Searching for anomalies in the Cosmic Microwave Background

Nicolaas Ervik Groeneboom
Institute of Theoretical Astrophysics
University of Oslo
Norway

March, 2010

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No.962

ISSN 1501-7710

Cover: Inger Sandved Anfinsen.
Printed in Norway: AiT e-dit AS.

Produced in co-operation with Unipub.
The thesis is produced by Unipub merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Acknowledgments

I wish to mention key people in a chronological order, making sure that credit is given where credit is due. First, I would like to thank an old buddy from ages past, Håkon Bergsaker, for convincing me to apply as a student at the University of Oslo in 2003. I would never have worked as tirelessly were it not for the undergraduate physics students organization "Fysikkforeningen" and the excellent community we had there. I need mention my closest and most important friends and colleagues from this period: Øystein Rudjord, Ingeborg Ligaarden, Jostein Riiser Kristiansen, Tuva Hope, Lillian Smestad, Hanne-Torill Mevik, Knut Berg, Kosovare Olluri and Anders Nermoen being some. During my first year, it was Jostein who convinced me to follow more advanced physics courses. Later, he advised me to have a chat with prof. Øystein Elgarøy at the astrophysics department, and I soon became his master student. I was therefore saved from pure mathematics by computational cosmology in 2006. Øystein proved to be an excellent supervisor, and before I knew it, the head of institute Per B. Lilje managed to obtain funding for a PhD position. I became the first PhD student of Hans Kristian Eriksen, to whom I give my greatest regards: his keen intellect, his authentic interest, inclusive kindness and immense production output has proved to be an extremely valuable combination. Without Hans Kristian, my thesis would have been close to nothing. In addition, I would like to thank the rest of the cosmology crew at the Institute of Theoretical Astrophysics for their 3-year support, and especially all the Øysteins and Josteins, Kristin and Magnus in addition to Håkon, Frode and David.

Contents

Part I

Modern Cosmology

1 Introduction

A couple of words before you start reading. This thesis is structured into two parts, where the first contains an introduction to CMB analysis together with a quick summary of my work. The second part contains the academic papers that I have produced, and represents three years of hardship. This introduction is indented to be readable by everyone, from family members to cosmologists, from cats and celery to fungi. Here, the history of cosmology is presented through an empirical perspective. Section 2 is more technical, and requires prior knowledge of mathematical concepts. This is intended to be readable for everyone with a background in physics, and might as well be most interesting for Master students or aspiring PhD students. Section 3, 4, 5 and 6 dive into the technical details of my work, and is probably most interesting for people who are already well into the field of cosmological data analysis. Section 7 contains a quick and easy-to-understand summary explaining what each of my papers are about. Finally, section 8 concludes the thesis.

1.1 Creation myths

Every civilization in human history has had a say on the dynamics of the universe. Alas, all previous civilizations have been somewhat incorrect in their assumptions. However cool it might sound, we know today that Viracocha never emerged from Lake Titicaca and made the moon, the sun and the stars. The Maoir god Maui did not fish the island Te Ika of New Zealand out of the sea using his grandmother's jawbone. And all animals are not created from clarified butter as Hindu myth boldly claims, but are evolved through slight mutations from generation to generation. But how can we be so certain that the old myths are not true, that they have no root in reality? Well, lack of empirical evidence is a major reason. Today, we have simpler and more concise theories that not only explain the dynamics in the universe, but can also give predictions that can be compared with observations. In this way, scientific cosmology made creation myths obsolete.

But science also differs greatly from myths, as it contains an intrinsic rule of feedback. Whereas a creation myth tries to explain the origin of the universe through supernatural means, science does not try to explain everything at once. Science also reports back to it self when better data is acquired, such that theories either perish or are strengthened. New theories are constantly being forged, and compete fiercely with older, more established theories. The surviving theories are usually the ones that in simplest means can explain the greatest variation of phenomena. Modern scientific cosmology only became possible when theoretical frameworks and empirical data were readily available. However, neither the ancient Greeks, old Arabic scholars nor Christian medieval philosophers spoke mathematics, the

language of nature. And, communicating in a strange language one doesn't master is not particularly fruitful. In order to unravel the mysteries of our universe, humanity first needed to develop mathematics.

1.2 Physical theories

While mathematical structures are universally true, proved rigorously by theorems, lemmas and propositions, the trueness of a physical theory will forever stay uncertain. This is because all physical theories are approximations to reality, and are never a complete description of the system it mimics. If such a theory happens to neatly describe a physical phenomena, it does not rule out the possibility that there might exist other equally correct theories describing the same system, but with a different take on the interpretation.

A physical theory therefore has a space of validity. A complicated theory

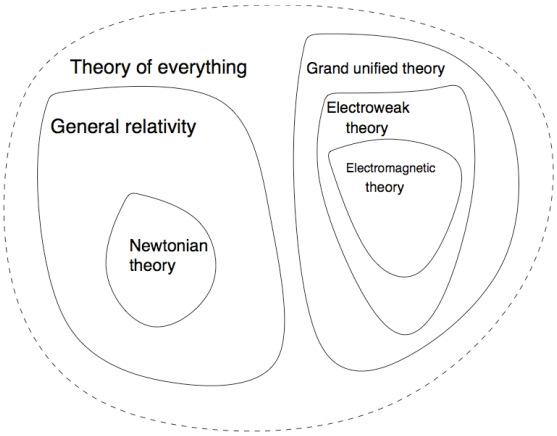


Figure 1: Grouped theories. Is theory of everything (TOE) bounded, or is it contained in another larger theory, theory-of-absolutely-everything?

that has a large space of validity might for certain limits of observables converge to simpler sub-theories with smaller spaces of validity. As an example, consider the theory of general relativity, which has a larger space of validity than the more static Newtonian gravity. Newton's theory isn't *incorrect*, but rather inaccurate when describing systems operating with velocities close to the speed of light. But for non-relativistic systems, Newton's theory of gravity is perfectly adequate. It can be shown that general relativity converges to the Newtonian limit for low-velocity systems. Eventually, general relativity breaks down for extremely small systems, which marks the boundary of

its space of validity. However, the physical interpretation of the two theories differ greatly.

A physical theory is not something that can exist independently from empirical evidence. Good science is waging a constant war between theory and empirical data, where every unsuccessful attack by observations on a model strengthens it. However, when data manages to overturn a theory, it may become obsolete and wither away. In this sense, data are always correct, wheras a theory can never be completely verified. Isaac Newton was the first person who effectively came up with a theory that describes the universe in a convenient, mathematical way. While the theory of Newton is a limited description of the universe as a whole, it was a vast improvement to previous physical theories. With Newtonian mechanics, it was suddenly possible to predict the movement of heavenly bodies, and show that the mechanics were in no way different than those on Earth. However, Newtonian physics had some severe limitations, the most important being that the speed of light was not constant. During the end of the 19th century, experiments confirmed that light always propagates at about 300 000 km/s, regardless of the reference frame. Several people developed new theories by flexing previous ideas, such that space and time become independent. If the light speed is constant in all reference frames, something else needs to bend. This "something else" is space and time, now unified into a single entity called space-time. No one managed to express this idea as clearly as Einstein. In 1915, he published his arguably most important work on the theory of General Relativity. General Relativity is a vast improvement to Newtonian physics, and corresponds to Newtons theory when going to down-to-earth limits, as presented in Figure 1. The framework also solved several problems with the nature of gravity and space-time, doing so by proposing a simple and elegant solution. The core of general relativity can be expressed with the following phrase:

> *Space and time is closely coupled, and is curved by matter. Matter moves in straight lines through curved space-time.*

The implications were profound. Not only is time and space closely coupled, but can "bend" according to the existence of matter. And matter does not "feel" gravity, but moves in a straight line through a curved space-time corresponding to orbits in 3D-space. This is much in the same way an airplane follows a straight line on the curved surface of Earth (a 2D sphere), corresponding to great circles in 3D-space.

1.3 The expanding universe

Now that Einstein had the framework of General Relativity, he used it to calculate some simplified properties of the universe. One of the major results shocked him: the equations indicated that the universe cannot be static, or

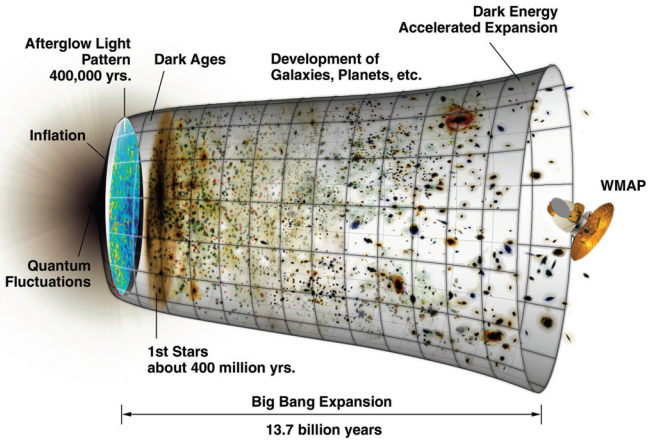


Figure 2: An illustration of the CMB and the evolution of the universe. Courtesy of the LAMBDA WMAP site (http://lambda.gsfc.nasa.gov/).

unchangeable. The universe has to either expand or contract, depending on its contents. Homogeneous matter alone cannot be stable, it either clumps together or moves apart. This discovery repulsed Einstein, who had strong philosophical beliefs that the universe was unchangeable, that it had always existed as it is today. This belief made him introduce a term in the equations, something he dubbed "the cosmological constant". The constant forced the universe to remain static, but at a cost: the solutions were highly unstable. Some 15 years after Einstein, Hubble discovered through supernova data that the universe indeed is expanding. This is possible even though the universe might be infinitely large. If space is expanding now, it must earlier have been smaller and denser. Going all the way back, there must have been a time where it was nearly infinitely compact. But how do we confirm whether this scenario has happened? By observational predictions and empirical evidence, naturally.

1.4 The Cosmic Microwave Background

If the universe used to be smaller and denser, it must have been hotter. In fact, going far enough back in time, it should have been so hot and dense that photons (light) and matter crashed together constantly, acting as one huge plasma fluid. As the universe expanded and cooled, it eventually became transparent for the photons. This means they no longer collided with particles, and were free to escape. This would have happened about 400 000 years after the big bang, where the escaping photons defined a sort of

photographic "snapshot" of what the universe looked like back then. In 1964, radio astronomers Penzias and Wilson [19] discovered these remnants from the early universe, indicating that it used to be hotter and denser. The photons from this era have since been stretched by the expansion of the universe, turning them into low-energy microwave radiation. In addition, this radiation penetrates the whole universe, having nearly the same temperature in every direction. Also, the microwave background has shown to be the most-precisely measured black body spectrum in nature. The radiation was named the Cosmic Microwave Background (CMB). Suddenly, the field of cosmology went from being a purely theoretical to having observational data. This was the birth of a completely new field in science, and the implications were astonishing. The CMB is probably the most important cosmological data set we have today.

The temperature of the CMB is today about 2.73 Kelvin, but there exist small additional fluctuations that are in the order of milikelvin, as depicted in Figure 4. A schematic presentation of the CMB and the evolution of the universe is presented in Figure 2. In itself, these small fluctuations in the temperature are not particularly interesting. However, the real value resides

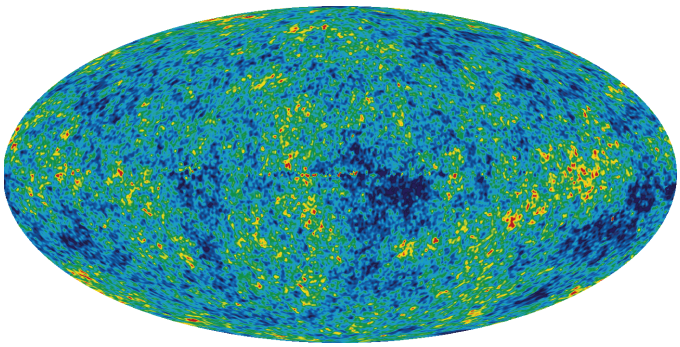


Figure 3: The best-fit combined WMAP anisotropic temperature fluctuations map. Courtesy of the LAMBDA WMAP site (http://lambda.gsfc.nasa.gov/).

in their statistical properties, as we will see in the following section.

1.5 The weird stuff

Today, we have a solid theoretical framework that effectively describes the dynamics of the universe from right after the Big Bang and up until today. Theories have a set of initial conditions, something that defines and tunes the physical effects. That is, the theory has a set of free parameters. As it

happens, it is possible to implement the equations that govern the universe numerically, and hence we can simulate what a CMB map should look like for various parameters. The statistical properties of the simulated data is then compared with the real data, and from this it one can determine matching theoretical parameters.

The current best-fit cosmological model is dubbed the ΛCDM model. Λ (Lambda) represents the cosmological constant that was introduced by Einstein, and describes a yet-unknown energy component that currently dominates the universe. This energy component is not directly visible, but its effect can be seen in several independent experiments (supernovae, WMAP). Since we cannot "see" the component, it is nicknamed "Dark Energy". From supernovae experiments and the WMAP experiment, we know that about 72 % of the stuff in the universe is dark energy. Dark energy has some interesting properties, for instance it possesses negative pressure, which explains the accelerated expansion of the universe. Another property is that it remains constant in space, so if you increase space-time, there will be more dark energy. On the other hand, CDM is an acronym for Cold Dark Matter, representing non-relativistic massive particles that do not interact electromagnetically. This means that CDM does not clump like ordinary matter, and is "invisible". About 23 % of the energy in the universe is Cold Dark Matter, and it has never been observed directly, but only indirectly. The remaining 4 % of the universe represents normal light, gas, stars and planets. Hopefully, dark matter will be confirmed at the CERN LHC experiment. If not, things surely will get interesting.

2 Modern cosmology

This section gives a quick and dirty summary of classic theoretical cosmology. The focus of this thesis has been the modern approaches of data analysis, and will be thoroughly dealt with in the subsequent chapters. In order to understand how the universe works, one first needs to be familiar with the equations that govern space-time and matter. The sentence from the previous section that describes how space-time is curved is called the Einstein's field equations, and looks like this;

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu}. \tag{1}$$

Here R is the Ricci scalar, and represents the curvature of space-time together with the Ricci tensor $R_{\mu\nu}$, while T the energy-momentum tensor and represents matter and energy. G is the gravitational constant, while $g_{\mu\nu}$ is the metric. It is impossible to solve the exact field equations for a general universe, so several constraints are needed. By assuming that matter in the universe is homogeneously (evenly spaced out) and isotropically (looks the

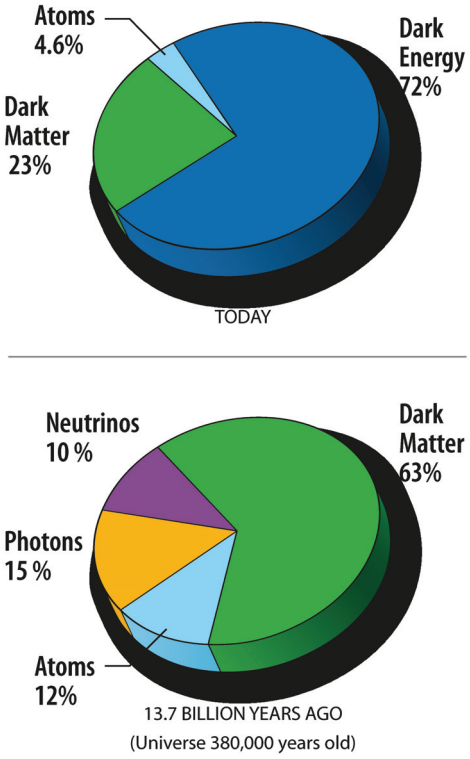


Figure 4: The general results from the WMAP experiment. Courtesy of the LAMBDA WMAP site (http://lambda.gsfc.nasa.gov/).

same in all directions) distributed on large scales, it is possible to develop the Friedmann-Robertson-Walker metric:

$$ds^2 = -dt^2 + a(t)^2\Big(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\Big) \tag{2}$$

where $d\Omega$ is a solid angle and k the curvature. This equation describes 3D space enclosed in 4D-space as either a spherical, hyperbolic or flat surface. When assuming that the energy components of the universe follow a non-viscous ideal fluid, the pressure is proportional to energy density, $p = w\rho$. Einsteins field equations then gives rise to the Friedmann equations:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{3}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\rho(1+3w) \tag{4}$$

where the Hubble parameter $H = \dot{a}/a$ measures the expansion rate of the universe. From these equations, it is possible to calculate the general properties of the universe by assuming various energy components such as radiation, matter and dark energy. For instance, if we assume the universe is dominated by a cosmological constant with $w = -1$, it is straight-forward to calculate that space-time will expand exponentially. This kind of model is called a de-Sitter universe, and we believe that the dominating energy component today is due to a cosmological constant. However, a cosmological constant-alike scenario might also have happened in the early stages of the universe, namely during inflation.

2.1 Inflation

There are some problems with the standard big bang model. For instance, we observe today a universe that is nearly geometrically flat, without there being any apparent reason for it to be so. A second problem is that the universe has inhomogeneities on large scales - a universe that started out perfectly homogeneous, should continue to stay so. And third, we observe the CMB to exhibit nearly equal temperature at every directions on the sky. But within the standard model of cosmology, these areas have never been in causal contact and have therefore no reason to have the same temperature. One solution to these problem is to introduce a supplementary model called inflation. If one assumes that there existed a scalar field in the very early universe that had some special properties similar to the cosmological constant, it is possible to get the universe to go through an extremely rapid expansion right after Big Bang. Even though this sounds a bit ad-hoc, there are some very good reasons to believe this to have happened:

- Inflation solves the problem of isotropy and temperature equality by postulating that the observable universe today was once squeezed well within the particle horizon, enabling causal contact.
- The geometrical flatness problem is also solved; any wrinkles in the early universe are smoothed by the extreme expansion of space.
- Inflation predicts that quantum fluctuations in the early universe were blown up to horizon-size perturbations, giving rise to inhomogeneities that seeded the structure growth during later ages.
- However, there are some major flaws with the theory. A scalar field has never been directly observed in nature. In addition, the energy levels needed to reproduce or detect the inflaton scalar field can only be reproduced during Big Bang energy densities.

One of the most promising candidates for indirect observation if inflation is via polarization CMB measurements. However, the signal will be very weak, and we will have to wait for future experiments in order to test the inflation hypothesis fully.

2.2 Three CMB Experiments

In 1989, the Cosmic Background Explorer (COBE) satellite was launched and became the first satellite to detect the anisotropies in the cosmic microwave background. This made it possible to perform comparison between cosmological data and early universe-theory. In addition, the FIRAS instrument showed a perfect fit of the CMB and the theoretical curve for a black body at a temperature of 2.7 K. However, the data were heavily contaminated by noise, and it was not possible to map out the angular power spectrum to scales higher than $\ell \sim 30$, corresponding to ~ 1 degrees. In 2001, the Wilkinson Microwave Anisotropy Probe (WMAP) [2, 15]) was launched. For eight years, it has been extensively mapping out the anisotropies in the Cosmic Microwave Background, both the temperature and polarization. The most important results from the experiment are depicted in Figure 4.

The PLANCK satellite was successfully launched in 2009, and is currently undertaking observations of the CMB anisotropies, both temperature and polarization. This experiment will provide us with an extremely detailed CMB data set, and will most likely conclude our investigation of the temperature fluctuations in the microwave background.

An image of the detail levels in the three different experiments are shown in Figure 5. However, working with PLANCK data requires much more finesse than previous CMB experiments. The resolution of the data is extremely high, and when performing likelihood analyses, computational costs are much higher. In addition, higher resolution implies more difficulties, as one needs to seriously take into account high-ℓ systematics. These include point sources, correlated noise and various foregrounds. A computational brute-force approach to solving these problems is therefore out of the question.

2.3 Anomalies in the CMB

From the concordance model of ΛCDM, the universe is supposed to be homogeneous on large scales. This is reflected in the cosmic microwave background, where the data should exhibit statistical isotropy. Measuring deviances from isotropy is therefore a good test to whether the ΛCDM is a good model for describing the universe. Since the WMAP data has been publicly available, there have been reported several anomalies that disagree with the assumption of isotropy. Here are the most important ones:

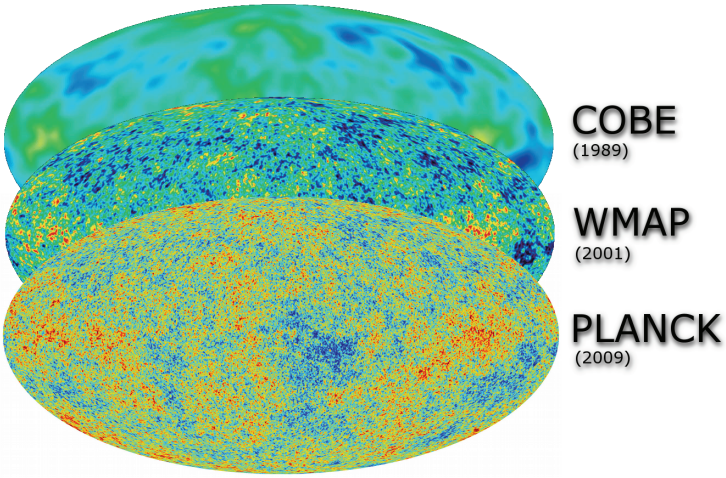


Figure 5: The three CMB experiments and their temperature maps: COBE (behind), WMAP (middle) and PLANCK (front).

- The $\ell = 2$ quadrupole amplitude is low compared to ΛCDM predictions.
- The $\ell = 2$ quadrupole and $\ell = 3$ octupole are both planar and strongly aligned.
- The north ecliptic hemisphere seems to have more of power than the southern, which contradicts isotropy.
- There exists "spots" in the CMB that are much colder than statistically likely withing the ΛCDM framework, and contradicts isotropy.
- There is a strong signal in the CMB that corresponds to a rotation feature. However, this is most likely due to systematic effects, and is one of the things I have been working on in my thesis.

3 CMB analysis

Theoretical physicists all over the world have been working on developing a mathematical framework for describing the evolution of various energy components in the universe from after inflation and up to today. These equations are commonly referred to as the Boltzmann equations, and are a set of coupled differential equations with free parameters. The free parameters tune

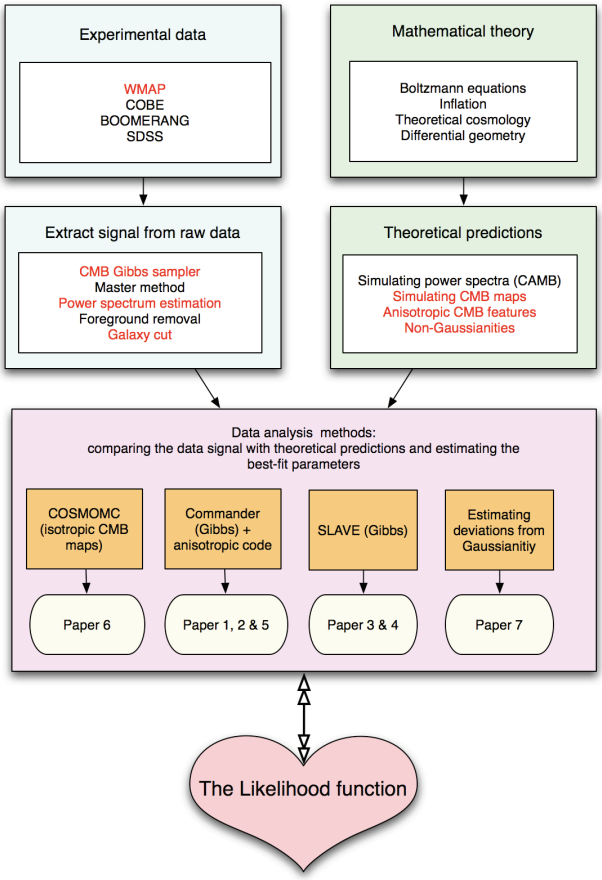


Figure 6: A flowchart of cosmological data analysis combined with my work. Red text highlights topics that are featured in my thesis.

the fractions of components in the universe, while the equations describe how they develop and behave. Examples of parameters are the dark matter component Ω_{CDM}, the dark energy Ω_Λ and the spectral index n_s which measures deviance from primordial scale invariance. Having a theoretical framework enables the possibility of forecasting, that is simulating specific data sets given a theory. The ultimate goal is then to compare the simulated data with experimental observations. This way, one can effectively sort out which theoretical solutions best fit the data. In CMB analysis, there al-

ready exists several Boltzmann codes (eg. CAMB [17]) that predict what an angular power spectrum should look like given the cosmological parameters. There also exist codes that uses WMAP data and compares with the theoretical framework through MCMC-methods (COSMOMC [18]).

3.1 Systematics

Statistically speaking, the only thing we can be sure is completely true when doing data analysis is the existence of the data itself. A theory might be flawed, but the data are always data. However, even if we have the data, we do not necessarily know what information it contains, because data are usually contaminated by uncertainties, errors and noise. These "bad" things are called systematics. When measuring the cosmic microwave background, there are many systematics that contaminate the data. The three most important systematics are as follows:

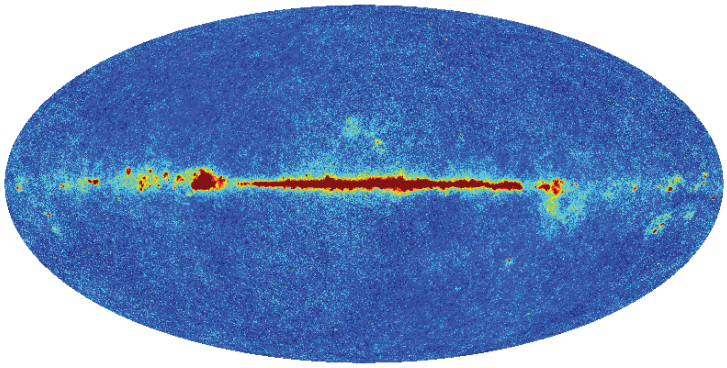


Figure 7: The Q1 WMAP raw data with various systematics: noise, the galaxy and foregrounds being the foremost.

1. Our galaxy covers about 25 % of the CMB map, and needs to be removed. The galaxy is clearly visible in Figure 7.

2. The WMAP instruments are not perfect, and much of the data visible in Figure 7 is noise.

3. There are many other closer objects out in space that produce unwanted effects in the CMB that needs to be removed, such as unidentified point sources (galaxies, clusters).

With a theoretical framework numerically implemented, it is possible to produce completely "perfect" simulations, free from these contaminations. However, in order to compare with real-life contaminated data, one needs to either contaminate the simulated data to resemble realistic data or "clean" the real data and obtain a estimated signal. Both these processes involve loss of information, and is schematically depicted in Figure 8. In this thesis, this estimation process is performed by the CMB Gibbs sampler.

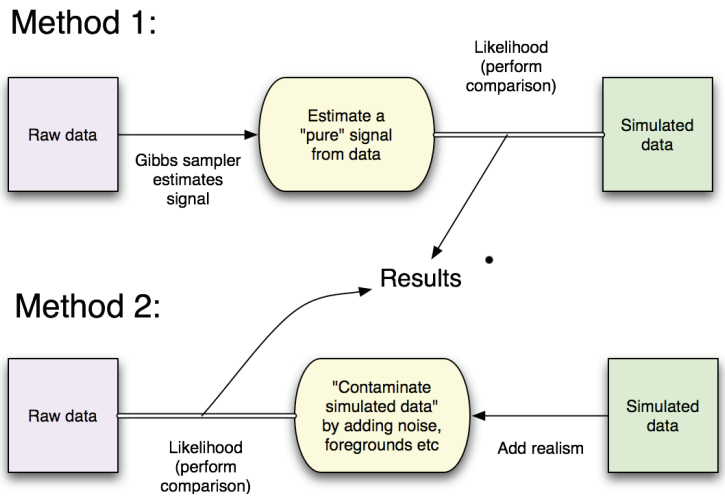


Figure 8: Two ways of comparing simulated data with experimental: either contaminate the simulated data, making it realistic, or try to estimate the "real" signal from the data.

3.2 The CMB Gibbs sampler

Data $\hat{\mathbf{d}}$ can be modeled as follows:

$$\hat{\mathbf{d}} = \mathbf{A}\mathbf{d} + \mathbf{n}, \tag{5}$$

where $\mathbf{d}$ is the signal we are interested in, $\mathbf{A}$ is the instrument beam and $\mathbf{n}$ noise. The standard algorithm today is called `MASTER` [16], and estimates the the power spectrum $\langle \hat{C}_\ell \rangle$ and the standard deviation ΔC_ℓ. However, this method is an "approximation" to a full likelihood that can be expressed as follows:

$$P(C_\ell|d) = \frac{1}{\sqrt{|\mathbf{C}|}} e^{-\frac{1}{2}\mathbf{d}^T(\mathbf{C})^{-1}\mathbf{d}}. \tag{6}$$

where $\mathbf{C}$ is the combined signal and noise covariance matrix. While it is fully possible to use MCMC-methods to sample from this distribution, the calculation of the $(\mathbf{C})^{-1}$-matrix scales as $O(N_{\rm pix}^3)$, where $N_{\rm pix}$ is the number of pixels in a map. In the isotropic case, the MASTER method is usually sufficient in order to estimate the C_ℓs, but it should be noted that it is only an approximation to a full likelihood analysis. Also, when investigating anisotropic universe models, one needs to work with full CMB maps. Calculating the likelihood from equation 6 using the $O(N_{pix}^3)$ method is impossible in the high-ℓ regime, and alternative solutions are needed. This is where the Gibbs sampler helps, by reducing the computational cost to a $O(N_{pix}^{1.5})$ operation in the uniform noise case.

The conditional distributions

The idea behind the CMB Gibbs sampler is to estimate the CMB sky, $\mathbf{s}$ and the angular power spectrum C_ℓ together with the covariance parameters by computing $P(\theta, \mathbf{s}|\mathbf{d})$, and then subsequently marginalize over $\mathbf{s}$ and C_ℓ. Specifically, the algorithm works as follows: first choose any initial guess, $(\theta, \mathbf{s})^0$. Then alternately sample from each of the *conditional* distributions,

$$\theta^{i+1} \leftarrow P(\theta|\mathbf{s}^i, \mathbf{d}) \tag{7}$$

$$\mathbf{s}^{i+1} \leftarrow P(\mathbf{s}|\theta^{i+1}, \mathbf{d}). \tag{8}$$

The theory of Gibbs sampling then guarantees that the joint samples $(\theta, \mathbf{s})^i$ will, after some burn-in period, be drawn from the desired joint distribution. The remaining step is then simply to formulate sampling algorithms for each of the two conditionals, $P(\theta|\mathbf{s}, \mathbf{d})$ and $P(\mathbf{s}|\theta, \mathbf{d})$. This method is depicted in Figure 3.2, and is computationally superior to conventional MCMC methods.

3.3 Gibbs sampling details

We saw in the previous section how the joint posterior $P(C_\ell, s|d)$ are broken into two conditional distributions $P(C_\ell|s, d)$ and $P(s|C_\ell, d)$. Gibbs sampling requires knowledge of these conditional distributions, and luckily, both are well-known. It can be shown that $P(C_\ell|s, d)$ follows an inverse Gamma function, and is rather straight forward to implement. However, sampling from $P(s|C_\ell, d)$ is more complicated, but it is possible to show that the requirements are met when s are drawn by the following rule:

$$s = (C^{-1} + N^{-1})^{-1}(N^{-1}\hat{d} + N^{-\frac{1}{2}}\omega_1 + C^{-\frac{1}{2}}\omega_2). \tag{9}$$

Here, ω_1 and ω_2 are independent random $N(0, 1)$ variables. This random map s has mean $\hat{s}$ and variance $(C^{-1} + N^{-1})$, exactly what we are interested in obtaining. Equation 9 can be broken into two separate parts:

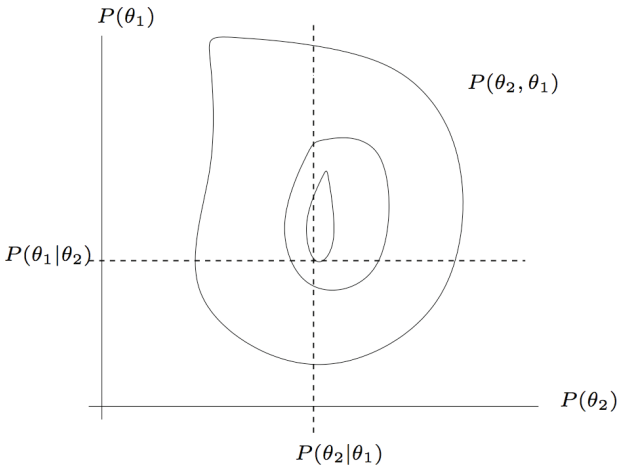


Figure 9: Conditional sampling implies alternating between sampling from $P(\theta_1|\theta_2)$ and $P(\theta_2|\theta_1)$, fixing the other parameter.

the Wiener filter $(C^{-1} + N^{-1})^{-1}(N^{-1}d)$ and the fluctuation map $(C^{-1} + N^{-1})^{-1}(N^{-\frac{1}{2}}\omega_1 + C^{-\frac{1}{2}}\omega_2)$. In Figure 10, each of these maps are depicted. The Wiener filter map estimates the fluctuations outside the galaxy sky cut, where they are heavily constrained by the known data given cosmic variance and noise. However, within the sky cut, only large-scale fluctuations are possible to statistically pin down, where small-scale fluctuations are repressed. The fluctuation map simulates the small-scale fluctuations within the unknown sky cut, and are constrained by cosmic variance and noise effects. Outside the sky cut, the fluctuation map is constrained by the data, allowing for only small variations in temperature fluctuations. The sum of these two maps make up the full CMB estimated signal sample.

3.4 The likelihood

Comparing a model with parameters θ with a CMB data set $\mathbf{d}$ is done by evaluating the likelihood function. Since the CMB is assumed to follow the distribution of a Gaussian field, the likelihood can be expressed as follows:

$$-2 \ln \mathcal{L} = \mathbf{d^T C}(\theta)^{-1}\mathbf{d} - \ln |\mathbf{C}(\theta)| \tag{10}$$

where $\mathbf{C}(\theta)$ is the model-based covariance matrix and $\mathbf{d}$ the data vector. As presented in Figure 8, it is usual to keep the data without systematics, and therefore estimate the signal data vector $\mathbf{d}$ from the original contaminated

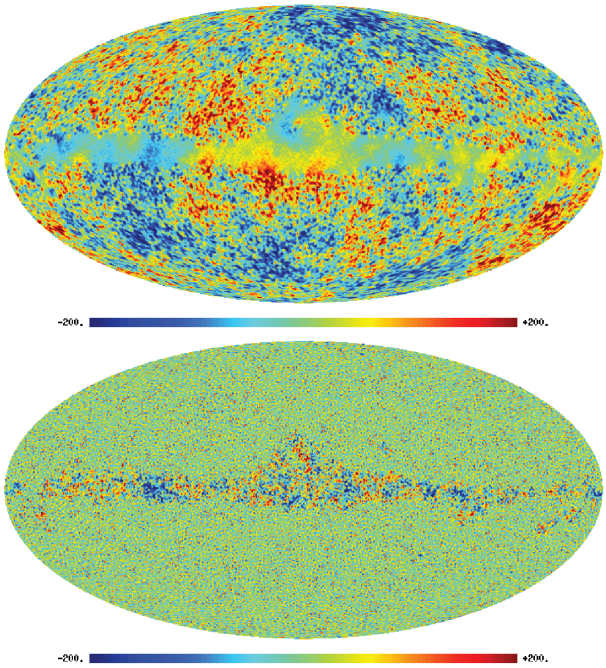


Figure 10: The two maps that together compose the full estimated signal: the fluctuation map (bottom) and the Wiener filter (top). Note that within the sky cut, the Wiener filter successfully estimates the large-scale structures while the fluctuation map produces random small-scale fluctuations.

data $\hat{\mathbf{d}}$. In this thesis, this estimation process is performed by the CMB Gibbs sampler.

3.5 The SLAVE framework

In order to uncover the details of the CMB Gibbs sampler, I started developing a software framework from scratch. The project was dubbed "Slave", and consists of two parts [12]. First, the Gibbs sampler itself, and then the post-processing tools for performing the necessary statistics. The details can be found in paper 4. Here, I present a quick summary of what the framework includes. As mentioned previously, the CMB Gibbs sampler alternates sampling the angular power spectrum C_ℓ from $P(C_\ell|\mathbf{d},\mathbf{s})$ and a signal $\mathbf{s}$ from $P(\mathbf{s}|\mathbf{d})$. Slave is run command-line using a parameter file that is described in paper 4, and instantly starts producing samples. An example of a set of

chains from the angular power spectrum C_ℓ estimation is presented in Figure 11. SLAVE outputs MCMC samples as text-based files. A post-processing utility is then run in order to round up all the samples and chains to a binary file, for efficiency and simpler file handling. All statistics is then performed using the combined binary file.

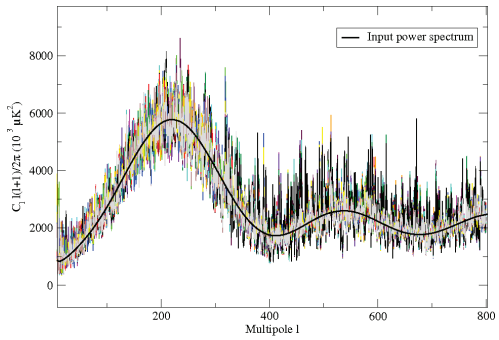


Figure 11: A set of C_ℓ samples together with the theoretical input power spectrum (black line). Note how the distribution is Γ for low-ℓ and Gaussian for high ℓ. Noise also contributes for high ℓ, increasing the power on smaller scales.

3.6 The Blackwell-Rao (BR) estimator

One of the most important uses of the Gibbs sampler is estimating the best-fit angular power spectrum C_ℓ^b from the samples of C_ℓs. The C_ℓs in the angular power spectrum for low ℓ follows an inverse Γ-distribution, and will for higher ℓs due to the central limit theorem converge to a Gaussian distribution. Therefore, it is not possible to simply calculate the average peak of the C_ℓ for each ℓ, and a better method is needed. One method is called the Blackwell-Rao estimator. By using prior knowledge of the distributions of the C_ℓs, we can build an analytical expression for the distribution for each C_ℓ given the signal power spectrum σ_ℓ, or $P(C_\ell|\sigma_l)$. Note that since the power spectrum only depends on the data through the signal and thus σ_ℓ, then

$$P(C_\ell \,|\, \mathbf{s}, \mathbf{d}) = P(C_\ell \,|\, \mathbf{s}) = P(C_\ell \,|\, \sigma_\ell). \tag{11}$$

It is therefore possible to approximate the distribution $P(C_\ell \,|\, \mathbf{d})$ as such:

$$P(C_\ell \,|\, \mathbf{d}) \approx \frac{1}{N_G} \sum_{i=1}^{N_G} P(C_\ell \,|\, \sigma_\ell^i) \tag{12}$$

where N_G is the number of Gibbs samples in the chain. This method of estimating the $P(C_\ell \,|\, \mathbf{d})$ is called the Blackwell-Rao estimator. Now, for a Gaussian field,

$$P(C_\ell \,|\, \sigma_\ell) \propto \prod_{\ell=0}^{\infty} \frac{1}{\sigma_\ell} \Big(\frac{\sigma_\ell}{C_\ell}\Big) e^{\frac{2\ell+1}{2}\frac{\sigma_\ell}{C_\ell}}. \tag{13}$$

Taking the logarithm, we obtain a nice expression:

$$\ln P(C_\ell|\sigma_l) = \sum \Big(\frac{2\ell+1}{2}\Big[-\frac{\sigma_\ell}{C_\ell} + \ln\Big(\frac{\sigma_\ell}{C_\ell}\Big)\Big] - \ln\sigma_l\Big) \tag{14}$$

which is straight-forward to implement numerically. The best-fit BR-estimated power spectrum is obtained by choosing the maximum likelihood value of C_ℓ for each ℓ. An example of the distribution of C_ℓs together with the BR-estimated likelihoods for various ℓs are shown in Figure 12. In paper

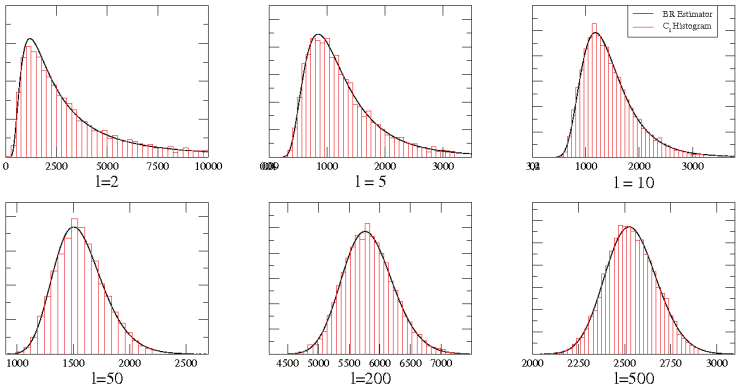


Figure 12: The histogram of the C_ℓs for various ℓ (red), together with the BR-estimated likelihood (black line). The best-fit power spectrum is obtained by choosing the peak value for each of the BR-estimated likelihoods.

5, we improved the standard Blackwell-Rao estimator by introducing a new likelihood approximation called the Gaussianized Blackwell-Rao (GBR) estimator [21]. This estimator is derived by transforming the observed marginal power spectrum distributions obtained by the CMB Gibbs sampler into standard univariate Gaussians, and then approximate their joint transformed distribution by a multivariate Gaussian. The method is exact for full-sky coverage and uniform noise, and an excellent approximation for sky cuts and scanning patterns relevant for WMAP and Planck data. This method is computationally superior to the previous BR-method. When we applied this tool to the 5-year WMAP temperature data and re-estimated the an-

gular temperature power spectrum, we obtained new cosmological parameters for the standard six-parameter ΛCDM model. The new spectrum was shown to be in excellent agreement with the official WMAP spectrum, but we also found slight differences in the cosmological parameters. Most importantly, the spectral index of scalar perturbations is $n_s = 0.973 \pm 0.014$, 1.9σ away from unity and 0.6σ higher than the official WMAP result, $n_s = 0.965 \pm 0.014$. This suggests that an exact likelihood treatment is required to higher l's than previously believed, reinforcing and extending our conclusions from the 3-year WMAP analysis.

3.7 Noise mis-estimation

Traditionally, the noise properties used in the Gibbs sampler have been assumed known to infinite precision. However, in paper 3 we relax this assumption and introduce a new free parameter, α, that scales the noise covariance matrix [10]. This way, the noise covariance matrix reads $N = \alpha\mathbf{N}^{\rm fid}$. Thus, if there is no deviation between the assumed and real noise levels, then α should equal 1.

The details

The full joint posterior, $P(\mathbf{s}, C_\ell, \alpha \,|\, \mathbf{d})$, now includes an additional amplitude parameter α. This is expressed as

$$P(\mathbf{s}, C_\ell, \alpha \,|\, \mathbf{d}) = P(\mathbf{d} \,|\, \mathbf{s}, \alpha) \cdot P(\mathbf{s}, C_\ell) \cdot P(\alpha) \tag{15}$$

where the first term is the likelihood,

$$P(\mathbf{d} \,|\, \mathbf{s}, \alpha) = \frac{e^{-\frac{1}{2}(\mathbf{d}-\mathbf{s})(\alpha\mathbf{N})^{-1}(\mathbf{d}-\mathbf{s})}}{\sqrt{|\alpha\mathbf{N}|}}, \tag{16}$$

the second term is a CMB prior, and the third term is a prior on α. Note that the latter two are independent, given that these describe two a-priori independent objects. We typically adopt a Gaussian prior centered on unity on α, $P(\alpha) \sim N(1, \sigma_\alpha^2)$. We normally choose a very loose prior, such that the posterior is completely data-driven. The conditional distribution for α can now be expressed as:

$$P(\alpha \,|\, \mathbf{s}, C_\ell, \mathbf{d}) \propto \frac{e^{-\frac{\beta}{2\alpha}}}{\alpha^{n/2}} \cdot P(\alpha) \tag{17}$$

where $n = N_{\rm pix}$ and $\beta = (\mathbf{d}-\mathbf{s})\mathbf{N}^{-1}(\mathbf{d}-\mathbf{s})$ is the χ^2. Note that the χ^2 is already calculated within the Gibbs sampler, as it is used to validate that the input noise maps and beams are within a correct range for each Gibbs iteration. Sampling from this distribution within the Gibbs sampler therefore represent a completely negligible extra computational cost. For

the Gaussian prior with unity mean and standard deviation σ_α, we find that

$$P(\alpha \,|\, \mathbf{s}, C_\ell, \mathbf{d}) \propto \frac{e^{-\frac{1}{2}\left(\frac{\beta}{\alpha} + \frac{(\alpha-1)^2}{\sigma_\alpha^2}\right)}}{\alpha^{n/2}} \tag{18}$$

For large degrees of freedom n, the inverse gamma function converges to a Gaussian distribution with mean $\mu = b/(k+1)$, where we have defined $k = n_{\rm pix}/2 - 1$, and variance $\sigma^2 = b^2/((k-1)(k-1)(k-2))$. A good approximation is therefore letting α_{i+1} be drawn from a product of two Gaussian distributions, which itself is a Gaussian, with mean and standard deviation

$$\mu = \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \tag{19}$$

$$\sigma = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{20}$$

This sampling step has been implemented in `SLAVE` and we have successfully tested it on simulated maps. With $N_{\rm side} = 512$ and $l_{\rm max} = 1300$ and full sky coverage, we find $\alpha = 1.000 \pm 0.001$. With such high resolution, the standard deviation on α is extremely low.

We have then applied this method to the 5-year WMAP data, and re-estimated the noise levels of both the raw and foreground-reduced sky maps. In doing so, we found that the predicted noise levels for the raw maps are in acceptable agreement with the predictions, while the noise levels in the foreground-reduced maps are $0.5 - 1.0\%$ higher than the estimate initially provided by the WMAP team on LAMBDA. The explanation for this effect has after the publication of paper 3 been found by the WMAP team simply to be an error in the results provided on LAMBDA: The quoted values were derived from the 3-year analysis instead of the 5-year analysis. However, the correct values were used in their cosmological analysis for the 5-year data, and no results are therefore compromised by this error. Thus, the method has already been demonstrated to be both accurate and useful on a practical example. Further, it carries virtually no extra computational cost within a Gibbs sampler, since all required quantities are already computed within this algorithm.

3.8 Conclusion

We have now concluded the statistical part of the thesis, corresponding to papers 3, 4 and 5. We have also explained some details about the CMB Gibbs sampler, and summarized the most important new features that we have developed. While the CMB Gibbs sampler has not previously been a necessity in CMB data analysis, new high-resolution data from from upcoming CMB experiments makes it vital to reduce computational cost. This is where the Gibbs sampler is superior to conventional MCMC methods. We now move

on to the physically based topics in this thesis, and start by reviewing the ACW rotational anomaly we detected in the WMAP data. This analysis would have been near-impossible to perform without the Gibbs sampler.

4 Anisotropic universe models

A firm prediction of inflation is that the observed universe should be nearly isotropic on large scales. The question of isotropy has received considerable attention during recent years, due to unexpected signatures observed in the WMAP sky maps. These data appear to exhibit several significant and distinct signatures of violation of statistical isotropy. First, de Oliveira-Costa et al. [4] found a striking alignment between the two largest harmonic modes in the temperature anisotropy sky, the quadrupole and the octupole. Second, Vielva et al. [23] pointed out the presence of a very large cold spot in the southern Galactic sky, apparently incompatible with ΛCDM-based simulations. Finally, Eriksen et al. [7] found a significantly anisotropic distribution of power between two hemispheres. The tools developed in the present paper may be able to constrain specific models relevant for these observations. In particular, we use these methods to estimate the anisotropy parameters in the ACW model from the 5-year WMAP temperature data.

Two specific examples are those presented by Ackerman et al. [1] (ACW) and Erickcek et al. [6]. The first model considers violation of rotational invariance in the early universe, while the second model describes the effects on the observed perturbation distribution due to a large-scale curvaton field.

The introduction of anisotropic models poses several problems in terms of data analysis. The definition of a proper likelihood function may be non-trivial for a general case, although many models can be described as multivariate Gaussians with non-diagonal covariance matrices. Both of the models mentioned above are examples of this. Yet, even in these relatively simple cases, the numerical evaluation of the likelihood is computationally unfeasible due to the sheer size of the relevant covariance matrix.

4.1 The ACW-model

ACW considered breaking of rotational invariance by generalizing the spectrum of primordial density perturbations $P(k)$ to include a preferred direction, $\hat{\mathbf{n}}$, as well as wave-number k,

$$P(\mathbf{k}) = P(k)(1 + g_*(\hat{\mathbf{k}} \cdot \hat{\mathbf{n}})^2). \tag{21}$$

Here g_* is a real number and $\hat{n} = \hat{n}(\theta, \phi)$. Further, the correlations in the coefficients $a_{\ell m}$ (the covariance matrix) was expressed as

$$\langle a_{\ell m} a^*_{\ell m} \rangle = C_{\ell m, \ell' m'} = C_\ell \delta_{\ell \ell'} \delta_{m m'} + g_* \Delta_{\ell m, \ell' m'}. \tag{22}$$

where the Δ is a geometric matrix containing Clebsch-Gordon coefficients, which breaks $SU(2)$-invariance. The Δ couples different modes:m to $m' \pm 1, 2$ and ℓ to $\ell \pm 2$, correlating "local" scales along a preferred axis. In other words, the ACW model predicts that the isotropic underlying temperature map contains an additional linear anisotropic contribution, which will "smear" out (correlate) small scales along the equatorial plane of a preferred direction. This example is illustrated in Figure 4.1.

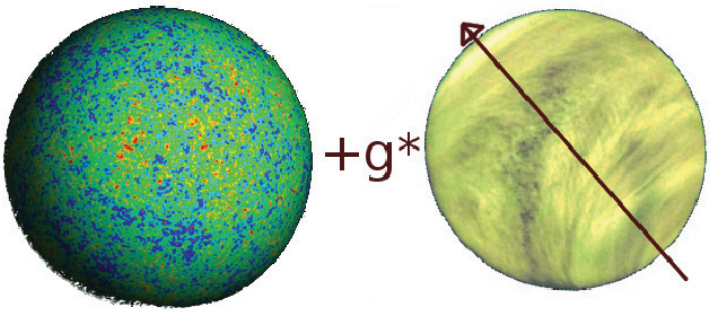


Figure 13: A rough example of the ACW-model: the underlying isotropic CMB map contains an additional linear structure with coupling g_*. This rotational structure mimic "clouds" that are stretched along the equatorial plane, while the poles contain less stretching.

4.2 Computational difficulties

The heart of the problem is calculating the expression $\mathbf{d}^T\mathbf{C}^{-1}\mathbf{d}$. To perform a Cholesky decomposition and solving this equation is mathematically straight-forward, but for the WMAP data, the angular resolutions are well beyond $\ell \sim 600$. The covariance matrix is in general a $\ell^2 \times \ell^2$ matrix, containing roughly 10^{11} elements. When working with strictly isotropic models, the covariance matrix is diagonal and without correlations, and the matrix operations are trivial, scaling as ℓ^2. The covariance matrix is not diagonal, but it is also not dense and has a well-defined shape in harmonic space. This is good news, and means that what needs to be done is implementing a framework for sparse matrices. When this was done, we performed several anisotropic analyses on simulated data. An example of a realization of the the anisotropic Δ-term can be seen in Figure 14, with a random direction $(\theta, \phi) = (1.0, 1.0)$. Circles are drawn on the rotational axis for illustrative purposes.

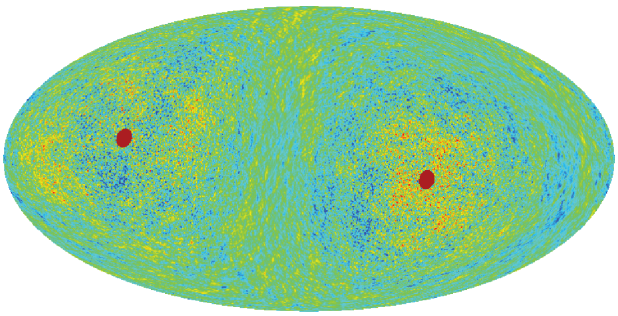


Figure 14: A realization of the ACW signal alone. Note the "smearing" of local modes in the equatorial plane, and the absence of smearing at the directional poles.

4.3 Anisotropic CMB Gibbs sampler flow

The next big step is to use the framework not only to produce simulated anisotropic data, but compare with experimental data. The first step was implementing the likelihood described in equation 10. The steps for developing a full code for analyzing real experimental data was as follows:

1. Do a brute-force grid analysis on a low-resolution ($N_{\mathrm{side}=32}$) simulated anisotropic map, and reproduce the original input parameters (g_*, θ, ϕ).

2. Upgrade the brute-force analysis to a MCMC method on the sphere. This requires the use of the Metropolis-Hastings algorithm, where (g_*, θ, ϕ) are chosen from an uniform cylinder in $\mathbb{R} \times \mathbb{S}^2$-parameter space.

3. Go to realistic resolutions such as ($N_{\mathrm{side}=512}$) and reproduce the correct input parameters.

4. Implement the anisotropic ACW MCMC sampler into the Gibbs framework "Commander" to enable sampling from maps that has noise, sky cut and foreground contaminations. The flowchart is shown in Figure 15.

5. Simulate "realistic" maps with correlated noise, WMAP instrumental beam and galaxy cut. Use the combined Gibbs/Anisotropic MCMC sampler to reproduce input parameters.

6. When simulated data is completely "realistic" and the sampler successfully reproduces the anisotropic input parameters, do "the old

switcheroo": replace the simulated data with real WMAP data.

Figure 15 depicts the general flow chart of the anisotropic Gibbs sampler used in paper 1 and 2.

4.4 Results and interpretations

The details of the implementation and analyses are described by [8] and [9]. The final results from all analyses show that there exist a 9σ presence of the ACW model in the 5-year WMAP W-band data. This is a rather strong detection, and definitely needed further investigation. However, the direction of the signal is located very close to the ecliptic poles, and is in alignment with the solar system. Therefore, the probability that this is a cosmological signal is minuscule, and is most likely a product of systematic effects from the satellite. However, the signal might have an impact on various cosmological parameters, so its origin is important to be resolved. We considered several systematics candidates that share similarities with the ACW signal, but have yet to find and source for the signal. Three types of systematics we have investigated are shown in Figure 16.

An independent analysis [14] also confirmed the anomaly, and claims that it is most likely not of cosmological origin. The WMAP team also acknowledged the existence of the signal in their 7-year papers [3], but do not discuss the matter further.

5 Early universe modifications

Another topic in my thesis has been quantum-gravity corrections to the equations governing early inflation. In paper 6, we investigate whether general quantum gravity corrections to the primordial power spectrum are visible in the CMB data [11]. The effects contribute to all angular scales, but exhibit a rather chaotic behavior. It can be shown that, assuming a cutoff scale Λ, the HZ scale-invariant primordial angular power spectrum can be expressed as:

$$P(\epsilon, \xi, k) = P_0(k)\Big(1 - \xi\Big(\frac{k}{k_n}\Big)^{-\epsilon} \sin\Big[\frac{2}{\xi}\Big(\frac{k}{k_n}\Big)^{\epsilon}\Big]\Big) \tag{23}$$

where ϵ and η are the slow roll parameters and k the wave number. This modification is due to quantum effects in a highly curved space, and is a first-order approximation only. The two slow roll parameters were implemented into CAMB, together with the modified primordial power spectrum. A plot of the corrections are shown in Figure 17. It should be clear that the effects are in general minuscule, and are mostly well within the cosmic variance. CosmoMC was then set up to include these parameters, and countless CPU hours were spent in order to find out whether traces of these effects are visible in the 3-year WMAP data.

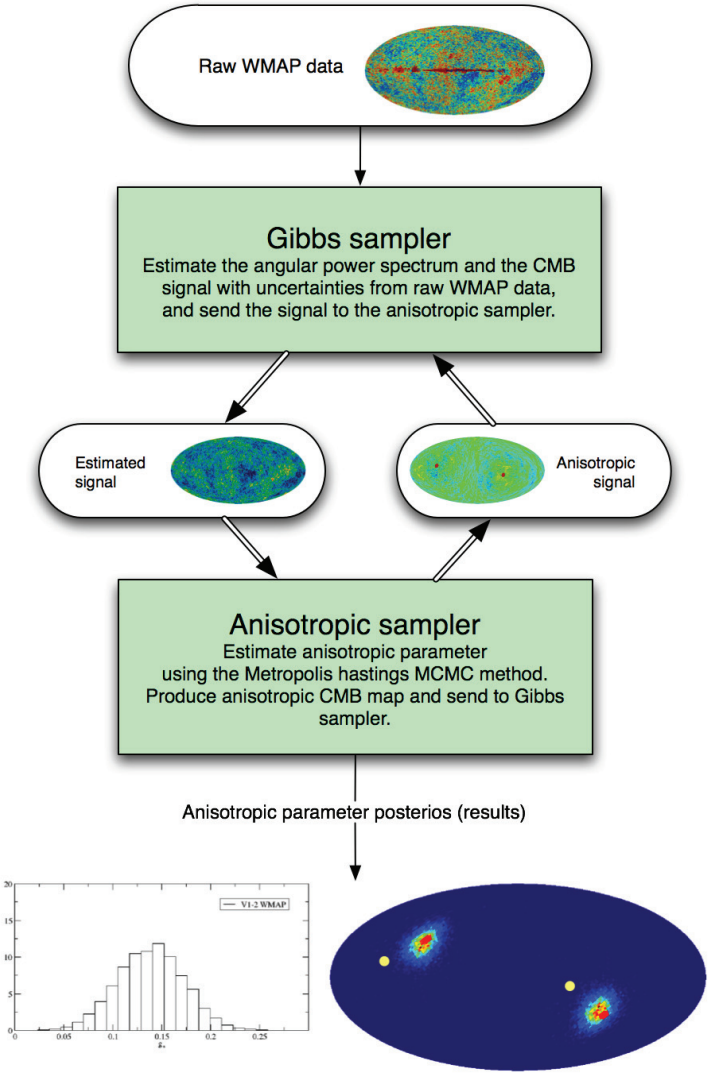


Figure 15: The flow chart of the anisotropic Gibbs sampler

5.1 Problems with the model..

However, there were some serious problems with the model. The modifications in general goes as $x \sin(\frac{1}{x})$, which shows a quite erratic behavior for

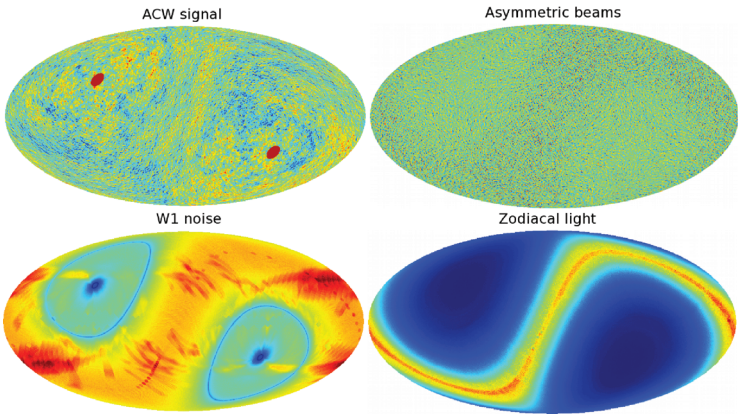


Figure 16: Systematics in the WMAP data that resembles the ACW model, but does not contribute.

low values of $x \to 0$. This is illustrated in Figure 17, together with the intrinsic errors in the ΛCDM power spectrum and the uncertainty of cosmic variance. On this case, even small variations in the parameters will result in major changes in the likelihood, and induce oscillations in the angular power spectrum. This way, the likelihood function ceases to be smooth and Gaussian, and becomes riddled with local minima. Chains transversing this likelihood landscape tends to get stuck in any of these local minima, and conventional MCMC methods break down. In order to illustrate this problem, we calculated a brute force two-dimensional likelihood surface of the two transplanckian parameters depicted in Figure 17. This extremely complex likelihood scenario, together with the overshadowing uncertainty of cosmic variance, makes the detection of these transplanckian effects near-impossible.

6 Alternative searches for non-Gaussianity

In paper 7 [22], we developed an independent framework for estimating deviations from Gaussianity in CMB data based on the methods established by [5] and [13]. The methods used are model-independent, and do not share any obvious connections with non-Gaussianity frameworks of physical origin. By counting the fraction of lakes, hills and saddles in simulated Gaussian maps while increasing the temperature threshold, we have built a distribution for what is expected for Gaussian maps. We then compared experimental data to this distribution, determining the deviation from the

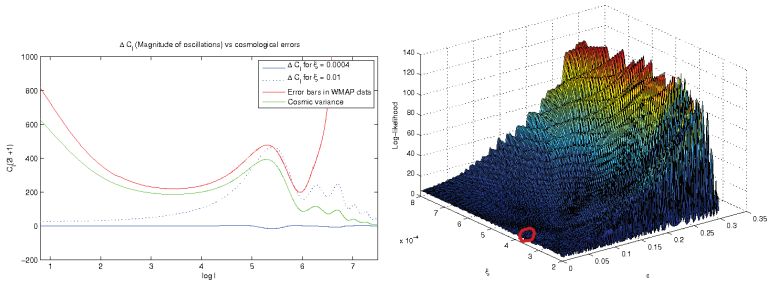


Figure 17: Left: The small transplanckian oscillations (blue), the WMAP error bars (red) and the cosmic variance (green). Right: the 2D likelihood surface of the trans-Planckian parameters. Note that this is riddled with local minima, and almost nearly impossible to transverse for MCMC chains.

Gaussian assumption. We then considered a combined V + W full-sky data set with the extended KQ85 and KQ75 mask, and found evidence of a 2σ deviation from Gaussianity on scales around 3°. For scales at 1° and 5°, we found only a 1σ presence. We also analyzed the north and south galactic/ecliptic hemispheres, but discovered no deviation from Gaussianity greater than 2σ. We went in-depth in order to investigate the 2σ signal, and analyzed the co-added Q band, V band and W band maps independently, and found that the deviation is equal in all frequencies. We continued by performing an analysis on each of the hemispheres centered around a pixel on a Healpix map with $N_{\text{side}} = 2$ using the combined V+W data. We produced directional maps for hills, lakes, saddles using the three scales, and found no evidence for a preferred direction in either of the maps.

Finally, we calculated the combined χ^2 from all our results, which resulted in an overall agreement with Gaussianity. It will be interesting to see whether any deviations from Gaussianity will be evident in the upcoming data from the PLANCK experiment.

7 Paper summary

In this section, I have summarized three years of my work in a convenient manner. A schematic description of the papers can be seen in Figure 6. Paper 1 and 2 are twin papers, and concerns a specific anisotropic universe model. Paper 3 and 4 dive into the field of CMB Gibbs sampling, and are connected. Paper 5 also concerns an advanced topic within Gibbs sampling, but the topic is more distant from the previous papers. In paper 6, we investigate whether traces of quantum gravity effect from inflation can be

seen in the WMAP data, while paper 7 concerns the possible existence of non-Gaussianity in the CMB.

Paper 1: Detection of a new anomaly in the WMAP data

We implement a numerical framework for testing whether traces of a specific anisotropic universe model is evident in the 5-year WMAP data. The model is named "ACW" and introduces a direction along which the CMB can rotate. When searching WMAP data for a possible trace of an ACW signal, we find a strong signature corresponding to a $3-4\sigma$ detection.

Paper 2: Continuing investigating the anomaly

We continue the investigation of the ACW signal in the 5-year WMAP data, and show that the signal is not affected by several types of systematics. However, after fixing an initial error in the model equations, the direction is now aligned with the ecliptic plane, or the plane of the satellite. The now 9σ signal is therefore most likely due to systematics, but we are unable to resolve which type of systematics.

Paper 3: A new method for estimating noise in CMB data

In this paper, we develop a new near cost-free method for estimating noise levels in CMB data when using the Gibbs sampler. When applied to the 5-year WMAP data using the SLAVE Gibbs sampler, we found that the noise levels in general were $0.5-1\%$ off. After the paper review, the WMAP team explained they had provided incorrect RMS levels, demonstrating the efficiency of the method.

Paper 4: Introducing SLAVE: A thorough guide to the CMB Gibbs sampler

This paper serves as a self-contained guide to the details of the CMB Gibbs sampler. I developed my own CMB Gibbs sampler in C++, and went through the whole process of deriving the conditional distributions and post-processing tools like the Blackwell-Rao power spectrum estimator.

Paper 5: An improved method for low-ℓ power spectrum estimation

In this paper, we developed an improved Blackwell-Rao method for obtaining the power spectrum using a Gibbs sampler for low ℓs. When applying this to cosmological parameter estimation, we find a minor 0.5σ shift in the spectral index n_s.

Paper 6: Can specific curvature effects from the early universe be seen in the CMB today?

Quantum gravity effects are expected to have modified the Bunch-Davies vacuum during inflation, and could have induced oscillations in the primordial power spectrum. We implemented the model in the CAMB framework, and searched for evidence in the 3-year WMAP data. However, we ultimately discovered that the oscillations are most likely too small and too chaotic to ever be observed in the CMB.

Paper 7: Estimating deviations from Gaussianity in the CMB by counting hills and lakes

Here, we use a well-known method for estimating deviations from statistical Gaussianity in CMB data by counting the fraction of hills, lakes and saddles on the sphere. By comparing results from Gaussian simulations, we are able to test deviances from assumed Gaussianity in the 5-year WMAP data. The results show that, when using a full covariance matrix, there really isn't much non-Gaussianities of this type in the WMAP data.

8 Conclusions

In this thesis, I have mainly been investigating whether traces of alternative universe models exist in the cosmic microwave background data. Surprisingly, we detected a 9σ anomaly in the WMAP data that resembles an ACW-rotational effect. However, we have also shown that the anomaly i most likely due to systematic effects, as it is aligned with the ecliptic plane, or the rotational axis of the satellite. An independent analysis performed by another group [14] confirmed the anomaly in WMAP data, and also claims that it is most likely not of cosmological origin. Even the WMAP team acknowledged the existence of the signal in their 7-year papers [3], but conclude that the effect is probably due to asymmetric beams.

We have not yet been able to explain the origin of the systematic effect, and this will be pursued in future work. Even though the signal contributes to about 5% of the anisotropies in the CMB, is uncertain whether the effect gives any contribution to cosmological parameters. New results also confirm the hypothesis that the effect is of systematic origin, as the signal is not evident in large-scale structure data [20].

We have also analyzed whether traces of trans-Planckian effects are visible in the angular power spectrum. However, there were some serious computational problems with this model, and it is currently not feasible to continue investigations of this kind. Also, we have analyzed whether the CMB fluctuations deviate from non-Gaussianity, but find only a very weak signal.

A large part of this thesis concerns the development of statistical tools for analyzing CMB data. In order to uncover the details of the CMB Gibbs sampler, I have developed a framework called "Slave" for estimating the CMB signal and angular power spectrum. During the development, we also implemented a new method for testing noise levels in CMB data. When applied to the 5-year WMAP data, we found a relatively small deviation in noise power from what is expected. During the review of the paper, it was discovered that the deviation was due to incorrect values provided by the WMAP team. We also developed a new method for low ℓ power spectrum estimation, and applied it to the 5-year WMAP data. We found a small shift in the spectral index, significant at 0.5σ. However, the most important result is the improvement of computational power needed to perform a full-scale likelihood analysis.

Bibliography

[1] Ackerman, L., Carroll, S. M., & Wise, M. B. 2007, Phys. Rev. D, 75, 083502

[2] Bennett, C. L., et al. 2003, The Astrophysical Journal Supplementary, 148, 1

[3] Bennett, C. L., et al. 2010, arXiv:1001.4758

[4] de Oliveira-Costa, A., Tegmark, M., Zaldarriaga, M., & Hamilton, A. 2004, Phys. Rev. D, 69, 063516

[5] Doré, O., Colombi, S., & Bouchet, F. R. 2003, Monthly Notices of the Royal Astronomical Society, 344, 905

[6] Erickcek, A. L., Kamionkowski, M., & Carroll, S. M. 2008, PRL, submitted, [arXiv:0806.0377]

[7] Eriksen, H. K., Hansen, F. K., Banday, A. J., Górski, K. M., & Lilje, P. B. 2004a, Astrophys. J., 609, 1198

[8] Groeneboom, N. E., & Eriksen, H. K. 2009, Astrophys. J., 690, 1807

[9] Groeneboom, N. E., Ackerman, L., Kathrine Wehus, I., & Eriksen, H. K. 2009, preprint [arXiv:0911.0150]

[10] Groeneboom, N. E., Eriksen, H. K., Gorski, K., Huey, G., Jewell, J., & Wandelt, B. 2009, The Astrophysical Journal Letters, 702, L87

[11] Groeneboom, N. E., & Elgarøy, Ø. 2008, Phys. Rev. D, 77, 043522

[12] Groeneboom, N. E. 2009, arXiv:0905.3823

[13] Hansen, F. K., Cabella, P., Marinucci, D. & Vittorio, N. 2004, ApJ, 607, L67

[14] Hanson, D., & Lewis, A. 2009, Phys. Rev. D, 80, 063004

[15] Hinshaw, G., et al. 2007, The Astrophysical Journal Supplementary, 170, 288

[16] Hivon, E., Górski, K. M., Netterfield, C. B., Crill, B. P., Prunet, S., & Hansen, F. 2002, Astrophys. J., 567, 2

[17] Lewis, A., Challinor, A., & Lasenby, A. 2000, Astrophys. J., 538, 473

[18] Lewis, A., & Bridle, S. 2002, Phys. Rev. D, 66, 103511

[19] Penzias, A. A., & Wilson, R. W. 1965, Astrophys. J., 142, 419

[20] Pullen, A. R., & Hirata, C. M. 2010, arXiv:1003.0673

[21] Rudjord, Ø., Groeneboom, N. E., Eriksen, H. K., Huey, G., Górski, K. M., & Jewell, J. B. 2009, Astrophys. J., 692, 1669

[22] Rudjord, Ø., Groeneboom, N. E., Hansen, F. K., & Cabella, P. 2010, arXiv:1002.1811

[23] Vielva, P., Martínez-González, E., Barreiro, R. B., Sanz, J. L., & Cayón, L. 2004, Astrophys. J., 609, 22

Part II

Papers

Paper I

Bayesian Analysis of Sparse Anisotropic Universe Models and Application to the Five-Year WMAP Data

Groeneboom, Nicolaas E.; Eriksen, Hans Kristian

APJ, Volume 690, Issue 2, pp. 1807-1819 (2009)

Paper II

Bayesian analysis of an anisotropic universe model: systematics and polarization

Groeneboom, Nicolaas E.; Ackerman, Lotty;
Kathrine Wehus, Ingunn; Eriksen, Hans Kristian
arXiv:0911.0150

Paper III

Bayesian Analysis of White Noise Levels in the Five-Year WMAP Data
Groeneboom, N. E.; Eriksen, H. K.; Gorski,
K.; Huey, G.; Jewell, J.; Wandelt, B.
APJL, Volume 702, Issue 1, pp. L87-L90 (2009)

Paper IV

A self-contained guide to the CMB Gibbs sampler
Groeneboom, N. E.
eprint arXiv:0905.3823

Draft version May 23, 2009
Preprint typeset using LaTeX style emulateapj v. 10/09/06

A SELF-CONTAINED GUIDE TO THE CMB GIBBS SAMPLER

Nicolaas E. Groeneboom[1]

(Dated: May 23, 2009)
Draft version May 23, 2009

ABSTRACT

We present a consistent self-contained and pedagogical review of the CMB Gibbs sampler, focusing on computational methods and code design. We provide an easy-to-use CMB Gibbs sampler named `SLAVE` developed in C++ using object-oriented design. While discussing why the need for a Gibbs sampler is evident and what the Gibbs sampler can be used for in a cosmological context, we review in detail the analytical expressions for the conditional probability densities and discuss the problems of galactic foreground removal and anisotropic noise. Having demonstrated that `SLAVE` is a working, usable CMB Gibbs sampler, we present the algorithm for white noise level estimation. We then give a short guide on operating `SLAVE` before introducing the post-processing utilities for obtaining the best-fit power spectrum using the Blackwell-Rao estimator.

Subject headings: cosmic microwave background — cosmology: observations — methods: numerical

1. INTRODUCTION

In recent years, increased resolution in the measurement of the cosmic microwave background (CMB) have driven the need for more accurate data analysis techniques. During the early years of CMB experiments, data was so sparse and noise levels so high that error bars in general overshadowed the observed signal. With the COBE experiment, (Smoot et al. 1992) posteriors were mapped out by brute force, and the statistical methods employed were simplistic. This was sufficient, as advanced statistical methods weren't needed for analyzing crude data. However, all this changed with the Wilkinson Microwave Anisotropy Probe (WMAP) experiment (Bennett et al. 2003; Hinshaw et al. 2007). Suddenly, cosmological data became much more detailed, vastly improving our knowledge of the universe, but also introduced new problems. Which parts of the signal were pure CMB, and which were not? The need for knowledge about instrumental noise, point sources, dust emission, synchrotron radiation and other contaminations were required in order to estimate the pure CMB signal from the data. And, how does one properly deal with the the sky cut, the contamination from our galaxy? Even harder, how does one maximize the probability that the resulting signal really is the correct CMB signal? A new era of cosmological statistics emerged.

An important event was the introduction of Bayesian statistics in cosmological data analysis. Bayesian statistics differs from the frequentist thought by quantizing ignorance: what one knows and not knows are intrinsic parts of the analysis. The goal of any Bayesian analysis is to go from the prior $P(\theta)$, or what is known about the model, to the posterior $P(\theta|\text{data})$, the probability of a model given data. This is summarized via Bayes' famous theorem:

$$P(\theta|\text{data}) = \frac{P(\text{data}|\theta)P(\theta)}{P(\text{data})}. \quad (1)$$

The posterior $P(\theta|\text{data})$ tells us something about how well a model θ fits the data, and is obtained by multiplying the prior $P(\theta)$, our assumption of the model, with the likelihood $P(\text{data}|\theta)$, the probability that the data fits the model.

The need for Bayesian statistics becomes evident when considering that we only have data from one single experiment to analyze. Bayesian statistics merges with frequentist statistics for large number of samples. And, in a cosmological context, we are stuck with only one sample, a sample that we are constantly measuring to higher accuracies. This sample is one realization of the underlying universe model, and we are unable to obtain data from another sample.

In a standard Metropolis-Hastings (MH) Monte Carlo Markov chain-approach (MCMC), one samples from the joint distribution by letting chains of "random walkers" transverse the parameter space. The posterior is obtained by calculating the normalized histogram of all the samples in the chains. The posterior will eventually resemble the underlying joint distribution, or the likelihood surface. This is a simple and easy-to-understand approach, but not without drawbacks. For one, each MH step is required to test the likelihood value of the chain at the current position in parameter space up against a new proposed position. Many of these steps will be rejected, and this is where the computational costs usually reside. The Gibbs sampler provides something new: one never needs to reject samples, and every move becomes accepted and usable for building the posterior. This is done by assuming that we have prior knowledge of the conditional distributions. These are then sampled from, each in turn yielding accepted steps.

However, the main motivation for introducing the CMB Gibbs sampler is the drastically improvement in scaling. With conventional MCMC methods, one needs to sample from the joint distribution, which results in an $\mathcal{O}(n^3)$ operation. For a white noise case, the Gibbs sampler splits the sampling process into independently sampling from the two conditional distributions, which together yields a $\mathcal{O}(n^{1.5})$ operation. In other words, the Gibbs sampler enables sampling the high-ℓ regime much more effective than previous MCMC methods.

Electronic address: leuat@irio.co.uk
[1] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, N-0315 Oslo, Norway

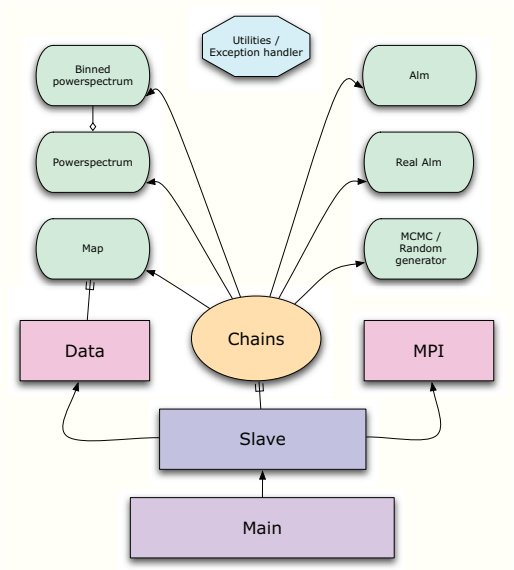


FIG. 1.— C++ class diagram of the SLAVE framework.

The problem of estimating the cosmological signal **s** from the full signal by Gibbs sampling was first addressed in Jewell et al. (2004), Wandelt et al. (2004) and Eriksen et al. (2004b). The ultimate goal of the Gibbs sampler is to estimate the CMB signal s from the data d, eliminating noise n, convolution A, all while including the sky cut. Today, a great number of papers have employed the Gibbs sampler since the introduction of the method (Eriksen et al. 2008a,b; Dunkley et al. 2008; Cumberbatch et al. 2009; Groeneboom et al 2008; Groeneboom et.al. 2009a; Eriksen et al. 2006; Rudjord et al. 2009; Jewell et al. 2009; Dickinson et al. 2007; Chu 2005; Dickinson et al. 2009; Larson et al. 2007).

In this paper, we review the basics of the CMB Gibbs sampler, and provide a simple, intuitive non-parallelized CMB Gibbs software bundle named SLAVE. SLAVE is written in C++, and employs object-oriented design in order to simplify mathematical implementation. The OOP design of SLAVE is presented in figure 1. For instance, assuming A, B and C are instances of the "real alm" class (they contain a set of real $a_{\ell m}$s), operator overloading enables us to directly translate the expression $A = (B + C)^{-1}$ by writing

```
A = (B+C).Invert();
```

This yields fast code that closely resembles equations, without having optimized too much for parallel computing, multiple data sets and other complexities.

1.1. *The Master algorithm*

One method of likelihood-estimator for obtaining the best-fit power spectrum for masked CMB data is given by the MASTER algorithm (Hivon et al. 2002). While Gibbs sampling estimates the full CMB signal s, the MASTER method only estimates the power spectrum. This method does not allow for variations in the estimated signal, except for the natural variations from simulating different realizations from the same power spectrum. However, the master algorithm estimates the power spectrum with cost scaling as $\mathcal{O}(n^3)$, which is slow for high-ℓ operations.

1.2. *What do I need the CMB Gibbs sampler for?*

Often, people misunderstand the concepts behind the CMB Gibbs sampler, and what the Gibbs sampler can be used for. In this section, we try to explain in simple terms when you should consider employing the CMB Gibbs sampler.

Assume that you have a theoretical universe model $M(\theta)$, where $\theta = \{\theta_i\}$ is a set of cosmological parameters. This model might give rise to some additional gaussian effects in the CMB map, either as fluctuations, altered power, anisotropic contributions, dipoles, ring structures or whatever. You now wish to test whether existing CMB data contains traces of your fabulous new model, and how significant those traces are. Or maybe you are just interested in ruling out the possibility that this model could be observed at all.

In any case, you need to implement some sort of numerical library that generates CMB maps based on your model. These maps will be "pure", in the sense that you have complete control over its generation process and systematics. Assume that your model has 1 free parameter. You could now loop over the 1-dimensional parameter space and calculate the χ^2 between a pure CMB signal map and the map from your model. This would have to be done for each step in parameter space, before obtaining the minimum. Even better, you could implement a Monte Carlo Markov chain framework, letting random walkers traverse a likelihood surface, yielding posteriors. This would enable support for a larger number of parameters, and is superior to the slow brute force approach.

In real-life however, things are not this simple. Data from any CMB experiment is contaminated by noise and foregrounds, most notably our own galaxy. This means that estimating the signal s from the data is not trivial - one needs to "rebuild", or make an assumption of what the fluctuations are within the sky cut and noise limits. This implies that it really isn't possible to obtain "the correct" CMB map, all we can know is that there exist a statistical range of validity where a simulated map agrees with the true CMB signal. Therefore, the consideration that that the estimated CMB signal s is a statistical random variable and not a fixed map should be included in the analysis. Hence, if you have implemented the MASTER method mentioned in section 1.1, you should test your model map against a set of realizations from the MASTER-estimated signal power spectrum.

This is where the Gibbs sampler enters the stage. As previously mentioned, the Gibbs sampler will estimate the CMB signal given data, and not only the power spectrum. The Gibbs sampler also ensures that every step in parameter space is always valid, so one never needs to discard samples. And even better, each of these independent steps provide an operation cost for obtaining samples that are much lower than more conventional MCMC methods. In order to test whether your model m fits the data, you therefore include the uncertainty in data by varying the signal. For example:

```
initialize Cl
do
  s = the CMB signal given the
       power spectrum Cl
  m = the CMB signal of your model given
       the estimated CMB signal s
  Cl = the CMB power spectrum given m
  save s, m and Cl
repeat until convergence
```

In the end, you calculate the statistical properties of s, m and Cl. Your model parameters have now been estimated, and the process included the intrinsic uncertainties in the signal. This method is not the most rapid - but it will always yield correct results.

2. THE CMB GIBBS SAMPLER

Throughout this paper, we assume that the data can be expressed as

$$d = As + n \tag{2}$$

where s is the CMB signal, A the instrument beam and n uncorrelated noise.

The `MASTER` algorithm estimates the the power spectrum $\langle \hat{C}_\ell \rangle$ and the standard deviation ΔC_ℓ. However, this method is a approximation to a full likelihood that can be expressed as follows:

$$P(C_\ell|d) = \frac{1}{\sqrt{|S+N|}} e^{-\frac{1}{2}d^T(S+N)^{-1}d}. \tag{3}$$

where S and N are the signal and noise covariance matrices, respectively. While it is fully possible to use MCMC-methods to sample from this distribution, the calculation of the $(S+N)^{-1}$-matrix scales as n^3, where n is the size of the $n \times n$ matrix. This is therefore an extremely slow operation, and is not feasible for large ℓs. If we demand that we sample the sky signal s as well, the joint distribution becomes $P(C_\ell, s|d)$. This might seem unnecessary complicated, as one most of the time doesn't need the signal s. But when feeding this distribution through the Gibbs sampler - that is, calculating the conditional distributions $P(C_\ell|s,d)$ and $P(s|C_\ell,d)$, we find that sampling from both are computationally faster than sampling from the full distribution in equation 3. The derivations of the conditional distributions are presented in section 3.

2.1. *Review of the Metropolis-Hastings algorithm*

The Gibbs sampler is a special case of the Metropolis-Hastings algorithm. We therefore review the basics of Monte Carlo Markov (MCMC) chain methods. The Metropolis-Hastings algorithm is a MCMC method for sampling directly from a probability distribution. This is done by letting "random walkers" transverse a parameter space, guided by the likelihood function, the probability that the data fits the model for the given parameter configuration. If a proposal step yields a likelihood greater than the current likelihood, then random walker accepts the step immediately. If the likelihood is less, then the walker will with a certain probability step "down" the likelihood surface. Eventually, the histogram of all the random walkers will converge to the posterior, the full underlying distribution.

Assume you have a model with n parameters, $\theta = \{\theta_k\}$ and you wish to map out a joint distribution from $P(\theta)$. Usually, one calculates the ratio R between the posteriors at the two steps $P(\theta^{i+1})$ and $P(\theta^i)$, such that

$$R = \frac{P(\theta^{i+1})}{P(\theta^i)} \cdot \frac{T(\theta^i|\theta^{i+1})}{T(\theta^{i+1}|\theta^i)} \tag{4}$$

where $T(\theta^i|\theta^{i+1})$ is the proposal distribution for going left or right. If the proposal distribution is symmetric (i.e. the probability of going left-right is equal for all θ_k), then $T(\theta^i|\theta^{i+1}) = T(\theta^{i+1}|\theta^i)$ such that:

$$R = \frac{P(\theta^{i+1})}{P(\theta^i)} \tag{5}$$

The MH acceptance rule now states: if R is larger than 1, accepted the step unconditionally. If $R > 1$, then accept the step if a random uniform variable $x = U(0,1) < R$.

2.2. *Review of the Gibbs algorithm*

Assume you have a model with two parameters, θ_1 and θ_2, and you wish to map out a joint distribution from $P(\theta_1, \theta_2)$. Now, also presume that you have prior knowledge of the conditional distributions, $P(\theta_1|\theta_2)$ and $P(\theta_2|\theta_1)$. A general proposal density is not necessary symmetric, and one must therefore consider the asymmetric proposal term as described in equation 4. However, we now define the proposal density T for θ_2 to be the conditional distributions:

$$T(\theta_1^{i+1}, \theta_2^{i+1}|\theta_1^i, \theta_2^i) = \delta(\theta_1^{i+1} - \theta_1^i) P(\theta_2^{i+1}|\theta_1^i). \tag{6}$$

In words, the proposal is only considered when $\theta_1^{i+1} = \theta_1^i$, which means that θ_1 is fixed while θ_2 can vary. If so, the acceptance is then given as the conditional distribution $P(\theta_2^{i+1}|\theta_1^i)$, which we must have prior knowledge of. The reason for choosing such a proposal density becomes clear when investigating the Metropolis Hastings acceptance rate:

$$R = \frac{P(\theta_2^{i+1}, \theta_1^{i+1})}{P(\theta_2^i, \theta_1^i)} \cdot \frac{T(\theta_1^i, \theta_2^i|\theta_1^{i+1}, \theta_2^{i+1})}{T(\theta_1^{i+1}, \theta_2^{i+1}|\theta_1^i, \theta_2^i)} \tag{7}$$

Using the conditional sampling proposal (6) one obtains

$$R = \frac{P(\theta_2^{i+1}|\theta_1^{i+1}) P(\theta_1^{i+1})}{P(\theta_2^i|\theta_1^i) P(\theta_1^i)} \cdot \frac{P(\theta_2^i|\theta_1^{i+1})}{P(\theta_2^{i+1}|\theta_1^i)} \frac{\delta}{\delta} \tag{8}$$

We now enforce the delta-function such that $\theta_1^{i+1} = \theta_1^i$. This sampling from the conditional distributions is the crucial step in the Gibbs sampler, such that all terms cancel out:

$$R = 1. \tag{9}$$

This implies that all steps are valid, and none are ever rejected. Hence one alternates between sampling from the known conditional distributions, where each step is independently accepted and can be performed as many times as needed.

3. THE CONDITIONAL DISTRIBUTIONS

In section 2.2, it was explained how the Gibbs sampler requires previous knowledge about the underlying conditional distributions. The CMB Gibbs sampler will alternate between sampling power spectra C_ℓ and CMB signal s, where each proposed step will always be valid. In order to enable sampling from the joint distribution,

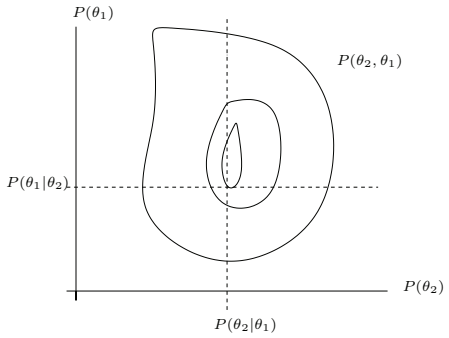


FIG. 2.— Conditional sampling implies alternating between sampling from $P(\theta_1|\theta_2)$ and $P(\theta_2|\theta_1)$, fixing the other parameter.

we therefore need to derive the analytical properties of the conditional distributions:

$$P(C_\ell|s,d) \qquad \text{and} \qquad P(s|C_\ell,d). \tag{10}$$

The derivations described here were first presented in Jewell et al. (2004), Wandelt et al. (2004) and Eriksen et al. (2004b).The full, joint distribution is expressed as

$$P(C_\ell,s|d) \propto P(d|C_\ell,s)P(C_\ell,s) \tag{11}$$
$$= P(d|C_\ell,s)P(s|C_\ell)P(C_\ell) \tag{12}$$

where $P(C_\ell)$ is a prior on C_ℓ, typically chosen to be flat. The first term, $-2\ln P(d|C_\ell,s)$, is nothing but the χ^2. The χ^2 measures the goodness-of-fit between model and data, leaving only fluctuations in noise. As $n = d - s$ is distributed accordingly to a Gaussian with mean 0 and variance N, we find that

$$P(d|C_\ell,s) \propto e^{-\frac{1}{2}(d-s)^t N^{-1}(d-s)}. \tag{13}$$

As we now assume that the signal s is known and fixed, the data d becomes redundant and $P(C_\ell|s,d) = P(C_\ell|s) \propto P(s|C_\ell)$. We therefore first need to obtain an expression for $P(C_\ell|s,d)$.

3.1. *Deriving* $P(C_\ell|s,d)$

Assuming that the CMB map consists of Gaussian fluctuations, we can express the conditional probability density for a power spectrum C_ℓ given a sky signal s as follows:

$$P(C_\ell|s,d) = \frac{e^{-\frac{1}{2}s^T C^{-1} s}}{\sqrt{|C|}} \tag{14}$$

where $C = C(C_\ell)$ is the covariance matrix. We now perform a transformation to spherical harmonics space, where $s = \sum_{\ell m} a_{\ell m} Y_{lm}$ and $C_{ij} = \sum_i \sum_j Y^i_{\ell' m'} C_{\ell' m',\ell m} Y^j_{\ell m}$. Then equation (14) transforms to

$$s^T C^{-1} s = \sum_{\ell m} \sum_{\ell' m'} a^*_{\ell m} Y^*_{\ell m} Y_{\ell' m'} C^{-1} Y^*_{\ell m} Y_{\ell' m'} a_{\ell' m'}. \tag{15}$$

As the spherical harmonics are orthogonal, they all cancel out and leave delta functions for $\delta_{\ell\ell'}\delta_{mm'}$ such that

$$s^T C^{-1} s = \sum_{\ell m} a^*_{\ell m} C^{-1}_\ell a_{\ell m} = \sum_{\ell m} a^*_{\ell m} \frac{1}{C_\ell} a_{\ell m}. \tag{16}$$

We now define a power spectrum $\sigma_\ell = \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2$ such that

$$s^T C^{-1} s = \sum_\ell (2\ell+1)\frac{\sigma_l}{C_\ell}. \tag{17}$$

Similarly, the determinant is given as the product of the diagonal matrix C, which for each l has $2\ell+1$ values of C_ℓ. The determinant is thus $|C| = \prod_\ell C_\ell^{2\ell+1}$. Expression (14) can now be written as

$$P(C_\ell|s) = \prod_\ell \frac{e^{-\frac{(2\ell+1)}{2}\frac{\sigma_\ell}{C_\ell}}}{\sqrt{C_\ell^{2\ell+1}}} \tag{18}$$

which by definition means that the $C\ell$'s are distributed as an inverse Gamma function. In the computational section, we will discuss how to draw random variables from this distribution.

3.2. *Deriving* $P(s|C_\ell,d)$

Again, we begin with the full, joint distribution:

$$P(C_\ell,s|d) \propto P(d|C_\ell,s)P(C_\ell|s). \tag{19}$$

We now know from equation 18 and 13 that the joint distribution can be expressed as

$$P(C_\ell,s|d) \propto e^{-\frac{1}{2}(d-s)^t N^{-1}(d-s)} \prod_\ell \frac{e^{-\frac{2\ell+1}{2}\frac{\sigma_\ell}{C_\ell}}}{C_\ell^{\frac{2\ell+1}{2}}} \tag{20}$$

omitting the prior $P(C_\ell)$. Again, note that it would be nearly impossible to sample directly from the full distribution. We now investigate what happens with equation 20 when C_ℓ becomes a fixed quantity. As the C_ℓs in the denominator vanishes, we use equation 14 to obtain

$$P(s|C_\ell,d) \propto e^{-\frac{1}{2}(d-s)^T N^{-1}(d-s)} e^{-\frac{1}{2}s^T C^{-1} s}. \tag{21}$$

We now introduce a residual variable $r = d - s$, such that r roughly consist of noise. As noise was uncorrelated, we can expect that r follows a Gaussian distribution with zero mean and N variance. Also, if s is known, then C_ℓ is redundant. We complete the square, and introduce $\hat{s} = (S^{-1} + N^{-1})^{-1} N^{-1} d$. Equation (21) can now be rewritten as

$$P(s|C_\ell,d) \propto e^{-\frac{1}{2}(s-\hat{s})^T (C^{-1}+N^{-1})(s-\hat{s})}. \tag{22}$$

Hence $P(s|C_\ell,d)$ is a Gaussian distribution with mean $\hat{s}$ and covariance $(C^{-1}+N^{-1})^{-1}$. In the computational section, we will discuss how to draw random variables from this distribution.

4. NUMERICAL IMPLEMENTATION

In its utter simplicity, the mechanics of the Gibbs sampler can be summarized as follows:

```
load data
initialize s and cl
loop number of chains
  s = generate from p(s | cl, d)
  cl = generate from p(cl | s, d)
  save s and cl
end loop
```

We now present the computational methods for drawing from $P(s|C_\ell,d)$ and $P(C_\ell|s,d)$.

4.1. $P(C_\ell|s,d)$

We show that equation 18 is an inverse Gamma distribution. A general gamma-distribution is proportional to

$$P_\Gamma(x;k,\theta) \propto x^{k-1}e^{-\frac{x}{\theta}}. \quad (23)$$

Equation 18 can be expressed as

$$P(C_\ell|s) = C_\ell^{-\frac{2l+1}{2}} e^{-\beta/C\ell} \quad (24)$$

where $\beta = \frac{2l+1}{2}\sigma_i$. If we now perform a substitution $y = 1/C_\ell$, we see that

$$P(y|s) = y^{\frac{2l+1}{2}} e^{-\beta y} \cdot y^{-2} \quad (25)$$

where the last term is the Jacobian. Hence

$$P(y|s) = y^{\frac{2l-1}{2}-1} e^{-\beta y} \quad (26)$$

which is a gamma-distribution for $k = \frac{2l-1}{2}$. We now show that this particular distribution also happens to be a special case of the χ^2 distribution:

$$\chi(x;k) = x^{k'/2-1}e^{-\frac{x}{2}}. \quad (27)$$

Letting $z = 2\beta y$ and ignoring the constants, we find that

$$P(z|s) = z^{k-1}e^{-z/2} \quad (28)$$

such that if $k' = 2k = 2l-1$, z is distributed according to a χ^2 distribution with $2l-1$ degrees of freedom. A random variable following such a distribution can be drawn as follows:

$$z_\chi = \sum_{i=0}^{2l-1} |N_i(0,1)|^2 \quad (29)$$

where $N_i(0,1)$ are random Gaussian variables with mean 0 and variance 1. Since $z = 2\beta y = 2\beta/C_\ell$, we find that

$$C_\ell = (2l+1)\sigma_i/z_\chi. \quad (30)$$

Numerically, one can implement this as

```
for each l
  z = 0
  for i=0 to 2l-1
    z = z+ rand_gauss()^2
  end
  C(l) = (2l+1)*sigma(l)/z
end
```

An example of this method can be found in the SLAVE libraries, within class "powerspectrum" method "draw_gamma".

4.2. $P(s|C_\ell,d)$

From equation 22, it is easy to see that $P(s|C_\ell,d)$ is a Gaussian distribution with mean $\hat{s}$ and variance $(C^{-1}+N^{-1})^{-1}$. Instead of deriving a method for drawing a random variable from this distribution, we present the solution and show that this solution indeed has the necessary properties (Jewell et al. 2004). Let

$$s = (C^{-1}+N^{-1})^{-1}(N^{-1}d + N^{-\frac{1}{2}}\omega_1 + C^{-\frac{1}{2}}\omega_2) \quad (31)$$

where ω_1 and ω_2 are independent, random $N(0,1)$ variables. We now show that the random variable s indeed has mean $\hat{s}$ and variance $(C^{-1}+N^{-1})$. First,

$$\langle s\rangle = (C^{-1}+N^{-1})^{-1}(N^{-1}\langle d\rangle + N^{-\frac{1}{2}}\langle\omega_1\rangle + C^{-\frac{1}{2}}\langle\omega_2\rangle). \quad (32)$$

As $\langle\omega_1\rangle = \langle\omega_2\rangle = 0$,

$$\langle s\rangle = (C^{-1}+N^{-1})^{-1}N^{-1}\langle d\rangle = \hat{s} \quad (33)$$

by definition.

The covariance is then

$$\langle (s-\hat{s})(s-\hat{s})^T\rangle. \quad (34)$$

Note that in the term $s-\hat{s}$, we have $(C^{-1}+N^{-1})^{-1}(N^{-1}d - N^{-1}d) = 0$, so we are only left with the terms with the random variables ω:

$$\langle (s-\hat{s})(s-\hat{s})^T\rangle = (C^{-1}+N^{-1})^{-2}\cdot \langle (N^{-\frac{1}{2}}\omega_1 + C^{-\frac{1}{2}}\omega_2)(\omega_1^T N^{-\frac{T}{2}} + \omega_2^T C^{-\frac{T}{2}})\rangle$$

But, as ω_1 and ω_2 are independently drawn from a $N(0,1)$ distribution, then $\langle\omega_i\omega_j\rangle = \delta_{ij}I$, and we end up with

$$\langle (s-\hat{s})(s-\hat{s})^T\rangle = (C^{-1}+N^{-1})^{-1} \quad (35)$$

which shows that a random variable drawn using equation 31 has the desired properties of being drawn from $P(s|C_\ell,d)$.

Having implemented a "real alm" class in SLAVE with operator overloading, it is possible to directly translate equation 31 into code:

```
omega1.gaussian_draw(0, 1, rng);
omega2.gaussian_draw(0, 1, rng);
calculate_CNI();
S = CNI* (NI*D + NI.square_root()*omega1
   + CI.square_root()*omega2);
```

where the code has been slightly optimized: both C^{-1}, N^{-1} and $(C^{-1}+N^{-1})^{-1}$ has been pre-calculated for efficiency. Note that this is only possible to do when assuming full-sky coverage with constant RMS noise. If the noise isn't constant on the sky, then N is a dense off-diagonal matrix, nearly impossible to calculate directly for large ℓ. However, it is still possible to perform the calculation in pixel space, but this requires that we assume N to be an operator instead of a matrix. We will address this issue in section 4.6.

We have now presented the main simplified Gibbs-steps for calculating $P(s|C_\ell,d)$ and $P(C_\ell|s,d)$, without convolution, uniform noise and no sky cut. Sampling from these two distributions is then done alternating between the two Gibbs steps, and the chain output - s and C_ℓ - are saved to disk during each step.

We now investigate the behavior of these fields, as each have special properties.

4.3. *Field properties*

Equation 31 can be broken into two separate parts: the Wiener filter $(C^{-1}+N^{-1})^{-1}(N^{-1}d)$ and the fluctuation map $(C^{-1}+N^{-1})^{-1}(N^{-\frac{1}{2}}\omega_1 + C^{-\frac{1}{2}}\omega_2)$. In figure 3, each of these maps are depicted. The Wiener filter map determines the fluctuations outside the sky cut - where they are heavily constrained by the known data, given cosmic variance and noise. However, within the sky cut, large-scale fluctuations are possible to pin down statistically while small-scales are repressed. The fluctuation map determines the small-scale fluctuations within the unknown sky cut, and are constrained by cosmic variance and noise effects. Outside the sky cut, the fluctuation map is constrained by the data, yielding very low small-scale fluctuations. The sum of these two parts make up the full CMB signal sample.

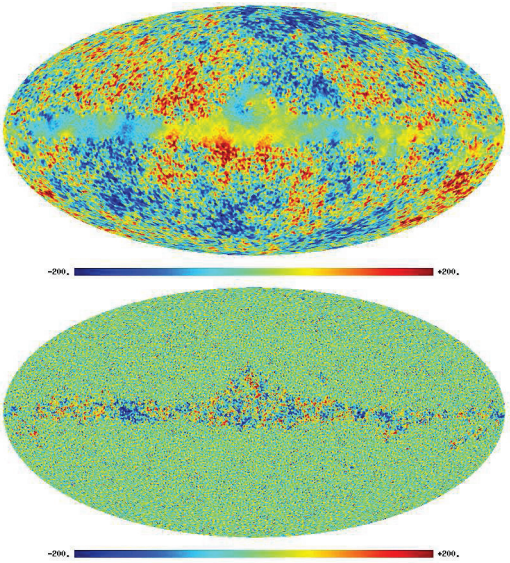


FIG. 3.— The two maps that together compose the full signal: the fluctuation map (bottom) and the Wiener filter (top). Note that within the sky cut, the Wiener filter successfully estimates the large-scale structures while the fluctuation map produces random small-scale fluctuations.

4.4. *Verifying the sample signal: the χ^2 test*

When the signal is being sampled, it is vital to check that the input parameters/data maps are correctly set up. For instance, if you use `SLAVE` to start a large job, say, estimating the CMB signal s for a $n_{pix} = 512$ map, it can be very frustrating when realizing that one of the input parameters were incorrect, for instance beam convolution or noise RMS. The software will continue to run without errors, but the resulting output files will be incorrect. We therefore adopt a simple and useful method for verifying that the estimated CMB signal s for each Gibbs step really is close to what one would expect.

The trick lies with the noise. As $d = As + n$, then $n = d - As$. Uniform white noise is assumed to be $N(0, \sigma^2_{\rm RMS})$-distributed, so

$$N(0,1) \sim \frac{d - As}{\sigma_{\rm RMS}}. \tag{36}$$

A χ^2 distribution is nothing but a sum of squared Gaussian distributions. Hence

$$\chi^2_{n_{\rm pix}} \sim \sum_{n_{\rm pix}} (\frac{d - As}{\sigma_{\rm RMS}})^2 \tag{37}$$

and the χ^2 should be close to the number of pixels in the map plus minus $\sqrt{2n}$. Usually, when an incorrect parameter is used, the χ^2 comes out far away from the expected value.

Calculating the χ^2 is not particularly time-consuming, but it has other uses as well: the χ^2 is used in the estimation of noise, as presented in section 5.

4.5. *Convolution*

A thing we did not address in the previous section was the inclusion of the instrumental beam convolution A. Including this in equation 31, we obtain

$$(C^{-1} + A^T N^{-1} A)s = AN^{-1}d + AN^{-\frac{1}{2}}\omega_1 + C^{-\frac{1}{2}}\omega_2. \tag{38}$$

In `SLAVE`, the beam is loaded directly from a fits file, or generated as a Gaussian beam given a full width half-maximum (FWHM) range. The beam is then multiplied with the corresponding pixel window, and stored in the $a_{\ell m}$-object A throughout the code.

4.6. *The sky cut*

Until now, we have only assumed full-sky data sets contaminated by constant noise. However, in order to be able to investigate real data, we need to take into account both the foreground galaxy and anisotropic noise. The galaxy contributes to almost 20% of the WMAP data, and needs to be removed with a mask. This means that the usable pars of the maps becomes anisotropic, giving rise to correlations in the spherical harmonics $a_{\ell m}$s. In other words, all the previously diagonal and well-behaved matrices now have off-diagonal elements, which for large $\ell_{\rm max}$ is an impossible feat to perform for dense matrices.

One way to get around these problems is to perform the calculations containing the sky cut mask in pixel space. This means that every time one needs to take into account the sky cut, one transforms from harmonic to pixel space, performs the operation including the sky cut before transforming back to harmonic space. While this operation in itself is trivial, equation 31 provides a few other problems:

$$(C^{-1} + A^T N^{-1} A)s = AN^{-1}d + AN^{-\frac{1}{2}}\omega_1 + C^{-\frac{1}{2}}\omega_2. \tag{39}$$

The right-hand side can easily be calculated, letting N^{-1} be an operator acting on d and ω_1, switching from spherical harmonics to pixel space and back. However, the left-hand side is troublesome - one cannot solve this equation explicitly. First, we need to rewrite 39 a bit:

$$(1 + C^{\frac{1}{2}} A^T N^{-1} A C^{\frac{1}{2}})(C^{-\frac{1}{2}} s) = \tag{40}$$

$$C^{\frac{1}{2}} A N^{-1} d + C^{\frac{1}{2}} A N^{-\frac{1}{2}} \omega_1 + \omega_2 = b \tag{41}$$

The first thing one should note about equation 41 is that the left-hand term is proportional to $(1 + S/N)$, where the diagonal parts are just the signal-to-noise ratios of the corresponding mode. Another nice feature about this form is that the variance of the signal is kept constant, that is, $\mathrm{Var}(s) \sim \ell^{-2}$, but $\mathrm{Var}(C^{-1/2}s) \sim I$. Hence we obtain better numerical stability. In order to solve the equation $(1 + S/N)x = b$, we implement a direct-from-textbook Conjugate Gradient (CG) algorithm presented on page 40 in Shewchuk (1994). The code looks like this:

```
b = L*( A*NI(D) + A*NI(map_work2,true)) + omega2;
MI = setup_preconditioner();
x = mult_by_A(x);
r =  b - x;
d = MI*r;
r0 = r.norm_L1(r);
do {
  Ad = mult_by_A(d);
  alpha = r.dot(MI*r) / (d.dot(Ad));
  x = x + d*alpha;
  rn = r - Ad*alpha;
```

```
    beta = rn.dot(MI*rn) / (r.dot(MI*r));
    d = MI * rn + d*beta;
    r = rn;
    norm = r.norm_L1(r);
  }
  while (norm>r0*epsilon);
  S = L*x;
```

C++ enables the CG algorithm to be translated almost directly from mathematical syntax to code. Here, the sky cut mask is taken into account in the NI-method - one only needs the mask when multiplying with the inverse noise matrix. The only other "initial condition" is the preconditioner. The preconditioner cannot affect the result, that is, it has nothing to do with the estimated signal s. The preconditioner only affects the number of iterations needed for the equation $Ax = b$ to be solved, and corresponds to a "best guess" of A. Without going into details, the standard preconditioner in `SLAVE` is proportional to $(1 + S/n)$, but there exists many other suggestions for better pre-conditioners, yielding quicker convergence. See Eriksen et al. (2004) or Smith et al. (2007) for more examples.

When the CG search has completed, the signal S has been obtained, including the sky cut and anisotropic noise.

4.7. *Low signal-to-noise regime*

A final thing we need to take into account is the low signal-to-noise regime. When the noise starts dominating the signal, the estimated s will fluctuate wildly on small scales. In addition, the deconvolution will add to this effect, blowing up noise to extreme values. In itself, this isn't a bad thing as we really cannot say exactly what is going in this regime, but it will affect the overall correlations between chains. In order to reduce this effect, we present a simple way to bin multipoles together on large l, reducing noise variance.

Let $N_\ell = \sigma^2_{\rm RMS} 4\pi/n_{pix}$ be the noise RMS in harmonic space. The variance is then given as

$$Var(N_\ell) = \frac{2}{2l+1} N_l^2. \tag{42}$$

For a single binned set with n multipoles ranging from $\ell_{\rm low}$ to $\ell_{\rm high}$, the average value of the power spectrum is given as

$$D_\ell = \frac{1}{n} \sum_{\ell_{\rm low}}^{\ell_{\rm high}} C_\ell. \tag{43}$$

Similarly for the noise power spectrum,

$$N_b = \frac{1}{n} \sum_{\ell_{\rm low}}^{\ell_{\rm high}} N_\ell. \tag{44}$$

Thus, the variance of the noise is given as

$$\sigma_N^2 = {\rm Var}(N_b) = \frac{1}{n^2} \sum_{\ell_{\rm low}}^{\ell_{\rm high}} {\rm Var}(N_\ell). \tag{45}$$

Obviously, σ_N is reduced as the number of multipoles in the bin n is increased. We now select bins such that the noise variance in a single bin is always less than three times the value of the angular power spectrum, or $\sigma_n < 3D_\ell$.

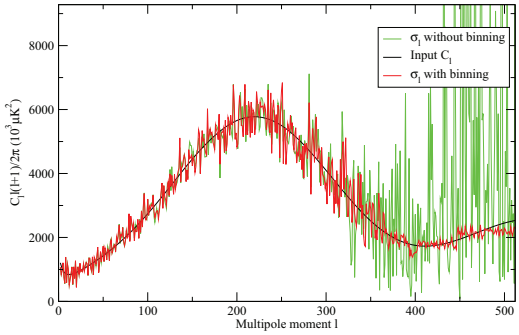


FIG. 4.— Examples of two estimated σ_ℓ without binning (green) and with binning (red). If the C_ℓs are produced from the binned σ_ℓs, the fluctuations in the low S/N-regime become less volatile. The input power spectrum is depicted in black.

The only affected part of the code is where one determines $P(C_\ell|s,d)$. Instead of generating a power spectrum C_ℓ given a set of σ_l, the calculation is now performed via a binning class that calculates the binned power spectrum C_b. That is,

$$P(C_b|\sigma) = \prod_{\ell_{\rm low}}^{\ell_{\rm high}} \left(\frac{e^{-\frac{2l+1}{2}\frac{\sigma_\ell}{C_b}}}{C_b^{\frac{2\ell+1}{2}}}\right). \tag{46}$$

Absorbing the product into the exponential, this becomes

$$P(C_b|\sigma) = \frac{e^{-\frac{1}{2C_b}\sum_\ell (2l+1)\sigma_\ell}}{C_b^{\frac{1}{2}\sum_\ell (2\ell+1)}}. \tag{47}$$

We now sample the signal with flat bins in $\ell(\ell+1)/(2\pi)$, not in ℓ.

5. GENERALIZING THE MODEL: NOISE ESTIMATION

In this section, we give a direct example of how one could extend the data model to the `SLAVE` Gibbs sampler. We derive the necessary conditional distribution, explain how this was integrated, and present some results from Groeneboom et.al. (2009a), where a full analysis of the noise levels in the WMAP data was performed using the `SLAVE` framework.

Traditionally, the noise properties used in the Gibbs sampler (e.g., Eriksen et al. 2004) have been assumed known to infinite precision. In this section, however, we relax this assumption, and introduce a new free parameter, α, that scales the fiducial noise covariance matrix, $N^{\rm fid}$, such that $N = \alpha N^{\rm fid}$. Thus, if there is no deviation between the assumed and real noise levels, then α should equal 1. The full analysis of the 5-yr WMAP data was presented in Groeneboom et.al. (2009a), with interesting results. For the foreground-reduced 5-year WMAP sky maps, we find that the posterior means typically range between $\alpha = 1.005 \pm 0.001$ and $\alpha = 1.010 \pm 0.001$ depending on differencing assembly, indicating that the noise level of these maps are underestimated by 0.5-1.0%. The same problem is not observed for the uncorrected WMAP sky maps.

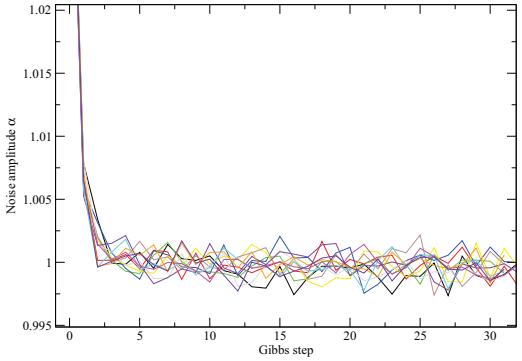


FIG. 5.— Even when assuming a large initial value, the noise amplitude α will quickly converge to the correct value.

The full joint posterior, $P(s, C_\ell, \alpha \,|\, d)$, now includes the amplitude α. We can rewrite this as follows:

$$P(s, C_\ell, \alpha \,|\, d) = P(d \,|\, s, \alpha) \cdot P(s, C_\ell) \cdot P(\alpha) \tag{48}$$

where the first term is the likelihood,

$$P(d \,|\, s, \alpha) = \frac{e^{-\frac{1}{2}(d-s)(\alpha N)^{-1}(d-s)}}{\sqrt{|\alpha N|}}, \tag{49}$$

the second term is a CMB prior, and the third term is a prior on α. Note that the latter two are independent, given that these describe two a-priori independent objects. In this paper, we adopt a Gaussian prior centered on unity on α, $P(\alpha) \sim N(1, \sigma_\alpha^2)$. Typically, we choose a very loose prior, such that the posterior is completely data-driven.

The conditional distribution for α can now be expressed as

$$P(\alpha \,|\, s, C_\ell, d) \propto \frac{e^{-\frac{\beta}{2\alpha}}}{\alpha^{n/2}} \cdot P(\alpha) \tag{50}$$

where $n = N_{\rm pix}$ and $\beta = (d-s)N^{-1}(d-s)$ is the χ^2. (Note that the χ^2 is already calculated within the Gibbs sampler, as it is used to validate that the input noise maps and beams are within a correct range for each Gibbs iteration. Sampling from this distribution within the Gibbs sampler represent therefore a completely negligible extra computational cost.) For the Gaussian prior with unity mean and standard deviation σ_α, we find that

$$P(\alpha \,|\, s, C_\ell, d) \propto \frac{e^{-\frac{1}{2}(\frac{\beta}{\alpha} + \frac{(\alpha-1)^2}{\sigma_\alpha^2})}}{\alpha^{n/2}} \tag{51}$$

For large degrees of freedom, n, the inverse gamma function converges to a Gaussian distribution with mean $\mu = b/(k+1)$, where we have defined $k = n_{\rm pix}/2 - 1$, and variance $\sigma^2 = b^2/((k-1)(k-1)(k-2))$. A good approximation is therefore letting α_{i+1} be drawn from a product of two Gaussian distributions, which itself is a Gaussian, with mean and standard deviation

$$\mu = \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \tag{52}$$

$$\sigma = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{53}$$

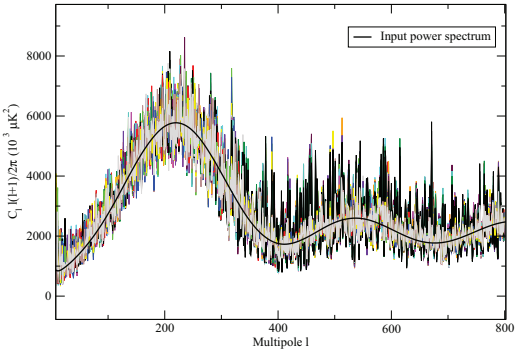


FIG. 6.— A typical plot of the C_ℓs obtained from a SLAVE run. Note that the input power spectrum is presented in black, and that the noise RMS for this particular run is very low.

This sampling step has been implemented in SLAVE and we have successfully tested it on simulated maps. With $N_{\rm side} = 512$ and $l_{\rm max} = 1300$ and full sky coverage, we find $\alpha = 1.000 \pm 0.001$. The chains for the noise amplitude α are shown in figure 5. Note that with such high resolution, the standard deviation on α is extremely low, and any deviation from the exact $\alpha = 1.0$ will be detected.

6. RUNNING SLAVE

In this section, we quickly review how to use SLAVE. For a more detailed usage, please see the SLAVE documentation (when the framework will be released).

SLAVE requires the HEALPIX (Górski et al. 2005) CXX-libraries installed. Please see the HEALPIX documentation on this topic. SLAVE is run command-line, and requires a parameter file as command-line parameter. The most important options in the parameter file are listed in table 1.

6.1. *Post-processing*

After the Gibbs sampler has been cooking for a while, it is time to investigate the results. The main output of SLAVE are the estimated power spectra C_ℓ's and the signals s. However, as the signal is assumed to be statistically isotropic, we instead output the signal power spectra σ_ℓ defined as:

$$\sigma_\ell \equiv \frac{1}{2\ell+1} \sum_{m=-\ell}^{m=\ell} |s_{\ell m}|^2. \tag{54}$$

The text-files may be plotted directly through software such as XMGRACE, as presented in figure 6. In addition, SLAVE outputs the σ_ℓ's as a binary file for each chain. These binary files can be combined through the main post-processing software utility for SLAVE called SLAVE_PROCESS. This software will combine the binary chains into a single file, in addition to removing burn-in samples. To combine the sigmas into one file, type

```
slave_process 1 [no_chains] [no_samples]
                [burnin] [output sigma_l file]
```

TABLE 1
SLAVE PARAMETER TABLE

General parameters		
seed	int	Initial random seed
verbosity	int	Text output level (0=none)
healpix_dir	string	HEALPIX home directory
output_sigmas	bool	Output σ_ℓ or not
output_cls	bool	Output C_ℓs or not
output_directory	string	Output file directory
output_chisq	bool	Output the χ^2 or not
output_beam	bool	Output the beam or not
output_beam_file	string	Beam output filename
Operations		
method	string	Analysis type: brute force_fullsky or CG (normal)
CG_convergence	double	CG Convergence criteria (type 10^{-6})
preconditioner	string	Pre-conditioner type: none, static or 3j
init_powerspectrum_power	double	Initialized flat power spectrum value
init_powerspectrum_use_file	bool	Use file instead of flat power spectrum
init_powerspectrum_file	string	Initial power spectrum file
samples	int	Number of Gibbs samples to produce
burnin	int	Number of burn-in samples to reject
Data		
datasets	int	Number of data sets (only 1 allowed yet..)
data_nsideN	int	$n_{\rm side}$ for data set $N = \{1, 2, 3, \dots\}$
data_mapN	string	FITS map for data set $N = \{1, 2, 3, \dots\}$
data_rmsN	string	FITS rms map for data set $N = \{1, 2, 3, \dots\}$
data_maskN	string	FITS mask for data set $N = \{1, 2, 3, \dots\}$
beam_fileN	string	FITS beam for data set $N = \{1, 2, 3, \dots\}$
lmax	int	$\ell_{\rm max}$ for the analysis
constant_rms	bool	Use constant rms or not
constant_rms_value	double	Value of constant rms
gaussian_beam	bool	Use a Gaussian beam or not
gaussian_beam_fwhm	double	Value of Gaussian beam
Noise estimation parameters		
enable_noise_amplitude_sampling	bool	Enable noise estimation or not
noise_sampling_sigma	double	The noise prior sigma
noise_amplitude_filename	string	Output noise filename
noise_alpha_init_val	double	Initial value for α
Binning		
use_binning	bool	Enable binning of power spectrum
binning_powerspectrum	string	Power spectrum used for binning
bins_filename	string	Text output the bins

NOTE. — The SLAVE parameter names and usage may have changed when the first version is released.

6.2. *C_ℓ likelihoods*

The first important step is to verify that the output C_ℓs follow the desired inverse-Gamma distribution for low ℓ, but converges to Gaussians for larger ℓ. The SLAVE processing utility SLAVE_PROCESS can generate a set of C_ℓs from the σ_ℓs and output the corresponding values for a single ℓ. It is then straight-forward to use a graphical utility such as XMGRACE to obtain the histogram. Such histograms are plotted together with the analytical likelihoods in figure 7. Note the good match between the histogram of the C_ℓs and the likelihoods obtained from the Blackwell-Rao estimator. The analysis for producing these plots was performed on simulated high-detail data, in order to verify the validity of the BR-estimator.

To save the cls for a specific ℓ, type

```
./process 4 [sigma_l file] [l] [generate no cls]
          [output textfile]
```

6.3. *The Blackwell-Rao estimator*

Our primary objective is obtaining the best-fit power spectrum from the estimated signal power spectra. If the C_ℓs were completely distributed according to a Gaussian, one would only need to select the maximum of the distribution for each C_ℓ. However, as we saw in equation 18, this is not the case, and we need a better way to obtain the likelihood $\mathcal{L}(C_\ell)$ for each ℓ.

Luckily, we can obtain an analytical expression of the likelihood for the C_ℓs via the Blackwell-Rao (BR) estimator, as presented in Chu et al. (2005). By using prior knowledge of the distributions of the C_ℓs, we can build an analytical expression for the distribution for each C_ℓ given the signal power spectrum σ_ℓ, or $P(C_\ell|\sigma_l)$.

Note that since the power spectrum only depends on the data through the signal and thus σ_ℓ, then

$$P(C_\ell \,|\, s, d) = P(C_\ell \,|\, s) = P(C_\ell \,|\, \sigma_\ell). \qquad (55)$$

It is therefore possible to approximate the distribution

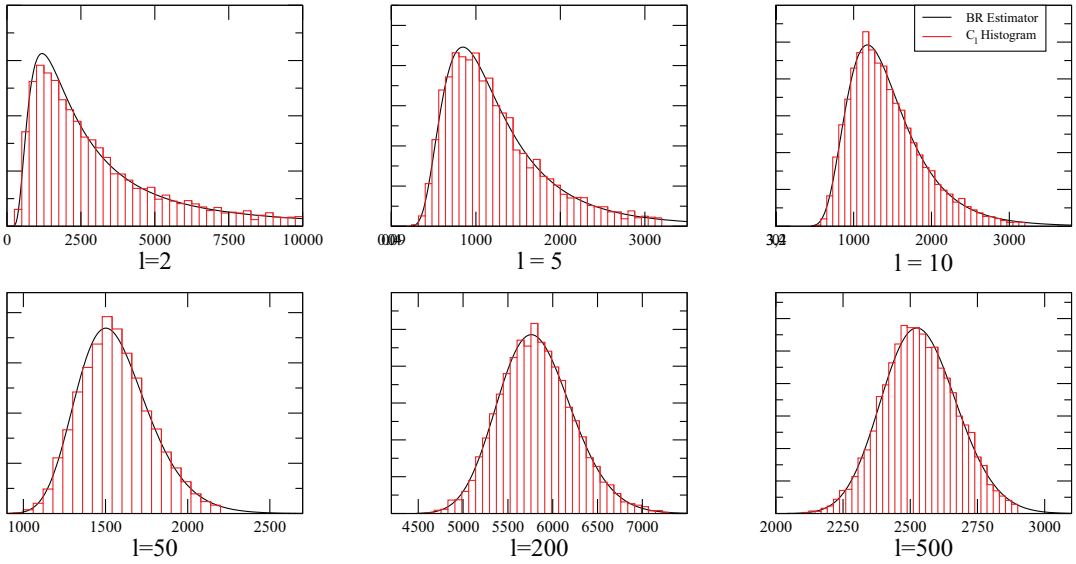


FIG. 7.— The histograms of the C_ℓs (red) and the BR-estimated likelihoods (black) for various ℓ. Note how the distribution converges to a Gaussian for larger multipoles ℓ. The analysis has been performed on simulated WMAP-like data.

$P(C_\ell \,|\, d)$ as such:

$$P(C_\ell \,|\, d) = \int P(C_\ell, s \,|\, d)\, ds \tag{56}$$

$$= \int P(C_\ell \,|\, s, d) P(s \,|\, d)\, ds \tag{57}$$

$$= \int P(C_\ell \,|\, \sigma_\ell) P(\sigma_\ell \,|\, d)\, D\sigma_\ell \tag{58}$$

$$\approx \frac{1}{N_G} \sum_{i=1}^{N_G} P(C_\ell \,|\, \sigma_\ell^i) \tag{59}$$

where N_G is the number of Gibbs samples in the chain. This method of estimating the $P(C_\ell \,|\, d)$ is called the Blackwell-Rao estimator. Now, for a Gaussian field,

$$P(C_\ell \,|\, \sigma_\ell) \propto \prod_{\ell=0}^{\infty} \frac{1}{\sigma_\ell} \Big(\frac{\sigma_\ell}{C_\ell}\Big) e^{\frac{2\ell+1}{2} \frac{\sigma_\ell}{C_\ell}}. \tag{60}$$

Taking the logarithm, we obtain a nice expression

$$\ln P(C_l|\sigma_l) = \sum \Big(\frac{2\ell+1}{2} \Big[-\frac{\sigma_\ell}{C_\ell} + \ln\Big(\frac{\sigma_\ell}{C_\ell}\Big) \Big] - \ln\sigma_l \Big) \tag{61}$$

which is straight-forward to implement numerically. To output the BR-estimated likelihood for one ℓ, type

```
./process 3 [sigma_l file] [l]
            [output likelihood]
```

6.4. *Power spectrum estimation*

The best-fit BR-estimated power spectrum is obtained by choosing the maximum likelihood value of C_ℓ for each ℓ. To do so, type

```
./process 2 [sigma_l file]
            [output power spectrum file]
```

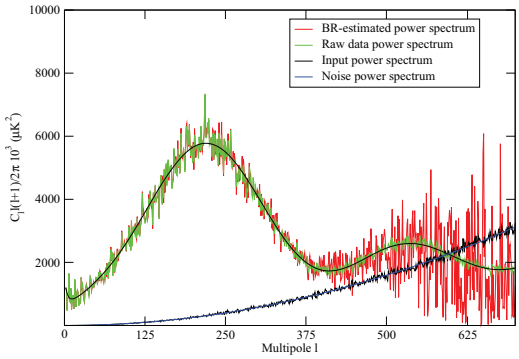


FIG. 8.— The BR-estimated power spectrum (red) versus the simulated input data power spectrum (green). Note that these two power spectra agree on large scales. The noise power spectrum is also shown (blue).

An example of a BR-estimated power spectrum can be seen in figure 8. In addition, both the input-and noise power spectra are shown. Note how the BR-estimated power spectrum is exact on small scales (low ℓ), while the convolution and noise dominated on higher scales.

7. CONCLUSIONS

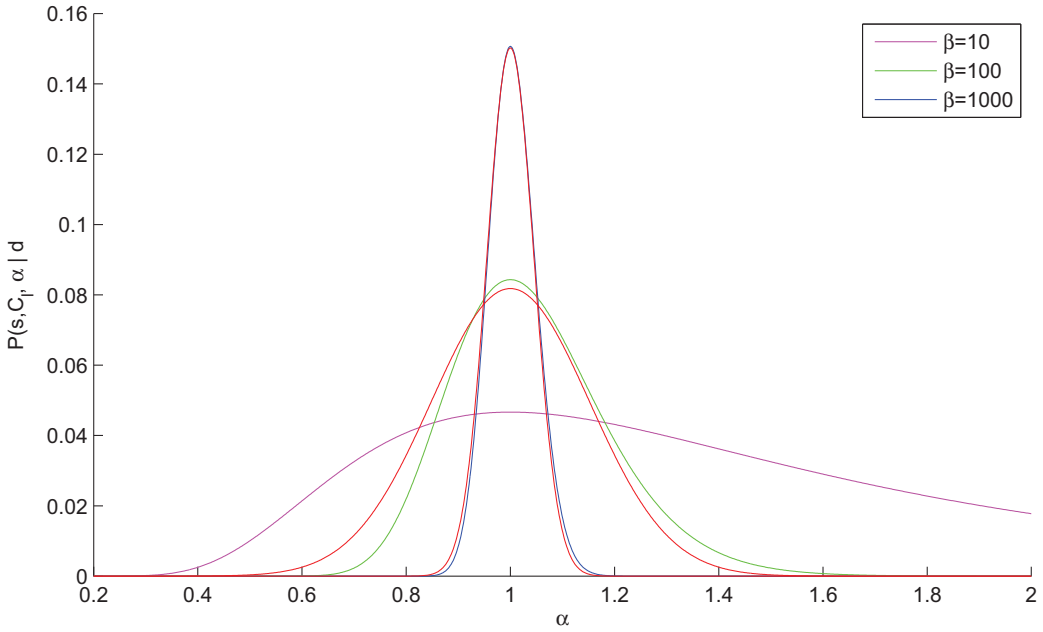
We have presented a self-contained guide to a CMB Gibbs sampler, having focused on both deriving the conditional probability distributions and code design. We described in detail how one can draw samples from the conditional distributions, and saw how the Gibbs sampler is numerically superior to conventional MCMC methods, scaling as $\mathcal{O}(n^{1.5})$. We have also introduced a new object-oriented CMB Gibbs framework, which employs the existing `HEALPix` (Górski et al. 2005) C++

package. We presented a small guide to the usage of `SLAVE`, including post-processing tools and the Blackwell-Rao estimator for obtaining the likelihoods and the best-fit power spectrum. We also reviewed a new way of estimating noise levels in CMB maps, as presented in Groeneboom et.al. (2009a). The software package `SLAVE` will hopefully be released when it is completed during 2009, and will run on all operating systems supporting the GNU C++ compiler. Please see `http://www.irio.co.uk` for release details and information.

Nicolaas E. Groeneboom acknowledges financial support from the Research Council of Norway. Nicolaas especially wishes to thank Hans Kristian Eriksen, but also Jeffrey Jewell, Kris Gorski, Benjamin Wandelt and the whole "Gibbs team" at Jet Propulsion Laboratories (JPL) for useful discussions, comments and input. The computations presented in this paper were carried out on Titan, a cluster owned and maintained by the University of Oslo and NOTUR. We acknowledge use of the `HEALPix` [2] software (Górski et al. 2005) and analysis package for deriving the results in this paper. We acknowledge the use of the Legacy Archive for Microwave Background Data Analysis (LAMBDA). Support for LAMBDA is provided by the NASA Office of Space Science.

REFERENCES

Bennett, C. L., et al. 2003, ApJS, 148, 1
Chu, M., Eriksen, H. K., Knox, L., Górski, K. M., Jewell, J. B., Larson, D. L., O'Dwyer, I. J., & Wandelt, B. D. 2005, Phys. Rev. D, 71, 103002
Chu, I.-W. M. 2005, Ph.D. Thesis,
Cumberbatch, D. T., Zuntz, J., Kamfjord Eriksen, H. K., & Silk, J. 2009, arXiv:0902.0039
Dickinson, C., Eriksen, H. K., Jewell, J., Banday, A. J., Gorski, K. M., & Lawrence, C. R. 2007, Bulletin of the American Astronomical Society, 38, 883
Dickinson, C., et al. 2009, arXiv:0903.4311
Dunkley, J., et al. 2008, arXiv:0811.4280
Eriksen, H. K., et al. 2004b, ApJS, 155, 227
Eriksen, H. K., et al. 2004b, ApJS, 155, 227
Eriksen, H. K., et al. 2006, ApJ, 641, 665
Eriksen, H. K., Jewell, J. B., Dickinson, C., Banday, A. J., Górski, K. M., & Lawrence, C. R. 2008a, ApJ, 676, 10
Eriksen, H. K., Dickinson, C., Jewell, J. B., Banday, A. J., Górski, K. M., & Lawrence, C. R. 2008b, ApJ, 672, L87
Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. 2005, ApJ, 622, 759
Groeneboom, N. E and Eriksen, H. K, 2008, ApJ, 690.1807G
Groeneboom, N. E., et al, 2009 [arXiv:0904.2554]
Hinshaw, G., et al. 2007, ApJS, 170, 288
Hivon, E., Górski, K. M., Netterfield, C. B., Crill, B. P., Prunet, S., & Hansen, F. 2002, ApJ, 567, 2
Jewell, J. B., Eriksen, H. K., Wandelt, B. D., O'Dwyer, I. J., Huey, G., & Górski, K. M. 2009, ApJ, 697, 258
Jewell, J., Levin, S., & Anderson, C. H., 2004, ApJ, 609
Rudjord, Ø., Groeneboom, N. E., Eriksen, H. K., Huey, G., Górski, K. M., & Jewell, J. B. 2009, ApJ, 692, 1669
Larson, D. L., Eriksen, H. K., Wandelt, B. D., Górski, K. M., Huey, G., Jewell, J. B., & O'Dwyer, I. J. 2007, ApJ, 656, 653
Shewchuk, Jonathan R., 1994, Carnegie Mellon University
Smith, K. M., Zahn, O., & Doré, O. 2007, Phys. Rev. D, 76, 043510
Smoot et al., 1992, ApJ396, L1
Wandelt, Benjamin D. and Larson, David L. and Lakshminarayanan, Arun Phys. Rev. D70,8


0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
P(s,C_l, α | d
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
α
β=10
β=100
β=1000

Paper V

Cosmic Microwave Background Likelihood Approximation by a Gaussianized Blackwell-Rao Estimator

Rudjord, .; Groeneboom, N. E.; Eriksen, H. K.; Huey, Greg; Grski, K. M.; Jewell, J. B.

APJ, Volume 692, Issue 2, pp. 1669-1677 (2009).

Paper VI

Detection of trans-Planckian effects in the cosmic microwave background
Groeneboom, Nicolaas E.; Elgary, ystein
Physical Review D, vol. 77, Issue 4, id. 043522

PHYSICAL REVIEW D **77,** 043522 (2008)

Detection of trans-Planckian effects in the cosmic microwave background

Nicolaas E. Groeneboom* and Øystein Elgarøy†

Institute of Theoretical Astrophysics, University of Oslo, Box 1029, 0315 Oslo, Norway

(Received 13 November 2007; published 25 February 2008)

Quantum gravity effects are expected to modify the primordial density fluctuations produced during inflation and leave their imprint on the cosmic microwave background observed today. We present a new analysis discussing whether these effects are detectable, considering both currently available data and simulated results from an optimal CMB experiment. We find that the WMAP (Wilkinson Microwave Anisotropy Probe) data show no evidence for the particular signature considered in this work but give an upper bound on the parameters of the model. However, a hypothetical experiment shows that with proper data, the trans-Planckian effects should be detectable through alternate sampling methods. This fuzzy conclusion is a result of the nature of the oscillations, since they give rise to a likelihood hypersurface riddled with local maxima. A simple Bayesian analysis shows no significant evidence for the simulated data to prefer a trans-Planckian model. Conventional Markov chain Monte Carlo (MCMC) methods are not suitable for exploring this complicated landscape, but alternative methods are required to solve the problem. This, however, requires extremely high-precision data.

DOI: 10.1103/PhysRevD.77.043522 PACS numbers: 98.70.Vc, 98.80.Cq

I. INTRODUCTION

Quantum fluctuations in the scalar field responsible for driving inflation can give rise to a power spectrum of primordial density perturbations consistent with what is required to seed the large-scale structure (LSS) and the temperature anisotropies in the cosmic microwave background (CMB) radiation. This fact is probably one of the main reasons for inflation having become a part of the concordance model of cosmology today. The quantum fluctuations are imprinted during the early inflationary epoch and become classical fluctuations in the gravitational potential as the fluctuations leave the causal horizon. The standard way of calculating the power spectrum of the density fluctuations make use of the fact that space-time becomes Minkowskian in the distant past. However, it has been argued [1–9] that since a given length scale observable in the Universe today shrinks and becomes smaller than the Planck length if followed sufficiently far back in time, effects of quantum gravity will at some point play a role in setting up the perturbations, and may potentially leave an observable signature. The primordial density fluctuations in the early universe may have small oscillations superimposed as a result of quantum gravitational effects. These effects are due to nonnegligible curvature imposed by high energy densities in de Sitter space during the early stages of inflation. Several papers have suggested a generic shape of these effects [3,4,9] in the form of small oscillations superimposed on the standard, nearly scale-invariant primordial power spectrum of density fluctuations. A crucial questions is, of course, whether such a signature is observable in the Universe today. This question has been investigated in several papers [3–8,10]. The conclusion in [10] was pessimistic: the so-called trans-Planckian effects were found to be unobservable in practice. However, this conclusion has been called into question in [8]. One of the problems pointed out in that work is that the oscillations require more numerical accuracy than what is commonly employed in codes like CMBFAST [11] and CAMB [12].

In this paper, we revisit the question of the observability of trans-Planckian effects in the CMB, using a more accurate numerical treatment than in [10]. Both present data and a hypothetical optimal CMB experiment are considered. Furthermore, we mention alternative sampling algorithms to overcome the numerical problems encountered in exploring what turns out to be a particularly nasty likelihood hypersurface. Section II gives a brief description of the particular model we consider and the data sets we use. In Sec. III we discuss some technical details in the numerical computations undertaken. Section IV then contains results obtained with currently available data sets, and in Sec. V we investigate the possibilities with an ideal CMB data set. The exact likelihood function is described in Sec. VI, while a Bayesian evidence analysis is performed in Sec. VII. A discussion of the results and a suggestion for an improved approach to the search for trans-Planckian effects can be found in Sec. VIII, and we conclude in Sec. IX.

II. BACKGROUND

Neglecting the perturbations in the metric during inflation, it is viable to assume that the inflaton scalar field can be expressed as

$$\phi(\mathbf{x}, t) = \underbrace{\phi_0(t)}_{\text{classical evolution}} + \underbrace{\delta\phi(\mathbf{x}, t)}_{\text{fluctuations}}. \tag{1}$$

The fluctuations in a scalar field $\delta\phi$ in a Friedmann-

*nicolaag@astro.uio.no
†oystein.elgaroy@astro.uio.no

Lemaître-Robertson-Walker background satisfy the equation of motion:

$$\ddot{\delta\phi} + 3H\dot{\delta\phi} - \frac{1}{a}(\nabla^2)\delta\phi + m^2\delta\phi = 0, \quad (2)$$

where the dot denotes the derivative with respect to time and $m^2 \propto d^2V/d\phi^2$, where $V(\phi)$ is the inflaton potential. Performing a Fourier expansion of the fluctuations,

$$\delta\phi(x,t) = \frac{1}{\sqrt{N}}\sum_k u_k(t)e^{ikx}, \quad (3)$$

where N is a normalization constant, Eq. (2) becomes

$$\ddot{u}_k + 3H\dot{u}_k + \left(\frac{k^2}{a^2} + m^2\right)u_k = 0. \quad (4)$$

Assuming de Sitter space, this equation can be rewritten using a change of variable ($u_k \equiv \varphi_k/a$) and switching to conformal time ($d\eta = dt/a$). This removes the friction term, and the final equation of motion becomes

$$\varphi_k'' + \left(k^2 - \frac{2}{\eta^2}\right)\varphi_k = 0, \quad (5)$$

where $'$ denotes derivative with respect to conformal time η. A detailed derivation of this result can be obtained in [13]. The general solution for φ_k is

$$\varphi_k = \frac{A_k}{\sqrt{2k}}e^{-ik\eta}\left(1 - \frac{i}{k\eta}\right) + \frac{B_k}{\sqrt{2k}}e^{ik\eta}\left(1 + \frac{i}{k\eta}\right), \quad (6)$$

where A_k and B_k are Bogoliubov coefficients satisfying $|A_k|^2 - |B_k|^2 = 1$, η is the conformal time, and k is the wave mode. The usual choice of vacuum is the Bunch-Davies vacuum, which corresponds to $A_k = 1$ and $B_k = 0$. Space-time becomes Minkowskian in the infinite past ($\eta \to -\infty$), where it is feasible to impose initial conditions on the field equations.

In [4], different possibilities for suitable vacua are discussed before focusing on a specific choice, the adiabatic vacuum, for which $a_k(\eta_0)|0, \eta_0\rangle = 0$. This vacuum (for a given η_0) is a typical representative of vacua, and corresponds to a minimum uncertainty between field and its conjugate momentum [14]. When imposing initial conditions, Bogoliubov coefficients are constrained by the relation [4]

$$B_k = \frac{ie^{-2ik\eta_0}}{2k\eta_0 + i}A_k. \quad (7)$$

Note that the Bunch-Davies vacuum is restored in the infinite past: When $\eta_0 \to -\infty$, $A_k = 1$ and $B_k = 0$. A discussion of the implications of the choice of vacuum can be found in [15,16].

When choosing this zeroth order adiabatic vacuum [3], the primordial curvature spectrum can be expressed as [4]

$$P(k) = \left(\frac{H}{\dot{\phi}}\right)^2\left(\frac{H}{2\pi}\right)^2\left(1 - \frac{H}{\Lambda}\sin\left(\frac{2\Lambda}{H}\right)\right), \quad (8)$$

TABLE I. Explaining Eq. (9).

Parameter	Value	Description
ϵ	$\epsilon \ll 1$	The slow roll parameter
γ	Λ/M_{pl}	The trans-Planckian scale
M_{pl}	$1/\sqrt{8\pi G}$	The reduced Planck mass
ξ	$4 \cdot 10^{-4}\frac{\sqrt{\epsilon}}{\gamma}$	A parametrization of H/Λ
k_n	k_n	The largest measurable scale
$P_0(k)$	$\propto k^{n_s-1}$	The standard power spectrum

where Λ is the (Planck) energy cutoff scale and H the Hubble parameter. Note that the size of the corrections is *linear* [17] in H/Λ. In [18], it is argued that Eq. (8) represents a generic effect, regardless of the details in the (undetermined) trans-Planckian physics. In [4], Eq. (8) is further parametrized as

$$P(\epsilon, \xi, k) = P_0(k)\left(1 - \xi\left(\frac{k}{k_n}\right)^{-\epsilon}\sin\left[\frac{2}{\xi}\left(\frac{k}{k_n}\right)^{\epsilon}\right]\right) \quad (9)$$

with variables explained in Table I. Notice how the standard inflationary power spectrum P_0 is restored as the cutoff $\xi \propto \frac{1}{\Lambda} \to 0$. It is also interesting to note that for low cutoff values $\xi \propto \frac{1}{\Lambda} \to \infty$, the amplitude of the modulation converges to 2 [as $\lim_{x\to\infty} x\sin(2/x) = 2$]. The amplitude of the oscillations never reaches values greater than 2, and only severely impacts P_0 for large values of ξ. Physically, if the cutoff $\xi \propto \frac{1}{\Lambda}$ was enforced at low values (large ξ), it should easily be detectable in the cosmic microwave background, as will be presented in the following section. This is, however, not the case, and we know that the trans-Planckian physics leave imprints at $\xi \ll 1$, corresponding to a cutoff at high energies. The amplitude of the oscillations is therefore expected to be very small.

When evolving the modulated primordial power spectrum through CAMB, the trans-Planckian imprints for varying ξ in the full power spectrum can be seen in

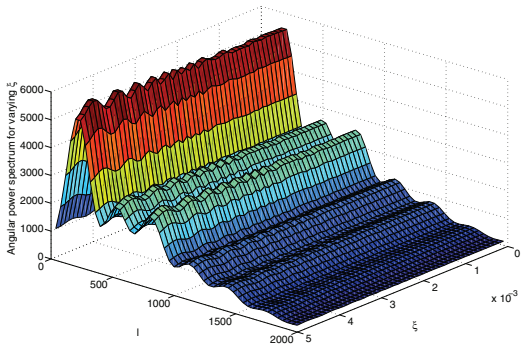


FIG. 1 (color online). The full power spectrum C_ℓ for varying ξ.

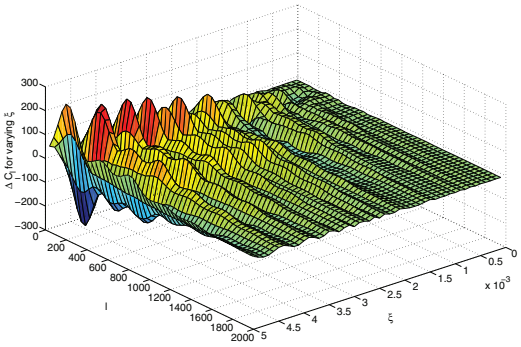


FIG. 2 (color online). The trans-Planckian modulations ΔC_ℓ for varying ξ.

Fig. 1. The oscillations singled out for varying ξ are presented in Fig. 2.

A. A brief summary of recent results

Reference [3] concludes that effects of trans-Planckian physics are possibly within the reach of cosmological observations. Equation (9) describes a a generic expression for how trans-Planckian effects would modify the primordial power spectrum. As seen in the previous section, these oscillations are caused by a nontrivial vacuum for the inflaton field [3,4]. As the oscillations are expected to contribute to the energy density, this could change the way the Universe expands, and in a worst-case scenario, the inflationary phase could be destroyed. The authors of [4] investigate this possibility and conclude that the back-reaction is under control and fully consistent with inflation, with a slow roll found to be completely dominated by the vacuum energy given the parameters suggested in [6].

1. WMAP data and trans-Planckian effects

A previous data analysis [19] concluded that no significant signals from trans-Planckian effects are visible in the CMB. Another analysis [6] claimed that there are some weak hints in the current data, and these indications have become slightly stronger with the WMAP3 (Wilkinson Microwave Anisotropy Probe three-year) data compared to earlier claims by the same authors [5–8]. The parameters implied by the data suggests oscillations in amplitude that are periodic in the logarithm of the scale of the CMB fluctuations, just as predicted from trans-Planckian physics.

In [5], the authors discuss the so-called cosmic variance outliers, i.e. points which lie outside the 1σ cosmic variance error. These outliers are considered interesting as the probability of their presence is very small [20]. The authors mention that it has been envisaged that the outliers could be a signature of new physics, even though the cosmic variance could be responsible for their presence. The conclusion of [5] is that there exists statistical justification for a presence of oscillations in the power spectrum. It should, however, be noted that the authors of [5] treat the *amplitude* of the oscillations as a free parameter, which ensures the fitting of the cosmic variance outliers with the trans-Planckian oscillations.

Another paper [10] was less optimistic and concluded that it is unlikely that a trans-Planckian signature of this type can be detected in CMB and large-scale structure data. However, [10] left some technical issues open, and we find it worthwhile to revisit this problem. We show that the current WMAP3 data gives weak constraints on the parameters of the trans-Planckian model using conventional Markov chain Monte Carlo (MCMC) methods.

2. Simulated data and trans-Planckian effects

The conclusion of [10] was that CMB and LSS data are in principle sensitive to trans-Planckian modulations in the primordial power spectrum, but that is practically impossible to make a positive detection even in future high-precision data. This has to do with the *nature* of the oscillations, as the value of the likelihood function is extremely sensitive to ϵ and ξ. But what "extremely sensitive" means was not explained in [10]. We will show explicit examples of how this sensitive likelihood function contemplates trouble, and why this renders the underlying MCMC method in CosmoMC of little use in estimating the trans-Planckian parameters.

III. NUMERICAL DETAILS

CAMB [12] was modified to include the primordial oscillations described in Eq. (9), adding the parameters ξ and ϵ. Based on a power spectrum generated using the modified version of CAMB, a TT data set was simulated by drawing from a χ^2 distribution with $2\ell + 1$ degrees of freedom. The data set and model are depicted in Fig. 3, and the choice of values for $\xi = 0.0004$ and $\epsilon = 0.01$ are motivated by the Horava-Witten model [21] of trans-Planckian physics [4]. In this model, unification occurs at the scale where a fifth dimension becomes visible [17], resulting in quite optimistic estimates of the energy scale where quantum gravity

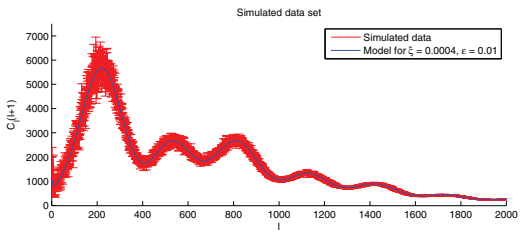


FIG. 3 (color online). Simulated TT data based on $\xi = 0.0004$ and $\epsilon = 0.01$.

effects are manifested. However, the oscillations for this low value of ξ corresponds to very small superimposed oscillations in the primordial power spectrum, as seen from Figs. 1 and 2. The remaining cosmological parameters used for creating the data set (e.g. Ω_b, τ, n_s) were chosen as ΛCDM best-fit values. We continued by creating a simulated TE, EE, and BB data set using the same methods described here.

The MCMC software package CosmoMC [22] was altered to include the parameters ξ and ϵ. Based on the likelihood for ideal noiseless simulated data described in [23], the WMAP-likelihood code of CosmoMC was adjusted to utilize the ideal data instead of the WMAP three-year data. The TT-only likelihood function is given by

$$-2\log\mathcal{L} = \sum_\ell \left(2\ell+1\right)\left[\ln\left(\frac{C_\ell}{\hat{C}_\ell}\right) + \frac{\hat{C}_\ell}{C_\ell} - 1\right], \quad (10)$$

where C_ℓ are the theoretical TT power spectra and $\hat{C}_\ell$ TT data sets. The same method for creating and using ideal data with CosmoMC was employed in [24]. The errors in this ideal data set are only limited by cosmic variance, so no experiment can give any better estimate of the power spectrum. Thus, if trans-Planckian effects are not detectable in the simulated data, they cannot be detectable with any future high-precision data. We continued by employing the full simulated TT, TE, EE, and BB data set using the likelihood presented in [25]:

$$-2\ln\mathcal{L} = \sum_\ell (2\ell+1)\left[\ln\left(\frac{C_\ell^{BB}}{\hat{C}_\ell^{BB}}\right) + \ln\left(\frac{C_\ell^{TT}C_\ell^{EE} - (C_\ell^{TE})^2}{\hat{C}_\ell^{TT}\hat{C}_\ell^{EE} - (\hat{C}_\ell^{TE})^2}\right)\right. \quad (11)$$

$$+ \frac{\hat{C}_\ell^{TT}C_\ell^{EE} + C_\ell^{TT}\hat{C}_\ell^{EE} - 2C_\ell^{TE}\hat{C}_\ell^{TE}}{C_\ell^{TT}C_\ell^{EE} - (C_\ell^{TE})^2} \quad (12)$$

$$\left. + \frac{\hat{C}_\ell^{BB}}{C_\ell^{BB}} - 3\right]. \quad (13)$$

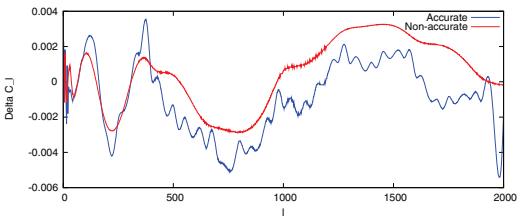


FIG. 4 (color online). The ratio of modulated power spectrum with low and high accuracy, all three parameters boosted.

Increasing accuracy

Previous papers [7,8] have pointed out that it is necessary to increase the accuracy of the calculated power spectrum used for estimating the trans-Planckian parameters. The results in [10] might therefore have been partly due to the inaccuracy in the version of CMBFAST [11] used in this work. The accuracy of the numerical calculations was boosted, which increased run time tenfold. The results can be seen in Fig. 4. However, the accuracy boosting had little or no effect on the WMAP parameter estimation, as will be explained in the following section.

IV. RESULTS FROM WMAP3 DATA

For small values of the trans-Planckian values, the constraints on both ξ and ϵ from the WMAP3 data turned out

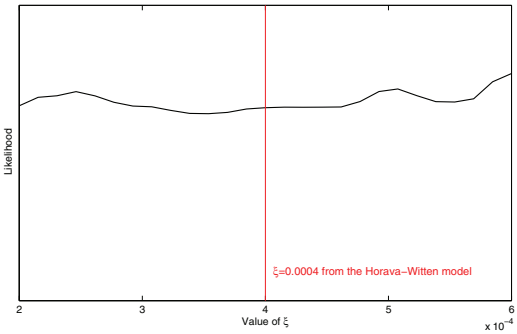


FIG. 5 (color online). Marginalized posteriror for ξ over a small interval surrounding the original input parameter $\xi = 0.0004$ from the Horava-Witten model (vertical line).

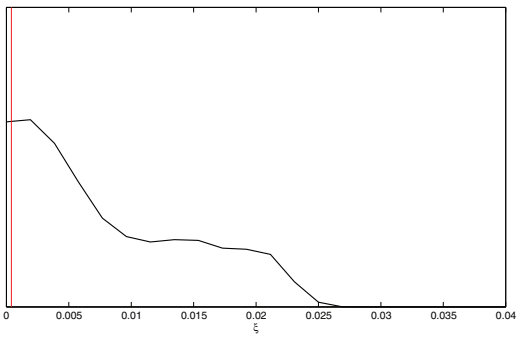


FIG. 6 (color online). Marginalized posteriror of ξ over a larger interval. The vertical line corresponds to $\xi = 0.0004$ from the Horava-Witten model. Notice how it is possible to estimate an upper bound: $\xi_{\max} \sim 0.02$. Above this threshold, there are occasional "bumps" in the posterior where the oscillation reaches zero [$\sin x = 0$, see Eq. (9)].

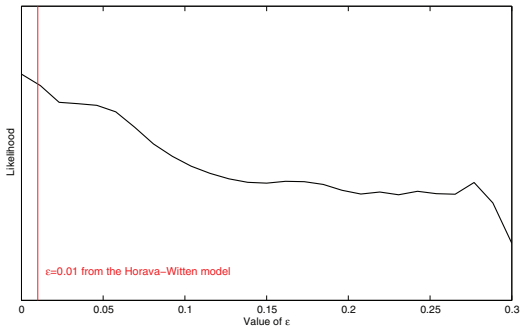


FIG. 7 (color online). Marginalized posterior for ϵ. The vertical line corresponds to $\epsilon = 0.01$ from the Horava-Witten model.

rather poor when running CosmoMC with all standard cosmological parameters free. This can be seen in Figs. 5 and 7, consistent with the results in [10]. For larger values of the trans-Planckian parameter ξ, the increasing amplitude of the primordial power spectrum effectively constrains ξ to the interval [0, 0.02] (recall from Sec. II that for large ξ, the amplitude converges to 2). This can be seen in Fig. 6. This corresponds to a cutoff $\Lambda \sim 0.002 M_{pl} \sim 10^{16}$GeV. Everything below this scale ($\xi > 0.02$) should therefore be ruled out by WMAP3 data.

For physically reasonable values of ξ and ϵ ($\ll 1$), the amplitude of the modulations are well below 1% of the total power spectrum. Also, the cosmic variance and the errors in the WMAP data are several orders of magnitude greater than the amplitude of the modulations in the power spectrum, as can be seen from Fig. 8. This explains why

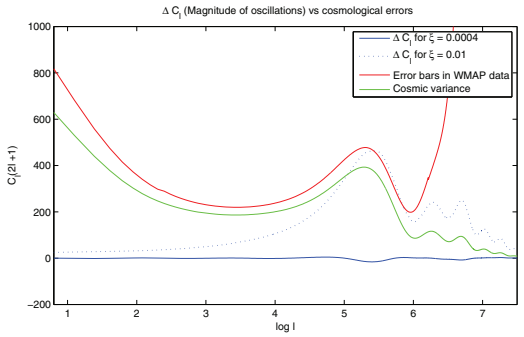


FIG. 8 (color online). The bottom line shows the oscillations ΔC_ℓ in the primordial power spectrum for $\xi = 0.0004$, $\gamma = 0.01$, and $\epsilon = 0.01$. The top line are the error bars in the WMAP data, while the middle line is the cosmic variance. Notice how the modulations in the low-ξ power spectrum are much smaller than either. The dotted line represents a larger $\xi \sim 0.01$, which should be possible to detect, and gives upper constraints on ξ.

any detection of trans-Planckian effects in the WMAP data fails for low values of ξ.

V. RESULTS FROM IDEAL CMB DATA

The WMAP3 data in CosmoMC was replaced by the optimal data set created from the power spectrum based on the Horava-Witten model. The ideal TT data are shown in Fig. 3.

A. Simulated TT data

When performing a parameter estimation of the trans-Planckian parameters ξ and ϵ using the TT-data setup, the Monte Carlo Markov chain method failed to converge to a stationary distribution. The resulting likelihoods for various ξ and ϵ are presented in Figs. 9 and 10, respectively. Notice how different initial conditions and varying step length affect the end result: different initial parameters for the same parameter results in different distributions. This suggests that the likelihood function which decides the path of the random walkers traversing the cosmological parameter space has some troublesome properties.

B. Simulated TT, TE, EE, and BB data

The marginalized posteriors can be seen in Fig. 11. Notice how τ is restored due to the inclusion of simulated TE and EE data. The trans-Planckian parameters now reflect the ruggedness of the likelihood function, and traces can also be seen in n_s and A_s. As in the TT case, the original trans-Planckian input parameters are still not possible to pin down.

Figures 11 and 12 show how the MCMC process fails in converging, as the marginalized posterior show random walkers ''trapped'' in local minima.

VI. THE EXACT LIKELIHOOD FUNCTION

In order to explain why the parameter estimation fails, we investigated the likelihood function in the case of noiseless data and different power spectra for fixed cosmological parameters except the trans-Planckian ξ and ϵ. This was done by grid integration, and no MCMC sampling was involved. The resulting likelihood surfaces for continuous variations in ξ and ϵ are seen in Figs. 13 and 15, respectively. A figure of the exact two-dimensional likelihood surface for both ξ and ϵ is presented in Fig. 16. Figures 13–16 are $-\log\mathcal{L}$ plots, so the best-fit value corresponds with the global minimum. The likelihood surfaces are riddled with local minima, effectively rendering the usual MCMC method useless: the random walkers become trapped in the local holes and will have problems converging to the correct stationary distribution. Take especially note of the two-dimensional parameter space presented in Fig. 16, a landscape reminiscent of that of an egg container.

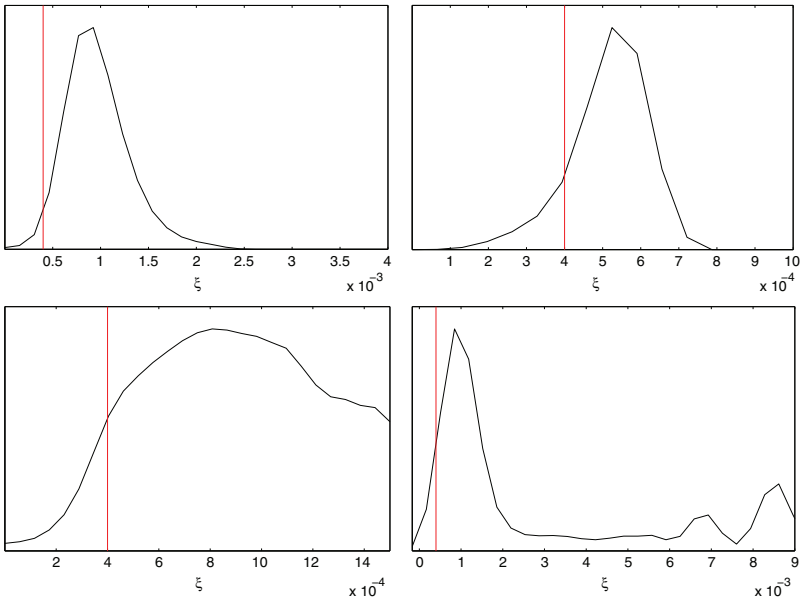


FIG. 9 (color online). Top left: Marginalized posterior for ξ with mean value at $0.978 \cdot 10^{-3}$. Top right: Marginalized posterior for ξ with mean value at $0.515 \cdot 10^{-3}$. Bottom left: Marginalized posterior for ξ with no peak, almost uniform distribution. Bottom right: Marginalized posterior for ξ with several peaks. Notice how the peaks do not match, where the vertical (red) line represents the Horava-Witten input parameter $\xi = 0.0004$ for the simulated data.

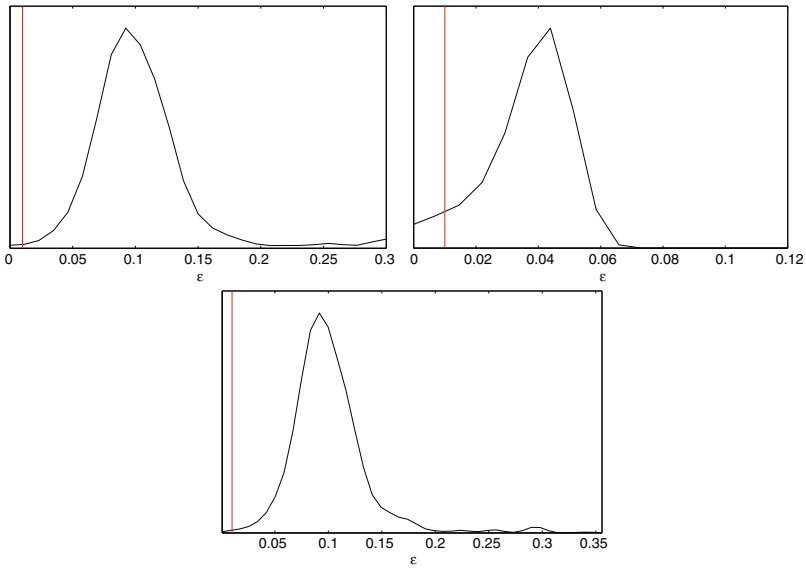


FIG. 10 (color online). Top left: Marginalized posterior for ϵ with mean value at 0.12. Top right: Marginalized posterior for ϵ with mean value at 0.037. Bottom left: Marginalized posterior for ϵ with mean peak at 0.101. Notice how the peaks do not agree for different step lengths, where the vertical (red) line represents the Horava-Witten input parameter $\epsilon = 0.01$ for the simulated data.

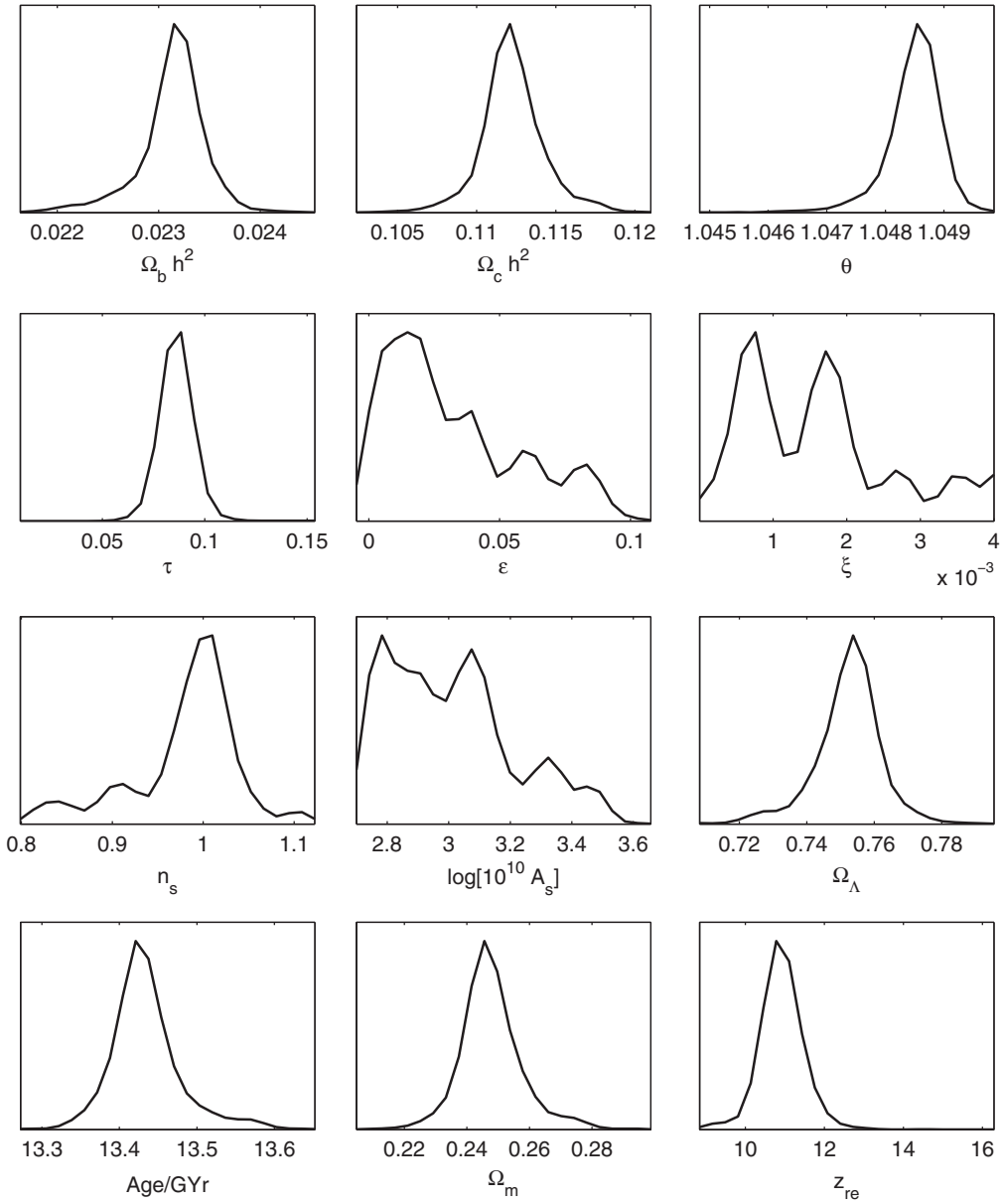


FIG. 11. Marginalized posteriors for the full *TT*, *TE*, *EE*, and *BB* likelihood analysis using Eq. (11).

The shape of the exact likelihood is a result of the *nature* of the oscillations: Eq. (9) gives rise to a particularly nasty behavior of the oscillations [26]. The problem with the likelihood function is not easily solved. In principle, one could perform a smoothing of the likelihood, something which works well in a low-dimensional parameter space, but is hopelessly inefficient for a realistic number of parameters.

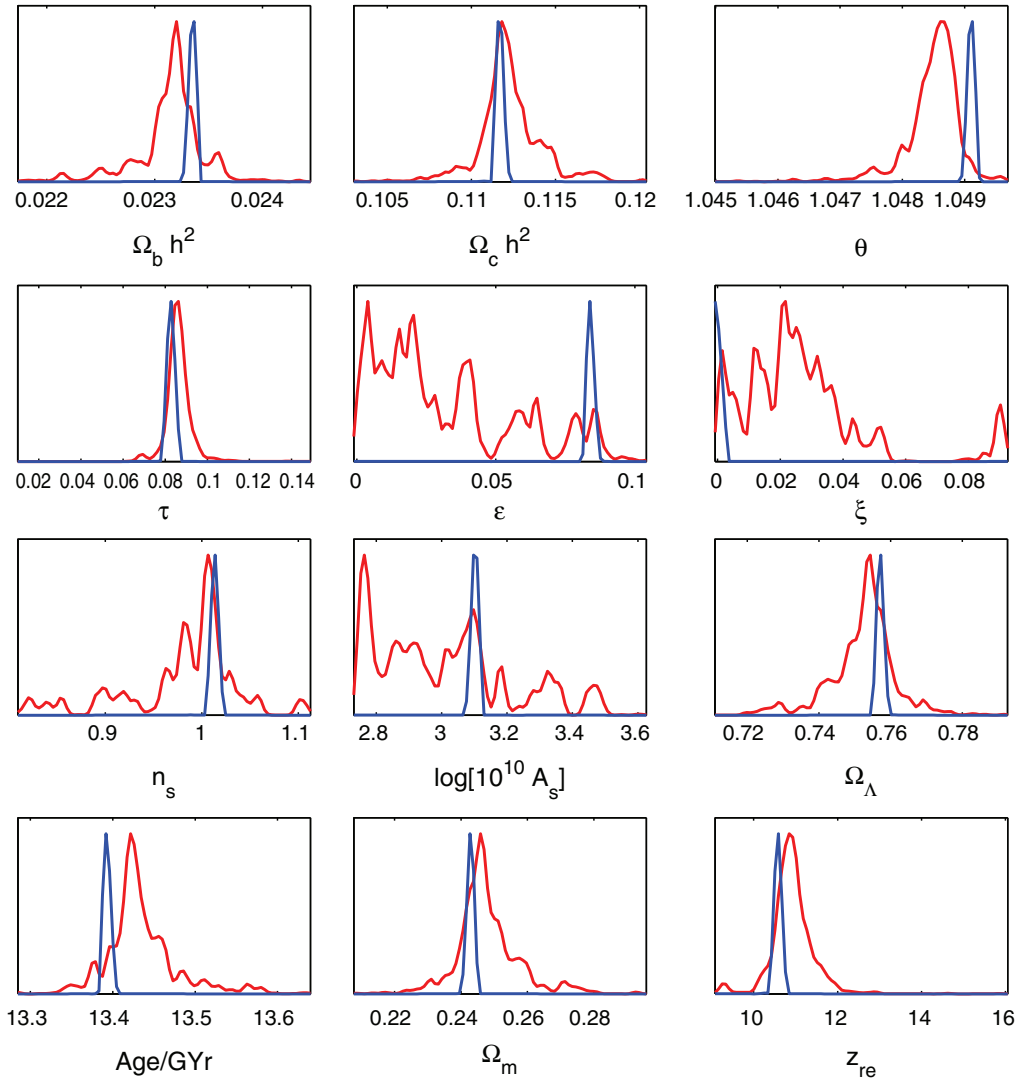


FIG. 12 (color online). Marginalized posteriors (wide distributions) with smaller bins than in Fig. 11 for the full *TT*, *TE*, *EE*, and *BB* likelihood analysis using Eq. (11). The likelihoods correspond with the narrow distributions. Notice how the posteriors completely failed to converge to a distribution.

Decoupling the amplitude

The amplitude and the frequency of the trans-Planckian oscillations are linked in the model we have considered so far. However, it is certainly possible that they might vary independently, and in [5–8], the authors operate with the amplitude of the trans-Planckian oscillations as a free amplitude. We repeated our analysis with frequency and amplitude decoupled. From Figs. 17 and 18 it should be clear that freeing the amplitude does not really add any significance to the likelihood surface: the likelihood still gives the correct preferred amplitude.

VII. BAYESIAN EVIDENCE

From the logarithmic likelihood in Figs. 13 and 15, it is obvious that it is impossible to extract the values of the parameters of our model from the simulated data.

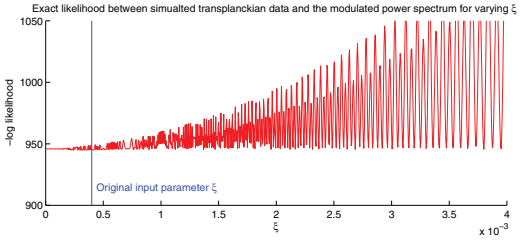


FIG. 13 (color online). $-\log\mathcal{L}$ as a function of ξ for simulated trans-Planckian data.

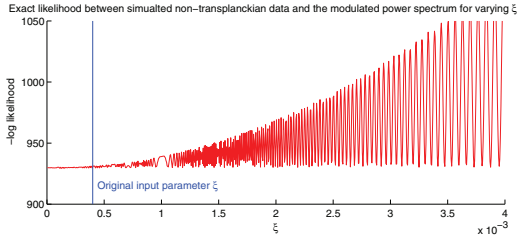


FIG. 14 (color online). $-\log\mathcal{L}$ as a function of ξ for simulated nontrans-Planckian data. Notice how the nontrans-Planckian data is preferred to the simulated trans-Planckian data set.

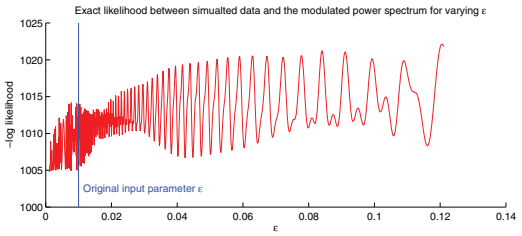


FIG. 15 (color online). $-\log\mathcal{L}$ between varying ϵ and simulated data.

However, one might still ask the question of whether the data prefer the trans-Planckian model to the standard power-law primordial power spectrum. In a Bayesian framework this question is answered by calculating the evidence ratio of the two models, see [27] for a nice introduction. Here we fix $\epsilon = 0.01$ and use the evidence ratio to see whether the data prefer the introduction of the additional parameter ξ. We assume that the trans-Planckian parameter ξ is *not* degenerate with any of the cosmological parameters, designated by the parameter vector θ. This might seem like a strong assumption, but we have checked that it is reasonable for the model and the parameters we consider in this paper. Also, the effects of primordial oscillations have been investigated in [28],

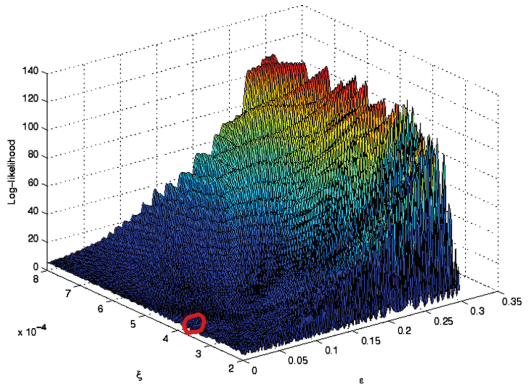


FIG. 16 (color online). $-\log\mathcal{L}$ for varying both ξ and ϵ. The (red) ring represents the original input parameters $\xi = 0.0004$, $\epsilon = 0.01$ for the simulated data set. Notice how a random walker would be trapped in any of the other local minima.

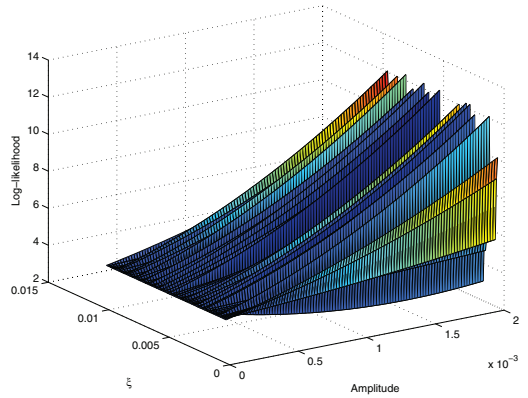


FIG. 17 (color online). $-\log$ likelihood for varying ξ and amplitude.

concluding that degeneracies of oscillations with standard cosmological parameters are virtually nonexistent.

Let θ denote the nontrans-Planckian cosmological parameters and $\hat{\theta}$ the best-fit ΛCDM parameters. The logarithmic evidence is given by

$$\Delta \ln E = \ln E - \ln E^0, \tag{14}$$

where E^0 is the evidence of the nontrans-Planckian model. Then

$$\Delta \ln E = \langle \mathcal{L} \rangle_{\pi(\theta,\xi)} - \langle \mathcal{L} \rangle_{\pi(\theta)} \tag{15}$$

$$= \langle \mathcal{L}(\theta, \xi) \rangle_{\pi(\xi)} - \mathcal{L}(\hat{\theta}), \tag{16}$$

where $\pi(\theta, \xi)$ is the prior over (θ, ξ) and $\pi(\theta)$ is the prior

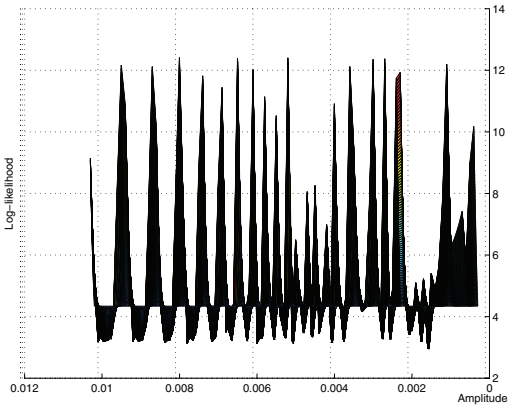


FIG. 18 (color online). $-\log$ likelihood for varying ξ and amplitude (different viewing angle).

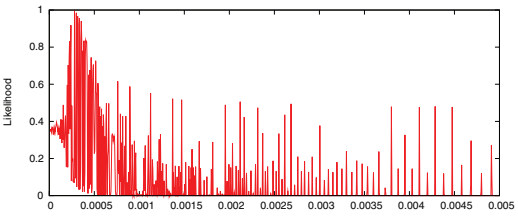


FIG. 19 (color online). Likelihood $\mathcal{L}(\xi)$ used in calculating the significant evidence.

over θ. Thus, $\langle\mathcal{L}\rangle_X$ denotes the mean likelihood over the prior space X. The likelihood can be found in Fig. 19, which is the exponential of the negative values in Fig. 13. We choose the prior to be $\xi \in [0, 0.005]$. This results in $\Delta \ln E \approx -1.5$ which shows that the trans-Planckian model is slightly *not* preferred. Hence even perfect CMB-measurements over the whole sky does not give any significant evidence for trans-Planckian effects. When constraining the prior to reside in $\xi \in [0, 0.0005]$, then $\Delta \ln E \approx 0.15$ and the trans-Planckian model is more or less equally preferable to the nontrans-Planckian model. This is hardly surprising, as a continuous constraint of the prior $Pr(\xi) = \lim_{\xi\to 0}[0, \xi]$ results in $\Delta \ln E = \ln E^0 - \ln E^0 = 0$, which corresponds with the standard nontrans-Planckian model. The evidence never grows larger than $\sim$0.15 for any prior constraint on ξ.

VIII. ALTERNATE MULTILEVEL SAMPLING

There exists a few multilevel sampling methods besides the standard MCMC/Metropolis-Hastings algorithm that can tame the ruffled likelihood through a series of optimizations. Reference [29] contains a detailed introduction to these methods, most notably simulated annealing, umbrella sampling, and the simulated tempering method. We proceed by a rough discussion on how some of these alternate sampling methods work when investigating the one-dimensional likelihood presented in Fig. 13.

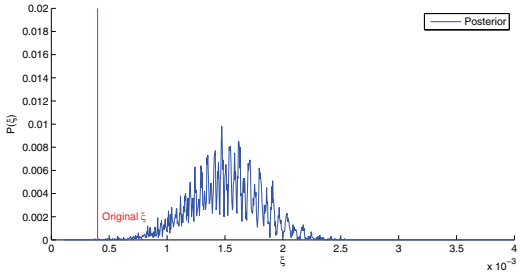


FIG. 20 (color online). The posterior of 10 000 MCMC walkers evaluated after 22 000 steps each in one dimension, with initial conditions $\xi_0 = 0.0015$. None of the walkers reached the correct value of $\xi = 0.0004$.

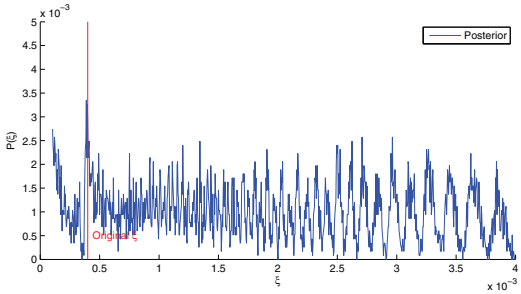


FIG. 21 (color online). The posterior of 10 000 MCMC walkers evaluated after 22 000 steps each in one dimension, with initial conditions $\xi_0 = 0.0015$. The posterior shows a clear contribution around the correct value of $\xi = 0.0004$.

A. Metropolis-Hastings method, 1D

Figure 20 shows the posterior of 10 000 walkers after $\sim$22 000 steps each, using the Metropolis-Hastings algorithm. The likelihood is shown in Fig. 13. In theory, the walkers should converge to a stationary distribution with the tallest peak located at $\xi = 0.0004$, but due to the ruffled likelihood this will take forever. The initial values for the walkers was here $\xi_0 = 0.0015$. This reflects what we have seen in Figs. 9–12: that the complicated likelihood surface prevents the posterior to converge to the correct shape.

B. Modified Metropolis-Hastings method, 1D

Figure 21 shows the posterior of 10 000 walkers after $\sim$20 000 steps each, using the Metropolis-Hastings algo-

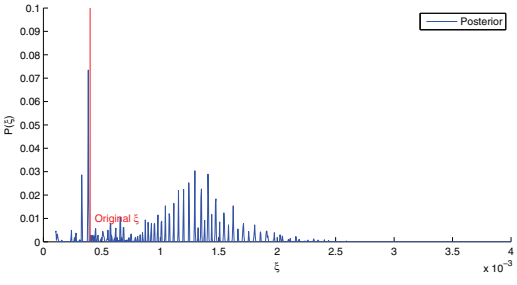


FIG. 22 (color online). The result of 10 000 MCMC walkers evaluated after 6 000 steps in one dimension, with initial conditions $\xi_0 = 0.0015$ (not a posterior). The temperature was slowly decreased until reaching $T = 0$, where the correct value $\xi = 0.0004$ was obtained.

rithm. In addition, for each step there is a 0.01% chance for a given walker to perform a "large" jump to a new, (uniformly) random location. The walkers will therefore tend to accumulate at the correct location (where the $-\log$ likelihood is lowest, $\xi = 0.0004$). This is a slow method and needs a fair amount of steps to give results. It is also not especially "robust," as the accumulation of walkers at the lowest point is dependent on fine-tuning the large-step rate (here, 0.01%).

C. Simulated annealing, 1D

Figure 22 shows the result of 10 000 walkers after ~6000 steps each, using the simulated annealing algorithm. A description of the simulated annealing (SA) method can be found in [29]. As in the previous examples, 10 000 random walkers were initialized at $\xi_0 = 0.0015$. The distribution spread out evenly as the SA temperature T was high and started converging to different minima as T was lowered. As $T \to 0$, a good portion of walkers ended up at $\xi = 0.0004$. This shows that the SA method can be successfully used to determine the correct minima for this type of likelihood landscape.

However, the implementation and investigating of these alternative methods in CosmoMC is a major undertaking which we hope to pursue in future work.

IX. CONCLUSION

We have shown that standard MCMC methods are not suitable for detecting trans-Planckian oscillations in the CMB. The 3-year WMAP3 data alone gives weak upper constraints on the trans-Planckian parameters. With the simulated noiseless data, the results are more ambiguous. In theory, the estimation of the trans-Planckian parameters *should* be possible, but the messy likelihood function renders conventional MCMC methods of little use. This is not due to the amplitude of the superimposed oscillations in the power spectrum alone, but rather reflects the chaotic behavior of the oscillations for varying trans-Planckian parameters ξ and ϵ. Small variations in the parameter space thus correspond to large variations in the likelihood function, as presented in Figs. 13 and 15. As in [10], we conclude that CMB data are highly sensitive to these modulations in the primordial power spectrum, but it is virtually impossible to perform a parameter estimation from WMAP3 data or high-precision simulated data using conventional MCMC methods. We have also shown, via Bayesian analysis, that even a high-precision measurement of the full CMB sky does not give any significant evidence for trans-Planckian effects. It should, however, be possible to implement the SA method in CosmoMC when using simulated high-precision data in order to detect the trans-Planckian parameters.

We have in this paper shown that it is very unlikely that CMB data could support the establishment of trans-Planckian effects, even for models where quantum gravity is relevant on relatively low energy scales, like in the Horava-Witten model. There are several reasons for this lamentable conclusion:

(i) A low quantum gravity scale, for example, the value $\xi = 0.0004$ chosen in our simulations, gives a very low amplitude of the oscillations in the primordial power spectrum. Hence, cosmic variance diffuses the (assumed noiseless) signal.
(ii) Evidence can maximally rise above 0.15, showing that the trans-Planckian model is only very weakly preferred when assuming a Horava-Witten universe.
(iii) Even large (and unphysical) values of ξ will result in good fits for singular points due to the behavior of the oscillations (they periodically reach zero when $\sin x = 0$).
(iv) The conventional MCMC methods in CosmoMC fails, and we have no current computational methods implemented for solving the problem of transversing the extremely complicated likelihood surface.

ACKNOWLEDGMENTS

Ø. E. acknowledges support from the Research Council of Norway, Project No. 162830. We thank H. K. Eriksen and J. R. Kristiansen for valuable discussions.

[1] J. Martin and R. H. Brandenberger, Phys. Rev. D **63**, 123501 (2001).
[2] R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, Phys. Rev. D **64**, 103502 (2001).
[3] U. H. Danielsson, Phys. Rev. D **66**, 023511 (2002).
[4] U. H. Danielsson, arXiv:astro-ph/0606474.
[5] J. Martin and C. Ringeval, Phys. Rev. D **69**, 083515 (2004).
[6] J. Martin and C. Ringeval, J. Cosmol. Astropart. Phys. 08 (2006) 009.
[7] J. Martin and C. Ringeval, Phys. Rev. D **69**, 127303 (2004).
[8] J. Martin and C. Ringeval, J. Cosmol. Astropart. Phys. 01 (2005) 007.
[9] L. Bergström and U. H. Danielsson, Phys. Rev. D **66**, 023511 (2002).
[10] Ø. Elgarøy and S. Hannestad, Phys. Rev. D **68**, 123513 (2003).
[11] U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).
[12] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. **538**, 473 (2000).
[13] S. Dodelson, *Modern Cosmology* (Academic Press, New York, 2003).
[14] D. Polarski and A. Starobinsky Classical Quantum Gravity **13**, 377 (1996).
[15] V. Bozza, M. Giovannini, and G. Veneziano, J. Cosmol. Astropart. Phys. 05 (2003) 001.
[16] D. Campo, J. Niemeyer, and R. Parentani, Phys. Rev. D **76**, 023513 (2007).
[17] N. Kaloper, M. Kleban, A. Lawrence, and S. Shenker, Phys. Rev. D **66**, 123510 (2002).
[18] N. Kaloper and M. Kaplinghat, Phys. Rev. D **68**, 123522 (2003).
[19] R. H. Brandenberger, AAPPS Bull. **11**, 20 (2001) [arXiv:astro-ph/0208103v1].
[20] D. N. Spergel *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **170**, 377 (2007).
[21] P. Horava and E. Witten, Nucl. Phys. **B460**, 506 (1996).
[22] A. Lewis and S. Bridle, Phys. Rev. D **66**, 103511 (2002).
[23] L. Verde *et al.*, Astrophys. J. **148**, 195 (2003).
[24] D. F. Mota, J. R. Kristiansen, T. Koivisto, and N. E. Groeneboom, arXiv:0708.0830.
[25] R. Easther, W. H. Kinney, and H. Peiris, J. Cosmol. Astropart. Phys. 05 (2005) 009 (2005).
[26] Two movies describing how continuous variations in ϵ and ξ affect the superimposed oscillations can be downloaded from http://www.irio.co.uk/projects/thesis.
[27] P. Mukherjee, D. Parkinson, and A. R. Liddle, Astrophys. J. **638**, L51 (2006).
[28] J. Hamann, L. Covi, A. Melchiorri, and A. Slosar, Phys. Rev. D **76**, 023503 (2007).
[29] J. S. Liu, *Monte Carlo Strategies in Scientific Computing* (Springer, New York, 2002).

Paper VII

Non-Gaussianities in the local curvature of the 5-year WMAP data
Rudjord, Ø.; Groeneboom, N. E. ; Hansen, F . K.; Cabella, P.
arXiv:1002.1811
