 209] make no assumptions regarding  $n_{\text{iso}}$  and conclude that models with  $n_{\text{iso}} \simeq 2$  provide the best fit to the data. Since their bounds on  $\alpha$  and  $\gamma$  are marginalized over a range of  $n_{\text{iso}}$  values that are unreachable in the curvaton scenario, these constraints are not applicable to our model.

There are analyses that specifically target the curvaton scenario, but they assume that the curvaton generates all of the primordial fluctuations (i. e.,  $\mathcal{A}^2 \ll \mathcal{T}_{\zeta S}^2 \mathcal{B}^2$ ) [208, 18]. In this case,

$\xi = 1$ , and the isocurvature and adiabatic fluctuations are completely correlated or anti-correlated, depending on the sign of  $\mathcal{T}_{SS}$ . Furthermore, Eq. (6.44) shows that  $\alpha \gtrsim 0.9$  if  $\xi = 1$  and  $\mathcal{T}_{SS}^2 \gtrsim R^2$ . Since this high value for  $\alpha$  is thoroughly ruled out, these analyses of isocurvature perturbations in the curvaton scenario disregard the possibility that  $\tilde{B}R \ll 1$  and assume that most of the dark matter is created by curvaton decay. In this case,  $\mathcal{T}_{SS}$  is given by Eq. (6.28) and the derived upper bound on  $\alpha$  ( $\alpha < 0.0041$  from Ref. [18]) implies that  $R > 0.98$ . Since we require  $R \ll 1$ , we can conclude that we will be restricted to mixed-perturbation scenarios in which both the curvaton and the inflaton contribute to the adiabatic perturbation spectrum and  $\xi < 1$ .

Finally, some analyses constrain completely uncorrelated ( $\gamma = 0$ ) isocurvature and adiabatic perturbations (a.k.a. axion-type isocurvature) with  $n_{\text{iso}} = 1$  [210, 18]. These constraints are relevant to our models, however, because we will see that  $\xi = |\gamma|$  must be small to create an asymmetry that vanishes on small scales. (The discussion in the previous paragraph also foreshadows the fact that  $\xi \ll 1$  will be necessary to obtain  $R \ll 1$ .) We will therefore use the bound on  $\alpha$  derived for uncorrelated adiabatic and isocurvature in our analysis. WMAP5 data alone constrains  $\alpha < 0.16$  at 95% confidence, but the upper bound on  $\alpha$  is significantly reduced if BAO and SN data are used to break a degeneracy between  $\alpha$  and  $n_{\text{ad}}$  [18]. With the combined WMAP5+BAO+SN dataset, the 95% C.L. upper bound on  $\alpha$  is

$$\alpha < 0.072, \quad (6.46)$$

with a best-fit value of  $n_{\text{ad}} \simeq 1$ . Ref. [210] found a similar bound:  $\alpha < 0.08$  at 95% C.L.

The other observable effect of isocurvature fluctuations that we must consider is non-Gaussianity. Following Ref. [211], we define  $f_{\text{NL}}^{(\text{iso})}$  through

$$S_{m\gamma} = \eta + f_{\text{NL}}^{(\text{iso})} (\eta^2 - \langle \eta^2 \rangle), \quad (6.47)$$

where  $\eta$  is drawn from a Gaussian probability spectrum. This is analogous to the definition of  $f_{\text{NL}}$  for adiabatic perturbations [196]. For isocurvature perturbations from the curvaton,

$$S_{m\gamma} = \mathcal{T}_{SS} S_{\sigma\gamma} = \mathcal{T}_{SS} \left[ 2 \frac{\delta\sigma_*}{\bar{\sigma}_*} + \left( \frac{\delta\sigma_*}{\bar{\sigma}_*} \right)^2 \right], \quad (6.48)$$

and we can set  $\eta = 2\mathcal{T}_{SS}\delta\sigma_*/\bar{\sigma}_*$ . Thus we see that

$$f_{\text{NL}}^{(\text{iso})} = \frac{1}{4\mathcal{T}_{SS}} \quad (6.49)$$

for the curvaton model. Given the current upper bound on  $\alpha$ ,  $f_{\text{NL}}^{(\text{iso})} \simeq 10^4$  produces a CMB bispectrum that is equal in magnitude to the CMB bispectrum if  $f_{\text{NL}} \simeq 20$  for purely adiabatic perturbations [211]. Since the current upper limit on  $f_{\text{NL}}$  from the CMB is  $f_{\text{NL}} \lesssim 100$  [199, 59,

18, 61, 62], we see that the non-Gaussianity of the isocurvature fluctuations is undetectable for  $\mathcal{T}_{SS} \gtrsim 10^{-5}$ . Recall from Section 6.2.2 that the curvaton also introduces non-Gaussianity in the adiabatic perturbations;  $f_{\text{NL}}$  for mixed perturbations from the inflaton and curvaton is given by [186, 198]

$$f_{\text{NL}} = \frac{5\xi^2}{4R}. \quad (6.50)$$

## 6.4 A power asymmetry from curvaton isocurvature

In Section 6.2.2 we proposed that the hemispherical power asymmetry in the CMB could result from a large-amplitude superhorizon fluctuation in the initial value of the curvaton field  $\sigma_*$ , as depicted in Fig. 6.4. The difference between  $\bar{\sigma}_*$  on one side of the surface of last scatter and its average value in the observable Universe,  $\Delta\bar{\sigma}_*$ , will introduce a power asymmetry  $\Delta C_\ell$  in the CMB through Eq. (6.40). The CMB power spectrum depends on  $\bar{\sigma}_*$  through  $\mathcal{B}^2$ , as given by Eq. (6.36), and through  $\mathcal{T}_{\zeta S}$  and  $\mathcal{T}_{SS}$ , which are functions of  $R$  and depend on  $\bar{\sigma}_*$  through Eq. (6.8). Differentiating Eq. (6.40) with respect to  $\sigma_*$  gives

$$\Delta C_\ell = 2 \frac{\Delta\bar{\sigma}_*}{\bar{\sigma}_*} \mathcal{B}^2 \left[ \frac{R^2}{9} \hat{C}_\ell^{\text{rad}} - \left( \mathcal{T}_{SS}^2 - 2\mathcal{T}_{SS}R \frac{d\mathcal{T}_{SS}}{dR} \right) \hat{C}_\ell^{\text{iso}} + \frac{R^2}{3} \frac{d\mathcal{T}_{SS}}{dR} \hat{C}_\ell^{\text{cor}} \right], \quad (6.51)$$

where we have used  $\mathcal{T}_{\zeta S} = R/3$  for the curvaton scenario.

If the curvaton also generates isocurvature perturbations, the power asymmetry will be scale-dependent due to the differences between  $\hat{C}_\ell^{\text{rad}}$ ,  $\hat{C}_\ell^{\text{iso}}$ , and  $\hat{C}_\ell^{\text{cor}}$  shown in Fig. 6.3. We extract this scale-dependence by defining  $K_\ell$  through

$$\left| \frac{\Delta C_\ell}{C_\ell} \right| \equiv 2 \frac{\Delta\bar{\sigma}_*}{\bar{\sigma}_*} K_\ell. \quad (6.52)$$

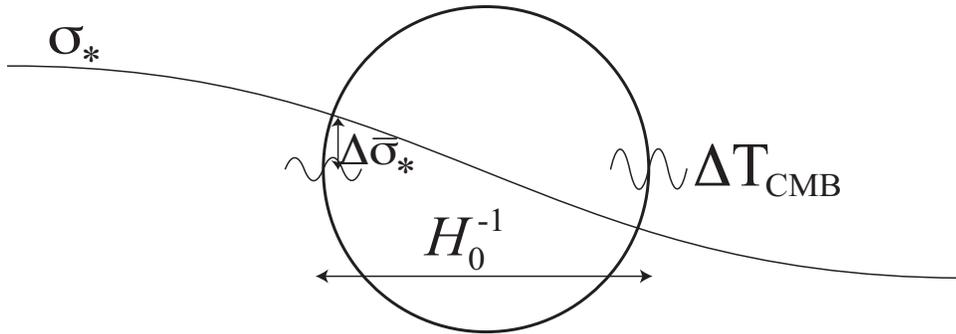


Figure 6.4: Measurements of temperature fluctuations in the CMB show that the rms temperature-fluctuation amplitude is larger in one side of the sky than in the other. We propose that this asymmetry is generated by a large-amplitude fluctuation in the initial value of the curvaton field  $\sigma_*$ . The fluctuation in  $\sigma_*$  across the observable Universe is  $\Delta\bar{\sigma}_*$ .

The dipolar modulation parameter  $A$  used by Refs. [44, 46] and defined in Eq. (6.1) describes the asymmetry in the amplitude of temperature fluctuations, so for small  $A$ ,  $A \simeq (1/2)(\Delta C_\ell/C_\ell)$ . The modulation is assumed to be scale-invariant and is measured for  $\ell \leq \ell_{\max}$ . To relate the scale-dependent power asymmetry described by  $K_\ell$  to  $A$ , we assume that, for all  $\ell$  between 2 and  $\ell_{\max}$ , each of the  $2\ell + 1$  modes contributes equally to the measured asymmetry. Since there are  $(\ell_{\max} - 1)(\ell_{\max} + 3)$  modes in total,

$$A = \frac{\Delta\bar{\sigma}_*}{\bar{\sigma}_*} \sum_{\ell=2}^{\ell_{\max}} \frac{2\ell + 1}{(\ell_{\max} - 1)(\ell_{\max} + 3)} K_\ell \equiv \frac{\Delta\bar{\sigma}_*}{\bar{\sigma}_*} \tilde{A}. \quad (6.53)$$

We note that  $\tilde{A}$  does not depend on the amplitude of the superhorizon fluctuation; it is determined by  $\mathcal{T}_{SS}$ ,  $R$ , and  $\xi$ . Since  $\Delta\bar{\sigma}_*$  cannot be larger than  $\bar{\sigma}_*$ ,  $\tilde{A}$  is the largest asymmetry that can be produced by a particular curvaton scenario. Unless otherwise noted, we set  $\ell_{\max} = 64$  to match Ref. [46]. As mentioned previously, Ref. [46] found that  $A = 0.072 \pm 0.022$  for  $\ell \lesssim 64$ , yet the isotropic distribution of quasars constrains  $A \lesssim 0.012$  for  $k \simeq 1.3h - 1.8h \text{ Mpc}^{-1}$  [90].

In the following subsections we will examine  $K_\ell$  for the two limiting cases discussed in Section 6.3.1. First, we will consider scenarios in which most of the dark matter is created during curvaton decay and  $\mathcal{T}_{SS} \simeq 1 - R$ . Then we will consider scenarios in which the curvaton's contribution to the dark matter is negligible and  $\mathcal{T}_{SS} = \kappa R$  with  $-1 \lesssim \kappa \lesssim 1/R$ . In both cases, we will see that  $K_\ell$  decreases rapidly when  $\ell \gtrsim 10$ . We will also find that the superhorizon curvaton fluctuation required to generate the observed asymmetry must have a large amplitude:  $\Delta\bar{\sigma}_*/\bar{\sigma}_* \gtrsim 1/2$ . We therefore must consider how this large-amplitude superhorizon fluctuation will create large-scale anisotropies in the CMB through the Grishchuk-Zel'dovich effect [85].

We saw in Chapter 5 that a superhorizon adiabatic fluctuation does not induce a dipolar anisotropy in the CMB due to a cancellation between the intrinsic dipole and the Doppler dipole, but this is not the case for superhorizon isocurvature fluctuations [212, 213]. After matter-domination, the evolution of the potential  $\Psi$  and the fluid velocity's dependence on  $\Psi$  are the same for adiabatic and isocurvature initial conditions [171]. Therefore, the induced Doppler dipole and the anisotropy from the integrated Sachs-Wolfe effect will be the same for adiabatic and isocurvature fluctuations if the surface of last scatter is taken to be in the matter-dominated era. The only difference between the CMB dipole induced by an adiabatic perturbation and the CMB dipole induced by an isocurvature perturbation arises from the Sachs-Wolfe anisotropy; for adiabatic perturbations  $(\Delta T/T)_{\text{SW}} = \Psi_{\text{dec}}/3$ , while  $(\Delta T/T)_{\text{SW}} = 2\Psi_{\text{dec}}$  for isocurvature perturbations, where  $\Psi_{\text{dec}}$  is evaluated at the time of decoupling. Since we know that the integrated Sachs-Wolfe effect and the Doppler dipole exactly cancel the Sachs-Wolfe anisotropy for adiabatic perturbations, the residual dipolar anisotropy for isocurvature fluctuations must be  $5\Psi_{\text{dec}}/3$ .

If  $S_0$  is the initial matter isocurvature fluctuation set deep in the radiation-dominated era, then

$\Psi_{\text{dec}} = -S_0/5$  [171]. We are considering dark-matter isocurvature fluctuations, so we have  $S_0 = S_{m\gamma}\Omega_{\text{cdm}}/(\Omega_{\text{cdm}} + \Omega_{\text{b}})$ , where  $S_{m\gamma}$  is given by Eq. (6.48). We treat the superhorizon fluctuation in the curvaton field as a sine wave:  $\delta\sigma_* = \sigma_{\vec{k}}\sin(\vec{k}\cdot\vec{x})$ , where  $k \ll H_0$ . By choosing this form for  $\delta\sigma$ , we have placed ourselves at the node of the sine wave, but the constraints we derive on  $\delta\sigma$  are not strongly dependent on this choice, as we saw in Section 5.4. To lowest order in  $kx_{\text{dec}}$ , where  $x_{\text{dec}}$  is the comoving distance to the last scattering surface, the dipolar moment generated by the superhorizon curvaton fluctuation is

$$|a_{10}| = \frac{1}{3}\sqrt{\frac{4\pi}{3}}(kx_{\text{dec}})\frac{\Omega_{\text{cdm}}}{\Omega_{\text{cdm}} + \Omega_{\text{b}}}\left(2\mathcal{T}_{SS}\frac{\sigma_{\vec{k}}}{\bar{\sigma}_*}\right), \quad (6.54)$$

where we have chosen axes that are aligned with the asymmetry ( $\hat{z} = \hat{k}$ ). The variation in  $\sigma$  across the surface of last scattering is  $\Delta\bar{\sigma}_* = \sigma_{\vec{k}}(kx_{\text{dec}})$ , and it is constrained by the dipolar anisotropy:

$$\mathcal{T}_{SS}\left(\frac{\Delta\bar{\sigma}_*}{\bar{\sigma}_*}\right) \lesssim 0.9\mathcal{D}, \quad (6.55)$$

where  $\mathcal{D}$  is the largest value of  $|a_{10}|$  that is consistent with observations of the CMB dipole. The observed dipolar anisotropy has amplitude  $\Delta T/T \simeq 10^{-3}$  [214], but it is not aligned with the asymmetry, and at least a portion of it is attributable to our proper motion [215, 216]. We therefore take  $\mathcal{D} = 10^{-4}$  to generate a conservative upper bound.

The superhorizon fluctuation in the curvaton field will also generate a quadrupolar anisotropy in the CMB. The induced quadrupole is higher-order in  $\Delta\bar{\sigma}_*/\bar{\sigma}_*$  because it originates from the quadratic term in  $S_{\sigma\gamma}$  (see Eq. 6.48). Nevertheless, the upper bound on  $\Delta\bar{\sigma}_*/\bar{\sigma}_*$  from the CMB quadrupole is similar to the bound from the dipole because observations of the CMB quadrupole are not contaminated by our proper motion. The CMB quadrupole is the sum of contributions from the superhorizon isocurvature perturbation and the superhorizon adiabatic perturbation ( $\zeta = \mathcal{T}_{\zeta S}S_{\sigma\gamma}$  during radiation domination) generated by the curvaton field. In the coordinate system defined by the asymmetry,

$$|a_{20}| = \frac{1}{3}\sqrt{\frac{4\pi}{5}}(kx_{\text{dec}})^2\left(\frac{\sigma_{\vec{k}}}{\bar{\sigma}_*}\right)^2\left[\delta_2^{\text{ad}}\left(\frac{2R}{5}\right) + \delta_2^{\text{iso}}\frac{\Omega_{\text{cdm}}}{\Omega_{\text{cdm}} + \Omega_{\text{b}}}\left(\frac{2\mathcal{T}_{SS}}{5}\right)\right], \quad (6.56)$$

where  $\delta_2^{\text{ad}}$  is derived by analyzing the Sachs-Wolfe effect, the integrated Sachs-Wolfe effect, and the fluid velocity at the surface of last scatter generated by a superhorizon adiabatic perturbation (see Chapter 5). In the limit that decoupling occurs after matter-domination,  $\delta_2^{\text{ad}} = 0.338$ , and  $\delta_2^{\text{iso}} = 5/3 + \delta_2^{\text{ad}}$  since only the contribution from the Sachs-Wolfe effect is different for isocurvature

initial conditions. It follows that the upper-bound on  $\Delta\bar{\sigma}_*/\bar{\sigma}_*$  from the CMB quadrupole is

$$(0.34 R + 1.7 \mathcal{T}_{SS}) \left( \frac{\Delta\bar{\sigma}_*}{\bar{\sigma}_*} \right)^2 \lesssim 4.7 Q, \quad (6.57)$$

where  $Q$  is the largest value of  $|a_{20}|$  that is consistent with observations of the CMB quadrupole. As in Chapter 5, we take  $Q = 3\sqrt{C_2} \lesssim 1.8 \times 10^{-5}$  as a  $3\sigma$  upper limit.

#### 6.4.1 Case 1: The curvaton creates most of the dark matter.

If most of the dark matter is created when the curvaton decays, then  $\mathcal{T}_{SS} \simeq 1 - R$ , as in Eq. (6.28).

In this case, Eqs. (6.40) and (6.51) imply

$$\begin{aligned} \frac{\Delta C_\ell}{C_\ell} &\simeq 2 \frac{\Delta\bar{\sigma}_*}{\bar{\sigma}_*} \left( \frac{\mathcal{B}^2 R^2}{9} \right) \left[ \frac{\hat{C}_\ell^{\text{ad}} - \frac{9}{R^2} \hat{C}_\ell^{\text{iso}} - 3\hat{C}_\ell^{\text{cor}}}{(\mathcal{A}^2 + \frac{\mathcal{B}^2 R^2}{9}) \hat{C}_\ell^{\text{ad}} + \frac{\mathcal{B}^2 R^2}{9} \left( \frac{9}{R^2} \hat{C}_\ell^{\text{iso}} + \frac{3}{R} \hat{C}_\ell^{\text{cor}} \right)} \right], \\ &\simeq 2 \frac{\Delta\bar{\sigma}_*}{\bar{\sigma}_*} \xi \left[ \frac{\hat{C}_\ell^{\text{ad}} - \frac{9}{R^2} \hat{C}_\ell^{\text{iso}} - 3\hat{C}_\ell^{\text{cor}}}{\hat{C}_\ell^{\text{ad}} + \xi \left( \frac{9}{R^2} \hat{C}_\ell^{\text{iso}} + \frac{3}{R} \hat{C}_\ell^{\text{cor}} \right)} \right], \end{aligned} \quad (6.58)$$

where we have kept only the leading-order term in  $R$  in the coefficients of  $\hat{C}_\ell^{\text{iso}}$  and  $\hat{C}_\ell^{\text{cor}}$ . We can also assume that  $R\hat{C}_\ell^{\text{cor}} \ll \hat{C}_\ell^{\text{iso}}$  since Fig. 6.3 shows that  $\hat{C}_\ell^{\text{iso}} \simeq \hat{C}_\ell^{\text{cor}}$ . Finally, if  $R \lesssim 0.01$ , then  $R^2\hat{C}_\ell^{\text{ad}} \ll \hat{C}_\ell^{\text{iso}}$  for  $\ell \lesssim 1500$ , and we may neglect  $\hat{C}_\ell^{\text{ad}}$  in the numerator. With these simplifications, we have  $\Delta C_\ell/C_\ell = -2(\Delta\bar{\sigma}_*/\bar{\sigma}_*)K_\ell$  where

$$K_\ell \simeq \frac{\frac{9\xi}{R^2} \hat{C}_\ell^{\text{iso}}}{\hat{C}_\ell^{\text{ad}} + \frac{9\xi}{R^2} \hat{C}_\ell^{\text{iso}}}. \quad (6.59)$$

This approximate expression for  $K_\ell$  is useful because it only depends on  $\xi/R^2$ . It is accurate to within 1% for  $\ell \leq 1500$  if  $R \leq 0.01$  and accurate to within 0.1% if  $R \leq 0.001$ . However, it does not have the appropriate limit for  $\ell \rightarrow \infty$ ; since  $\hat{C}_\ell^{\text{iso}}/\hat{C}_\ell^{\text{ad}} \rightarrow 0$  in this limit, Eq. (6.58) shows that  $K_\ell \rightarrow \xi$ , but the approximate form goes to zero.

Figure 6.5 shows the approximate form of  $K_\ell$ , given by Eq. (6.59), for four values of  $\xi/R^2$ : 0.005, 0.006, 0.007, and 0.0086. We see that  $K_\ell$  increases on large scales as  $\xi/R^2$  increases. On smaller scales, we see that  $K_\ell$  is not very sensitive to changes in  $\xi/R^2$ . Thus, to obtain the desired asymmetry on large scales and nearly no asymmetry on small scales, we just need to increase  $\xi/R^2$ ! Unfortunately, the upper bound on  $\alpha$  places an upper bound on  $\xi/R^2$ :

$$\alpha < 0.072 \implies \frac{\xi}{R^2} < 0.0086. \quad (6.60)$$

The solid curve in Fig. 6.5 corresponds to  $\xi/R^2 = 0.0086$  and is therefore the maximal  $K_\ell$  curve that

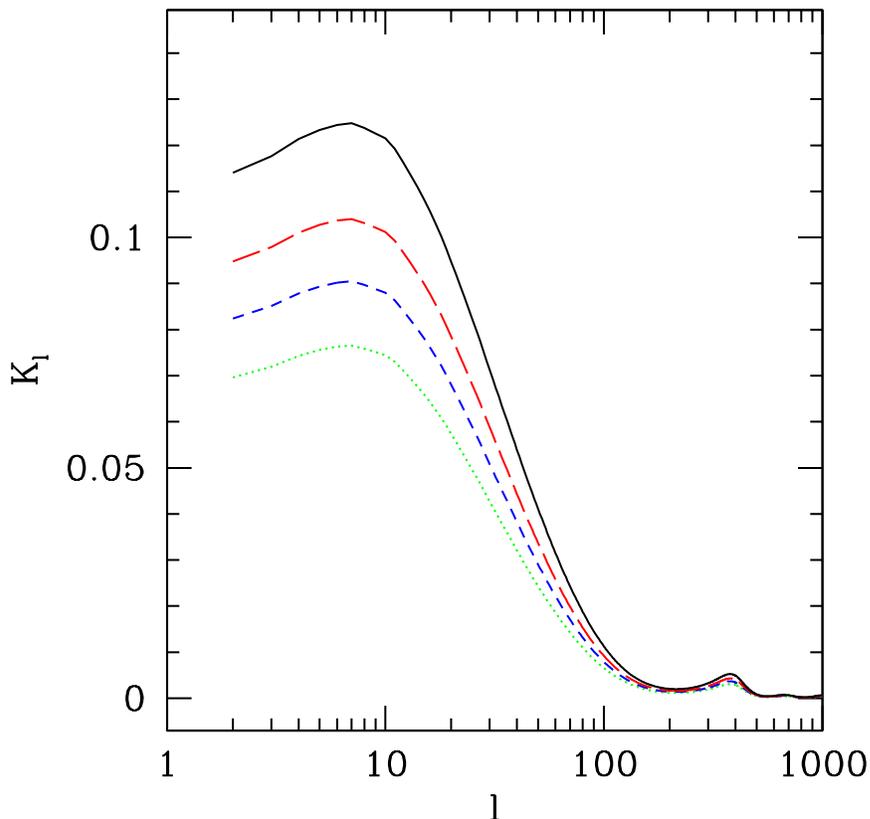


Figure 6.5:  $K_\ell$  for scenarios in which most of the dark matter comes from curvaton decay. The power asymmetry is given by  $\Delta C_\ell/C_\ell = -2(\Delta\bar{\sigma}_*/\bar{\sigma})K_\ell$ . The solid curve corresponds to  $\xi/R^2 = 0.0086$ , which saturates the current bound on power from isocurvature perturbations. The lower curves have  $\xi/R^2 = 0.007$  (long-dashed), 0.006 (short-dashed) and 0.005 (dotted). For descending values of  $\xi/R^2$ , these curves correspond to asymmetry amplitudes  $\tilde{A} = 0.055, 0.045, 0.039$ , and 0.033.

is consistent with the current limits on the isocurvature contribution to the CMB power spectrum. We also note that satisfying the upper bound on  $\alpha$  requires  $\xi$  to be much smaller than  $R$ , and we have assumed that  $R \ll 1$ . The adiabatic and isocurvature fluctuations are therefore uncorrelated.

Figure 6.5 also shows that  $K_\ell$  peaks for  $\ell \simeq 10$  and decreases rapidly as  $\ell$  increases from 10 to 100. Furthermore, the asymmetry nearly vanishes for larger  $\ell$ , so it will be easy to satisfy the quasar constraint. The desired scale-dependence comes at a cost though; the smaller values of  $K_\ell$  at  $\ell \gtrsim 20$  dilute the scale-averaged asymmetry  $A$ . For  $\xi/R^2 = 0.0086$ , the maximal asymmetry is given by  $\tilde{A} = 0.055$ . Thus we see that saturating the upper bound on isocurvature power ( $\alpha$ ) and setting  $\Delta\bar{\sigma}_* = \bar{\sigma}_*$  leads to an asymmetry that is almost  $1\sigma$  below the observed value. Moreover, the curvaton creates most of the dark matter in this scenario; since  $\Delta\bar{\sigma}_*/\bar{\sigma} \simeq 1$  is required to generate sufficient asymmetry, this model requires that the dark matter density varies by a factor of unity across the observable Universe! Unsurprisingly, such a large isocurvature fluctuation is not consistent with the large-scale homogeneity of the CMB. Since  $\mathcal{T}_{SS} \simeq 1$  in this scenario, the CMB

dipole constrains  $\Delta\bar{\sigma}_*/\bar{\sigma}_* \lesssim 10^{-4}$  from Eq. (6.55). We conclude that the curvaton cannot generate the observed power asymmetry if the dark matter is created during curvaton decay.

### 6.4.2 Case 2: The curvaton's contribution to the dark matter is negligible

We now turn our attention to the opposite scenario, in which the curvaton's contribution to the dark matter density is insignificant. In this case, Eq. (6.29) tells us that  $\mathcal{T}_{SS} \simeq \kappa R$ , with  $-1 \lesssim \kappa \lesssim 1/R$ . We will see, however, that this model can generate the observed asymmetry only if  $\kappa \lesssim 1.4$ . Thus, we will be considering scenarios in which the curvaton generates adiabatic and isocurvature fluctuations that are equal in magnitude ( $\mathcal{T}_{SS}^2 \simeq \mathcal{T}_{\zeta S}^2$ ), in stark contrast to the scenarios considered in the previous section. We anticipate that generating comparable adiabatic and isocurvature fluctuations from the curvaton will be advantageous for two reasons. First, the asymmetry can be partially contained in the adiabatic perturbations, which will make it easier to generate the observed asymmetry without violating the current bounds on isocurvature power. Second, the superhorizon isocurvature perturbation generated by  $\Delta\bar{\sigma}_*$  will be proportional to  $R$  and can therefore be reduced by decreasing  $R$ . The downside is that it will be difficult to make the asymmetry sufficiently scale-dependent to satisfy the quasar bound because the adiabatic perturbations are asymmetric as well.

For  $\mathcal{T}_{SS} = \kappa R$ , the power asymmetry generated by the superhorizon curvaton perturbation is given by  $\Delta C_\ell/C_\ell = 2(\Delta\bar{\sigma}_*/\bar{\sigma}_*)K_\ell$  where, from Eqs. (6.40) and (6.51), we have

$$K_\ell = \xi \left[ \frac{\hat{C}_\ell^{\text{ad}} + 9\kappa^2 \hat{C}_\ell^{\text{iso}} + 3\kappa \hat{C}_\ell^{\text{cor}}}{\hat{C}_\ell^{\text{ad}} + \xi (9\kappa^2 \hat{C}_\ell^{\text{iso}} + 3\kappa \hat{C}_\ell^{\text{cor}})} \right]. \quad (6.61)$$

We see that  $K_\ell \rightarrow \xi$  as  $\ell \rightarrow \infty$  as expected; on small scales, the only source of asymmetry is the adiabatic power from the curvaton. We can therefore anticipate that the quasar constraint will place an upper bound on  $\xi$ . We also see that all the isocurvature contributions to the power asymmetry are proportional to  $\kappa$  or  $\kappa^2$ , and this implies that the necessary scale-dependence of  $K_\ell$  will place a lower limit on  $|\kappa|$ .

Differentiating  $K_\ell$  with respect to  $\xi$  and  $|\kappa|$  reveals that increasing  $\xi$  or  $|\kappa|$  increases  $K_\ell$ , unless  $-0.2 \lesssim \kappa \lesssim 0$ , in which case the  $\hat{C}_\ell^{\text{iso}}$  and  $\hat{C}_\ell^{\text{cor}}$  terms partially cancel on large scales, leaving  $K_\ell$  nearly scale-invariant. Unfortunately, the upper limit in isocurvature power places an upper limit on  $|\kappa|$  and  $\xi$ :

$$\alpha < 0.072 \implies \kappa^2 \xi < 0.0086. \quad (6.62)$$

If we differentiate  $K_\ell$  with respect to  $|\kappa|$  while keeping  $\kappa^2 \xi$  fixed, we find that increasing  $|\kappa|$  decreases  $K_\ell$  for  $\kappa > -0.3$  and increases  $K_\ell$  for  $\kappa < -0.3$ . Furthermore, the  $|\kappa| \rightarrow \infty$  limit of  $K_\ell$ , with fixed

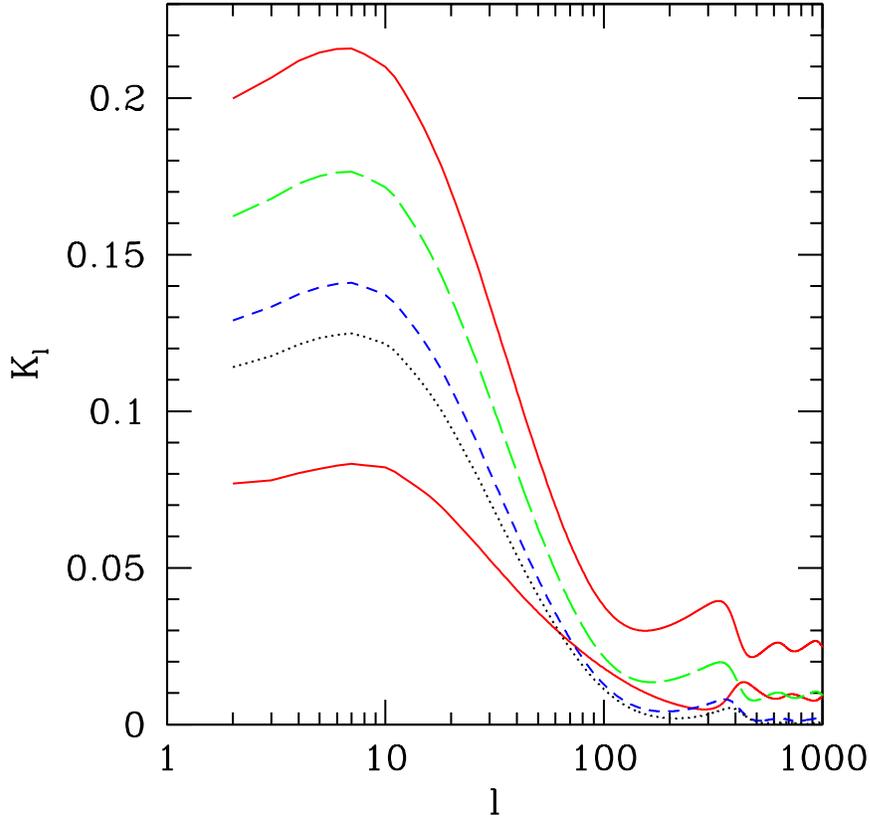


Figure 6.6:  $K_\ell$  for scenarios in which the curvaton's contribution to the dark matter density is negligible and  $\mathcal{T}_{SS} = \kappa R$ . The power asymmetry is given by  $\Delta C_\ell / C_\ell = 2(\Delta\bar{\sigma}_*/\bar{\sigma})K_\ell$ . All of the curves have  $\kappa^2\xi = 0.0086$ , which saturates the upper limit on isocurvature power. The top three curves have  $\kappa = 0.6$  (top, solid),  $\kappa = 1$  (long-dashed), and  $\kappa = 3$  (short-dashed). The maximal scale-averaged asymmetries possible for these curves are  $\tilde{A} = 0.11$ ,  $\tilde{A} = 0.081$ , and  $\tilde{A} = 0.062$ . The dotted curve is the limit as  $\kappa \rightarrow \infty$ , and it has  $\tilde{A} = 0.055$ . The bottom solid curve has  $\kappa = -1$  and  $\tilde{A} = 0.043$ .

$\kappa^2\xi$ , is

$$\lim_{|\kappa| \rightarrow \infty; \text{fixed } \kappa^2\xi} K_\ell = \frac{9(\kappa^2\xi)\hat{C}_\ell^{\text{iso}}}{\hat{C}_\ell^{\text{ad}} + 9(\kappa^2\xi)\hat{C}_\ell^{\text{iso}}}. \quad (6.63)$$

Figure 6.6 shows  $K_\ell$  with  $\kappa^2\xi = 0.0086$  for various values of  $\kappa$ . We see that as  $\kappa$  increases from zero, the curves rapidly approach the dotted curve, which is Eq (6.63). If we could decrease  $\kappa$  toward  $-\infty$ , the curves would approach this limit from below, but in Section 6.3.1 we found that  $\kappa \gtrsim -1$ , which corresponds the lower solid curve.

The observed asymmetry is  $A = 0.072 \pm 0.022$  for  $\ell \lesssim 64$ , which requires  $K_\ell \gtrsim 0.08$  on average over this  $\ell$  range. Comparing Figs. 6.5 and 6.6 reveals that it is much easier to generate the required asymmetry if the curvaton's contribution to the dark matter is insignificant because the peak in  $K_\ell$  is higher for  $\mathcal{T}_{SS} \simeq R$  than for  $\mathcal{T}_{SS} \simeq 1$ . If we saturate the upper bound on isocurvature power by setting  $\kappa^2\xi = 0.0086$ , then  $\tilde{A} \gtrsim 0.08$  if  $0 < \kappa \leq 1$ , and any positive value of  $\kappa$  has  $\tilde{A} \geq 0.055$ , which

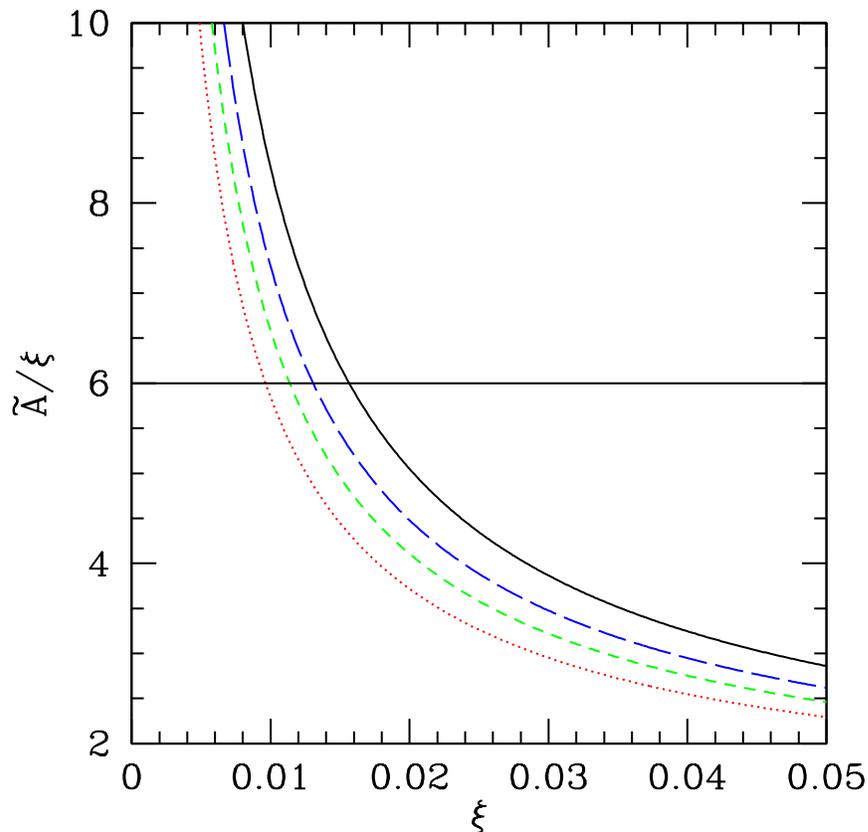


Figure 6.7: The ratio  $\tilde{A}/\xi$ , with  $\ell_{\max} = 64$ , as a function of  $\xi$  for four values of  $\kappa^2\xi$ :  $\kappa^2\xi = 0.0086$  (solid), 0.007 (long-dashed), 0.006 (short-dashed), and 0.005 (dotted). Since  $K_\ell \rightarrow \xi$  as  $\ell \rightarrow \infty$ , this ratio illustrates the fractional enhancement in the asymmetry on large scales ( $\ell \leq 64$ ) compared to small scales. Since we require the asymmetry to be about 6 times larger on large scales than on small scales, we see that we require  $\xi \lesssim 0.016$ .

is less than  $1\sigma$  below the observed value if  $\Delta\bar{\sigma}_*/\bar{\sigma}_* \simeq 1$ . Negative values of  $\kappa$  are less promising;  $\kappa = -1$  maximizes  $K_\ell$  for negative  $\kappa$ , and it gives  $\tilde{A} = 0.043$  for  $\ell_{\max} = 64$ . We will therefore only consider positive values for  $\kappa$  for the rest of the analysis.

Figure 6.6 also illustrates how  $\xi$  determines the small-scale value of  $K_\ell$ . As  $\xi$  decreases and  $\kappa^2$  increases,  $K_\ell$  decreases on small scales, and we see that  $\kappa = -1$  and  $\kappa = 1$  give the same small-scale value for  $K_\ell$ . Furthermore, the ratio of  $K_\ell$  on large scales to  $K_\ell$  on small scales decreases with increasing  $\xi$ . We want the asymmetry to go from  $A \simeq 0.072$  on large scales ( $\ell_{\max} = 64$  in the CMB) to  $A \lesssim 0.012$  on small scales ( $k \simeq 1.3h - 1.8h \text{ Mpc}^{-1}$ ). As mentioned above, the isocurvature perturbations' contribution to the total power is negligible on small scales, and any asymmetry is due solely to adiabatic perturbations from the curvaton, which implies that  $A = \xi(\Delta\bar{\sigma}_*/\bar{\sigma}_*)$  on these scales. A reduction in  $A$  from 0.072 on large scales to less than 0.012 on small scales therefore requires that  $\tilde{A}/\xi \gtrsim 6$ . This requirement places an upper bound on  $\xi$ , as shown in Fig. 6.7. For  $\kappa^2\xi = 0.0086$ , which saturates the upper bound on isocurvature power, we see that the required

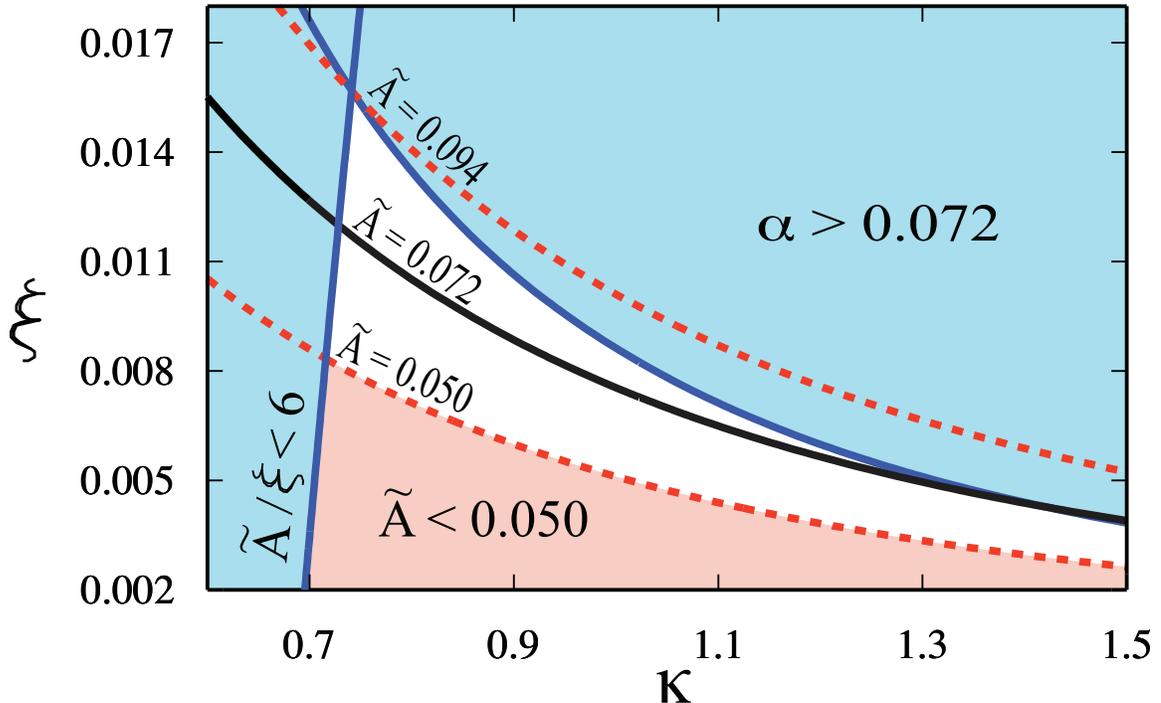


Figure 6.8: The  $\xi - \kappa$  parameter space for models in which the curvaton does not contribute significantly to the dark matter density. The shaded region in the upper-right corner is excluded by the upper bound on isocurvature power ( $\alpha < 0.072$ ), and the left shaded region is excluded by the scale-invariance of the resulting asymmetry ( $\tilde{A}/\xi < 6$ ). The dotted curves show where the maximal possible asymmetry  $\tilde{A}$  equals the observed asymmetry  $\pm 1\sigma$ ; the bottom shaded region cannot produce an asymmetry within  $1\sigma$  of the observed value.

enhancement on large scale is attained only if  $\xi \lesssim 0.016$ , and the upper bound on  $\xi$  decreases with decreasing  $\alpha$ . This upper limit implies that the curvaton contributes only a small fraction of the adiabatic power (although it is a much bigger fraction than in the  $\mathcal{T}_{SS} \simeq 1$  case). The adiabatic and isocurvature fluctuations are therefore nearly uncorrelated.

We now see that the required scale-dependence of the asymmetry limits its magnitude: for fixed  $\alpha$ , the asymmetry is maximized if  $\xi$  is large and  $\kappa$  is small, but increasing  $\xi$  makes the asymmetry more scale-invariant. Figure 6.8 summarizes the constraints on  $\kappa$  and  $\xi$ . We see that only a limited region of the  $\kappa$ - $\xi$  plane can produce asymmetries with  $A \geq 0.072$  while also satisfying the upper bound on isocurvature power ( $\alpha < 0.072$ ) and the quasar constraint ( $\tilde{A}/\xi > 6$ ). If future observations reveal that  $A \lesssim 0.055$ , however, then the allowed range of  $\kappa$  values has no upper bound. The minimum values of  $\alpha$  in the  $\tilde{A} \geq 0.072$  and  $\tilde{A} \geq 0.050$  allowed regions are 0.054 and 0.037, respectively. A moderate improvement in the bound on isocurvature power will therefore rule out this proposed origin of the power asymmetry.

The maximum value for  $\tilde{A}$  is obtained when both the upper bound on  $\alpha$  and the upper bound on  $\xi$  are saturated; as shown in Fig. 6.8,  $\tilde{A} = 0.094$  for  $\xi = 0.016$  and  $\kappa = 0.74$ . Since  $A = (\Delta\bar{\sigma}_*/\bar{\sigma}_*)\tilde{A}$ , we

see that  $A \gtrsim 0.05$ , which is  $1\sigma$  below the observed value, can only be obtained if  $(\Delta\bar{\sigma}_*/\bar{\sigma}_*) \gtrsim 1/2$ . With this lower bound on  $(\Delta\bar{\sigma}_*/\bar{\sigma}_*)$ , the CMB dipole and quadrupole constraints given by Eqs. (6.55) and (6.57) become upper bounds on  $R$  that are inversely proportional to  $\kappa$ . These two bounds are similar in the allowed region with  $\tilde{A} \geq 0.072$ , shown in Fig. 6.8 ( $0.6 \lesssim \kappa \lesssim 1.4$ ). For  $\kappa \simeq 1.4$ ,  $R \lesssim 0.0001$  is required to satisfy the CMB constraints. Since  $K_\ell$  does not depend on  $R$  if  $\mathcal{T}_{SS} = \kappa R$ , it will be possible to evade these constraints without changing the asymmetry. (Even though  $\xi$  depends on  $R$ , we can treat  $\xi$  and  $R$  as independent variables because  $\xi$  also depends on  $\epsilon_H$  and  $R$  does not.)

The upper limit on  $R$  does have consequences for the non-Gaussianity parameters, however. From Eq. (6.50) for  $f_{\text{NL}}$  we see that the upper bound on  $R$  and the constraint  $f_{\text{NL}} \lesssim 100$  implies that  $\xi \lesssim 0.1$ . Since this upper bound is much larger than the  $\xi$  values required to generate the necessary suppression of the power asymmetry on small scales, non-Gaussianity in the adiabatic perturbations is not a concern. The non-Gaussianity from the isocurvature perturbations is bounded from below by the CMB dipole constraint; Eq. (6.55) implies an upper bound on  $\mathcal{T}_{SS}$  that leads directly to a lower bound on  $f_{\text{NL}}^{\text{iso}}$  through Eq. (6.49). Thus the CMB dipole constraint implies that  $f_{\text{NL}}^{\text{iso}} \gtrsim 1390$ , independent of  $\kappa$ . Fortunately, this lower limit is well below current observational constraints [211].

## 6.5 Summary and discussion

A superhorizon perturbation in an inflationary field would introduce a preferred direction in the Universe, and in this chapter, we have investigated how such a perturbation could generate the hemispherical power asymmetry that has been observed in the CMB [42, 43, 44, 45, 46]. We found that the required superhorizon fluctuation in the inflaton field has too large an amplitude to be consistent with measurements of the CMB octupole. A superhorizon fluctuation in a subdominant scalar field, however, is a viable alternative. We determined that a superhorizon curvaton perturbation can generate the observed power asymmetry without introducing unacceptable anisotropy and non-Gaussianity in the CMB, provided that the curvaton is always a subdominant component of the Universe's energy density.

If the curvaton decays while all particle species are still in thermal equilibrium with radiation, then the fluctuations in the curvaton field create only adiabatic perturbations, and the resulting asymmetry is scale-invariant. Recent studies have revealed that the asymmetry is not scale-invariant; while  $\Delta C_\ell/C_\ell \simeq 0.14$  for  $\ell \lesssim 64$  [46], there are indications that this asymmetry does not extend to  $\ell \gtrsim 600$  [182, 45], and an analysis of quasar number counts found  $\Delta P(k)/P(k) \lesssim 0.024$  for  $k \simeq 1.3h - 1.8h \text{ Mpc}^{-1}$  [90].

With the aim of explaining this scale-dependence, we have considered how the asymmetry produced by a large-amplitude curvaton fluctuation changes if the curvaton decays after dark matter

freezes-out. In this scenario, the curvaton produces dark-matter isocurvature perturbations in addition to adiabatic perturbations, and both types of perturbations have asymmetric power. Since isocurvature fluctuations decay after entering the horizon, their contribution to the CMB power spectrum is much greater on large scales ( $\ell \lesssim 100$ ) than on smaller scales, and the magnitude of the asymmetry on small scales will decrease accordingly. (Alternatively, the power asymmetry from the superhorizon fluctuation in the curvaton field can be reduced on smaller scales by making the mixture of curvaton and inflaton perturbations scale-dependent, with the inflaton making a larger contribution to small-scale power. We investigate this possibility in Appendix C, and we find that it is not possible to suppress the asymmetry on small scales sufficiently to satisfy the quasar constraint without violating the upper bound on the running of the scalar spectral index or introducing unacceptable oscillations in the primordial power spectrum.)

There are two limiting cases if the curvaton decays after dark matter freezes out: the majority of the dark matter can be created when the curvaton decays, or the curvaton's contribution to the dark matter density may be insignificant. In the first scenario, the isocurvature fluctuations from the curvaton are much larger than the adiabatic perturbations from the curvaton, and all of the adiabatic power comes from inflaton fluctuations. Since only the isocurvature fluctuations are asymmetric, it is very difficult to generate the observed asymmetry without violating the current bound on power from isocurvature modes. It is necessary to introduce an order-unity variation in the curvaton density across the observable Universe, and since the curvaton creates the dark matter in this model, this would have profound observational consequences. For instance, the resulting large-amplitude isocurvature perturbation induces a dipolar anisotropy in the CMB that is far too large to be consistent with observations. We conclude that it is not possible to generate the observed asymmetry with a superhorizon curvaton fluctuation if the curvaton creates the dark matter.

The second scenario, in which the curvaton's contribution to the dark matter density is negligible, is far more promising. In this scenario, the curvaton produces adiabatic and isocurvature fluctuations of roughly equal amplitude. It is therefore slightly easier to generate the observed asymmetry, but the requirement that the asymmetry magnitude decrease by a factor of six between large and small scales limits the curvaton's contribution to the total adiabatic power to less than 1.6%. Consequently, the variation in the curvaton field across the observable Universe must be greater than 50% to generate the observed asymmetry. Fortunately, the amplitude of both the isocurvature mode and the adiabatic mode generated by this superhorizon variation in the curvaton is proportional to the fraction of the total energy contained in the curvaton at the moment of its decay ( $R$ ). We can therefore suppress any observational signature of the superhorizon curvaton fluctuation in the CMB without altering the asymmetry by decreasing the energy density of the curvaton. Decreasing  $R$  does increase the non-Gaussianity of the fluctuations created by the curvaton, but the resulting non-Gaussianity is well within the current observational bounds.

We conclude that a superhorizon fluctuation in the curvaton field is capable of generating the observed asymmetry in the CMB while satisfying the upper bound on asymmetry in the quasar population if the curvaton's contribution to the dark matter is negligible. Our model has two free parameters:  $\xi$  is the fraction of the adiabatic power that comes from the curvaton field, and  $\kappa$  determines the strength of the isocurvature perturbation created by the curvaton through  $S_{m\gamma} = \kappa R S_{\sigma\gamma}$ . Both  $\xi$  and  $\kappa$  depend on other features of the curvaton model:  $\xi$  depends on the slow-roll parameter  $\epsilon_H$ ,  $R$ , and the initial value of the curvaton field, while  $\kappa$  depends on the fraction of curvaton energy that is converted to dark matter and the dark matter density at curvaton decay. The asymmetry parameter  $A$  for the ILC WMAP5 CMB map is  $0.072 \pm 0.022$ . We find that a very narrow region of the  $\kappa$ - $\xi$  is capable of generating an asymmetry with  $A \gtrsim 0.072$ , as shown in Fig. 6.8, so our model requires a fair amount of fine-tuning. If future observations reveal that  $A \simeq 0.05$ , however, then the allowed region opens up considerably and includes  $\xi \simeq 0$  with  $\kappa \gg 1$ .

Finally, we note that this method of generating a scale-dependent power asymmetry through isocurvature perturbations produces an asymmetry with a specific spectrum. The magnitude of the resulting asymmetry peaks at  $\ell \simeq 10$ , rapidly decreases for  $\ell = 10 - 100$ , and is nearly gone for  $\ell \gtrsim 100$ . Throughout this paper we have considered the scale-averaged asymmetry parameter  $A$  with  $\ell_{\max} = 64$ ; the observed value for this  $\ell$  range in  $V$ -band is  $A = 0.080 \pm 0.021$  [46]. To probe the scale-dependence of the asymmetry, Ref. [46] also considered two other values of  $\ell_{\max}$  in  $V$ -band and found that  $A = 0.119 \pm 0.034$  for  $\ell_{\max} = 40$  and  $A = 0.070 \pm 0.019$  for  $\ell_{\max} = 80$ . Despite the rapid fall-off of the asymmetry generated by our model for  $\ell \gtrsim 10$ , it is consistent with these nearly scale-invariant results. For instance, if  $\kappa = 0.75$  and  $\xi = 0.013$  (a point near the middle of the allowed region in Fig. 6.8), then  $A = 0.113(\Delta\bar{\sigma}_*/\bar{\sigma}_*)$  for  $\ell_{\max} = 40$ ,  $A = 0.080(\Delta\bar{\sigma}_*/\bar{\sigma}_*)$  for  $\ell_{\max} = 64$ , and  $A = 0.065(\Delta\bar{\sigma}_*/\bar{\sigma}_*)$  for  $\ell_{\max} = 80$ .

There are also indications that the asymmetry is present, at least to some extent, out to  $\ell \simeq 600$  [45]. Unfortunately, the asymmetry parameterization employed in Ref. [45] cannot be directly related to the  $A$  parameter, so it is difficult to interpret these results. An analysis analogous to Refs. [44, 46] out to higher  $\ell$  values is required to determine whether the scale-dependence of  $A$  predicted by our model is consistent with observations. Our model also predicts that at least 5.4% (3.7%) of the primordial power comes from isocurvature fluctuations if  $A \gtrsim 0.072$  (0.05). Future searches for isocurvature fluctuations will therefore provide an additional test of our proposed origin of the CMB hemispherical power asymmetry. Finally, the asymmetry in the primordial power should produce anisotropic signatures in the CMB polarization and temperature-polarization correlations that may be observed by future CMB satellites [202, 203].

**Acknowledgments**

The author thanks Marc Kamionkowski, Sean Carroll, and Chris Hirata for their contributions to the research described in this chapter. The author also thanks K. Górski and H. K. Eriksen for useful discussions. This work was supported by DoE DE-FG03-92-ER40701 and the Gordon and Betty Moore Foundation.

## Appendix A

# A review of $f(R)$ gravity's equivalence to scalar-tensor gravity<sup>1</sup>

The action for the scalar-tensor theory that is equivalent to  $f(R)$  gravity is

$$S = \frac{1}{2\kappa} \int d^4x \sqrt{-g} [f(\phi) + f_\phi(\phi)(R - \phi)] + S_m, \quad (\text{A.1})$$

where  $f_\phi(\phi) \equiv df/d\phi$  and  $S_m$  is the matter action. The field equation for  $\phi$  is  $\phi = R$  if  $d^2f/d\phi^2 \neq 0$ . Since the relation between  $\phi$  and  $R$  is purely algebraic, it can be resubstituted into the action to reproduce the action for  $f(R)$  gravity given by Eq. (3.25). After the conformal transformation  $g_{\mu\nu}^E \equiv f_\phi(\phi)g_{\mu\nu}$ , the action becomes that of general relativity with a minimally coupled scalar field:

$$S = \frac{1}{2\kappa} \int d^4x \sqrt{-g_E} \left( R_E - \frac{3}{2f_\phi(\phi)^2} g_E^{\mu\nu} [\nabla_{E\mu} f_\phi(\phi)] [\nabla_{E\nu} f_\phi(\phi)] - \frac{1}{f_\phi(\phi)^2} [\phi f_\phi(\phi) - f(\phi)] \right) + S_m. \quad (\text{A.2})$$

Introducing a canonical scalar field  $\varphi$  such that  $f_\phi(\phi) = \exp(\varphi\sqrt{2\kappa/3})$ , Eq. (A.2) can be rewritten as

$$S = \int d^4x \sqrt{-g_E} \left( \frac{1}{2\kappa} R_E - \frac{1}{2} (\nabla_E \varphi)^2 - V(\varphi) \right) + S_m,$$

where the potential is defined by

$$V(\varphi) \equiv \frac{\phi(\varphi) f_\phi[\phi(\varphi)] - f[\phi(\varphi)]}{2\kappa f_\phi[\phi(\varphi)]^2}. \quad (\text{A.3})$$

The absence of the kinetic term in Eq. (A.1) implies the Brans-Dicke parameter of  $f(R)$  gravity theories is  $\omega = 0$  [115]. From an analysis of Brans-Dicke gravity, if the scalar degree of freedom can propagate on scales much larger than the Solar System, we can conclude that  $\gamma = (1 + \omega)/(2 + \omega) = 1/2$  [115].

In the frame where  $\varphi$  is canonical (the Einstein frame)  $\varphi$  has the equation of motion

$$\square_E \varphi = \frac{dV}{d\varphi} + \sqrt{\frac{\kappa}{6}} f'(\phi)^{-2} T^M, \quad (\text{A.4})$$

where the prime denotes differentiation with respect to  $\phi$ . When we re-express Eq. (A.4) in terms of

---

<sup>1</sup>This appendix was published in *Solar System constraints to general  $f(R)$  gravity*, Takeshi Chiba, Tristan L. Smith, and Adrienne L. Erickcek; Phys. Rev. D **75**, 124014 (2007). Reproduced here with permission, copyright (2007) by the American Physical Society.

$f'(\phi)$  and the usual metric  $g_{\mu\nu}$ , we recover Eq. (3.31). Therefore, we stress that this reformulation contains *no new dynamics* compared to the expressions used in this paper. The two formulations are entirely equivalent.

In order to derive the mass  $m_\varphi$ , we let  $\varphi = \varphi_0(t) + \varphi_1(r)$  and  $T^{\text{M}} = T^{\text{cos}} + T^{\text{s}}$  so that  $\varphi_0(t)$  satisfies Eq. (A.4) with  $T^{\text{cos}}$ . We then expand to linear order in the perturbation  $\varphi_1$ , writing Eq. (A.4) in terms of the physical metric  $g_{\mu\nu}$ . We find

$$\square\varphi_1 = f'(\phi_0) \left( \left. \frac{d^2V}{d\varphi^2} \right|_{\varphi_0} - \frac{2}{3} \kappa \frac{T^{\text{cos}}}{[f'(\phi_0)]^2} \right) \varphi_1 + \sqrt{\frac{\kappa}{6}} \frac{T^{\text{s}}}{f'(\phi_0)}, \quad (\text{A.5})$$

where  $\phi_0$  denotes the background field value for the  $\phi$  field. Using Eq. (A.3) to evaluate  $d^2V/d\varphi^2$ , we have

$$m_\varphi^2 = \frac{f'(\phi_0)}{3} \left[ \frac{1}{f''(\phi_0)} + \frac{\phi_0}{f'(\phi_0)} - \frac{4f(\phi_0)}{[f'(\phi_0)]^2} - 2\kappa \frac{T^{\text{cos}}}{[f'(\phi_0)]^2} \right]. \quad (\text{A.6})$$

Finally, we may rewrite  $m_\varphi^2$  as Eq. (3.38) since  $\phi_0 = R_0$ . We conclude that if  $m_\varphi^2 r^2 \ll 1$  then  $\gamma = 1/2$  as discussed in Ref. [72].

## Appendix B

# The Cancellation of the CMB Temperature Dipole<sup>1</sup>

### B.1 Dipole cancellation in a $\Lambda$ CDM Universe

In this appendix, we show how the intrinsic CMB dipole induced by a superhorizon adiabatic perturbation is exactly cancelled by the dipole arising from the Doppler effect in a  $\Lambda$ CDM Universe with negligible radiation. Specifically, we will assume that decoupling occurred long after matter-radiation equality so that the evolution of the gravitational potential is given by Eq. (5.14).

First, we derive an alternate expression for the dipole produced by the ISW effect. For a superhorizon mode with  $\Psi(\tau, \vec{x}) \simeq \Psi_{\vec{k}}(\tau)[\vec{k} \cdot \vec{x}]$ , the dipolar component of the ISW anisotropy is given by

$$\begin{aligned} \left[ \frac{\Delta T}{T}(\hat{n}) \right]_{\text{ISW}} &= 2 \int_{\tau_{\text{dec}}}^{\tau_0} \frac{d\Psi_{\vec{k}}}{d\tau} \vec{k} \cdot \hat{n} [\tau_0 - \tau] d\tau, \\ &= (\vec{k} \cdot \hat{n} x_{\text{dec}}) \Psi_{\vec{k}}(\tau_{\text{dec}}) \left[ -2 + \frac{2}{x_{\text{dec}}} \int_{\tau_{\text{dec}}}^{\tau_0} \frac{\Psi_{\vec{k}}(\tau)}{\Psi_{\vec{k}}(\tau_{\text{dec}})} d\tau \right], \end{aligned} \quad (\text{B.1})$$

where we integrated by parts to obtain the second line. In subsequent expressions, we will omit the factor of  $(\vec{k} \cdot \hat{n} x_{\text{dec}}) \Psi_{\vec{k}}(\tau_{\text{dec}}) \simeq \Psi(\tau_{\text{dec}}, \hat{n} x_{\text{dec}})$ .

Transforming the integral over  $\tau$  into an integral over  $a$  and using Eq. (5.14) for  $\Psi_{\vec{k}}(a)$  gives

$$\left[ \frac{\Delta T}{T} \right]_{\text{ISW}} = -2 + \left[ \frac{\frac{9}{10} \Psi_{\text{p}}}{\Psi_{\vec{k}}(\tau_{\text{dec}})} \right] \frac{5\Omega_{\text{M}}}{H_0 x_{\text{dec}}} \int_{a_{\text{dec}}}^1 \frac{G(a)}{a^3} da, \quad (\text{B.2})$$

where  $G(a)$  is defined by

$$G(a) \equiv H_0^3 \int_0^a \frac{da'}{[a' H(a')]^3} = \int_0^a \frac{da'}{[a' \sqrt{\Omega_{\text{M}}(a')^{-3} + \Omega_{\Lambda}}]^3}. \quad (\text{B.3})$$

Integrating Eq. (B.2) by parts yields

$$\left[ \frac{\Delta T}{T} \right]_{\text{ISW}} = -2 + \left[ \frac{\frac{9}{10} \Psi_{\text{p}}}{\Psi_{\vec{k}}(\tau_{\text{dec}})} \right] \frac{5\Omega_{\text{M}}}{H_0 x_{\text{dec}}} \left[ \frac{G(a_{\text{dec}})}{2a_{\text{dec}}^2} - \frac{G_0}{2} + \frac{1}{2} \int_{a_{\text{dec}}}^1 \frac{H_0^3 da}{a^5 H^3(a)} \right], \quad (\text{B.4})$$

<sup>1</sup>This appendix was published in *Superhorizon perturbations and the cosmic microwave background*, Adrienne L. Erickcek, Sean M. Carroll, and Marc Kamionkowski; Phys. Rev. D **78**, 083012 (2008). Reproduced here with permission, copyright (2008) by the American Physical Society.

where  $G_0 \equiv G(a=1)$ . Since decoupling occurs long before matter- $\Lambda$  equality,  $\Psi_{\vec{k}}(\tau_{\text{dec}}) \simeq [9/10]\Psi_{\text{p}}$ . However, we wish to show that the dipole cancellation applies more generally, and so we do not assume that the Universe is matter-dominated at decoupling.

The next step is crucial for the upcoming cancellation, and it relies on a special feature of  $G(a)$ . Focusing on the last term of Eq. (B.4), we see that

$$\begin{aligned} \int_{a_{\text{dec}}}^1 \frac{H_0^3 da}{a^5 H^3(a)} &= \int_{a_{\text{dec}}}^1 \frac{da}{a^5 [\Omega_{\text{M}} a^{-3} + \Omega_{\Lambda}]^{3/2}}, \\ &= \frac{2}{3\Omega_{\text{M}}} \int_{a_{\text{dec}}}^1 \frac{1}{a} \frac{d}{da} \left[ \frac{1}{\sqrt{\Omega_{\text{M}} a^{-3} + \Omega_{\Lambda}}} \right] da, \\ &= \frac{2}{3\Omega_{\text{M}}} \left[ 1 - \frac{1}{a_{\text{dec}}} \frac{H_0}{H(a_{\text{dec}})} + H_0 x_{\text{dec}} \right]. \end{aligned}$$

Using this result to eliminate the integral in Eq. (B.4) gives

$$\begin{aligned} \left[ \frac{\Delta T}{T} \right]_{\text{ISW}} &= -2 + \frac{5}{3} \left[ \frac{\frac{9}{10} \Psi_{\text{p}}}{\Psi_{\vec{k}}(\tau_{\text{dec}})} \right] \\ &+ \frac{1}{H_0 x_{\text{dec}}} \left[ \frac{\frac{9}{10} \Psi_{\text{p}}}{\Psi_{\vec{k}}(\tau_{\text{dec}})} \right] \left[ \frac{H_0 D(a_{\text{dec}})}{H(a_{\text{dec}}) a_{\text{dec}}^2} - D_0 + \frac{5}{3} - \frac{5H_0}{3a_{\text{dec}} H(a_{\text{dec}})} \right], \end{aligned} \quad (\text{B.5})$$

where  $D(a) \equiv (5/2)\Omega_{\text{M}}[H(a)/H_0]G(a)$  so that  $\Psi(a) = [9/10]\Psi_{\text{p}}D(a)/a$  in Eq. (5.14), and  $D_0 \equiv D(a=1)$ . The first two terms in this expression are exactly cancelled by the anisotropy produced by the SW effect, which was given in Eq. (5.3). We will now show that the last term in the ISW anisotropy will be cancelled by the Doppler effect.

The Doppler dipole is

$$\left[ \frac{\Delta T}{T}(\hat{n}) \right]_{\text{Doppler}} = \left( \vec{k} \cdot \hat{n} x_{\text{dec}} \right) \Psi_{\vec{k}}(\tau_{\text{dec}}) [\mathcal{V}(\tau_0) - \mathcal{V}(\tau_{\text{dec}})]. \quad (\text{B.6})$$

Long after matter-radiation equality ( $y \rightarrow \infty$ ), with  $\Psi(a) = [9/10]\Psi_{\text{p}}D(a)/a$ , Eq. (5.20) for  $\mathcal{V}(\tau)$  becomes

$$\mathcal{V}(\tau) = - \left[ \frac{\frac{9}{10} \Psi_{\text{p}}}{\Psi_{\vec{k}}(\tau_{\text{dec}})} \right] \frac{2a(\tau)^2}{3\Omega_{\text{M}} H_0 x_{\text{dec}}} \left[ \frac{H(\tau)}{H_0} \right] \frac{dD}{da} \Big|_{a(\tau)}. \quad (\text{B.7})$$

From

$$\frac{dD}{da} = -\frac{3\Omega_{\text{M}}}{2a^3} \left[ \frac{H_0^2}{H^2(a)} \right] \left[ \frac{D(a)}{a} - \frac{5}{3} \right] \quad (\text{B.8})$$

it follows that

$$\mathcal{V}(\tau_0) - \mathcal{V}(\tau_{\text{dec}}) = \frac{1}{H_0 x_{\text{dec}}} \left[ \frac{\frac{9}{10} \Psi_{\text{p}}}{\Psi_{\vec{k}}(\tau_{\text{dec}})} \right] \left[ D_0 - \frac{5}{3} - \frac{H_0 D(a_{\text{dec}})}{H(a_{\text{dec}}) a_{\text{dec}}^2} + \frac{5H_0}{3a_{\text{dec}} H(a_{\text{dec}})} \right]. \quad (\text{B.9})$$

As promised, this contribution to the anisotropy cancels the last term in Eq. (B.5). Thus we see that the CMB temperature anisotropy terms proportional to  $\hat{k} \cdot \hat{n} x_{\text{dec}}$  arising from the SW, ISW

and Doppler effects sum to zero in a  $\Lambda$ CDM universe. Consequently, the CMB temperature dipole is comparable in magnitude to the temperature octupole since both are primarily sourced by the component of the temperature anisotropy that is proportional to  $(\hat{k} \cdot \hat{n}x_{\text{dec}})^3$ .

The analytical calculation presented in this Appendix does not apply if there is a significant amount of radiation at decoupling. The presence of radiation would change the evolution of  $\Psi$ , preventing us from relating the integral in Eq. (B.2) to  $x_{\text{dec}}$ . However, a numerical calculation confirms that the same dipole cancellation occurs when radiation is present, as discussed in Section 5.3 and illustrated in Fig. 5.2.

## B.2 Dipole cancellation in a universe with an exotic fluid

To demonstrate that the CMB dipole cancellation is not a special feature of universes containing only matter and radiation, we calculate  $\delta_1$  for adiabatic superhorizon perturbations in a flat universe that contains an  $X$  fluid with constant equation of state  $w = p_X/\rho_X$  and a cosmological constant. In this two-component universe, the Hubble parameter is

$$H^2(a) = H_0^2 \left[ \frac{\Omega_X}{a^{3(1+w)}} + \Omega_\Lambda \right], \quad (\text{B.10})$$

where  $\Omega_X$  is the present day ratio of  $\rho_X$  and the critical density and  $\Omega_\Lambda = 1 - \Omega_X$ . Of course, the existence of the CMB in this universe implies that there is some radiation that we have not included in Eq. (B.10). To justify this omission, we will only consider values of  $w$  that are greater than  $1/3$  so that the  $X$ -fluid energy density is always greater than the radiation density, even in the very early universe.

We now consider an adiabatic superhorizon perturbation in this universe. The perturbed fluid, which contained matter and radiation in Section 5.2, is now dominated by a single component: the  $X$  fluid. The overdensity of the  $X$  fluid in its rest frame is therefore related to its peculiar motion through Eqs. (5.8) and (5.9), with  $w$  being the equation of state parameter of the  $X$  fluid. The potential perturbation in a flat universe is directly related to the density perturbation [171]:

$$\Psi_{\vec{k}} = \frac{-4\pi G}{k^2} a^2 \rho \Delta_{\vec{k}}, \quad (\text{B.11})$$

where  $\rho$  is the density of the perturbed fluid. Setting  $\rho$  equal to the sum of the matter and radiation densities yields Eq. (5.10), but in this case the only component of the perturbed fluid is the  $X$  fluid, so  $\rho = \rho_X = \Omega_X [3H_0^2/(8\pi G)] a^{-3(1+w)}$ . We can use Eqs. (5.8), (5.9), and (B.11) to derive an

equation for containing only  $\Psi_{\vec{k}}(a)$ , just as we did in Section 5.2. The resulting equation is

$$0 = \Psi_{\vec{k}}''(a) + \frac{1}{a} \left[ 5 + 3w + \frac{d \ln H}{d \ln a} \right] \Psi_{\vec{k}}'(a) + \frac{1}{a^2} \left[ 3(1+w) + \frac{d \ln H}{d \ln a} - \frac{3(1+w)}{2} \left\{ \frac{\Omega_X}{a^{3(1+w)}} \right\} \frac{H_0^2}{H^2} \right] \Psi_{\vec{k}}. \quad (\text{B.12})$$

We may also use Eq. (B.11) to eliminate  $\Delta_{\vec{k}}$  from Eq. (5.8), which yields an equation for  $\vec{v}$ :

$$\vec{v}(\tau, \vec{x}) = -\frac{2a^{3w+2}}{3(1+w)} \frac{1}{H_0 \Omega_X} \frac{H(a)}{H_0} \left[ \vec{\nabla} \Psi + \frac{d}{d \ln a} \vec{\nabla} \Psi \right]. \quad (\text{B.13})$$

In the very early universe, the  $X$  fluid's energy density is much greater than the vacuum energy, and we may neglect  $\Omega_\Lambda$  in Eq. (B.10) for the Hubble parameter. In that case, the terms proportional to  $\Psi_{\vec{k}}$  in Eq. (B.12) sum to zero, and we are left with

$$\Psi_{\vec{k}}''(a) + \frac{1}{a} \left( \frac{7}{2} + \frac{3}{2}w \right) \Psi_{\vec{k}}'(a) = 0. \quad (\text{B.14})$$

A constant value of  $\Psi_{\vec{k}}$  is the only non-decaying solution to this equation. Therefore,  $\Psi$  is constant in the early universe, and we may set  $\Psi'$  equal to zero as an initial condition when numerically solving Eq. (B.12). Moreover,  $\Psi_{\vec{k}}$  is always constant if  $\Omega_X = 1$ .

Now that we have expressions for  $\Psi(a)$  and  $\vec{v}$ , the only remaining component of the dipole anisotropy is the SW effect. Eq. (5.3) only applies to universes that were initially radiation-dominated, so we need to derive the analogous expression for a universe that is initially  $X$ -dominated. Given that the perturbations are adiabatic, the perturbation to the radiation density will be proportional to the density perturbation in the  $X$  fluid:

$$\frac{\Delta T}{T} = \frac{1}{4} \delta_\gamma = \frac{1}{3(1+w)} \delta_X, \quad (\text{B.15})$$

where  $\delta_\gamma$  and  $\delta_X$  are the fractional density perturbations in the radiation and the  $X$  fluid, respectively, in conformal Newtonian gauge. The superhorizon limit ( $k \rightarrow 0$ ) of the temporal Einstein equation, with  $\dot{\Psi} = 0$ , implies that  $\delta_X = -2\Psi_{\text{p}}$  at very early times. We can then use the adiabatic condition to obtain the primordial temperature anisotropy:

$$\frac{\Delta T}{T}(\tau_{\text{p}}) = -\frac{2}{3(1+w)} \Psi_{\text{p}}. \quad (\text{B.16})$$

The Boltzmann equations for superhorizon perturbations still imply that  $\Delta T/T = \Psi(\tau)$  plus a constant, and that constant is determined by Eq. (B.16). The final expression for the SW effect is

$$\frac{\Delta T}{T}(\tau_{\text{dec}}) + \Psi(\tau_{\text{dec}}) = \Psi(\tau_{\text{dec}}) \left[ 2 - \frac{5+3w}{3(1+w)} \frac{\Psi_{\text{p}}}{\Psi(\tau_{\text{dec}})} \right], \quad (\text{B.17})$$

which is equivalent to Eq. (5.3) if the  $X$  fluid is radiation ( $w = 1/3$ ).

We now have all the components necessary to numerically evaluate the observed CMB temperature anisotropy following the same procedure as described in Section 5.3. The SW anisotropy  $\mathcal{S}$  is given by Eq. (B.17). The ISW anisotropy may be obtained by numerically solving Eq. (B.12) and using that solution to evaluate  $\mathcal{I}_0 - \mathcal{I}_1$ . Finally, Eq. (B.13) gives the Doppler anisotropy  $[\mathcal{V}(\tau_0) - \mathcal{V}(\tau_{\text{dec}})]$ . Combining these terms, we find that  $\delta_1 = 0$  for any value of  $w \geq 1/3$  and any value of  $\Omega_X$ .

In the case that  $\Omega_X = 1$ , the dipole cancellation is easy to see analytically. Since  $\Psi$  is constant, there is no ISW effect. The Doppler anisotropy is given by

$$\mathcal{V}(\tau_0) - \mathcal{V}(\tau_{\text{dec}}) = \frac{2}{3(1+w)H_0x_{\text{dec}}} \left[ -1 + \frac{H(a_{\text{dec}})}{H_0} a_{\text{dec}}^{3w+2} \right]. \quad (\text{B.18})$$

Since the  $X$ -fluid is the only energy in the universe, the Hubble parameter is simply  $H(a) = H_0 a^{-3(1+w)/2}$ , and the comoving distance to the surface of last scattering is

$$H_0 x_{\text{dec}} = \frac{2}{1+3w} \left[ 1 - a_{\text{dec}}^{(1+3w)/2} \right]. \quad (\text{B.19})$$

Inserting these expressions into Eq. (B.18) yields

$$\mathcal{V}(\tau_0) - \mathcal{V}(\tau_{\text{dec}}) = -\frac{1+3w}{3(1+w)}, \quad (\text{B.20})$$

which exactly cancels the SW anisotropy given by Eq. (B.17) with  $\Psi_p = \Psi(\tau_{\text{dec}})$ .

## Appendix C

# Attempts to Generate a Scale-Dependent Asymmetry without Isocurvature<sup>1</sup>

In this appendix, we will attempt to make the power asymmetry generated by a superhorizon curvaton fluctuation scale-dependent without introducing isocurvature perturbations. As in Section 6.2, we define  $\xi$  to be the fraction of the total power that comes from the curvaton ( $\sigma$ ):

$$\xi(k) = \frac{\mathcal{P}_{\zeta,\sigma}(k)}{\mathcal{P}_{\zeta,\sigma}(k) + \mathcal{P}_{\zeta,\phi}(k)}, \quad (\text{C.1})$$

where  $\phi$  refers to the inflaton and  $\zeta$  is the curvature of uniform-density hypersurfaces defined by Eq. (6.16). The adiabatic power asymmetry amplitude is  $\Delta P(k)/P(k) = 2\xi(\Delta\bar{\sigma}/\bar{\sigma})$ . Since  $\bar{\sigma}$  has no scale dependence, the only way to make  $\Delta P(k)/P(k)$  dependent on  $k$  is to make  $\xi$  dependent on  $k$ . We need  $\Delta P(k)/P(k)$  to go from  $2A = 0.144$  on large CMB scales ( $\ell \lesssim 64$ ) [46] to less than 0.024 on quasar scales ( $k \simeq 1.5h \text{ Mpc}^{-1}$ ) [90].

We recall that the power from curvaton fluctuations is

$$\mathcal{P}_{\zeta,\sigma}(k) = \left(\frac{R}{3}\right)^2 \frac{H_{\text{inf}}^2}{\pi^2 \bar{\sigma}^2}, \quad (\text{C.2})$$

where  $H_{\text{inf}}$  is the Hubble parameter during inflation and  $R$  is three-fourths times the fraction of energy in the curvaton just prior to its decay. The power from inflaton fluctuations is

$$\mathcal{P}_{\zeta,\phi}(k) = \frac{GH_{\text{inf}}^2}{\pi\epsilon_H}, \quad (\text{C.3})$$

where  $\epsilon_H$  is related to the inflaton potential  $V(\phi)$  through  $\epsilon_H \simeq \epsilon_V \equiv (m_{\text{Pl}}^2/16\pi)[V'(\phi)/V(\phi)]^2$ . Inserting these expressions into Eq. (C.1) gives

$$\xi(k) = \left[1 + \frac{9\pi}{R^2\epsilon_H(k)} \left(\frac{\bar{\sigma}}{m_{\text{Pl}}}\right)^2\right]^{-1}, \quad (\text{C.4})$$

where  $m_{\text{Pl}} = G^{-1/2}$ . It will be useful to define

$$\tilde{\epsilon}(k) \equiv \frac{1}{9\pi} \left(\frac{\bar{\sigma}}{m_{\text{Pl}}}\right)^{-2} R^2\epsilon_H(k) \quad (\text{C.5})$$

---

<sup>1</sup>This appendix contains previously unpublished work by the author.

because then  $\xi$  has a simple form:

$$\xi(k) = \frac{\tilde{\epsilon}}{\tilde{\epsilon} + 1}. \quad (\text{C.6})$$

Note that even though  $R \ll 1$  and  $\epsilon_H \ll 1$ ,  $\tilde{\epsilon}$  may have any value:  $\tilde{\epsilon} \gg 1$  is possible if  $\bar{\sigma} \ll m_{\text{Pl}}$ , while  $\tilde{\epsilon} \ll 1$  can be obtained by decreasing  $R$  or  $\epsilon_H$ .

There are two ways that we can give  $\xi$  the scale-dependence necessary to generate the observed asymmetry in the CMB and still satisfy the quasar constraint on small-scale asymmetry: we can make  $\xi$  discontinuous by inserting a kink into  $V'(\phi)$ , or we can choose the spectral indices of  $\mathcal{P}_{\zeta,\sigma}(k)$  and  $\mathcal{P}_{\zeta,\phi}(k)$  in such a way that  $\xi(k)$  decreases smoothly as  $k$  increases. We will consider both approaches in this appendix.

### C.1 A discontinuity in $\xi(k)$ ?

First, we suppose that  $\xi$  is a step function:  $\xi = \xi_{\text{max}}$  for  $k < k_*$  and  $\xi = \xi_{\text{min}}$  for  $k > k_*$  where both  $\xi_{\text{max}}$  and  $\xi_{\text{min}}$  are constants. To achieve the necessary suppression of the asymmetry on small scales, we require that  $\xi_{\text{min}}/\xi_{\text{max}} \lesssim 1/6$  and  $k_*$  be located somewhere between large CMB scales ( $k \simeq 0.0033h \text{ Mpc}^{-1}$ ) and quasar scales ( $k \simeq 1.5h \text{ Mpc}^{-1}$ ) [90]. A step function in  $\xi$  requires a step function in  $\tilde{\epsilon}$ , which requires a step function in  $\epsilon_H$ . From  $\xi_{\text{min}}/\xi_{\text{max}} = 1/6$  and Eq. (C.6), we see that

$$\frac{\tilde{\epsilon}_{\text{min}}}{\tilde{\epsilon}_{\text{max}}} = \frac{\epsilon_{H,\text{min}}}{\epsilon_{H,\text{max}}} = \frac{1}{6 + 5\tilde{\epsilon}_{\text{max}}}. \quad (\text{C.7})$$

Thus we see that the size of the necessary discontinuity in  $\epsilon_H$  is determined by  $\xi_{\text{max}}$ . If the curvaton dominates on large scales ( $\xi \simeq 1$  and  $\tilde{\epsilon} \gg 1$ ) then the drop in  $\epsilon_H$  necessary to make the inflaton dominant on small scales is large. This drop is minimized if  $\xi_{\text{max}}$  is minimized, but  $\xi_{\text{max}} \gtrsim 0.07$  is required to generate the observed asymmetry.

Now we have to worry about the shape of the power spectrum. Inserting a downward step in  $\epsilon_H$  leaves the curvaton perturbation spectrum unaltered but it gives the inflaton perturbation spectrum an upward step. Therefore, we would expect that the total power on large scales would be smaller than the total power on small scales. In contrast, the primordial  $\mathcal{P}_{\zeta}(k)$  that fits CMB and large-scale-structure observations is nearly flat on all scales from  $k = 0.0001 \text{ Mpc}^{-1}$  to  $k = 0.2 \text{ Mpc}^{-1}$  [191]. Furthermore, the value for  $\sigma_8$ , the fluctuation amplitude at  $8h \text{ Mpc}^{-1}$ , derived from the CMB is consistent with the measurements from weak lensing observations [217], so there can no major change in the primordial power spectrum between the scales probed by the CMB and those probed by weak lensing. We conclude that the total primordial power spectrum must be nearly scale-invariant.

The total primordial power spectrum may be expressed in terms of  $\xi$  and the curvaton power:

$$\mathcal{P}_\zeta(k) = \frac{1}{\xi(k)} \left(\frac{R}{3}\right)^2 \frac{H_{\text{inf}}^2}{\pi^2 \sigma^2}. \quad (\text{C.8})$$

The ratio of power on large scales ( $k_{\text{CMB}} < k_*$ ) to the power on small scales ( $k_{\text{Q}} > k_*$ ) is

$$\frac{\mathcal{P}_\zeta(k_{\text{CMB}})}{\mathcal{P}_\zeta(k_{\text{Q}})} = \left[ \frac{H_{\text{inf}}^2(k_{\text{CMB}})}{H_{\text{inf}}^2(k_{\text{Q}})} \right] \left( \frac{\xi_{\text{min}}}{\xi_{\text{max}}} \right). \quad (\text{C.9})$$

To compensate for the injection of inflaton power on scales smaller than  $k_*$ , we must introduce a discontinuity in  $V(\phi)$  at the same  $\phi$  value as the discontinuity in  $V'(\phi)$ . Since  $H_{\text{inf}}^2 \propto V(\phi)$ , we see that

$$\frac{V(\phi_{\text{CMB}})}{V(\phi_{\text{Q}})} = \frac{\xi_{\text{max}}}{\xi_{\text{min}}} \gtrsim 6, \quad (\text{C.10})$$

where  $k_{\text{CMB,Q}} = aH_{\text{inf}}(\phi_{\text{CMB,Q}})$ , is required to keep the primordial power spectrum scale-invariant.

Thus we see that it is possible to hide the kick that the power spectrum gets when the inflaton takes over on small scales by introducing a drop in the inflationary energy scale that leads to a total reduction in power in both the curvaton and inflaton fluctuations. As the inflaton rolls across the discontinuity in the power spectrum, the value of  $V(\phi)$  must drop by at least a factor of 6, and the potential on the lower side must be significantly flatter than the potential on the upper side; from  $\epsilon \propto (V'(\phi)/V)^2$  and Eq. (C.7), we conclude that

$$\frac{V'(\phi_{\text{Q}})}{V'(\phi_{\text{CMB}})} = \frac{1}{6} \sqrt{\frac{1}{6 + 5\epsilon_{\text{max}}}} \quad (\text{C.11})$$

for  $\xi_{\text{min}}/\xi_{\text{max}} = 1/6$ . Note that there is no lower bound on  $\epsilon_H$  on either side of the potential break;  $V(\phi)$  can be as flat as we want it to be on large scales provided that it is even flatter on small scales. Therefore, we don't have to worry about constraints from the tensor-scalar ratio or the scalar spectral index.

There is another concern, however. Even though the discontinuities in  $V(\phi)$  and  $V'(\phi)$  conspire to preserve the flatness of the total power spectrum, the momentary interruption of slow-roll inflation that occurs as the inflaton field crosses the break could induce oscillations in the power spectrum localized around  $k_*$ . Ref. [183] analyzes the effects of a step in the mass  $m$  of a quadratic  $V = m^2 \phi^2/2$  potential, and Ref. [184] generalizes this analysis to steps in other inflaton potentials. They restrict their analyses to breaks in the potential with  $\Delta V/\bar{V} \leq 0.2$  because they do not want the inflaton's kinetic energy to exceed its potential energy after the inflaton crosses the break. Even with this constraint, the smoothness of the observed power spectrum is very restrictive; CMB and large-scale-structure observations constrain  $\Delta V/\bar{V} \lesssim 10^{-3}$  at 99% confidence for  $k_* \lesssim 0.1 \text{ Mpc}^{-1}$  [184].

The models considered in Refs. [183, 184] include a small change in the perturbation amplitude

across the step and do not include perturbations from a curvaton field, but their findings are still very discouraging for our proposal. They find that the amplitude of the oscillations far exceeds the change in  $\mathcal{P}_\zeta(k)$  across the break, so it is reasonable to expect that any break in the potential will induce large oscillations, even if  $\mathcal{P}_\zeta(k)$  is unaffected. Furthermore, the oscillations reach their maximum on scales that leave the horizon after  $\phi$  crosses the break. On these scales, the inflaton must dominate the power spectrum to suppress the asymmetry, so there is no hope of masking the oscillations in the inflaton spectrum with the curvaton spectrum. We could hope to hide the oscillations induced by our model by setting  $k_* \gtrsim 0.1 \text{ Mpc}^{-1}$ , but the potential change required by our model is so large that inflation may not resume after the inflaton crosses the break. We therefore conclude that the discontinuity in  $V(\phi)$  that is required by our model is highly unlikely to be consistent with observations.

## C.2 A smooth transition?

Next we consider the possibility that  $\xi$ , and therefore  $\epsilon_H$ , smoothly decrease as  $k$  increases. Given the difference  $k$  between large CMB scales and quasar scales, we would need

$$\frac{d \ln \xi}{d \ln k} \simeq \frac{\Delta \ln \xi}{\Delta \ln k} \simeq -\frac{1.8}{6.1} = -0.29 \quad (\text{C.12})$$

if  $(d \ln \xi / d \ln k)$  is to be roughly constant over the scales of interest ( $0.0033h \text{ Mpc}^{-1} \lesssim k \lesssim 1.5h \text{ Mpc}^{-1}$ ).

The spectral index for  $\xi$  is related to the spectral indices of the power spectrum for inflaton and curvaton fluctuations [88]:

$$\frac{d \ln \xi}{d \ln k} = \frac{d \ln \mathcal{P}_{\zeta,\sigma}}{d \ln k} - \xi \frac{d \ln \mathcal{P}_{\zeta,\sigma}}{d \ln k} - (1 - \xi) \frac{d \ln \mathcal{P}_{\zeta,\phi}}{d \ln k} \quad (\text{C.13})$$

$$= -2\epsilon_H - \xi(-2\epsilon_H) - (1 - \xi)(-4\epsilon_H + 2\eta_H) \quad (\text{C.14})$$

$$= -(1 - \xi)(2\eta_H - 2\epsilon_H) \quad (\text{C.15})$$

where  $\eta_H$  is the slow-roll parameter

$$\eta_H \equiv -\frac{\ddot{\phi}}{\dot{\phi}H}, \quad (\text{C.16})$$

and we note that  $\eta_H \simeq \eta_V - \epsilon_H$  [218]. Meanwhile, the spectral index for the total power spectrum is

$$n_s - 1 \equiv \frac{d \ln[\mathcal{P}_{\zeta,\sigma} + \mathcal{P}_{\zeta,\phi}]}{d \ln k}, \quad (\text{C.17})$$

$$= -2\epsilon_H - (1 - \xi)(2\epsilon_H - 2\eta_H), \quad (\text{C.18})$$

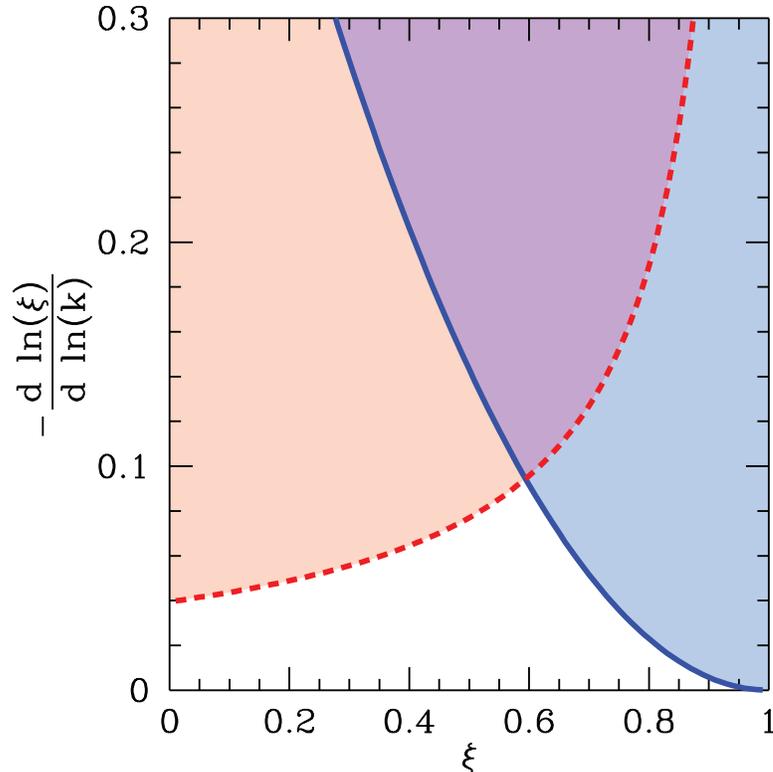


Figure C.1: The upper bounds on  $-\frac{d \ln \xi}{d \ln k}$ . The dashed curve is the upper bound from Eq. (C.20) with  $n_s - 1 \leq 0.002$  and  $r \leq 0.3$ . The solid curve follows from Eqs. (C.25) and (C.26) with  $\alpha_s \leq 0.047$  and  $r = 0.3$ .

and the tensor-scalar ratio is

$$r = 16\epsilon_H(1 - \xi). \quad (\text{C.19})$$

Therefore, the spectral index for  $\xi$  depends only on  $\xi$  and observable parameters:

$$-\frac{d \ln \xi}{d \ln k} = n_s - 1 + \frac{r}{8(1 - \xi)}. \quad (\text{C.20})$$

The  $3\sigma$  upper bound on  $n_s - 1$  is 0.002 on most scales, although the upper bound is somewhat higher on very large scales ( $k \lesssim 0.001 \text{ Mpc}^{-1}$ ) [191, 18]. The upper bound on  $r$  is  $\sim 0.3$  on large scales ( $k \simeq 0.002 \text{ Mpc}^{-1}$ ) [18]. These two constraints place an upper bound on the right-hand-side of Eq. (C.20), which is the dashed curve in Fig. C.1. Thus we see that  $\xi$  is going to have to be close to one if  $-\frac{d \ln \xi}{d \ln k}$  is going to be 0.29. However, we can make  $|\frac{d \ln \xi}{d \ln k}|$  as large as we want if we are willing to set  $\xi \simeq 1$  on large scales. Setting  $\xi \simeq 1$  on large scales comes at a cost though. To get a factor of six reduction in  $\Delta P(k)/P(k)$ , we'll need the curvaton to go from contributing nearly all of the primordial power to contributing less than 17% of the primordial power. To compensate, we have given a strong blue tilt to the inflaton perturbation spectrum so that  $n_s$  remains close to one on all scales.

Unfortunately, the transition from having the curvaton contribute all of the power to contributing only 17% leads to a significant running of the spectral index. To evaluate

$$\alpha_s \equiv \frac{dn_s}{d \ln k}, \quad (\text{C.21})$$

we will need the running of the slow-roll parameters

$$\frac{d \ln \epsilon_H}{d \ln k} = 2(\epsilon_H - \eta_H), \quad (\text{C.22})$$

$$\frac{d \ln \eta_H}{d \ln k} = \epsilon_H - \frac{\xi_H^2}{\eta_H}, \quad (\text{C.23})$$

where

$$\xi_H \equiv \frac{4\pi}{m_{\text{Pl}}^2} \sqrt{\frac{H'(\phi)H'''(\phi)}{H^2}} \quad (\text{C.24})$$

is a higher-order slow-roll parameter. From Eq. (C.18), it follows that

$$\alpha_s = \frac{d \ln \xi}{d \ln k} \left[ \frac{\xi}{1 - \xi} (1 - n_s) - \frac{r}{4} \frac{1}{(1 - \xi)^2} \right] + \frac{r}{8} \left( \eta_H - \frac{\xi_H^2}{\epsilon_H} \right). \quad (\text{C.25})$$

The  $3\sigma$  upper limit on  $\alpha_s$  from WMAP5 is 0.05, and this puts another upper bound on  $-d \ln \xi / d \ln k$ , as shown by the solid curve in Fig. C.1. In this figure, we set

$$\eta_H - \frac{\xi_H^2}{\epsilon_H} = 0.1. \quad (\text{C.26})$$

Decreasing the value of this combination of slow-roll parameters does not significantly change the upper bound on  $-d \ln \xi / d \ln k$ , but increasing it beyond 0.1 leads to a more restrictive constraint. The most permissive choice is to make this combination negative, but it needs to be less than -0.5 to noticeably shift the bound. Unfortunately, the upper bound on  $-d \ln \xi / d \ln k$  from  $\alpha_s$  decreases as  $\xi$  increases. When combined with the upper bound from Eq. (C.20) as in Fig. C.1, we see that it is not possible to have  $-d \ln \xi / d \ln k = 0.29$  for any range of  $\xi$  values.

It appears that it will not be possible to suppress the asymmetry on small scales by having  $\xi$  smoothly decrease as  $k$  increases. It is worth noting, however, that the dashed curve in Fig. C.1 follows from  $r \leq 0.3$ , and that this constraint only applies on large scales. On scales that were subhorizon at the surface of last scattering, the tensor perturbations are suppressed relative to their primordial values. In that case, the only constraint on  $r$  follows from the slow-roll condition  $\epsilon_H \ll 1$ . Instead of the dashed curve in Fig. C.1, we have

$$-\frac{d \ln \xi}{d \ln k} \ll 2 + (n_s - 1). \quad (\text{C.27})$$

So while a scale-invariant  $d \ln \xi / d \ln k = -0.29$  is not allowed, we could get sufficiently negative

values of  $d \ln \xi / d \ln k$  for  $\ell \gtrsim 100$  and  $\xi \lesssim 0.3$ .

To investigate this possibility, we derive a differential equation for  $\xi(k)$ . We start with

$$-\frac{d \ln \xi}{d \ln k} = n_s(k) - 1 + 2\epsilon_H(k). \quad (\text{C.28})$$

This equation is simply Eq. (C.20), but with  $\epsilon_H$  instead of  $r$ . From Eq. (C.6) it follows that

$$\epsilon_H(k) = \epsilon_0 \frac{\xi(k)}{1 - \xi(k)}, \quad (\text{C.29})$$

where  $\epsilon_0 \equiv \epsilon_H(k)/\tilde{\epsilon}(k)$  and is therefore independent of  $k$  and can take any value. Since there are measurements of  $n_s$  and its running, we will use these values for  $n_s(k)$ . The resulting differential equation for  $\xi(k)$  is

$$-\frac{d \ln \xi}{d \ln k} = n_s(k_0) - 1 + \alpha_s \ln \left( \frac{k}{k_0} \right) + 2\epsilon_0 \frac{\xi(k)}{1 - \xi(k)}. \quad (\text{C.30})$$

To maximize the variation in  $\xi$ , we use the upper bounds on  $\alpha_s$  and  $n_s(k_0)$  mentioned above:  $n_s(k_0) - 1 = 0.002$  and  $\alpha_s = 0.047$  for  $k_0 = 0.002 \text{ Mpc}^{-1}$ .

The only free parameters that remain are  $\epsilon_0$  and the ‘‘initial’’ value of  $\xi_s \equiv \xi(k_s)$ , where  $k_s$  is the smallest  $k$  value of interest and the starting point of the numerical integration of Eq. (C.30). Together,  $\epsilon_0$  and  $\xi_s$  determine  $\epsilon_H(k_s)$ . From Eq. (C.29), we see that  $\epsilon_H$  decreases as  $\xi(k)$  decreases; since  $\xi(k)$  is monotonically decreasing for all ( $k > k_0$ ),  $\epsilon_H(k_s)$  is the maximum value that  $\epsilon_H$  will attain in the scale range of interest. Since we wish to maximize  $-d \ln \xi / d \ln k$ , we want to choose the largest possible value for  $\epsilon_0$ . There are two upper bounds to consider. First,  $\epsilon_H \ll 1$  is required by the slow-roll approximation. If we insist that  $\epsilon_H \leq 1/2$ , then the maximum possible value for  $\epsilon_0$  is

$$\epsilon_0 = 0.5 \frac{1 - \xi_s}{\xi_s}. \quad (\text{C.31})$$

There is, however, an additional constraint. Since  $\epsilon_H(k)$  is a monotonically decreasing function of  $k$  for  $k \geq k_0$ , the value of  $\epsilon_H(k_s \geq k_0)$  sets a lower bound on the value of  $\epsilon_H(k_0)$ , provided that  $\epsilon_H$  is continuous. There is an upper bound on  $\epsilon_H(k_0)$  that follows from the measured upper bound on the tensor-to-scalar ratio  $r$ . This upper bound implies that the maximum value of  $\epsilon_0$  should be

$$\epsilon_0 = \frac{0.019}{\xi_s}, \quad (\text{C.32})$$

which is lower than Eq. (C.31) for  $\xi_s \lesssim 0.9$ .

We integrate Eq. (C.30) to obtain  $\xi(k)$  for different values of  $\xi_s$  with  $\epsilon_0$  given by both Eqs. (C.31) and (C.32). We set  $k_s = 0.002 \text{ Mpc}^{-1}$ , and for each value of  $\epsilon_0$ , we find the value of  $\xi_s$  that gives the smallest value of  $\xi_{\min}/\xi_{\max} = \xi(1.5h \text{ Mpc}^{-1})/\xi(0.0033h \text{ Mpc}^{-1})$  with  $h = 0.7$ . We find that it is possible to obtain  $\xi_{\min}/\xi_{\max} \lesssim 1/6$  only if we relax the condition that  $r \leq 0.3$ . In that case,

$\xi_{\min}/\xi_{\max} \lesssim 1/6$  if  $\xi_s \lesssim 0.8$  and  $\epsilon_0$  is given by Eq. (C.31). It is encouraging that this range includes values for  $\xi_s$  that are greater than 0.07 because this means that  $\xi$  on large scales can be large enough to generate the observed asymmetry. The downside is that this model has  $\epsilon_H \simeq 0.5$  on large scales; given that  $\xi_s \leq 0.8$ , this means that  $r \gtrsim 1.6$  on the largest scales for which this model applies. Since  $\epsilon_H$  is monotonically decreasing for  $k \geq k_0$ , and we set  $k_s = k_0$ , it is not possible to make  $\epsilon_H$  small enough to satisfy  $r \lesssim 0.3$  on the scales for which that bound applies ( $k \simeq k_0$ ) and then have it smoothly increase to the value necessary to give sufficient variation in  $\xi$ .

In summary, it is possible for  $\xi$  to be greater than 0.07 on large CMB scales and then smoothly decrease by a factor of six between large CMB scales and quasar scales in a way that keeps the total power spectrum flat enough to be consistent with observations. However, the required values of  $\xi$  and  $\epsilon_H$  on the largest scales are inconsistent with the upper bound on the tensor-scalar ratio on these scales. In order to satisfy this bound, we would have to discontinuously change the values of  $\xi$  and  $\epsilon_H$  to suppress the tensor-scalar ratio on large scales, and we saw in the previous section that such discontinuities are problematic.

## Bibliography

- [1] A. H. Guth, *Phys. Rev. D* **23**, 347 (1981).
- [2] A. J. Albrecht and P. J. Steinhardt, *Phys. Rev. Lett.* **48**, 1220 (1982).
- [3] A. D. Linde, *Phys. Lett. B* **108**, 389 (1982).
- [4] S. W. Hawking, *Phys. Lett. B* **115**, 295 (1982).
- [5] A. A. Starobinsky, *Phys. Lett. B* **117**, 175 (1982).
- [6] A. H. Guth and S. Y. Pi, *Phys. Rev. Lett.* **49**, 1110 (1982).
- [7] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, *Phys. Rev. D* **28**, 679 (1983).
- [8] S. Perlmutter *et al.* (Supernova Cosmology Project), *Astrophys. J.* **517**, 565 (1999), [astro-ph/9812133].
- [9] A. G. Riess *et al.* (Supernova Search Team), *Astron. J.* **116**, 1009 (1998), [astro-ph/9805201].
- [10] G. Bertone, D. Hooper, and J. Silk, *Physics Reports* **405**, 279 (2005), [arXiv:hep-ph/0404175].
- [11] R. R. Caldwell and M. Kamionkowski, arXiv e-print (2009), [0903.0866].
- [12] A. Einstein, *Ann. der Physik* **354**, 769 (1916).
- [13] W. H. Press and P. Schechter, *Astrophys. J.* **187**, 425 (1974).
- [14] C. Lacey and S. Cole, *Mon. Not. R. Astron. Soc.* **262**, 627 (1993).
- [15] R. A. Alpher, H. Bethe, and G. Gamow, *Phys. Rev.* **73**, 803 (1948).
- [16] A. Coc, E. Vangioni-Flam, P. Descouvemont, A. Adahchour, and C. Angulo, *Astrophys. J.* **600**, 544 (2004), [arXiv:astro-ph/0309480].
- [17] P. de Bernardis *et al.* (Boomerang), *Nature* **404**, 955 (2000), [astro-ph/0004404].
- [18] E. Komatsu *et al.* (WMAP), *Astrophys. J. Suppl. Ser.* **180**, 330 (2009), [arXiv:0803.0547].
- [19] S. Cole *et al.* (The 2dFGRS), *Mon. Not. R. Astron. Soc.* **362**, 505 (2005), [astro-ph/0501174].
- [20] S. Dodelson *et al.* (SDSS), *Astrophys. J.* **572**, 140 (2001), [astro-ph/0107421].
- [21] A. J. Benson, M. Kamionkowski, and S. H. Hassani, *Mon. Not. R. Astron. Soc.* **357**, 847 (2005), [astro-ph/0407136].

- [22] F. Zwicky, *Astrophys. J.* **86**, 217 (1937).
- [23] V. C. Rubin, W. K. J. Ford, and N. . Thonnard, *Astrophys. J.* **238**, 471 (1980).
- [24] A. Bosma, *Astron. J.* **86**, 1825 (1981).
- [25] R. Baum and W. Sheehan, *In Search of Planet Vulcan: The Ghost in Newton's Clockwork Universe* (Basic Books, 2003).
- [26] M. Milgrom, *Astrophys. J.* **270**, 365 (1983).
- [27] J. D. Bekenstein, *Phys. Rev. D* **70**, 083509 (2004), [arXiv:astro-ph/0403694].
- [28] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, *Astrophys. J. Lett.* **648**, L109 (2006), [arXiv:astro-ph/0608407].
- [29] G. R. Dvali, G. Gabadadze, and M. Porrati, *Phys. Lett. B* **485**, 208 (2000), [hep-th/0005016].
- [30] C. Deffayet, G. Dvali, and G. Gabadadze, *Phys. Rev. D* **65**, 044023 (2002), [arXiv:astro-ph/0105068].
- [31] K. Freese and M. Lewis, *Phys. Lett. B* **540**, 1 (2002), [arXiv:astro-ph/0201229].
- [32] N. Arkani-Hamed, S. Dimopoulos, G. Dvali, and G. Gabadadze, arXiv e-print (2002), [arXiv:hep-th/0209227].
- [33] S. M. Carroll, V. Duvvuri, M. Trodden, and M. S. Turner, *Phys. Rev. D* **70**, 043528 (2004), [astro-ph/0306438].
- [34] S. Capozziello, S. Carloni, and A. Troisi, *Recent Res. Dev. Astron. Astrophys.* **1**, 625 (2003), [astro-ph/0303041].
- [35] S. Nojiri and S. D. Odintsov, *Modern Phys. Lett. A* **19**, 627 (2004), [arXiv:hep-th/0310045].
- [36] N. A. Hamed, H. S. Cheng, M. A. Luty, and S. Mukohyama, *JHEP* **5**, 74 (2004), [arXiv:hep-th/0312099].
- [37] M. C. B. Abdalla, S. Nojiri, and S. D. Odintsov, *Class. Quant. Grav.* **22**, L35 (2005), [arXiv:hep-th/0409177].
- [38] R. M. Wald, *Phys. Rev. D* **28**, 2118 (1983).
- [39] M. Tegmark, A. de Oliveira-Costa, and A. J. Hamilton, *Phys. Rev. D* **68**, 123523 (2003), [arXiv:astro-ph/0302496].
- [40] A. de Oliveira-Costa, M. Tegmark, M. Zaldarriaga, and A. Hamilton, *Phys. Rev. D* **69**, 063516 (2004), [arXiv:astro-ph/0307282].

- [41] K. Land and J. Magueijo, *Phys. Rev. Lett.* **95**, 071301 (2005), [arXiv:astro-ph/0502237].
- [42] F. K. Hansen, A. J. Banday, and K. M. Górski, *Mon. Not. R. Astron. Soc.* **354**, 641 (2004), [arXiv:astro-ph/0404206].
- [43] H. K. Eriksen, F. K. Hansen, A. J. Banday, K. M. Górski, and P. B. Lilje, *Astrophys. J.* **605**, 14 (2004).
- [44] H. K. Eriksen, A. J. Banday, K. M. Górski, F. K. Hansen, and P. B. Lilje, *Astrophys. J. Lett.* **660**, L81 (2007), [arXiv:astro-ph/0701089].
- [45] F. K. Hansen, A. J. Banday, K. M. Gorski, H. K. Eriksen, and P. B. Lilje, arXiv e-print (2008), [0812.3795].
- [46] J. Hoftuft, H. K. Eriksen, A. J. Banday, K. M. Gorski, F. K. Hansen, and P. B. Lilje, arXiv e-print (2009), [0903.1229].
- [47] T. R. Jaffe, A. J. Banday, H. K. Eriksen, K. M. Górski, and F. K. Hansen, *Astrophys. J. Lett.* **629**, L1 (2005), [arXiv:astro-ph/0503213].
- [48] T. R. Jaffe, A. J. Banday, H. K. Eriksen, K. M. Górski, and F. K. Hansen, *Astron. Astrophys.* **460**, 393 (2006), [arXiv:astro-ph/0606046].
- [49] P. Bielewicz, H. K. Eriksen, A. J. Banday, K. M. Górski, and P. B. Lilje, *Astrophys. J.* **635**, 750 (2005), [arXiv:astro-ph/0507186].
- [50] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, *Mon. Not. R. Astron. Soc.* **367**, 79 (2006), [arXiv:astro-ph/0508047].
- [51] A. Bernui, T. Villela, C. A. Wuensche, R. Leonardi, and I. Ferreira, *Astron. Astrophys.* **454**, 409 (2006), [arXiv:astro-ph/0601593].
- [52] N. E. Groeneboom and H. K. Eriksen, *Astrophys. J.* **690**, 1807 (2009), [0807.2242].
- [53] J. Maldacena, *JHEP* **5**, 13 (2003), [arXiv:astro-ph/0210603].
- [54] P. Vielva, E. Martínez-González, R. B. Barreiro, J. L. Sanz, and L. Cayón, *Astrophys. J.* **609**, 22 (2004), [arXiv:astro-ph/0310273].
- [55] H. K. Eriksen, D. I. Novikov, P. B. Lilje, A. J. Banday, and K. M. Górski, *Astrophys. J.* **612**, 64 (2004), [arXiv:astro-ph/0401276].
- [56] M. Cruz, E. Martínez-González, P. Vielva, and L. Cayón, *Mon. Not. R. Astron. Soc.* **356**, 29 (2005), [arXiv:astro-ph/0405341].

- [57] E. Martínez-González, M. Cruz, L. Cayón, and P. Vielva, *New Astron. Rev.* **50**, 875 (2006).
- [58] C. R ath, P. Schuecker, and A. J. Banday, *Mon. Not. R. Astron. Soc.* **380**, 466 (2007).
- [59] A. P. S. Yadav and B. D. Wandelt, *Phys. Rev. Lett.* **100**, 181301 (2008), [0712.1148].
- [60] J. D. McEwen, M. P. Hobson, A. N. Lasenby, and D. J. Mortlock, *Mon. Not. R. Astron. Soc.* **388**, 659 (2008), [0803.2157].
- [61] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, *JCAP* **08**, 031 (2008), [0805.3580].
- [62] K. M. Smith, L. Senatore, and M. Zaldarriaga, arXiv e-print (2009), [0901.2572].
- [63] E. W. Kolb and M. S. Turner, *The Early Universe* (Basic Books, 1994), 2nd Ed.
- [64] L. Alabidi and J. E. Lidsey, *Phys. Rev. D* **78**, 103519 (2008), [0807.2181].
- [65] A. L. Erickcek, M. Kamionkowski, and A. J. Benson, *Mon. Not. R. Astron. Soc.* **371**, 1992 (2006), [astro-ph/0604281].
- [66] A. L. Erickcek, T. L. Smith, and M. Kamionkowski, *Phys. Rev. D* **74**, 121501 (2006), [astro-ph/0610483].
- [67] T. Chiba, T. L. Smith, and A. L. Erickcek, *Phys. Rev. D* **75**, 124014 (2007), [astro-ph/0611867].
- [68] T. L. Smith, A. L. Erickcek, R. R. Caldwell, and M. Kamionkowski, *Phys. Rev. D* **77**, 024015 (2008), [0708.0001].
- [69] A. L. Erickcek, S. M. Carroll, and M. Kamionkowski, *Phys. Rev. D* **78**, 083012 (2008), [0808.1570].
- [70] A. L. Erickcek, M. Kamionkowski, and S. M. Carroll, *Phys. Rev. D* **78**, 123520 (2008), [0806.0377].
- [71] A. L. Erickcek, C. M. Hirata, and M. Kamionkowski (2009), in preparation.
- [72] T. Chiba, *Phys. Lett. B* **575**, 1 (2003), [astro-ph/0307338].
- [73] A. Rajaraman, arXiv e-print (2003), [astro-ph/0311160].
- [74] T. Multamaki and I. Vilja, *Phys. Rev. D* **74**, 064022 (2006), [astro-ph/0606373].
- [75] V. Faraoni, *Phys. Rev. D* **74**, 023529 (2006), [gr-qc/0607016].
- [76] M. L. Ruggiero and L. Iorio, *JCAP* **0701**, 010 (2007), [gr-qc/0607093].
- [77] G. Allemandi, M. Francaviglia, M. L. Ruggiero, and A. Tartaglia, *Gen. Relativ. Gravit.* **37**, 1891 (2005), [gr-qc/0506123].

- [78] V. Kagramanova, J. Kunz, and C. Lammerzahl, *Phys. Lett. B* **634**, 465 (2006), [gr-qc/0602002].
- [79] W. Hu and I. Sawicki, *Phys. Rev. D* **76**, 064004 (2007), [0705.1158].
- [80] B. A. Campbell, M. J. Duncan, N. Kaloper, and K. A. Olive, *Nucl. Phys.* **351**, 778 (1991).
- [81] B. A. Campbell, M. J. Duncan, N. Kaloper, and K. A. Olive, *Phys. Lett. B* **251**, 34 (1990).
- [82] R. Jackiw and S. Y. Pi, *Phys. Rev. D* **68**, 104012 (2003), [gr-qc/0308071].
- [83] I. Ciufolini and E. C. Pavlis, *Nature* **431**, 958 (2004).
- [84] C. M. Will, *Phys. Rev. D* **67**, 062003 (2003), <http://einstein.stanford.edu>, [gr-qc/0212069].
- [85] L. P. Grishchuk and I. B. Zel'dovich, *Astron. Zh.* **55**, 209 (1978), [Translation: *Sov. Astron.* **22**, 125 (1978)].
- [86] S. Mollerach, *Phys. Rev. D* **42**, 313 (1990).
- [87] A. D. Linde and V. F. Mukhanov, *Phys. Rev. D* **56**, 535 (1997), [astro-ph/9610219].
- [88] D. H. Lyth and D. Wands, *Phys. Lett. B* **524**, 5 (2002), [arXiv:hep-ph/0110002].
- [89] T. Moroi and T. Takahashi, *Phys. Lett. B* **522**, 215 (2001), [Erratum-ibid. B 539, 303 (2002)], [arXiv:hep-ph/0110096].
- [90] C. M. Hirata (2009), in preparation.
- [91] M. Smoluchowski, *Physik. Zeit.* **17**, 557 (1916).
- [92] K. Menou, Z. Haiman, and V. K. Narayanan, *Astrophys. J.* **558**, 535 (2001), [astro-ph/0101196].
- [93] R. K. Sheth and G. Tormen, *Mon. Not. R. Astron. Soc.* **308**, 119 (1999), [astro-ph/9901122].
- [94] A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida, *Mon. Not. R. Astron. Soc.* **321**, 372 (2001), [astro-ph/0005260].
- [95] M. G. Haehnelt, *Mon. Not. R. Astron. Soc.* **269**, 199 (1994), [astro-ph/9405032].
- [96] J. S. B. Wyithe and A. Loeb, *Astrophys. J.* **590**, 691 (2003), [astro-ph/0211556].
- [97] A. Sesana, F. Haardt, P. Madau, and M. Volonteri, *Astrophys. J.* **611**, 623 (2004), [astro-ph/0401543].
- [98] A. Sesana, F. Haardt, P. Madau, and M. Volonteri, *Astrophys. J.* **623**, 23 (2005), [astro-ph/0409255].

- [99] K. J. Rhoads and J. S. B. Wyithe, *Mon. Not. R. Astron. Soc.* **361**, 1145 (2005), [astro-ph/0503210].
- [100] S. Koushiappas and A. Zentner, *Astrophys. J.* **639**, 7 (2006), [astro-ph/0503511].
- [101] L. Ferrarese and D. Merritt, *Astrophys. J. Lett.* **539**, L9 (2000), [astro-ph/0006053].
- [102] K. Gebhardt, D. Richstone, J. Kormendy, T. R. Lauer, E. A. Ajhar, R. Bender, A. Dressler, S. M. Faber, C. Grillmair, J. Magorrian, *et al.*, *Astron. J.* **119**, 1157 (2000), [astro-ph/9912026].
- [103] L. Ferrarese, *Current High Energy Emission Around Black Holes* (Singapore: World Scientific, 2002), ed. C.H. Lee, [astro-ph/0203047].
- [104] L. Ferrarese, *Astrophys. J.* **578**, 90 (2002), [astro-ph/0203469].
- [105] J. S. B. Wyithe and A. Loeb, *Astrophys. J.* **595**, 614 (2003), [astro-ph/0304156].
- [106] R. Barkana and A. Loeb, *Physics Reports* **349**, 125 (2001), [astro-ph/0010468].
- [107] T. Kitayama and Y. Suto, *Astrophys. J.* **469**, 480 (1996), [astro-ph/9604141].
- [108] N. N. Weinberg and M. Kamionkowski, *Mon. Not. R. Astron. Soc.* **341**, 251 (2003), [astro-ph/0210134].
- [109] D. J. Eisenstein and W. Hu, *Astrophys. J.* **496**, 605 (1998), [astro-ph/9709112].
- [110] P. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, Princeton, NJ, 1980), chap. II.10 Linear Perturbation Approximation, pp. 49–51, Princeton Series in Physics.
- [111] S. M. Carroll, W. H. Press, and E. L. Turner, *Annual Rev. Astron. Astrophys.* **30**, 499 (1992).
- [112] M. Colpi, L. Mayer, and F. Governato, *Astrophys. J.* **525**, 720 (1999), [astro-ph/9907088].
- [113] S. Kazantzidis, L. Mayer, M. Colpi, P. Madau, V. P. Debattista, J. Wadsley, J. Stadel, T. Quinn, and B. Moore, *Astrophys. J. Lett.* **623**, L67 (2005), [astro-ph/0407407].
- [114] A. A. Starobinsky, *Phys. Lett. B* **91**, 99 (1980).
- [115] C. Brans and R. H. Dicke, *Phys. Rev.* **124**, 925 (1961).
- [116] B. Bertotti, L. Iess, and P. Tortora, *Nature* **425**, 374 (2003).
- [117] S. S. Shapiro, J. L. Davis, D. E. Lebach, and J. S. Gregory, *Phys. Rev. Lett.* **92**, 121101 (2004).
- [118] D. N. Vollick, *Class. Quant. Grav.* **21**, 3813 (2004), [gr-qc/0312041].
- [119] E. E. Flanagan, *Phys. Rev. Lett.* **92**, 071101 (2004), [astro-ph/0308111].

- [120] S. M. Carroll *et al.*, Phys. Rev. D **71**, 063513 (2005), [astro-ph/0410031].
- [121] S. M. Carroll, I. Sawicki, A. Silvestri, and M. Trodden, New J. Phys. **8**, 323 (2006), [astro-ph/0607458].
- [122] K. Kainulainen, V. Reijonen, and D. Sunhede, Phys. Rev. D **76**, 043503 (2007), [gr-qc/0611132].
- [123] C.-G. Shao, R.-G. Cai, B. Wang, and R.-K. Su, Phys. Lett. B **633**, 164 (2006), [gr-qc/0511034].
- [124] J. A. R. Cembranos, Phys. Rev. D **73**, 064029 (2006), [gr-qc/0507039].
- [125] A. D. Dolgov and M. Kawasaki, Phys. Lett. B **573**, 1 (2003), [astro-ph/0307285].
- [126] G. J. Olmo, Phys. Rev. D **72**, 083505 (2005).
- [127] C. M. Will, *Theory and experiment in gravitational physics* (Cambridge, UK: University Press., 1993).
- [128] G. D. Birkhoff, *Relativity and Modern Physics* (Cambridge, MA: Harvard University Press, 1923).
- [129] X.-H. Jin, D.-J. Liu, and X.-Z. Li, arXiv e-print (2006), [astro-ph/0610854].
- [130] C. M. Will, *Theory and experiment in gravitational physics* (Cambridge, UK: Univ. Pr., 1993), pgs. 91-92.
- [131] T. Clifton and J. D. Barrow, Phys. Rev. D **72**, 103005 (2005), [gr-qc/0509059].
- [132] G. V. Bicknell, J. Phys. A: Math. Nucl. Gen. **7**, 1061 (1974).
- [133] J. Khoury and A. Weltman, Phys. Rev. Lett. **93**, 171104 (2004), [arXiv:astro-ph/0309300].
- [134] J. Khoury and A. Weltman, Phys. Rev. D **69**, 044026 (2004), [arXiv:astro-ph/0309411].
- [135] T. Faulkner, M. Tegmark, E. F. Bunn, and Y. Mao, Phys. Rev. D **76**, 063505 (2007), [arXiv:astro-ph/0612569].
- [136] S. Appleby and R. Battye, Phys. Lett. B **654**, 7 (2007), [0705.3199].
- [137] A. A. Starobinsky, Soviet Journal of Experimental and Theoretical Physics Letters **86**, 157 (2007), [0706.2041].
- [138] I. Navarro and K. Van Acoleyen, JCAP **0702**, 022 (2007), [gr-qc/0611127].
- [139] S. Deser, R. Jackiw, and S. Templeton, Ann. Phys. **140**, 372 (1982).

- [140] A. Lue, L.-M. Wang, and M. Kamionkowski, Phys. Rev. Lett. **83**, 1506 (1999), [astro-ph/9812088].
- [141] S. H.-S. Alexander, M. E. Peskin, and M. M. Sheikh-Jabbari, Phys. Rev. Lett. **96**, 081301 (2006), [hep-th/0403069].
- [142] S. H. S. Alexander and J. Gates, S. James, JCAP **0606**, 018 (2006), [hep-th/0409014].
- [143] S. Alexander and N. Yunes, Phys. Rev. Lett. **99**, 241101 (2007), [hep-th/0703265].
- [144] S. Alexander and N. Yunes, Phys. Rev. D **75**, 124022 (2007), [0704.0299].
- [145] S. M. Carroll, Phys. Rev. Lett. **81**, 3067 (1998), [astro-ph/9806099].
- [146] S. M. Carroll, G. B. Field, and R. Jackiw, Phys. Rev. D **41**, 1231 (1990).
- [147] K. Konno, T. Matsuyama, and S. Tanda, Phys. Rev. D **76**, 024009 (2007), [0706.3080].
- [148] D. Brouwer and G. M. Clemence, *Methods of Celestial Mechanics* (New York: Academic Press, 1961).
- [149] L. Iorio, Nuovo Cim. **116**, 777 (2001), [gr-qc/9908080].
- [150] T. L. Smith, *The gravity of the situation*, Ph.D. thesis, California Institute of Technology, Pasadena, CA (2008).
- [151] J. Lense and H. Thirring, Phys. Z. **19**, 711 (1918), translated by B. Mashhoon, F. W. Hehl, and D. S. Theiss, Gen. Rel. Grav. **16**, 711 (1984).
- [152] I. Ciufolini, A. Paolozzi, D. G. Currie, and E. C. Pavlis, *LARES/WEBER-SAT, frame-dragging and fundamental physics* (2004), prepared for International Workshop on Frontier Science: Physics and Astrophysics in Space, Frascati and Rome, Italy, 14-19 Jun 2004.
- [153] G. E. Pugh, Weapons System Evaluation Group, Research Memorandum No. 111, Department of Defense (USA) (1959).
- [154] L. I. Schiff, Phys. Rev. Lett. **4**, 215 (1960).
- [155] L. I. Schiff, Proc. Nat. Acad. Sci. **46**, 871 (1960).
- [156] G. Brunfiel, Nature **444**, 978 (2006).
- [157] G. F. Smoot *et al.* (COBE), Astrophys. J. Lett. **371**, L1 (1991).
- [158] G. Hinshaw *et al.* (WMAP), Astrophys. J. Suppl. Ser. **148**, 135 (2003), [arXiv:astro-ph/0302217].

- [159] M. S. Turner, *Phys. Rev. D* **44**, 3737 (1991).
- [160] A. Kashlinsky, I. I. Tkachev, and J. Frieman, *Phys. Rev. Lett.* **73**, 1582 (1994), [arXiv:astro-ph/9405024].
- [161] J. García-Bellido, A. R. Liddle, D. H. Lyth, and D. Wands, *Phys. Rev. D* **52**, 6750 (1995), [arXiv:astro-ph/9508003].
- [162] P. G. Castro, M. Douspis, and P. G. Ferreira, *Phys. Rev. D* **68**, 127301 (2003), [arXiv:astro-ph/0309320].
- [163] M. Bruni and D. H. Lyth, *Phys. Lett. B* **323**, 118 (1994), [arXiv:astro-ph/9307036].
- [164] C. Gordon, W. Hu, D. Huterer, and T. Crawford, *Phys. Rev. D* **72**, 103002 (2005), [arXiv:astro-ph/0509301].
- [165] C. Gordon, *Astrophys. J.* **656**, 636 (2007), [arXiv:astro-ph/0607423].
- [166] J. F. Donoghue, K. Dutta, and A. Ross, arXiv e-print (2007), [arXiv:astro-ph/0703455].
- [167] A. Kashlinsky, F. Atrio-Barandela, D. Kocevski, and H. Ebeling, *Astrophys. J. Lett.* **686**, L49 (2008), [0809.3734].
- [168] L. Mersini-Houghton and R. Holman, *JCAP* **0902**, 006 (2009), [0810.5388].
- [169] A. Kashlinsky, F. Atrio-Barandela, D. Kocevski, and H. Ebeling, *Astrophys. J.* **691**, 1479 (2009), [0809.3733].
- [170] R. Watkins, H. A. Feldman, and M. J. Hudson, *Mon. Not. R. Astron. Soc.* **392**, 743 (2009), [0809.4041].
- [171] W. Hu and N. Sugiyama, *Phys. Rev. D* **51**, 2599 (1995), [arXiv:astro-ph/9411008].
- [172] W. Hu and N. Sugiyama, *Astrophys. J.* **444**, 489 (1995), [arXiv:astro-ph/9407093].
- [173] R. K. Sachs and A. M. Wolfe, *Astrophys. J.* **147**, 73 (1967).
- [174] P. J. Peebles and D. T. Wilkinson, *Phys. Rev.* **174**, 2168 (1968).
- [175] M. Kamionkowski and L. Knox, *Phys. Rev. D* **67**, 063001 (2003), [arXiv:astro-ph/0210165].
- [176] H. Kodama and M. Sasaki, *Progress of Theoretical Physics Supplement* **78**, 1 (1984).
- [177] C. M. Hirata and U. Seljak, *Phys. Rev. D* **72**, 083501 (2005), [arXiv:astro-ph/0503582].
- [178] G. Efstathiou, *Mon. Not. R. Astron. Soc.* **348**, 885 (2004), [arXiv:astro-ph/0310207].
- [179] A. Linde and V. Mukhanov, *JCAP* **4**, 9 (2006), [arXiv:astro-ph/0511736].

- [180] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, *Phys. Rev. D* **28**, 679 (1983).
- [181] B. A. Bassett, S. Tsujikawa, and D. Wands, *Reviews of Modern Physics* **78**, 537 (2006), [arXiv:astro-ph/0507632].
- [182] B. Lew, *JCAP* **08**, 017 (2008), [0803.1409].
- [183] L. Covi, J. Hamann, A. Melchiorri, A. Slosar, and I. Sorbera, *Phys. Rev. D* **74**, 083509 (2006), [arXiv:astro-ph/0606452].
- [184] J. Hamann, L. Covi, A. Melchiorri, and A. Slosar, *Phys. Rev. D* **76**, 023503 (2007), [arXiv:astro-ph/0701380].
- [185] T. Moroi and T. Takahashi, *Phys. Rev. D* **66**, 063501 (2002), [arXiv:hep-ph/0206026].
- [186] D. H. Lyth, C. Ungarelli, and D. Wands, *Phys. Rev. D* **67**, 023503 (2003), [arXiv:astro-ph/0208055].
- [187] D. H. Lyth and D. Wands, *Phys. Rev. D* **68**, 103516 (2003), [arXiv:astro-ph/0306500].
- [188] S. Gupta, K. A. Malik, and D. Wands, *Phys. Rev. D* **69**, 063513 (2004), [arXiv:astro-ph/0311562].
- [189] F. Ferrer, S. Räsänen, and J. Väliviita, *JCAP* **10**, 10 (2004), [arXiv:astro-ph/0407300].
- [190] M. Lemoine and J. Martin, *Phys. Rev. D* **75**, 063504 (2007), [arXiv:astro-ph/0611948].
- [191] L. Verde and H. Peiris, *JCAP* **7**, 9 (2008), [arXiv:0802.1219].
- [192] S. Weinberg, *Phys. Rev. D* **70**, 083522 (2004), [arXiv:astro-ph/0405397].
- [193] K. A. Malik, D. Wands, and C. Ungarelli, *Phys. Rev. D* **67**, 063516 (2003), [arXiv:astro-ph/0211602].
- [194] D. Langlois and F. Vernizzi, *Phys. Rev. D* **70**, 063522 (2004), [arXiv:astro-ph/0403258].
- [195] L. Verde, L.-M. Wang, A. Heavens, and M. Kamionkowski, *Mon. Not. R. Astron. Soc.* **313**, L141 (2000), [astro-ph/9906301].
- [196] E. Komatsu and D. N. Spergel, *Phys. Rev. D* **63**, 063002 (2001), [arXiv:astro-ph/0005036].
- [197] K. A. Malik and D. H. Lyth, *JCAP* **0609**, 008 (2006), [astro-ph/0604387].
- [198] K. Ichikawa, T. Suyama, T. Takahashi, and M. Yamaguchi, *Phys. Rev. D* **78**, 023513 (2008), [0802.4138].
- [199] E. Komatsu *et al.* (WMAP), *Astrophys. J. Suppl. Ser.* **148**, 119 (2003), [astro-ph/0302223].

- [200] D. Babich and M. Zaldarriaga, *Phys. Rev. D* **70**, 083005 (2004), [arXiv:astro-ph/0408455].
- [201] A. P. S. Yadav, E. Komatsu, B. D. Wandelt, M. Liguori, F. K. Hansen, and S. Matarrese, *Astrophys. J.* **678**, 578 (2008), [0711.4933].
- [202] A. R. Pullen and M. Kamionkowski, *Phys. Rev. D* **76**, 103529 (2007), [0709.1144].
- [203] C. Dvorkin, H. V. Peiris, and W. Hu, *Phys. Rev. D* **77**, 063008 (2008), [arXiv:0711.2321].
- [204] T. L. Smith, H. V. Peiris, and A. Cooray, *Phys. Rev. D* **73**, 123503 (2006), [astro-ph/0602137].
- [205] M. Lemoine, J. Martin, and G. Petit, *Phys. Rev. D* **78**, 063516 (2008), [arXiv:0802.1601].
- [206] H. Kurki-Suonio, V. Muhonen, and J. Väliviita, *Phys. Rev. D* **71**, 063005 (2005), [arXiv:astro-ph/0412439].
- [207] U. Seljak and M. Zaldarriaga, *Astrophys. J.* **469**, 437 (1996), [arXiv:astro-ph/9603033].
- [208] M. Beltrán, J. García-Bellido, J. Lesgourgues, and M. Viel, *Phys. Rev. D* **72**, 103515 (2005), [arXiv:astro-ph/0509209].
- [209] R. Keskitalo, H. Kurki-Suonio, V. Muhonen, and J. Väliviita, *JCAP* **9**, 8 (2007), [arXiv:astro-ph/0611917].
- [210] M. Beltrán, J. García-Bellido, and J. Lesgourgues, *Phys. Rev. D* **75**, 103507 (2007), [arXiv:hep-ph/0606107].
- [211] C. Hikage, K. Koyama, T. Matsubara, T. Takahashi, and M. Yamaguchi, arXiv e-print (2008), [0812.3500].
- [212] D. Langlois and T. Piran, *Phys. Rev. D* **53**, 2908 (1996), [arXiv:astro-ph/9507094].
- [213] D. Langlois, *Phys. Rev. D* **54**, 2447 (1996), [arXiv:gr-qc/9606066].
- [214] G. Hinshaw *et al.* (WMAP), *Astrophys. J. Suppl. Ser.* **180**, 225 (2009), [0803.0732].
- [215] P. Erdoğdu, O. Lahav, J. P. Huchra, M. Colless, R. M. Cutri, E. Falco, T. George, T. Jarrett, D. H. Jones, L. M. Macri, *et al.*, *Mon. Not. R. Astron. Soc.* **373**, 45 (2006), [arXiv:astro-ph/0610005].
- [216] G. Lavaux, R. B. Tully, R. Mohayaee, and S. Colombi, arXiv e-print (2008), [0810.3658].
- [217] J. Dunkley *et al.* (WMAP), *Astrophys. J. Suppl. Ser.* **180**, 306 (2009), [0803.0586].
- [218] A. R. Liddle, P. Parsons, and J. D. Barrow, *Phys. Rev. D* **50**, 7222 (1994), [arXiv:astro-ph/9408015].