Data Analysis of and Results from Observations of the Cosmic Microwave Background with the Cosmic Background Imager

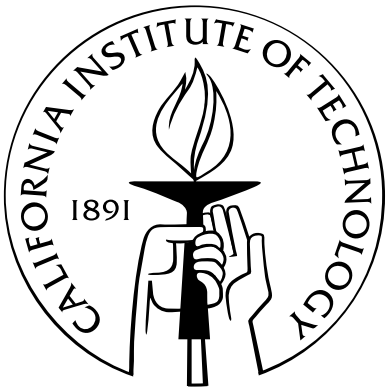
Thesis by

Jonathan LeRoy Sievers

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2004

(Defended September 30, 2003)

Acknowledgements

First off, I would like to thank the CBI team, since no CBI, no thesis. Tony Readhead worked wonders not only getting the CBI built, but gracefully navigating endless shoals in keeping it up and running. His unflagging good cheer brightened many a day. I firmly believe that Steve Padin, should he tire of astronomy, could have a long and fruitful career as an appliance faith healer. He made keeping a complex instrument working in a harsh site look easy, wich it certainly was not. I hope they have Carr's Table Waters at the pole! Thanks to Tim Pearson for the years of hard work in so many areas that made dealing with CBI data possible. His sharp eye caught many things that may have otherwise slipped by. Once my thesis made it by Tim, I figured it had to be OK. And Martin Shepherd deserves a special thanks. His code did much of the work in this thesis, without which I would probably still be languishing in the basement of Robinson. My programming style has been permanently ++ed by watching him at work. In addition to being great friends, Brian Mason, John Cartwright, and Pat Udomprasert helped keep me sane in Chile. Well, sort-of sane. But I shudder to think what might have happened otherwise. Brian, this meow's for you.

I would also like to thank the whole analysis team, both for their expertise, and their large computers. Dick Bond provided a steady hand in keeping the analyis moving. Steve Myers was critical in turning the data set into something useful in finite time, as well as quickly deflating dumb ideas. Carlo Contaldi, Simon Prunet, and Dmitri Pogosyan provided the computing and parameter expertise that let us get CBI results out the door. We owe a great debt to the CITA computing facilities, without which we would probably still be crunching away on the data with rhino. Ue-Li Pen was especially useful, both for his keen mind, and for tossing us the keys to octopus whenever we asked.

Thanks to the gang in Pasadena, too. They were a far more interesting, well-rounded, and fun group than one has any right to expect from a bunch of astronomers. The old guard were great at welcoming us and showing us the ropes when we were still wet behind the ears. Brad, Roy, Kern, everybody else (I know there are more, but this has to be in the mail in an hour if I want to graduate), thanks guys. Thanks also to John Yamasaki, both for his work on the CBI, and especially his youthful spirit. It is impossible not to enjoy one's own life with Yama around. Thanks to Kathy and Pete for providing a home away from home. Dave Vakil was a great friend and roomate (not one late charge at the house the whole time he was there!) as well as a fun bridge partner. Dave, I finally cracked 50% on the Lehmans! I would thank Alice Shapley, but I feel I owe retribution for sicking people on my poor, sensitive sides. I thought a foreign country would finally provide refuge, but alas, I was in err. Dr. Green Cloud made the office fun, as well as providing an endless source of the odd and obscure. From whom else could I have learned about the albino sea-cucumber? And with whom else could I have hitch-hiked across the Andes? Rob Simcoe and Pat Udomprasert have been fast friends since the day I showed up, belatedly, to grad school. They have become family over the years - just ask my siblings. Thanks to Amy Mainzer who kept my life from turning into a monotony of matrices. Her friendship and insight kept life in perspective and made me think about many things that needed thinking about. Thanks to her also for providing such a good home for the fish.

Finally, thanks to my family for their endless support and provision of entertainment. One couldn't ask for a more interesting gropu of folks with more widely varying skills. Mom, Dad, Sara, Amy, Katie, Chuck, you guys are the best.

Abstract

We present results from observations of the Cosmic Microwave Background (CMB) with the Cosmic Background Imager (CBI), a sensitive 13-element interferometer located high in the Chilean Andes. We also discuss methods of analyzing the data from the CBI, including an improved way of measuring the true power spectrum using maximum likelihood estiamtion. This improved method leads to a saving of a factor of two in memory usage, and an increase in speed of order the number of points in the spectrum. The initial results are discussed, in which the fall-off in power at ell $>$ 1000 (the "damping tail") was first observed. We also present the results from the first year of observations with the CBI, and discuss cosmological intepretations both alone and in concert with the results from other experiments. These provide tight constraints on cosmological parameters, including a Hubble constant of 69 +/- 4 km/s/Mpc, an age of the universe of 13.7 +/- 0.2 billion years, and a denisty of dark energy of 0.70 +/- 0.05 of the critical density of the universe. Finally, we discuss an alternate method of data compression, with great flexibility in what information is kept, while being computationally tractable. We then apply this method to the CBI data to constrain the potential emission from foreground contaminants contributing to the observed CMB radiation. We find that the data is consistent with zero foreground, with a maximum allowed foreground contribution between about 8% and 12% of the total signal (at an ell of 600 and frequency of 30 GHz), depending on the spectral index of foreground emission.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

About forty years ago, Arno Penzias and Robert Wilson discovered that the sky was filled with a highly uniform glow with an antenna temperature at 4 GHz of about 3 degrees (Penzias & Wilson, 1965). The radiation was immediately interpreted by Dicke et al. (1965) to be the thermal radiation from the formation of the universe that they themselves were searching for, now called the Cosmic Microwave Background (CMB). They recognized its cosmic importance, even using the CMB temperature and cosmic helium abundance to calculate the current physical baryon density $\Omega_B h^2$ to within an order of magnitude, using the techniques of Big Bang Nucleosynthesis (BBNS). The CMB was measured to be an almost perfect black-body (Mather et al., 1994) and perhaps the smoothest astronomical field known, uniform throughout the sky to a part in a thousand. Despite its smoothness, observations of minute fluctuations in the CMB have become one of the most important sources of information about the large-scale properties of the cosmos. This thesis will discuss observations of CMB anisotropies using the Cosmic Background Imager (CBI), a special purpose radio interferometer.

I will describe CBI observations, techniques used to analyze the data, and the results obtained. In Chapter 2, I describe the framework of Maximum Likelihood Estimation used to extract a power spectrum once the expected behavior of the data is calculated, including a new way of converging to the best-fitting power spectrum that can decrease the computational work by a factor of a few dozen. In Chapter 3, I describe the commissioning data taken by the CBI, the analysis techniques used, the resulting power spectrum, and the significance of that power spectrum. In Chapter 4, I

describe the first-year observations of the CBI, the analysis of those data (which was much more sophisticated than that of Chapter 3), and the ensuing power spectrum. In Chapter 5, I describe a new, fast technique for measuring the power spectrum that has considerable flexibility in the choice of information retained while approaching the theoretical minimum number of estimators required to compress the data set almost losslessly. This compression is important because CMB analysis strains available computing resources. This technique has been coded into a program called CBISPEC, which I then use to place limits on galactic foregrounds possibly present in the CBI observations. This is a task for which CBISPEC is well suited, but which is impossible with our other analysis tools. In Appendix A, I carry out a derivation of statistical noise properties used in Chapter 4. Finally, in Appendix B I briefly summarize work conducted with Patricia Udomprasert in applying optimal CMB weighting to CBI observations of galaxy clusters. This has the potential to substantially increase the accuracy with which the CBI can characterize cluster structure from a given dataset.

1.1 Origin of the Microwave Background

The CMB is understood today to be the remnant radiation from the big bang. The universe started as an extremely hot, dense plasma that expanded and cooled. This expansion and cooling has continued from the earliest fraction of a second after the big bang through the current day. When the universe was very young, the thermal radiation was locked in place relative to the baryons through Thomson scattering. There was some diffusion on small scales (Silk, 1968), but otherwise the photon density behaved like the plasma density. Finally, about 400,000 years after the big bang, protons and electrons combined to form neutral hydrogen atoms, a process called recombination. The photons could then free-stream, and they have been (mostly) unaltered since this epoch, aside from the overall cooling of the CMB due to the expansion of the universe. The spot where photons last scattered off of electrons is called the *surface of last scattering*. Because recombination happened quickly ($\delta z/z < 0.1$ (see, *e.g.*, White, 2001), we have in the CMB essentially a snapshot of the conditions of the universe at an age of 400,000 years. This is only about 3×10^{-5} of its current age,

or about the age of a day-old baby relative to a 90-year-old. At this early time, the universe was almost perfectly uniform. But it can't have been completely uniform, or else there would have been no seeds from which the structure we see today could have formed. For decades, people searched for anisotropies in the CMB without success. The first and by far the largest anisotropy measured was a dipole moment due to the Earth's motion, most notably in Fixsen et al. (1994) (see Lineweaver, 1997, for dipole history), but the primordial fluctuations were not detected until the COBE satellite (Smoot et al., 1992) measured fluctuations on 10° scales in 1992. Since then, the study of the CMB has been one of the most active fields in astronomy, with a whole host of experiments measuring the anisotropies with higher sensitivity and on smaller scales from ground-based, balloon-born, and satellite experiments.

The reason that measuring CMB anisotropies is of such interest is because the angular power spectrum of the anisotropies contains a wealth of detailed information about the properties and evolutionary history of the universe. The power spectrum is so useful because the fluctuations are both calculable and small. Once the earliest spectrum has been set (such as during inflation), the evolution of the fluctuations does not depend on exotic and uncertain physics. Because the fluctuations are small they remain in the linear regime, and so the messy non-linear physics that dominates the universe today (star formation, gas dynamics, supernovae, AGN's *etc.*) doesn't affect the expected spectrum. Care must be taken calculating the spectrum, especially the radiative transfer in the transition region between optically thick and optically thin. Though the calculations are complicated, they are not uncertain, and a number of packages that calculate the spectrum are in good agreement (Bond & Efstathiou, 1984, 1987; Vittorio & Silk, 1984, 1992; Fukugita et al., 1990; Hu et al., 1995; Lewis et al., 2000, many others). We use versions of the fast code CMBFAST (Seljak & Zaldarriaga, 1996) for all the model spectra used in this thesis.

1.2 Power Spectrum Basics

The primary goal of microwave background experiments is to measure the angular power spectrum of CMB fluctuations. There is potentially confusing terminology (most notably the fact that C_ℓ and

$\mathcal{C}_\ell$ are different quantities), so the notation used in the remainder of this work is defined here, and power spectrum concepts specific to the CMB are outlined.

Generally, power spectra are thought of in Fourier space, as being the expected variance of modes of a given wavelength. The fact that the sky is a sphere, rather than an infinite plane, requires modifications to the standard Fourier picture. For the particular case of the surface of a sphere, the temperature everywhere on the sky is expressed as the sum of spherical harmonics, rather than the sine and cosine waves of Fourier transforms:

$$\frac{\Delta T}{T} = \sum_{\ell} \sum_{m=-\ell}^{\ell} a_{\ell m} \Phi_{\ell m} (\theta, \phi) \tag{1.1}$$

Here the $\Phi_{\ell m}$ are the spherical harmonics, and the $a_{\ell m}$ are their amplitudes. With the $\Phi_{\ell m}$, ℓ more or less corresponds to the wavelength of the mode, and m is akin to its orientation. Since we expect the microwave background to have no preferred orientation on the sky, the $a_{\ell m}$ should be statistically independent of m, depending only on ℓ. Furthermore, we expect the CMB to be a Gaussian random field if the fluctuations arise during the era of inflation (White, Scott, & Silk, 1994), though other sources of structure formation, *e.g.* topological defects, will give rise to non-Gaussianity. This means that the $a_{\ell m}$ are independent of each other and have a Gaussian probability distribution with mean zero. Under these assumptions, *all* of the information contained in the CMB is contained in a set of coefficients C_ℓ such that

$$\left\langle a_{\ell m}^2 \right\rangle = \mathrm{C}_\ell \tag{1.2}$$

This is not usually the quantity quoted, however. To see the problem, picture a power spectrum where C_ℓ is constant and compare the variance on small scales to that on large scales. If we pick a patch size of interest, then it will feel power from some fractional width in ℓ, so a small patch at higher ℓ will feel more discrete values of ℓ than a large patch at lower ℓ. In addition, each ℓ feels $2\ell + 1$ individual $\Phi_{\ell m}$, and so the total number of $a_{\ell m}$ that contribute to the variance of a patch is proportional to ℓ^2. So, a flat power spectrum in C_ℓ will have sharply rising temperature fluctuations on smaller scales. Another way of thinking about it is that a spectrum flat in C_ℓ is a pure white-noise

spectrum with every mode statistically equivalent, so large-scale fluctuations average over more noise and hence will have smaller amplitudes than small-scale fluctuations. In order to make the numbers in the power spectrum more physically meaningful, the quantity $\mathcal{C}_\ell$ is often used, with the definition (Bond, 1996)

$$\mathcal{C}_\ell \equiv \frac{\ell(\ell+1)\mathrm{C}_\ell}{2\pi} \tag{1.3}$$

A flat spectrum in $\mathcal{C}_\ell$ will then have scale-invariant temperature fluctuations, equal on all lengths. Usually, $\mathcal{C}_\ell$ is scaled by the CMB temperature T_0 and plotted in $\mu\mathrm{K}^2$. This corresponds to the actual temperature variance on the sky of fluctuations with wavenumber ℓ. In general, the remainder of this work will refer to $\mathcal{C}_\ell$ and not C_ℓ.

1.3 Cosmological Effects on the Power Spectrum

The initial fluctuations are believed to have arisen from quantum uncertainty during the epoch of inflation, and hence to have a nearly scale-invariant spectrum, though the details depend on which particular flavor of inflation one uses (see, *e.g.*, Lyth & Riotto, 1999, for a review). Since the creation of the fluctuations, there are two broad classes of effects that determine the present day power spectrum—those processes that happened before recombination and those that happened after. The post-recombination effects include scattering off the reionized electrons in the modern universe (seen in Kogut et al., 2003), anisotropies introduced because of the time-varying potential along the flight path of a photon called the integrated Sachs-Wolfe effect, an overall size scaling in ℓ of the power spectrum set by the angular diameter distance to the surface of last scattering, and heating of CMB photons on small scales due to Compton scattering off hot gas in clusters, called the Sunyaev-Zeldovich effect. Before recombination, the photons were locked in place with the baryons, and so they carry the information about the state of the baryons at 400,000 years. The baryon/photon fluid underwent acoustic oscillations as overdense regions collapsed due to gravity, then expanded from pressure, while the dark matter continued to collapse. Because the fluctuations all started in phase at the big bang, the sound speed was uniform throughout the universe, and we see

a short period of time at the surface of last scattering, the phase of fluctuations at the surface of last scattering is only dependent on their wavelengths. So we expect to see the power rising as we go to smaller scales up until the length where the fluctuations are at their maximal compression (at $\ell \sim 200$ for a flat universe). As we move to smaller scales, the power will drop as the scale length moves towards modes that have completed their first compression and are expanding back to a density null (but a peak in the velocity). Then we will see modes that have compressed, re-expanded, and hit the point of maximal expansion, for another peak in the power spectrum. And so on down to ever smaller scales that have completed more and more oscillations by the surface of last scattering. So, we expect to see peaks and dips in the angular power spectrum of the CMB. The details are very sensitive to the exact conditions of the universe, though. Dark matter has no pressure, and so rather than oscillate it will continue to collapse, and try to pull the photon-baryon fluid with it through gravity. On small scales, photons will diffuse out of the fluctuations, reducing power exponentially in a process called Silk damping (Silk, 1968). On larger scales, photons are gravitationally redshifted by climbing out of the potential wells of the perturbations, called the (non-integrated) Sachs-Wolfe effect (Sachs & Wolfe, 1967). The effect is 1/3 that expected solely due to gravitational redshifting because time dilation at the surface of last scattering partially cancels the gravitational redshift, since it causes the photons to appear to come from a younger, hotter universe (*c.f.* Peacock, 1999).

As the fluid collapses, the more baryons there are driving the infall, the more pressure the photons have to exert before they can turn the collapse around, leading to an increase in power in the odd numbered (compression) peaks. Power on small scales is also reduced because of the finite thickness of the surface of last scattering. Instead of seeing a single fluctuation, as is the case for large-scale modes, a single point on the sky will have contributions from the number of small modes that can fit into the finite recombination thickness. Consequently, the average temperature anisotropy drops from purely geometric effects on small scales (in addition to the reduction from Silk damping). This can be used to test, *e.g.*, non-standard recombination theories (for instance, if the fine-structure constant α varies with time). Because the amplitude at the surface of last scattering is proportional to the initial amplitude of the fluctuations, we also expect to be able to see the imprint of the

primordial fluctuations in the microwave background. It is precisely because the evolution of the fluctuations is sensitive to so many fundamental parameters that detailed observations of the CMB fluctuations can determine many fundamental parameters.

I have found a few simple rules helpful when trying to understand the behavior of the power spectrum that will be illustrated in Figures 1.1 through 1.6, which plot sample power spectra. All spectra were calculated using CMBFAST. The unit of density used in cosmology is Ω, which is the fractional density of a component relative to the critical density required to make the universe flat. For matter densities, this is not the important density. Rather, the important density is the physical density at the surface of last scattering, which (absent the creation or destruction of particles) is the same as the physical density today, scaled by the relative volumes of the universe, $(1+z)^3$. Because the critical density depends on the Hubble constant like H_0^{-2}, a fixed physical density will be proportional to $H_0^2\Omega$. In keeping with astronomical tradition, the Hubble constant will be listed as $100h$km/s/Mpc. So, the physical density of the component of the universe x will be given as $\Omega_x h^2$, which is sometimes also written in the literature as ω_x. For these figures, unless explicitly varied, the baryon physical density $\Omega_B h^2$, the cold dark matter density $\Omega_{cdm} h^2$, and the total matter density $\Omega_m h^2 \equiv \Omega_B h^2 + \Omega_{cdm} h^2$ will be kept fixed, unless explicitly varied. The other cosmological parameters that specify the models are the spatial curvature of the universe Ω_k, the scalar power-law index of the primordial fluctuations n_s, the cosmological constant Ω_Λ, and the optical depth due to reionization τ_c. The Hubble constant is implicitly defined through the relation $\Omega_k + \Omega_\Lambda + \Omega_m = 1$. The fiducial model in the plots is $\Omega_k = 0$ (flat universe), $h = 69$, $\Omega_B h^2 = 0.023$, $\Omega_m h^2$=0.143, $\Omega_\Lambda = 0.699$, $n_s = 1.0$, and $\tau_c = 0$, with one parameter varied in each set. When Ω_k, $\Omega_B h^2$, $\Omega_m h^2$, and h were varied, Ω_Λ was varied to maintain $\Omega_k + \Omega_\Lambda + \Omega_m = 1$. Rather than the traditional normalization to COBE-DMR at low-ℓ, I normalize the plots to the value at the first peak. This is often more illustrative than the traditional normalization, for instance, in the $\mathcal{C}_\ell$ as a function of h plot.

There is a distinction between the power spectrum at the surface of last scattering and the power spectrum we observe today, because of effects along the line of sight. If a signal originates

at a given redshift, it cannot be coherent on scales larger than the horizon size at that redshift, so we expect the signature of events between the surface of last scattering and the present day to be primarily concentrated at low-ℓ, while the fluctuations intrinsic to the surface of last scattering to appear predominantly at high-ℓ. One such effect is from the reionization of the universe by stars at a comparatively recent redshift. When reionization happens, CMB photons will scatter off the newly free electrons. Since the scattering happens through large angles, it essentially leads to an average scattered component equal to the mean CMB temperature as seen by the scattering electron. That scattering will average out over scales smaller than the electron's horizon size, but not over larger scales. Since the electron density after reionization will fall like $(1+z)^3$, most of the scattering will happen near the redshift of reionization, so the effect on the spectrum will be roughly to reduce the amplitude on scales smaller than the horizon by $\exp(-\tau)$ while leaving the larger scales mostly untouched. This is indeed the case, as can be seen in Figure 1.3. Another important large-ℓ secondary anisotropy is the integrated Sachs-Wolfe effect, which is the heating or cooling of photons as they travel through a changing gravitational potential. If a potential weakens as a photon travels through it (*e.g.*, from a matter overdensity expanding with the Hubble flow), then the blueshift as the photons falls into the potential well will be larger than the redshift as the photon climbs out. This is the one place that the cosmological constant Λ can effect the CMB spectrum (other than an its effect on Ω_k, which doesn't change the shape of the spectrum), since larger values of Λ in a flat universe mean that the expansion is Λ-dominated earlier, and so the integrated Sachs-Wolfe contribution to the spectrum is larger in amplitude and happens on smaller scales. This effect is clearly seen in Figure 1.4, which keeps Ω_k and the physical matter densities $\Omega_B h^2$ and $\Omega_m h^2$ fixed while trading between h and Λ. As h increases, Ω_B and Ω_m decrease to keep the physical densities fixed, leading to a higher value of Λ to keep the universe flat. This shows up at very low-ℓ (about $\ell = 10$) as increased power, with the spectrum otherwise unchanged.

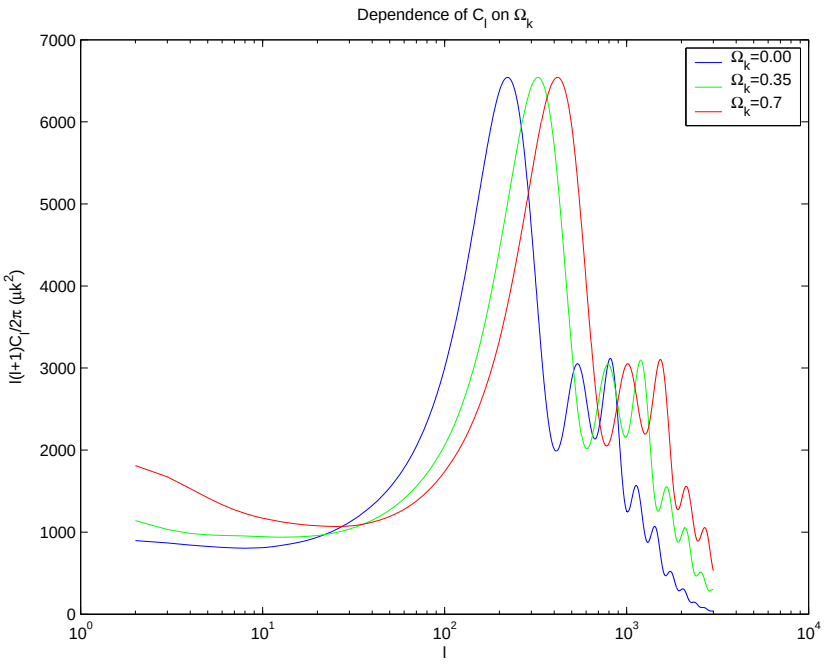


Figure 1.1 Dependence of $\mathcal{C}_\ell$ on Ω_k, the flatness of the universe while keeping the physical matter density fixed. The curvature of the universe doesn't affect the physical structure at the surface of last scattering, since the universe was highly matter+radiation dominated then. It can only affect the angular diameter distance D_A to the surface of last scattering, so the acoustic peaks are shifted to larger ℓ as the universe become less dense, without changing the structure of the peaks. Conveniently, D_A is sensitive predominantly to the overall spatial curvature of the universe, and only weakly sensitive to which individual constituents dominate. This is why the position of the first peak, which is really a direct measure of D_A, is so useful as a measure of the flatness of the universe. The low-ℓ structure is from the integrated Sachs-Wolfe effect as density perturbations along the line of sight in the intervening stretches of the universe evolve.

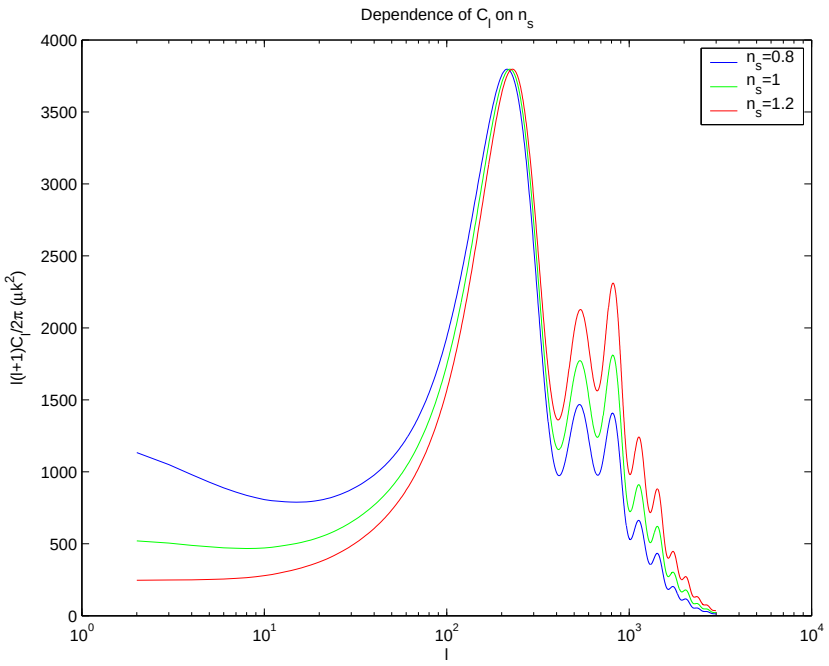


Figure 1.2 Dependence of $\mathcal{C}_\ell$ on n_s, the power law index of the primordial fluctuations. Inflationary theories predict a value slightly less than one. Measurement over very broad ℓ ranges increases the sensitivity to n_s. There has been a recent suggestion (Spergel et al., 2003) that the initial spectrum may have been more complicated than a simple power law.

1.4 Microwave Background Observations

The first detection of anisotropy in the CMB was that of the Differential Microwave Radiometer (DRM) on COBE (Smoot et al., 1992), which measured the power spectrum on scales of $\sim 10^\circ$. Ever since, there has been a flurry of activity in the field. The first generation of post-COBE experiments (*e.g.* Bond et al. (2000) for a list) concentrated on measuring the first acoustic peak, which for a flat universe is on angular scales of about a degree, or $\ell \sim 200$. Many experiments dectected anisotropies, but no single experiment succeeded in convincingly detecting a peak internally, though TOCO (Miller et al., 1999) came tantalizingly close. The combined set of experiments suggested the presence of a peak, but the heterogeneous nature of the data and the comparatively large errors of any single data set made the peak in the ensemble set somewhat questionable. The field changed dramatically with the first unambiguous detection of an acoustic peak by BOOMERANG (de Bernardis et al., 2000), followed shortly by MAXIMA (Hanany et al., 2000). The peak was just where it had been

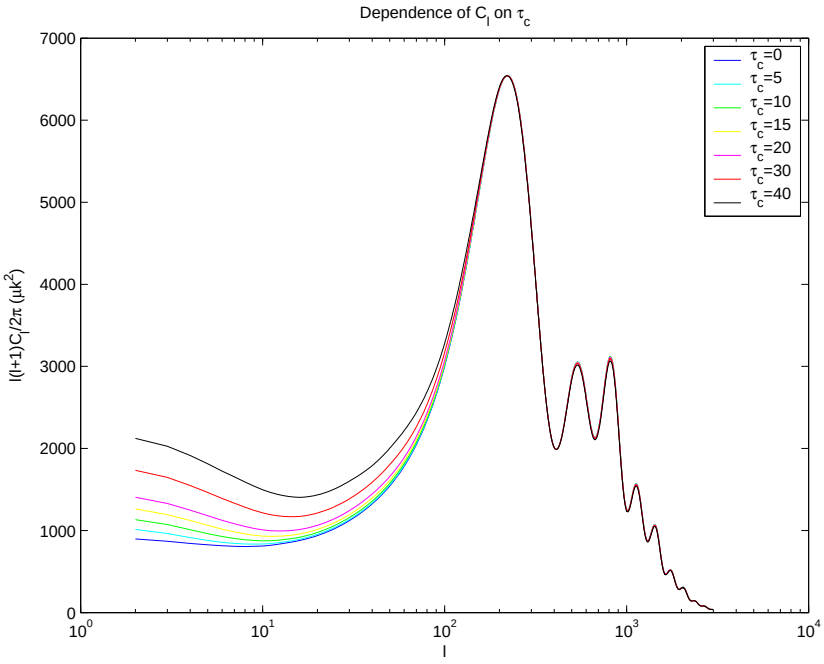


Figure 1.3 Dependence of $\mathcal{C}_\ell$ on τ_c, the optical depth in the local universe to the surface of last scattering. The assumption is that the universe reionized quickly at a given redshift and has remained largely ionized ever since. The CMB gets averaged out on scales smaller than apparent horizon size at recombination, but is largely untouched on larger scales. So, reionization picks out a special ℓ, and fluctuations smaller than that ℓ are suppressed relative to fluctuations larger than that ℓ. Since the plot is normalized so that the models are equal to each other at their peaks, this shows up as an amplification of power at small ℓ at τ_c increases. The higher τ_c is, the earlier the universe must have reionized to reach that optical depth, so the break in the spectrum will happen at larger ℓ (smaller scales) for higher values of τ_c, in addition to the relatively greater suppression at high ℓ.

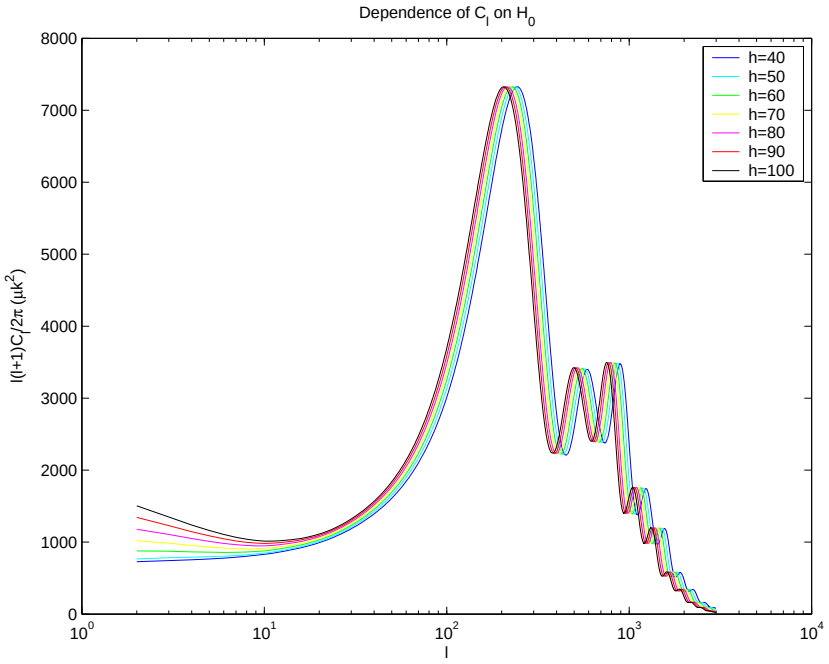


Figure 1.4 Dependence of $\mathcal{C}_\ell$ on H_0, the Hubble constant. This is an example of a degeneracy in the microwave background. If we keep the physical densities of matter components $\Omega_B h^2$ and $\Omega_m h^2$ fixed as we vary H_0, then the physical densities at recombination will also remain unchanged. This plot keeps $\Omega_B h^2$ and $\Omega_m h^2$ fixed, changing Λ to keep the universe flat for different values of H_0. The slight horizontal shifting for the different models is due to the degree to which D_A is sensitive to the constituents of the universe rather than just to its flatness. It is precisely this degeneracy between H_0 and D_A that makes the CMB, by itself, unable to measure Λ. There is a difference at low-ℓ because the Sachs-Wolfe effect is changed by the different expansion history, but it can be mimicked by other factors such as τ_c. The intrinsic cosmic variance on such large scales $\ell \sim 10$ also makes precision determinations of Λ solely through the CMB difficult.

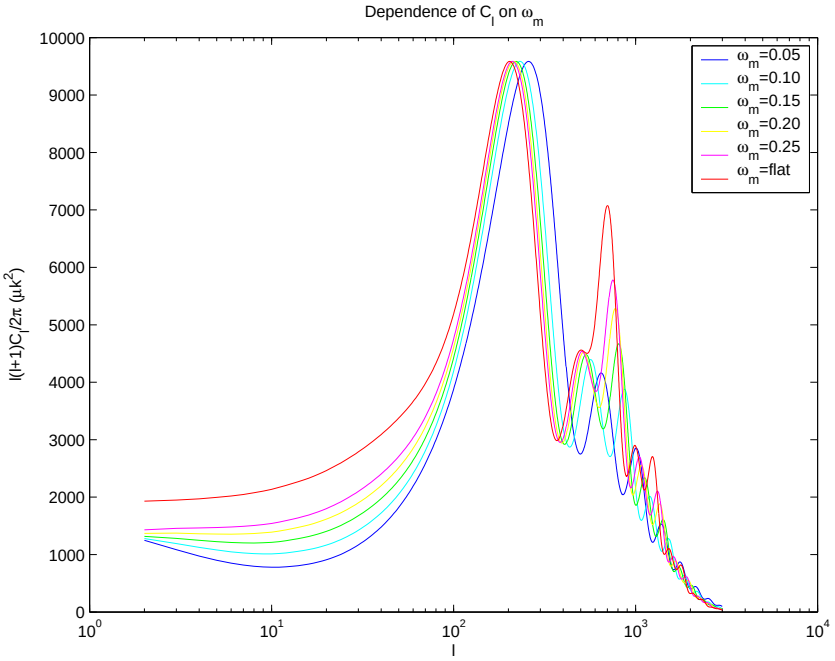


Figure 1.5 Dependence of $\mathcal{C}_\ell$ on $\Omega_m h^2$. Same as Figure 1.1, only varying the dark matter content while keeping the universe flat and h fixed.

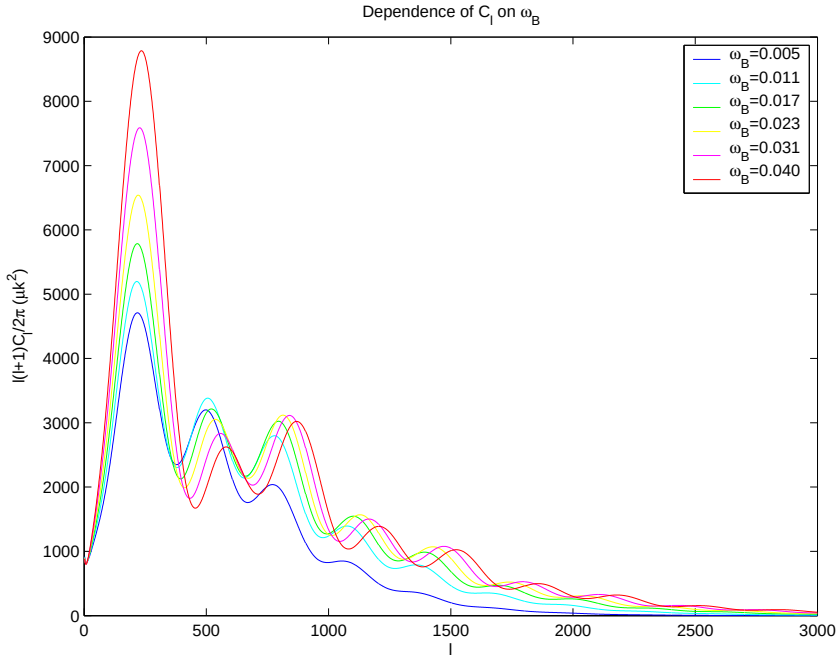


Figure 1.6 Dependence of $\mathcal{C}_\ell$ on $\Omega_B h^2$. This figure has been plotted on a linear scale without normalization to make the behavior in the second and third peaks easier to see. Note how the second peak amplitude drops and the third peak rises as $\Omega_B h^2$ increases. Also note that power is suppressed at higher ℓ by low values of $\Omega_B h^2$ because photons diffuse faster with fewer baryons to hold them in place, washing out power on small scales.

predicted to be if the universe were flat. In addition to measuring the first peak, BOOMERANG and MAXIMA probed smaller angular scales as well, beginning to unlock the information contained in the spectrum at higher ℓ. The BOOMERANG and MAXIMA spectra were joined in short order by spectra from CBI (Padin et al., 2001a; Mason et al., 2003; Pearson et al., 2003), DASI (Halverson et al., 2002), VSA (Scott et al., 2003; Grainge et al., 2003), ARCHEOPS (Benoît et al., 2003), and ACBAR (Runyan et al., 2003), as well as improved spectra from BOOMERANG (Netterfield et al., 2002; Ruhl et al., 2002) and MAXIMA (Lee et al., 2001). Of note are the first detection of the damping tail by CBI (Padin et al., 2001a), the first detection of the polarization signal of the CMB by DASI (Kovac et al., 2002), and a possible first detection of secondary anisotropy from the Sunyaev-Zeldovich effect by the CBI (Mason et al., 2003; Bond et al., 2002b), later joined on the same angular scales by BIMA (Dawson et al., 2002) and ACBAR. This second generation of ground-based or balloon-born experiments has been characterized by high signal-to-noise ratio (SNR) measurements of the CMB spectrum over large ranges of angular scales. This permits single experiments to trace out important structures in the power spectrum. In addition, the different power spectra are in good agreement (see, *e.g.*, Sievers et al., 2003), which gives one confidence in them. This second generation is being brought to completion by the WMAP satellite and its all-sky power spectrum (Hinshaw et al., 2003), which is very good to $\ell \sim 600$ and cosmic-variance limited to $\ell \sim 350$. It is worth stressing that where WMAP is cosmic variance limited, it has used all the information present in the full sky. No future experiment will be able to substantially improve the total-intensity spectrum through the first peak.

There will be two main thrusts in future microwave background observations. The first is to do an ever better job of measuring the power spectrum on smaller scales. There will be an improvement through the second and third peak region of the spectrum as WMAP continues observing. Upcoming experiments, such as Planck, the Atacama Cosmology Telescope, and the South Pole Telescope, will also improve the spectrum out to higher ℓ, with the hope of eventually finding galaxy clusters because of their imprint on the CMB through the Sunyaev-Zeldovich effect (see, *e.g.*, Komatsu & Seljak, 2002; Bond et al., 2002b, for current estimates of the effect). The other thrust is to measure the

polarization of the microwave background. DASI (Kovac et al., 2002) first measured the polarization power spectrum, though the spectrum is too noisy to have cosmologically useful information. Shortly thereafter, WMAP measured the cross-correlation spectrum of the polarization and total-intensity anisotropies on large scales (Kogut et al., 2003). This large-angle spectrum contains information about the optical depth to the surface of last scattering. The optical depth comes from free electrons after hydrogen has been ionized by the first sources of light in the universe. Because the scattering from electrons is polarized, and the radiation scattered is the CMB as seen by the scattering electrons, τ_c introduces a correlation between the total intensity and the polarization of the CMB. It is this that allowed WMAP to break the degeneracies in total intensity to measure τ_c and find that the universe reionized at $z = 20 \pm 10$. Future measurements will refine this number.

1.5 Interferometers

I give here a brief description of interferometers, along with some of the terminology used throughout this thesis. For more quantitative details as to the response of interferometers, especially with regards to CMB observations, see Chapter 3. Radio frequency interferometers are an important part of microwave background research. The CBI, along with DASI and the VSA, are radio interferometers. The remaining second-generation ground/balloon based CMB experiments use bolometers to map the total intensity of the CMB in maps. An interferometer consists of an array of collection devices (usually parabolic dishes, but sometimes feed horns as is the case with DASI and VSA), with a *receiver* at the focus of each dish sensitive to the incoming electric field. The receiver amplifies the electric field, then usually the signal is mixed down to lower frequencies and perhaps split into channels. The receiver outputs are then fed into the correlator which multiplies the signals from each pair of receivers and integrates the product. The fundamental measurement produced by an interferometer is this integrated signal product, called a *visibility*. Because incoming electric fields have amplitudes and phases, the visibilities need amplitudes and phases as well, which makes them complex, so each visibility really has two independent pieces of information. The *baseline* is the pair of antennas that were combined to give the visibility. The baseline is usually referred to either using

the antenna pair, or more commonly, using the separation vector of the dishes either in physical distance or in wavelengths. The vector position of the baseline in wavelength is known as the UV position of the visibility, and the total set of UV points observed by an interferometer is called the UV coverage. The areas in the UV plane covered by observations sets the total range of scales to which the interferometer is sensitive. The noise, usually dominated by thermal noise in the receivers, is independent for different visibilities. There can be correlated noise between different visibilities, but in a well-designed instrument it should be very small, with the receiver cross-talk in the range of -110 to -130 dB for the CBI's most closely-spaced dishes (Padin et al., 2000).

The response of a visibility to the signal on the sky depends both on the separation of the dishes and the details of the collecting element. Perhaps the easiest way to visualize the output of an interferometer is to run the signal backwards and think of the receiver as a transmitter. For the case of a single dish, there will be a single-aperture diffraction pattern on the sky that is the Fourier transform of the collecting aperture. The power pattern on the sky is called the *primary beam* and is the Fourier transform of the square of the electric field response of the dish. It typically has a large response in the center, with ripples extending out to large angles falling in amplitude. The surrounding ripples are called *sidelobes*. Sidelobes are undesirable because they make the interferometer respond to (usually unknown and possibly changing) sources far away from the position on the sky where the dish is pointed, called the *pointing center*. Consequently, there is often some sort of taper applied to the dish to make the sidelobes fall off more quickly, at the expense of a slight broadening of the main part of the beam and a reduction in sensitivity. In the CBI dishes, we use a Gaussian taper since the reduction in sidelobes is so important. For the case of a two-element baseline, a phase modulation gets applied to the primary beam because the radiation from the two receivers alternately goes in and out of phase, with the wavevector k of the visibility on the sky equal to the vector separation of the baseline in wavelengths, which is the UV coordinate of the baseline. Note that each element is sensitive to the electric field, so the product of two baselines will be sensitive to the square of the electric field, which is precisely the single dish power pattern, if the two primary beams are the same. So, running the radiation from the sky to

the receivers again, a visibility will be equal to the integral of primary beam times a plane wave on the sky times the sky signal. In Fourier space, the multiplication becomes a convolution, and we have that the visibility is equal to the Fourier transform of the sky convolved with the Fourier transform of the primary beam, sampled at the UV coordinate of the baseline (see Chapter 3 for quantitative details of the response). It is precisely this property that makes interferometers well suited for CMB observations: on small scales, the power spectrum C_ℓ is equivalent to the Fourier space power spectrum, which is exactly what an interferometer measures, modulo the smearing by the primary beam. Unlike the bolometer experiments where each pixel sample the entire range of ℓ up to the pixel size, interferometer data are localized in ℓ. Other advantages of interferometers are ease of measurement of the primary beam (notoriously difficult for balloon-born bolometers), stable calibration, and well-behaved noise properties since the visibilities have independent noises.

1.6 The Cosmic Background Imager

The Cosmic Background Imager (Padin et al., 2002) is a special purpose interferometer located in the Atacama desert of northern Chile. The site is both high and dry, making it an excellent place for centimeter-wavelength observations (though a non-negligible fraction of the time has been lost due to weather. See Figure 1.7). The CBI has 13 low-noise HEMT receivers, with a total system temperature of about 30 Kelvin, co-mounted on a 5.5 m rotating deck. The receivers accept a single circular polarization. During the observations described here, 12 receivers were set to measure left circular polarization, with the thirteenth receiver set to right circular polarization, in order to retain some polarization sensitivity. The polarization results are described elsewhere (Cartwright, 2002). The signals are downconverted and split into 10 1GHz channels between 1 and 2 GHz that are then combined using a high-speed analog correlator (Padin et al., 2001b). Rather than be locked into a single observing pattern, the dishes can be moved around the telescope mount in order to give the CBI maximum flexibility in its UV coverage. Each of the 10 channels per baseline is recorded separately, and since the fractional bandwidth is wide (R$\sim$3), each baseline covers a fractional width in UV space of about 30%. We also rotate the deck during observations to fill out the UV plane

Figure 1.7 The CBI site, which is also the future ALMA site, has been touted by many others as one of the driest, highest places in the world. The author is on the right.

tangentially without having to wait for the Earth to do the rotation for us. As a consequence, we have very dense UV coverage (see Figure 3.1 for a sample of the CBI UV coverage). The deck can be tilted to an angle of 42.75 degrees above the horizon, limiting the CBI observations to roughly $-70 < \delta < +24$, and limiting the length of time a single source can be tracked to < 6.5 hours, depending on the declination.

Prior to shipping the CBI to Chile, we assembled and tested it on the Caltech campus in Pasadena. The initial construction period was from early 1998 to August 1999. I worked on the CBI at that time, assembling and testing the receivers (see Figure 1.8). The construction was completed sufficiently for first light in Pasadena in January 1999, using three receivers. During the testing in Pasadena, we found that the CBI worked well, but that ground spillover in the sidelobes of the small dishes was substantial (see Section 3.2 for further discussion). After several months of testing, we disassembled the CBI and shipped it to Chile in August of 1999. Once there, it was transported to the site and reassembled, with first light on-site in December 1999. The first science observations of the microwave background were taken January 12 of 2000, and, apart from maintenance, repairs, and upgrades, the CBI has been taking data ever since (weather permitting). The first two years were devoted to total intensity measurements of the power spectrum, with the CBI switching in the fall of 2002 to predominantly polarization observations.

Figure 1.8 The author building the CBI receivers.

Chapter 2

Maximum Likelihood

Our task is to measure C_ℓ as accurately as we can. The conceptually simplest case is that of an all-sky map with no noise or contaminating signals, such as point sources or diffuse galactic foregrounds. In that case we could simply decompose the sky into its constituent modes and measure their variances. A real experiment is complicated by partial sky coverage (which can introduce apparent correlations between the $a_{\ell m}$), noise, point sources, galactic foregrounds, *etc.* But at its heart, CMB analysis is still nothing more complicated than measuring the variance of a data set.

2.1 Uncorrelated Likelihood

We can better understand how to measure the power spectrum by starting with the simple case of a single Gaussian random variable and then adding more and more complexity to the problem. For a single Gaussian random variable x with zero mean and variance $V = \sigma^2$, the PDF is

$$PDF(x) = \frac{\exp(-\frac{x^2}{2V})}{\sqrt{2\pi V}} \tag{2.1}$$

This is the probability density that we would get a certain value for x given the underlying variance. This can also be thought of as the likelihood that we would have gotten the observed data point x if the underlying variance were V. This interpretation gives rise to the method of Maximum Likelihood estimation of the variance. Our estimated value of V is that which would have yielded the observed data set with the highest probability. As an aside, note that in Bayesian terms we are setting

$P(V|x) = P(x|V)$. This is equivalent to the standard Bayesian expression $P(V|x) = P(x|V)P(V)$ with a uniform prior on V, *i.e.*, all values of V are equally probable by assumption. While not important for maximum likelihood estimation, this does show how in principle we could include prior knowledge of likely power spectrum or cosmological parameter values.

For the case of a single value, the maximum likelihood estimator for the variance is set by maximizing the likelihood with respect to V. We usually work with the log of the likelihood rather than the likelihood itself, as the log likelihood is mathematically simpler to use.

$$\log(\mathcal{L}) = -\frac{1}{2}\frac{x^2}{V} - \frac{1}{2}\log 2\pi V \tag{2.2}$$

The derivative is

$$\frac{d\log(\mathcal{L})}{dV} = \frac{1}{2}\frac{x^2}{V^2} - \frac{1}{2V} \tag{2.3}$$

If we set that equal to zero and solve for V, we find the standard result $V = x^2$ – our estimate of the variance is equal to the actual variance of the data point. The extension to many independent, identically distributed data points is straightforward. Because they are independent, the joint likelihood is merely the product of the individual likelihoods. In log likelihood space, the joint log likelihood is the sum of the individual likelihoods. We typically ignore the additive constants to the log likelihood since they don't affect the position of the peak or the shape of the likelihood surface around that peak. The log likelihood is then

$$\log(\mathcal{L}) = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i^2}{V} - \log(V)\right) \tag{2.4}$$

We can again maximize with respect to V to get

$$\frac{d\log(\mathcal{L})}{dV} = \frac{1}{2V^2}\sum_{i=1}^{n}(x_i^2 - V) = 0 \tag{2.5}$$

Again, this has a familiar solution $V = \sum x_i^2/n = \bar{x_i^2}$, our estimate of the variance is just the average variance of the data set. We can also rewrite the derivative as follows by pulling out a factor of V

from inside the sum

$$\frac{d\log(\mathcal{L})}{dV} = \frac{1}{2V}\sum_{i=1}^{n}\left(\frac{x_i^2}{V} - 1\right) = 0 \tag{2.6}$$

Note that the definition χ_i^2 is x_i^2/V, hence the maximum of the likelihood is the point where the average value of χ^2 is equal to one.

Real data usually have many contributions to their variance (signal, noise...), of which we may only be interested in fitting for a single one. Also, each data point can have a different expected response under a certain model. If we have a simple experiment that takes uncorrelated noisy data, then the expected variance of a data point is $V_i = qS_i + N_i$, where N_i is the (Gaussian) variance due to noise of the i^{th} data point, q is an overall amplitude we wish to measure, and S_i is the response of the i^{th} data point to a unit amplitude q. In principle, we could have a more sophisticated dependence on the parameter q which would complicate derivatives, but in practice that is a sufficiently flexible model for the CMB variance. In this case, we wish to maximize the likelihood as we vary q

$$\log(\mathcal{L}) = \sum -\frac{1}{2}\frac{x_i^2}{qS_i + N_i} - \frac{1}{2}\log(qS_i + N_i) \tag{2.7}$$

$$\frac{d\log(\mathcal{L})}{dq} = \sum \frac{1}{2}\frac{x_i^2}{(qS_i + N_i)^2}S_i - \frac{1}{2(qS_i + N_i)}S_i = 0 \tag{2.8}$$

This has a solution where

$$\sum \frac{qS_i}{qS_i + N_i}\left(\frac{x_i^2}{qS_i + N_i} - 1\right) = 0 \tag{2.9}$$

(with an extra factor of q multiplied on both sides). We are still setting the average value of χ^2 equal to one, but this time subject to a set of weights. Note that the total signal variance (or square of the signal amplitude) is qS_i, so the i^{th} weight is $\frac{Signal}{Signal+Noise}$. So, the condition at the maximum is

$$\sum \frac{qS}{qS + N}\left(\chi^2 - 1\right) = 0 \tag{2.10}$$

Note that as we change our model (by changing q), in addition to χ^2 changing, the weights also change. This is why maximum likelihood is non-linear. The weight is $(Signal/Noise)$ for small

signals and asymptotically approaches one for signals much larger than the noise. This means that once we have reasonably well determined a data point, a better measurement of that point does not significantly improve our estimate of q — we are better served by measuring more data points. This is known as the cosmic variance limit, and is the reason why CMB experiments try to cover as much sky as possible (more x_i's). The extension to many signal components is straightforward—maximum likelihood continues to try to set the weighted values of χ^2 equal to one.

2.2 Correlated Power Spectrum

Experimental data are typically correlated, and so the simple techniques of the preceding section are not directly applicable to real life situations. Fortunately, they can be extended to correlated data. First, note that the log likelihood for uncorrelated data can be written as a set of matrix operations

$$\log(\mathcal{L}) = -\frac{1}{2}\boldsymbol{x}^T\Lambda^{-1}\boldsymbol{x} - \frac{1}{2}\log(|\Lambda|) \tag{2.11}$$

with Λ the diagonal matrix whose elements Λ_{ii} are simply the variances of the x_i. (A quick work on notation: In general in this thesis, bold quantities are vectors, capitalized Roman letters are matrices (or single elements of matrices if subscripted), and other quantities are italicized in equations.) Noting that the determinant of a diagonal matrix is the product of the diagonal elements, and the inverse of a diagonal matrix is the same matrix with the elements along the diagonal inverted, the individual multiplications, divisions, etc., we carry out are identical for both the standard uncorrelated data representation and the matrix representation of the likelihood. We can then use machinery of matrix mathematics to transform the case of uncorrelated data into a realistic, correlated problem. To proceed, introduce an orthogonal matrix V (distinct from the uncorrelated variable variance V, to which we no longer refer). An orthogonal matrix has the propery that the i^{th} column dotted with the j^{th} column is $\delta_{i,j}$ – in other words, its transpose is its inverse. It is also true in general that the determinant of the product of two matrices is the product of their individual determinants, and that the determinant of the transpose is the same as the determinant of the original matrix. So, we

have

$$|\mathrm{V}^T\mathrm{V}| = |\mathrm{I}| = 1 \longrightarrow |\mathrm{V}|^2 = 1 \tag{2.12}$$

We can transform the uncorrelated likelihood using this matrix V while leaving the likelihood unchanged.

$$\log(\mathcal{L}) = -\frac{1}{2}\boldsymbol{x}^T\Lambda^{-1}\boldsymbol{x} - \frac{1}{2}\log(|\Lambda|) = -\frac{1}{2}\boldsymbol{x}^T\mathrm{V}\mathrm{V}^T\Lambda^{-1}\mathrm{V}\mathrm{V}^T\boldsymbol{x} - \frac{1}{2}\log(|\mathrm{V}^T\Lambda\mathrm{V}|) \tag{2.13}$$

The likelihood is identically unchanged because inserting $\mathrm{V}\mathrm{V}^T$ is simply multiplying by unity, and the determinant is multiplied by $|\mathrm{V}|^2$, which we have already shown to be one as well. We can now group terms using the definitions $\boldsymbol{\Delta} \equiv \mathrm{V}^T\boldsymbol{x}$ and $\mathrm{C} \equiv \mathrm{V}^T\Lambda\mathrm{V}$. The likelihood then becomes

$$\log(\mathcal{L}) = -\frac{1}{2}\boldsymbol{\Delta}^T\mathrm{C}^{-1}\boldsymbol{\Delta} - \frac{1}{2}\log(|\mathrm{C}|) \tag{2.14}$$

This is the standard expression for the likelihood of a theory under a particular data set that starts off most microwave background analysis papers. The meaning of V and Λ are now clear: they are the matrix of eigenvectors and their corresponding eigenvalues of the matrix C. Unfortunately, in general we cannot work in the diagonal space because as we change the theory, both the eigenvectors and eigenvalues change, and so a fixed transform does not remain diagonal. We need one more result before this becomes practically useful, namely, how do we compute C?

First, let us find the covariance of two data points. Using the definition of $\boldsymbol{\Delta}$, we have

$$\boldsymbol{\Delta}_i = \sum \mathrm{V}_{i,j}\boldsymbol{x}_j \tag{2.15}$$

and the expectation of the product of two $\boldsymbol{\Delta}_i$'s is

$$\langle \boldsymbol{\Delta}_i\boldsymbol{\Delta}_j \rangle = \left\langle \sum \mathrm{V}_{i,k}\boldsymbol{x}_k \sum \mathrm{V}_{j,l}\boldsymbol{x}_l \right\rangle \tag{2.16}$$

Since the $\boldsymbol{x}_i$ are independent, any term with k not equal to l has an expected value of zero. Also note that $\left\langle \boldsymbol{x}_i^2 \right\rangle = \Lambda_i$, leaving

$$\langle \boldsymbol{\Delta}_i \boldsymbol{\Delta}_j \rangle = \sum \mathrm{V}_{i,k} \mathrm{V}_{j,k} \Lambda_k \tag{2.17}$$

Now, what are the components of the transformed matrix C? Multiplying V on the left by Λ multiplies the rows of V by the corresponding element of Λ

$$\mathrm{V}_{i,k} \rightarrow \Lambda_k \tag{2.18}$$

We get the final answer for the element of C by multiplying by the inital V^T. The i, j^{th} element of the product of two matrices is the i^{th} row of the first times the j^{th} column of the second. Since the first matrix is the transpose of V, the i^{th} row of V^T is the i^{th} column of V. So, we have the following expression for the elements of C

$$\mathrm{C}_{i,j} = \sum \mathrm{V}_{i,k} \mathrm{V}_{j,k} \Lambda_k \tag{2.19}$$

But this is exactly the expectation value from Equation 2.17! So, in order to calculate the likelihood of a theory, we need only calculate the expected covariance of pairs of data points under that theory, and then calculate the likelihood using Equation 2.14. It is because the matrix C is made up of the data covariances that is is known as the *covariance matrix*. Because $\boldsymbol{\Delta}_i \boldsymbol{\Delta}_j = \boldsymbol{\Delta}_j \boldsymbol{\Delta}_i$, the covariance matrix is symmetric. The problem of measuring the power spectrum then falls into two fairly distinct parts: The first is calculating C for our data set $\boldsymbol{\Delta}$ for different theories, the second is how to efficiently find the theory that maximizes the likelihood, as well as characterizing the likelihood surface around that peak. Because typical data sets can have upwards of hundreds of thousands of data points, and calculating the likelihood is an order n^3 operation, considerable care is required in both parts to make the problem computationally feasible. For instance, the CBI extended mosaics have ∼800,000 distinct real and imaginary data points. A 2 GHz processor would

then take of order $(8\times10^5)^3/2\times10^9 \sim 10$ years to invert the matrix, and would require ~ 5 terabytes of memory to store it! Clearly, great care must be taken when creating C to make it as small as possible, and then one must work with it as efficiently as possible.

2.3 Likelihood Gradient

It is now time to find the Maximum Likelihood spectrum. One often sees the likelihood that a given spectrum would give rise to an observed *complex* data set written as (*e.g.* White et al., 1999)

$$\mathcal{L}(\mathcal{C}_\ell) = \frac{1}{\pi^n |\mathrm{C}|} \exp\left(-\mathbf{\Delta}^\dagger \mathrm{C}^{-1} \mathbf{\Delta}\right) \tag{2.20}$$

The missing factors of two relative to Equation 2.14 are because each visibility is really two independent points, one real and one imaginary, combined. The rest of this section will use the form of Equation 2.14 with the understanding that all complex measurements have been split into two real data.

Our task is to vary $\mathcal{C}_\ell$, which changes the covariance matrix C, until we have reached the maximum of the likelihood. We restrict ourselves to models of the form

$$\mathrm{C} = \sum_B q_B \mathrm{W}_B + \mathrm{N} \tag{2.21}$$

where N is our generalized noise matrix (it could have contributions from thermal noise, correlated noise between visibilities, galactic foregrounds, point sources, ground pickup *etc.*), the q_B are the band powers describing the CMB power spectrum, and the W_B describe the response of the data to those band powers, equivalent to $\frac{d\mathrm{C}}{dq_B}$. We will sometimes refer to the W_B as window matrices (since they are the matrices consisting of the visibility window functions, discussed in Section 3.3 and elsewhere). By restricting ourselves to this form, we can again use the technique of Section 2.2 where we calculate the gradient in the case of uncorrelated data and then transform it to the correlated case. In the next two sections I discuss how to efficiently reach the peak of the likelihood. Provided

the multidimensional search method used is relatively efficient, simply varying the q_B is not a bad way of reaching the peak, and in fact is what we use in Chapter 3. Because to measure the likelihood we need only factor C into the triangular matrix L such that $\mathrm{LL}^T = \mathrm{C}$ (a Cholesky factorization. See below for how to obtain the likelihood), a single calculation of the likelihood can be very much faster than iterations of more sophisticated methods that converge in fewer steps. For instance, using the LAPACK linear algebra library (Anderson et al., 1999) on a Pentium IV, factoring C is about six times faster than inverting it. To see how to get the likelihood from factoring, note that what we really need is $\mathrm{C}^{-1}\boldsymbol{\Delta}$ and $\log|\mathrm{C}|$. To get the determinant, we need merely multiply the diagonal elements of L, and to get $\mathrm{C}^{-1}\boldsymbol{\Delta}$, we solve the system of equations $\mathrm{C}\boldsymbol{y} = \boldsymbol{\Delta}$ which is done in $\mathcal{O}\left(n^2\right)$ time once C is factored.

We can do better than that, though, especially if we are fitting many bins. If we could characterize the likelihood surface around a point, in addition to being able to converge to the maximum more quickly (through, for instance, Newton-Raphson iteration), we could also directly estimate quantities of interest such as errors. Many authors have advocated calculating or approximating the gradient and curvature of the likelihood (Bond et al., 1998; Borrill, 1999*e.g.*), then using Newton-Raphson iteration to find the zero of the gradient. In order to do this, we need to be able to calculate gradients and curvatures of the likelihood. I show here the calculation of the gradient, with the curvature discussed in Section 2.4.

Recall the formula for the derivative of the likelihood of uncorrelated data under these assumptions, Equation 2.8. First let us analyze the second term, originating from the log of the determinant of C

$$-\sum \frac{1}{2(qS_i + N_i)} S_i \tag{2.22}$$

The denominator is the total variance Λ_i^{-1} (inverse since it's in the denominator), while the coefficient is the change in Λ_i with respect to the parameter in question q. So, we would like a matrix operation that will multiply those two sets of numbers and sum them. Fortunately there is such an operation—the trace of a matrix. The trace is the sum of the diagonal elements of a matrix, and has the nice

property that it is the sum of the eigenvalues, and hence is unchanged when we rotate the matrix. So, we can write the term as follows

$$-\sum \frac{1}{2(qS_i + N_i)} S_i = -\frac{1}{2} \sum \frac{\Lambda_{i,q}}{\Lambda_i} = -\frac{1}{2} Tr\left(\Lambda_{,q} \Lambda^{-1}\right) \tag{2.23}$$

where $\Lambda_{,q}$ is the derivative of Λ with respect to the band power q. We can now rotate from Λ to C since the trace is unaffected, giving the general expression

$$-\frac{1}{2} Tr\left(\mathrm{C}_{,q} \mathrm{C}^{-1}\right) \tag{2.24}$$

The first term, which is the χ^2 of the data

$$\sum \frac{1}{2} \frac{x_i^2}{(qS_i + N_i)^2} S_i \tag{2.25}$$

is rather more interesting since there are *two* ways it can be transformed into matrix notation, both of which are useful. It is reasonably straightforward to process it in the diagonal case and then rotate, but is not trivial because some care must be taken when rotating multiple matrices that do not have the same eigenvectors. Instead, I will proceed directly from the matrix description $-\frac{1}{2} \mathbf{\Delta}^T \mathrm{C}^{-1} \mathbf{\Delta}$. We will need the derivative of the inverse of a matrix, which is as follows

$$\frac{d}{dq}\left(\mathrm{C}^{-1}\mathrm{C}\right) = \frac{d\mathrm{C}^{-1}}{dq}\mathrm{C} + \mathrm{C}^{-1}\frac{d\mathrm{C}}{dq} = 0 \tag{2.26}$$

where it is equal to zero because the initial product is the identity matrix (by definition of the inverse), whose derivative is clearly zero. We can then solve for the derivative of the inverse

$$\frac{d}{dq}\left(\mathrm{C}^{-1}\right) = -\mathrm{C}^{-1}\frac{d\mathrm{C}}{dq}\mathrm{C}^{-1} \tag{2.27}$$

We can use this to calculate the derivative (Bond et al., 1998)

$$\frac{d}{dq}\left(\mathbf{\Delta}^T \mathrm{C}^{-1}\mathbf{\Delta}\right) = -\mathbf{\Delta}^T \mathrm{C}^{-1}\mathrm{C}_{,q}\mathrm{C}^{-1}\mathbf{\Delta} = -\mathbf{\Delta}^T \mathrm{C}^{-1}\mathrm{W}_q\mathrm{C}^{-1}\mathbf{\Delta} \tag{2.28}$$

where the final step is because of the parameterization of the spectrum, Equation 2.21. This form has appeared in the literature before (Oh et al., 1999; Borrill, 1999). Since the data vector is constant, it has no derivative.

The other expression for the derivative comes from noting that we can rewrite the first term in the likelihood $Tr\left(\mathbf{\Delta}\mathbf{\Delta}^T \mathrm{C}^{-1}\right)$. An element by element comparison with the standard formula shows that the operations are identical. We can then take the derivative using Equation 2.27, yielding

$$\frac{d}{dq}Tr\left(\mathbf{\Delta}\mathbf{\Delta}^T \mathrm{C}^{-1}\right) = -Tr\left(\mathbf{\Delta}\mathbf{\Delta}^T \mathrm{C}^{-1}\mathrm{C}_{,q}\mathrm{C}^{-1}\right) \tag{2.29}$$

Combining these with Equation 2.24 and evaluating the $\mathrm{C}_{,q}$ gives the final numerically equivalent expressions for the gradient of the likelihood

$$\frac{d\log\left(\mathcal{L}\right)}{dq} = \frac{1}{2}\mathbf{\Delta}^T \mathrm{C}^{-1}\mathrm{W}_q\mathrm{C}^{-1}\mathbf{\Delta} - \frac{1}{2}Tr\left(\mathrm{W}_q\mathrm{C}^{-1}\right) \tag{2.30}$$

$$\frac{d\log\left(\mathcal{L}\right)}{dq} = \frac{1}{2}Tr\left(-\mathbf{\Delta}\mathbf{\Delta}^T \mathrm{C}^{-1}\mathrm{W}_q\mathrm{C}^{-1} + \mathrm{W}_q\mathrm{C}^{-1}\right) \tag{2.31}$$

We are now in a position to see the different utilities of the two expressions. The first is important because it is fast to calculate, once we have the inverse. The χ^2 term requires only matrix times vector operations, which are fast. The determinant term looks like it should require an n^3 operation, but because we take the trace, we need only calculated the diagonal elements of the product, which is an n^2 operation. In fact, the trace of a product can be performed very quickly indeed for symmetric matrices. The jj^{th} element of AB $= \sum_i A_{ij}B_{ji}$, and the trace is the sum of that over i. If the matrices are symmetric, $B_{ij} = B_{ji}$, and the trace is simply $\sum_i \sum_j A_{ij}B_{ij}$. If the matrices are stored, as is usually the case, in a contiguous stretch of memory, then we are simply taking the dot product of an n^2 long vector. This is an extremely efficient way of accessing computer memory for

the trace, especially on multiprocessor machines (Sievers, 2004, *in prep*).

The usefulness of the second expression becomes clear if we introduce an extra factor of CC^{-1} into the determinant term, giving

$$\frac{d\log\left(\mathcal{L}\right)}{dq} = \frac{1}{2}Tr\left(\left(\boldsymbol{\Delta}\boldsymbol{\Delta}^T - \mathrm{C}\right)\mathrm{C}^{-1}\mathrm{W}_q\mathrm{C}^{-1}\right) \tag{2.32}$$

We can see that we reach the maximum of the likelihood, where the gradient is zero, at the point where the matrix formed by the data $\boldsymbol{\Delta}\boldsymbol{\Delta}^T$ "most closely" matches the covariance matrix C. In addition, we can see how the gradient will respond to the addition of an expected signal, which usually requires a matrix to describe rather than a vector. This is the key to understanding the contribution to the power spectrum from other signals, discussed in Section 2.5. Unfortunately, calculating the gradient using this expression is computationally expensive, requiring n_{bin} matrix-matrix multiplications. We can get one matrix multiplication for free because of the trace, but we have to pay for the others. Since we need the derivative for *each* bin, this requires a factor of order the number of bins more work to calculate the gradient using this formula rather than Equation 2.30. When the number of bins becomes large (for the CBI, we have typically around 20), this factor can be the difference between being able to run on a typical desktop machine and having to run on a supercomputer, or the difference between being able to run on a supercomputer and not being able to extract a power spectrum at all.

2.4 Likelihood Curvature

We could use the gradient to get to the likelihood maximum, but it would be nice to have a curvature matrix as well, so we know how far to follow the gradient. We can converge very quickly indeed using Newton-Raphson iteration

$$\boldsymbol{q} \rightarrow \boldsymbol{q} - \mathcal{F}^{-1}\frac{d\log\left(\mathcal{L}\right)}{dq} \tag{2.33}$$

where $\mathcal{F}$ is the second derivative matrix, defined below in Equation 2.34. This is the fundamental algorithm we use to find the set of q_B that give the best fitting spectrum, and once we have $\mathcal{F}$ and $\frac{d\log(\mathcal{L})}{dq}$ for a model, we can update the q_B to get a better fitting model. Fortunately, it turns out that we can get an approximate curvature matrix, which will also work in Newton's method, for only marginally more computational effort than the exact gradient. Let us differentiate both Equations 2.30 and 2.32. Recall that we have by definition restricted ourselves to the class of covariance matrices expressable by Equation 2.21,

$$\mathrm{C} = \sum_B q_B \mathrm{W}_B + \mathrm{N}$$

This means that the only contributions to derivatives come from differentiating C itself, and all other factors are constant. We can differentiate (2.30) to get two equivalent expressions for the curvature matrix

$$\frac{d^2\log(\mathcal{L})}{dq_B dq_{B'}} \equiv \mathcal{F} = -\boldsymbol{\Delta}^T \mathrm{C}^{-1}\mathrm{W}_B\mathrm{C}^{-1}\mathrm{W}_{B'}\mathrm{C}^{-1}\boldsymbol{\Delta} + \frac{1}{2}Tr\left(\mathrm{W}_B\mathrm{C}^{-1}\mathrm{W}_{B'}\mathrm{C}^{-1}\right) \tag{2.34}$$

$$\mathcal{F} = -Tr\left(\left(\boldsymbol{\Delta}\boldsymbol{\Delta}^T - \mathrm{C}\right)\mathrm{C}^{-1}\mathrm{W}_B\mathrm{C}^{-1}\mathrm{W}_{B'}\mathrm{C}^{-1}\right) - \frac{1}{2}Tr\left(\mathrm{W}_B\mathrm{C}^{-1}\mathrm{W}_{B'}\mathrm{C}^{-1}\right) \tag{2.35}$$

We now have some choices we can make as to how to proceed from here. An early suggestion in Bond et al. (*e.g.* 1998) was to note that at the maximum of the likelihood the first term in (2.35) is approximately zero, and so we can approximate the curvature matrix by

$$\mathcal{F} \simeq \mathrm{F} \equiv \frac{1}{2}\mathrm{Tr}\left(\mathrm{W_B}\mathrm{C}^{-1}\mathrm{W_{B'}}\mathrm{C}^{-1}\right) \tag{2.36}$$

This approximation F to the curvature matrix is called the Fisher matrix. It is the expected curvature averaged over many data sets if the current model were true. Calculating the Fisher matrix requires us to both create and store $\mathrm{C}_B\mathrm{C}^{-1}$ for every band, which is requires n_{bin} matrix-matrix multiplications. The program MADCAP (Borrill, 1999), used in de Bernardis et al. (2000), uses Equation 2.34 to calculate the exact curvature $\mathcal{F}$ rather than the Fisher matrix. The first term in

Equation 2.34 is quick to calculate, as it is simply a series of matrix times vector operations. Let us label this term $2\mathcal{D}$. The second term is again the Fisher matrix, only with the opposite sign. So, $\mathcal{F}$ takes about as much effort to calculate as F. So, we have two ways of writing the curvature, one of which is approximate

$$\frac{d^2 \log(\mathcal{L})}{dq_B dq_{B'}} = 2\mathcal{D} - \mathrm{F} \simeq \mathrm{F} \tag{2.37}$$

So it must be true that $\mathcal{D} \simeq \mathrm{F}$, and we have then the key result that

$$\mathcal{F} \simeq \mathcal{D} \tag{2.38}$$

This is a new way of measuring the curvature (Sievers, 2004, *in prep.*) that greatly increases the speed of measuring the spectrum and halves the memory requirements. Why does this do so? Because, with a single inversion of the covariance matrix we can use this equation, along with Equation 2.30 to calculate both the exact gradient and approximate curvature of the likelihood surface! This increases the execution speed by a factor of the number of bins, which for modern experiments is often a few dozen. It is also a more accurate description of the curvature than the Fisher matrix, which has been used successfully for years (including in Mason et al. (2003) and Pearson et al. (2003)). To see this note that

$$\mathcal{F} = 2\mathcal{D} - \mathrm{F} = \mathcal{D} + (\mathcal{D} - \mathrm{F}) = \mathrm{F} + 2(\mathcal{D} - \mathrm{F}) \tag{2.39}$$

So the correction we need to apply to F in order to get $\mathcal{F}$ is twice as large as that required by $\mathcal{D}$. This means the algorithm converges to the maximum of the likelihood in fewer iterations. To calculate F one needs to store the set of matrix products $\mathrm{C}^{-1}\mathrm{W}_B$. This doubles the storage/memory requirements. Because these products are never calculated using $\mathcal{D}$, they don't need to be stored. Practically speaking, using $\mathcal{D}$ means that one can do the analysis in Pearson et al. (2003) on a desktop PC in thirty minutes that took several hours to do using F on a 32 CPU Alpha supercomputer (GS320 with 733 MHz alpha CPUs). While this method had not yet been developed at the time of

our first-year papers, it has since been adopted into our analysis pipeline and will be used for all upcoming spectrum measurements. Also note that we could continue to differentiate $\mathcal{D}$ to be able to approximate the likelihood over successively larger areas. Since, when we are far from the maximum, the error in the step is predominantly due to the third derivative rather than the difference between $\mathcal{D}$ and $\mathcal{F}$, we may be able to converge in fewer steps, though I have yet to investigate this in detail.

Incidentally, the errors in the band powers are easy to estimate when we have an (approximate) curvature matrix. To reasonably high accuracy for most experiments, the error on q_B is simply that of the Gaussian approximation to the likelihood surface, $\mathcal{F}_{BB}^{-1}$ (see, *e.g.*, Press et al., 1992). There are also higher accuracy approximations available for more detailed work (Bond et al., 2000), and one can always map out the likelihood surface by direct evaluation, but for the CBI these give very similar results to the errors (for further discussion, see Sievers et al., 2003).

2.5 Band Power Window Functions

It is very useful to understand how the power spectrum responds to a change in the expected signal. This is used to estimate both the band power spectrum from a real spectrum and the shift in the band power spectrum due to other non-CMB signals. The situation in which these are most familiar is that of the response of the power spectrum parameterized in bins to that of a real power spectrum, known as the band power window functions. This is distinct from the response of observed data to a power spectrum, known as the visibility window functions, as dicussed by Knox (1999) who shows how to calculate the window functions for an experiment with a single bin. The generalization to the window functions when there are many bins is given here. We have parameterized the power spectrum as a set of bins with a uniform power level in each bin. We could just as easily have picked a shape other than flat—the important point is that the shape of the bin is not allowed to change. Needless to say this is not how a real model power spectrum behaves, so in order to test cosmological models we need to know how to transform from a model power spectrum to a binned one. In other words we would like to have the set of coefficients such that

$$\langle q_B \rangle = \sum_\ell \phi_{B\ell} \mathcal{C}_\ell \tag{2.40}$$

where $\mathcal{C}_\ell$ is the true power spectrum and $\phi_{B\ell}$ are the window functions describing the response of q_B to the true power spectrum. Unfortunately no such set of coefficients exists valid for all $\mathcal{C}_\ell$ because maximum likelihood is a non-linear method—the shift in the power spectrum from adding twice a signal is not exactly twice the shift from adding the original signal. We can, however, come up with such a set of coefficients if we restrict ourselves to the region around the maximum where the curvature is well described by $\mathcal{F}$. In order to do this we need the new expected gradient of the likelihood when we add in the new signal. If W_ℓ is the expected covariance from our new signal, then on average we have

$$\boldsymbol{\Delta}\boldsymbol{\Delta}^T \rightarrow \mathrm{W}_\ell + \boldsymbol{\Delta}\boldsymbol{\Delta}^T \tag{2.41}$$

We can then use Equation 2.32 to estimate the new derivative

$$\frac{d \log(\mathcal{L})}{dq_B} = \frac{1}{2} Tr\left(\left(\mathrm{W}_\ell + \boldsymbol{\Delta}\boldsymbol{\Delta}^T - \mathrm{C}\right) \mathrm{C}^{-1} \mathrm{W}_B \mathrm{C}^{-1}\right) \tag{2.42}$$

If we are at the maximum, then the $\boldsymbol{\Delta}\boldsymbol{\Delta}^T - \mathrm{C}$ part of the gradient is equal to zero, and we are left with the expected gradient due to the new signal

$$\frac{d \log(\mathcal{L})}{dq_B} = \frac{1}{2} Tr\left(\mathrm{W}_\ell \mathrm{C}^{-1} \mathrm{W}_B \mathrm{C}^{-1}\right) \tag{2.43}$$

The expected shift in the band powers can then be calculated by doing a Newton-Raphson iteration

$$\langle dq_B \rangle = \frac{1}{2} \mathcal{F}^{-1}_{BB'} Tr\left(\mathrm{W}_\ell \mathrm{C}^{-1} \mathrm{W}_{B'} \mathrm{C}^{-1}\right) d\mathcal{C}_\ell \tag{2.44}$$

Now we have used no properties unique to the CMB to understand the response of the q_B to W_ℓ. This means that we could substitute *any* expected signal and see how the q_B responds. For instance, we can calculate the expected contribution to the power spectrum from a population of faint radio

point sources that are statistically isotropic. If the covariance describing the point sources is S_{iso}, then their effect on the power spectrum is

$$\langle dq_B \rangle = \frac{1}{2}\mathcal{F}^{-1}_{BB'} Tr\left(\mathrm{W}_\ell \mathrm{C}^{-1}\mathrm{S}_{iso}\mathrm{C}^{-1}\right) \tag{2.45}$$

We can also estimate our sensitivity to a fractional uncertainty of ϵ in our measured noise

$$\langle dq_B \rangle = \frac{1}{2}\mathcal{F}^{-1}_{BB'} Tr\left(\mathrm{W}_\ell \mathrm{C}^{-1}\boldsymbol{N}\mathrm{C}^{-1}\right)\epsilon \tag{2.46}$$

We use this algorithm in Mason et al. (2003) and Pearson et al. (2003) to measure the response of the CBI power spectrum to $\mathcal{C}_\ell$ as well as to errors in noise and source corrections.

It is worth a discussion of computational issues involved in measuring the filters, as they can easily far exceed the total computational effort required to measure the power spectrum itself. The best way to proceed depends on if one desires just a few filters (*i.e.* noise and source filters) or very many (for finely sampled window functions). If we desire many filters, then the fastest course of action is to calculate and store the set of matrices $\mathrm{C}^{-1}\mathrm{W}_B\mathrm{C}^{-1}$, and form the gradient vector by taking the trace of each of them multiplied by W_ℓ. This requires an expensive initial step of order $2n_B n^3$, which can easily be an order of magnitude more work than measuring the power spectrum. However each additional filter requires only an n^2 operation, so it is the most efficient way to calculate lots of filters. We can speed matters up considerably if we only require a few ($< n_B$) filters. First, note that the trace remains unchanged if we write it as

$$\mathrm{C}^{-1}\mathrm{S}\mathrm{C}^{-1}\mathrm{W}_B \tag{2.47}$$

for some matrix S whose filter we desire. This is clearly true if

$$Tr\left(\mathrm{A}\right) = Tr\left(\mathrm{B}^{-1}\mathrm{A}\mathrm{B}\right) \tag{2.48}$$

It is indeed generally true (see any linear algebra text), but I shall prove it for the specific case of symmetric matrices. If we decompose B into its eigenvalues and eigenvectors, we have

$$Tr\left(\mathrm{V}\Lambda^{-1}\mathrm{V}^{T}\mathrm{AV}\Lambda\mathrm{V}^{T}\right) \tag{2.49}$$

$\mathrm{V}^{T}\mathrm{AV}$ is just a rotation of A, and doesn't affect the trace, since a rotation doesn't change the eigenvalues. Similarly, the outer pair of V and V^{T} is also a rotation and doesn't affect the trace. So, if we rewrite the rotated A as A^{*}, then the trace is now

$$Tr\left(\Lambda^{-1}\mathrm{A}^{*}\Lambda\right) \tag{2.50}$$

We can carry out this multiplication element by element to get the ij^{th} element of the product is $\mathrm{A}^{*}_{i,j}\frac{\Lambda_i}{\Lambda_j}$. This will in general change all elements except those for which $i = j$ - in other words, the matrix changes *except* for the elements along the diagonal. Clearly, this leaves the trace unchanged. Now to get the filter from Equation 2.47, we need to calculate $\mathrm{C}^{-1}\mathrm{SC}^{-1}$, but then can take the trace of the set of n_B products quickly with only order n^2 operations. So we have a choice between doing 2 matrix multiplications per filter, or $2n_B$ matrix multiplications to get arbitrarily many filters.

Chapter 3

First CBI Results

We first used a simplified version of the formalism of Section 2 to analyze the first few months of CBI data, released in Padin et al. (2001a).

3.1 Early Observations

The data for Padin et al. (2001a), the CBI comissioning run, were taken between January and April of 2000 at the Llano de Chajnantor. The CBI was configured in a ring configuration (see Figure 3.1), designed to give good maintenance access to each receiver. The ring configuration also had reasonably uniform UV coverage (see Figure 3.2 for the distribution of the baseline lengths). Because very little was known about foreground radio emission at sub-degree scales and centimeter wavelengths at the time, we chose our initial fields with care. The target fields were selected to be low in IRAS 100 μm emission to avoid dust and possible anomalous galactic foregrounds (Leitch et al., 1997), low in synchrotron emission (Haslam et al., 1981, 1982), and low in NVSS radio point sources (Condon et al., 1998). The were also chosen to be far enough north ($\delta \sim -3^\circ$) to be observable by the OVRO 40 meter telescope so we could simultaneously monitor point sources with it. We measure all sources brighter than 6 mJy at 1.4 GHz with the 40 meter, reliably detecting those brighter than 8 mJy at 30 GHz, which we then subtract from our data. Because of ground spillover (see Section 3.2 for a more detailed discussion), the fundamental CBI observations are the differences of pairs of 8 minute observations of fields separated by 8 minutes in RA, with data taken every 8.4 seconds. The noise is calculated by measuring the scatter of the 8.4 second samples that go into each 8-minute

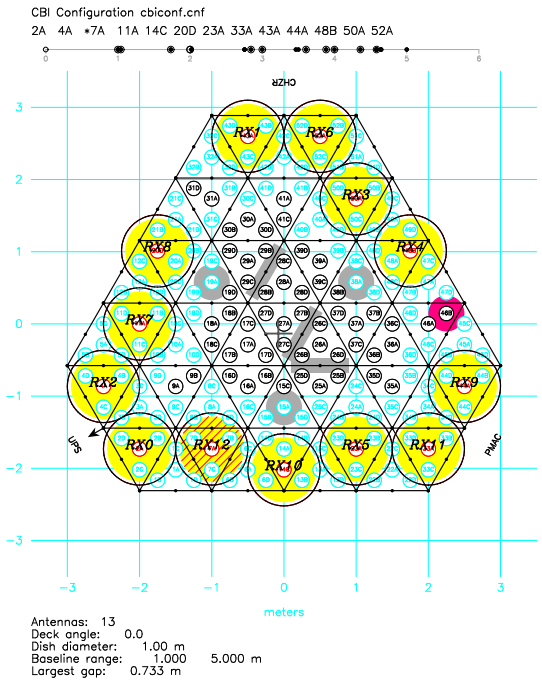


Figure 3.1 Antenna configuration for the commissioning run of the CBI. The dishes were placed in a ring around the outside for easy access. The ring configuration also provided a fairly uniform distribution of baseline lengths in the UV plane.

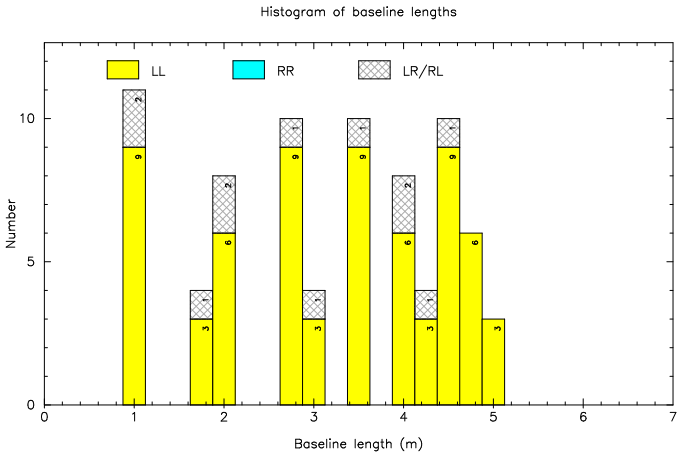


Figure 3.2 Distribution of baseline lengths during the commissioning run. The ring provided a fairly uniform distribution in baseline lengths. Since the CBI rotates the deck, a uniform distribution in length also leads to reasonably uniform sampling in the UV plane.

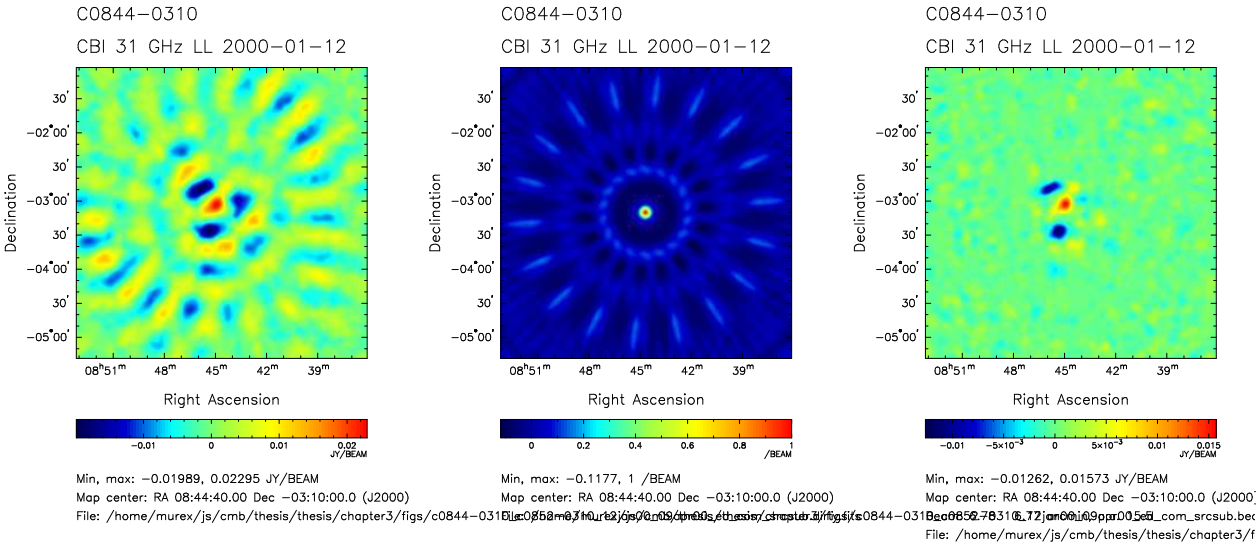


Figure 3.3 The 08 hour deep field. Left hand panel is the dirty map of the differenced data, center panel is the beam, right hand panel is the image cleaned to 1σ in the noise. The clean map has the signal in the center, as expected for on-sky sources in the primary beam, as opposed to ground, moon, weather, or instrumental artifacts. The cleaned map is not used for the analysis, as the beam effects are automatically included in the Maximum Likelihood pipeline.

scan. The sun is too bright in the CBI sidelobes to allow daytime observations, and the moon is too bright for observations within 60 degrees. Because the austral summer of 2000 was one of the wettest periods on record in the Atacama, we lost 50% of the nights to weather. This left us with a total of 58.5 hours on each of our 08 hour fields and 16.15 hours on each of our 14 hour fields. See Figures 3.3 and 3.4 for maps of the two fields.

3.2 Ground Spillover

Because the CBI has relatively small dishes ($\sim 100\lambda$ at 30 GHz), ground spillover was an issue. The signal from the ground comes principally from the horizon (where 3 K sky meets 300 K ground) moving through the fringes of the far sidelobes as the telescope tracks the sky. The 1 m baselines were the most corrupted by the ground, since they average over the fewest fringes, and had instantaneous ground signals typically of a few Jy on the short baselines. This is to be compared to the expected maximum of $\sim$50 mJy from the CMB. Since the ground will be, on average, uncorrelated with the CMB, it will eventually average out with enough observations, but the cost is extremely high. The

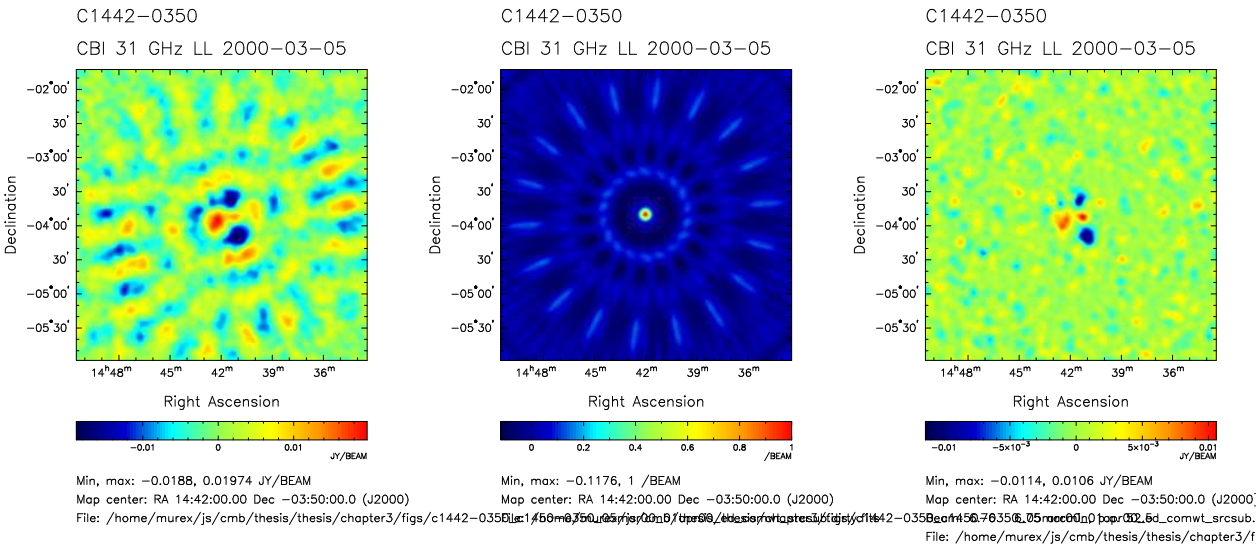


Figure 3.4 The 14 hour deep field. Same as Figure 3.3 for the 14 hour deep field.

noise in a set of observations over a period of time is

$$N_{tot} = \frac{N_i}{\sqrt{n}} \tag{3.1}$$

where N_{tot} is the total, final noise, N_i is the instantaneous noise, and n is the total number of independent observations. If the noise is correlated over time, the total number of observations is

$$n = \frac{t}{\tau_c} \tag{3.2}$$

where t is the total observing time and τ_c is the length of time over which the noise is correlated. For thermal noise, the correlation time is $\frac{1}{B}$ where B is the bandwidth, and the instantaneous noise is just the temperature. This gives the familiar formula

$$\Delta T = \frac{T}{\sqrt{Bt}} \tag{3.3}$$

The figure of merit for a noise source then is

$$\frac{T}{\sqrt{B}} \tag{3.4}$$

This number can be compared between different sources, and the one which has the highest value is the dominant source of noise. Note that the noise temperature T can be in units other than K, such as Jy. If we observe on scales much longer than the inverse bandwidth, as is the case for thermal noise, then this is just the noise in a second of observation. For the CBI thermal noise, this is about 6.5 Jy-s$^{1/2}$. On the short baselines, the ground spillover is comparable to the thermal noise at the CBI's data sampling rate of 8.4s (Figure 3.8). But the correlation time for the ground can be much longer than that—we frequently see phase ramps from the ground lasting many minutes, even hours, with a consequent effective bandwith of millihertz (see Figures 3.5 and 3.6). The ground noise then pretty easily reaches an effective system noise level of $\sim 2\text{Jy}/\sqrt{10^{-3}} \sim 60$ Jy. So the ground noise can be many times more important than the system noise, and since the observation time goes as noise squared, uncorrected ground signals can slow the data-taking by orders of magnitude. In addition, the exact statistics of ground noise are difficult to estimate reliably since they depend on the physical orientation of the telescope, the orientation of the baseline, the hour angle of the observed field, snow on the ground, the (possibly changing) correlation time of the ground signal, and so on. Since maximum likelihood effectively subtracts off the noise, any misestimate of the noise will shift the power spectrum, which would make any CBI result difficult to interpret.

A better way to combat the ground, rather than trying to beat it down by brute force, is to observe pairs of fields at the same declination and separated by a fixed difference in RA (in our case 8 minutes of time), rather than single fields. We observed the lead field for < 8 minutes, then slewed back to the trailing field and observed it for the same length of time, beginning 8 minutes after starting the observation of the lead field. In this way, the CBI moves through the same physical angles with respect to the ground for both the lead and trail observations. Since the pairs of observations observe the ground in identical ways, the ground signal should be identical in

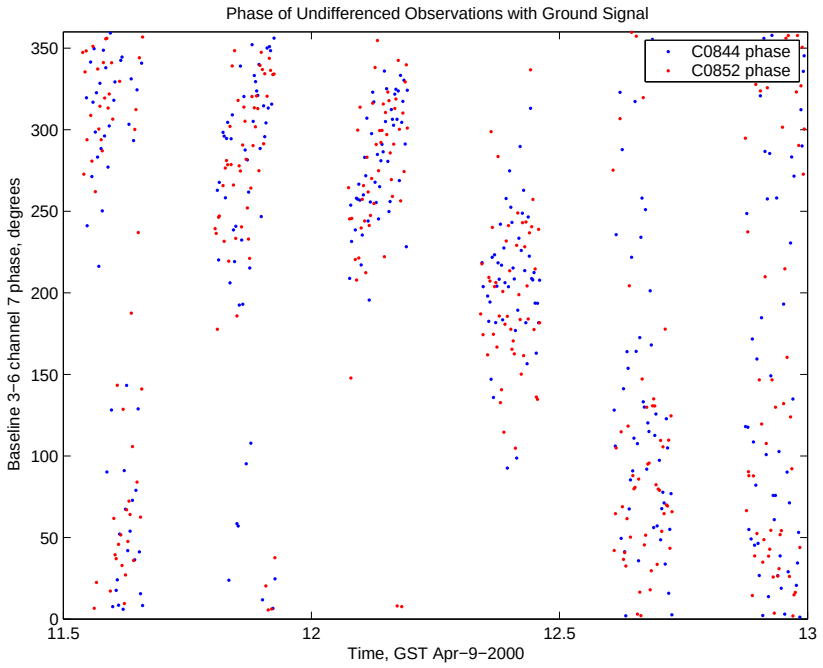


Figure 3.5 Phase of visibilities for a typical 1-meter baseline. There are two field here, the blue dots are the lead (c0844-0310) field, and the red dots are the trail (c0852-0310) field. The trail points have been shifted in time by 8 minutes (the length of a scan) so they lie on top of points in the lead taken with the same ground. As is clear in the plot, the phases are not random, which means they are set by the ground (thermal noise introduces no phase correlation and the sky signal is much weaker than the noise). If one extends the phase ramps to the next pair of 8-minute scans, one can see that the phase introduced by ground spillover remains intact for over an hour. While this particular set of data is somewhat more dominated by ground than is average, it is not at all atypical, and only slightly weaker phase ramps are the norm rather than the exception.

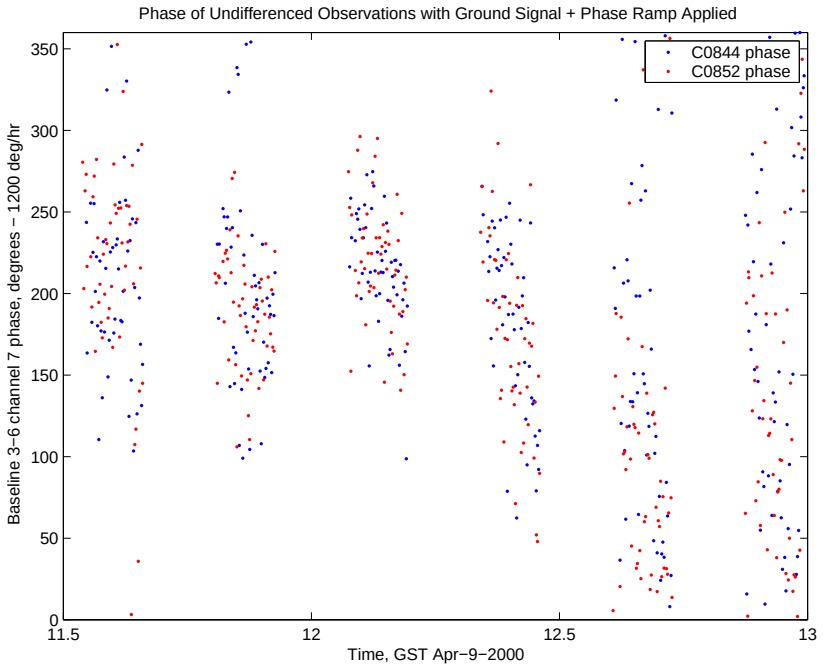


Figure 3.6 Same as Figure 3.5, but with a constant phase ramp of 1200 degrees/hour subtracted off. The purpose of this plot is to show the length of time over which the phases can remain coherent and predictable. In this case, the structure is intact for over an hour!

the two observations, modulo intrinsic changes in the ground signal on 8-minute scales.

Fortunately, the ground signal is quite stable both in theory and practice. The two most obvious sources of ground signal remaining in the differenced data are the signal from any changes in the ground signal over 8 minutes, and pointing errors causing the subtraction of slightly different ground signals. The signal strength expected to leak through the differencing from a changing ground should be something like the total ground signal times the fractional change in ground temperature over the course of 8 minutes. Typically, the air temperature will change by ~ 10 degrees over the course of an entire night, usually no more than a couple of degrees in an hour. So the ground probably isn't changing much faster than a few tenths of a degree in 8 minutes, for a *fractional* change in temperature of about one part per thousand. The effective ground noise in differenced observations should then be tens of mJy-s$^{1/2}$, rather than tens of Jy-s$^{1/2}$. This is highly sub-dominant to the thermal noise, and so doesn't present a problem. Pointing errors will also introduce errors in the ground subtraction, but again we expect them to be small. The CBI has a pointing accuracy of a

few arcseconds, and is especially reliable at returning to the same point from after a short track, which is the requirement for good ground subtraction (we don't care what ground we observe, as long as it's the same ground for the lead and trail fields). Near the equator, where our fields are, this is an error of a few arcsecond divided by 15, or a few tenths of a second of time. Because the ground changes on a time scale of about a few minutes (see Figure 3.6, and note that the phase change is 1200 degrees/hour, or a radian every three minutes), the fractional leakage of the ground signal from pointing errors should be something like the effective time uncertainty from the pointing error divided by the time it takes the ground to change (which is a different, shorter number than the coherence time of the ground since the ground can change coherently over a phase angle much greater than 2π). So, the ground noise leaked due to pointing errors should be something of order a few tenths of a second divided by a minute of the original ground signal, or a factor of a few hundred down from the original ground signal. This again is highly subdominant to the thermal noise. In practice, we also see no evidence of ground contamination in the differenced data sets (see Figures 3.7 and 3.8). To check in greater detail, we split the differenced data into various epochs and subtracted them, creating doubly differenced data sets, with zero expected signal. The noise level in the doubly differenced data sets is consistent with the expected thermal noise, indicating that there isn't a significant source of noise on long timescales leaking through, and that our noise measurements are accurate, once the statistics are done correctly (see Section 4.1).

While critical for rejecting the ground signal, the cost of the differencing is a factor of two in time. The variance of the differenced visibility is twice the variance of the individual visibilities (assuming they are widely enough separated so that their microwave background signals are mostly uncorrelated, which is the case), and the variance from the noise of the difference is the sum of the noises of the individual measurements. So, the expected variance doubles, and the noise variance doubles as well. This leaves the total signal-to-noise ratio unchanged, but required two data points instead of just one, hence the factor of two in time cost. The differencing also has the nice benefit that it rejects any instrumental signal that varies on timescales much slower than 8 minutes, including DC signals such as correlator offset. It is in fact possible to lose a smaller fraction of the data by

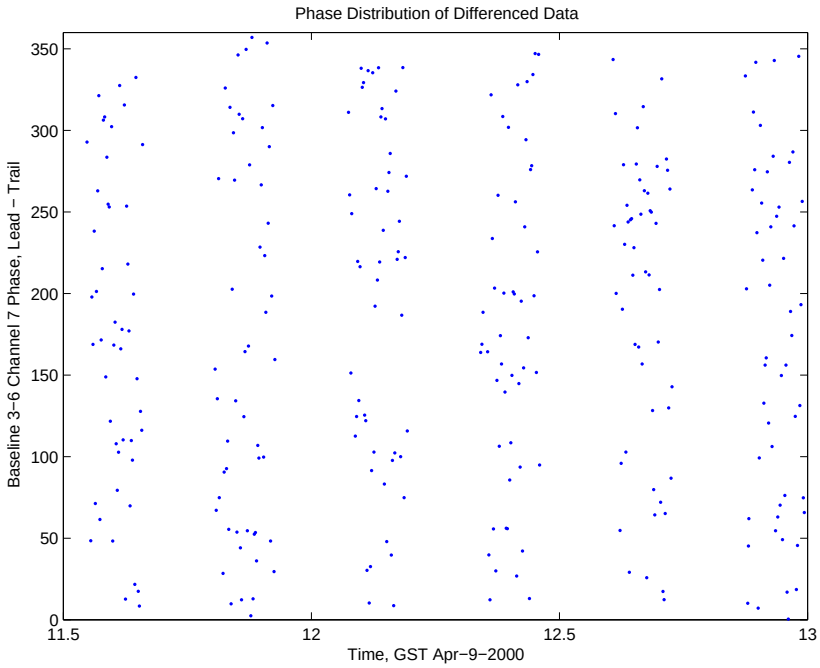


Figure 3.7 Same data as Figure 3.5, showing the phase distribution of the differenced (ground-free) data. The phase distribution is far more uniform. More sensitive statistical tests do not reveal any coherence introduced by the ground remaining in differenced data.

observing field triplets, quadruplets, quintuplets, or more, instead of pairs of fields. Each set of observations loses one mode to the ground, leaving $n-1$ good measurements of the CMB, for a total efficiency relative to undifferenced data of $1-\frac{1}{n}$ where n is the total number of fields observed with the same ground. Initially, we wanted to go as deep as possible over a small area in order to get a result quickly as well as to try to uncover any systematics, so we used the simple pair-wise differencing. Now that we have experience with the performance of the CBI and find it very stable, we are in fact using strips of fields for polarization observations, with $n=6$, for an efficiency increase of about 60%.

3.3 Analysis

There were several simplifying factors in the analysis of the Padin et al. (2001a) data. By far, the most significant was that the observations were all of a single field, which makes C much easier to calculate. We also approximated the primary beam with a Gaussian (see Figure 3.9 for the

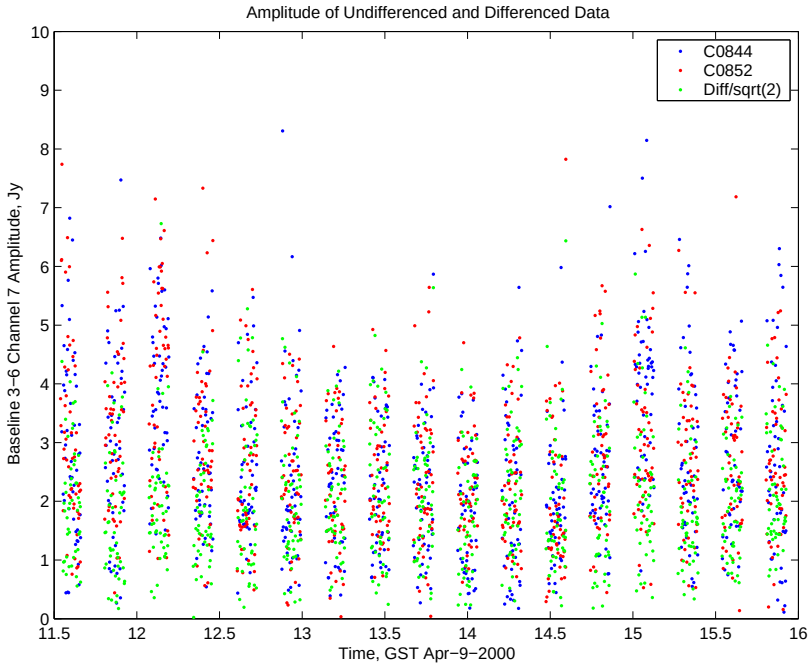


Figure 3.8 Same data as Figure 3.5, showing the amplitude distribution of differenced and undifferenced data. The differenced data have had a scaling of $\sqrt{1/2}$ applied to them, since their variance has been doubled by the differencing. For these data, the undifferenced data have a variance $> 70\%$ higher than would be expected from the variance of the differenced data. This excess variance is the relative strength of the ground vs. thermal noise on the 8.5 second sampling rate and is removed by differencing.

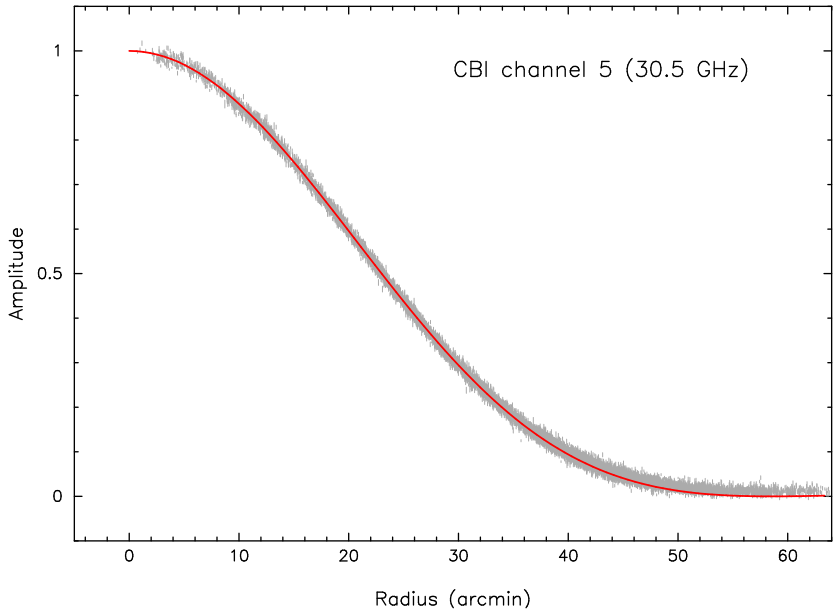


Figure 3.9 The CBI fitted beam. The dishes are modelled by a Gaussian taper in the illumination pattern with unknown width a. There is also a hole in the center of the illumination pattern due to blockage from the secondary. The beam on the sky is the Fourier transform of the dish autocorrelation pattern, which is equivalent to the square of the Fourier transform of the dish illumination. The beam is fit by varying the taper width a and minimizing χ^2 for a bright source, in this case TauA.

fit of the CBI beam to data by Timothy Pearson, and Figure 3.10 for the comparison of the fit beam to the Gaussian approximation) and ignored the very slight correlations introduced by our differencing scheme. Also, because of the small size of the data set, we could perform a maximum likelihood power spectrum extraction directly on the visibilities without having to shrink the size of the data set first. Because of this the biggest step in measuring the power spectrum is calculating the window matrices W_B. I shall outline our procedure below, starting from the initial response of interferometers.

3.3.1 Interferometer Response to a Random Temperature Field

The output visibility $V(\mathbf{u})$ of an interferometer is equal to the sky brightness integrated over the field of view, with an intensity modulation from the primary beam and a phase factor from the baseline separation $\mathbf{u}$

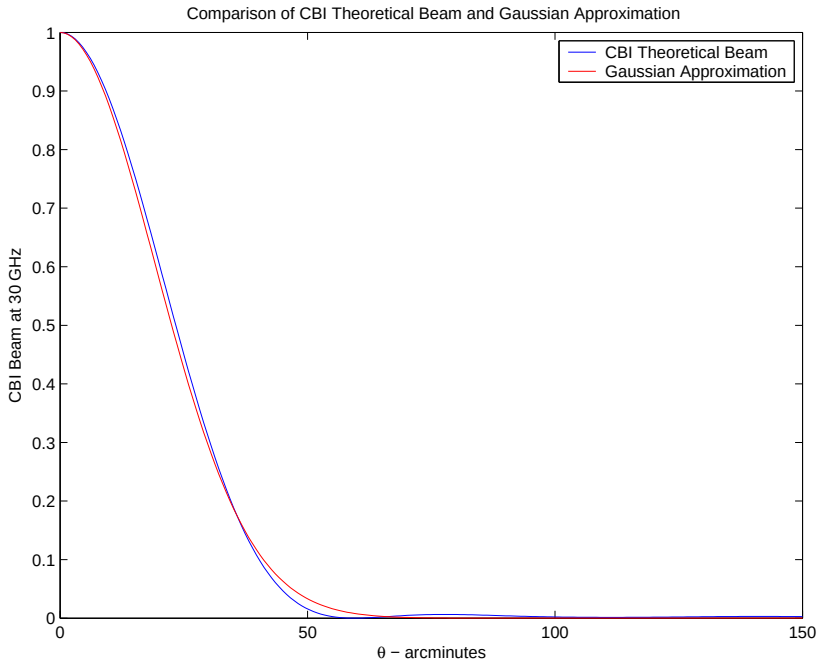


Figure 3.10 Comparison of CBI fit beam to the Gaussian approximation to it. This is the same CBI fit beam as in Figure 3.9, and the Gaussian has a FWHM of 45.1'. The fit is very good, and the Gaussian is much easier to work with computationally.

$$V(\mathbf{u}) = \int\int B_\nu(T(\mathbf{x}))A(\mathbf{x})\exp(2\pi i\boldsymbol{x}\cdot\boldsymbol{u})d^2\boldsymbol{x} \tag{3.5}$$

A is the square of the response of a receiver to the electric field (the primary beam), and $\mathbf{x}$ is position on the sky relative to the pointing center. The Planck function evaluated at the observing frequency ν is $B_\nu(T)$ for a radiation field of temperature T. It is convenient to convert the temperature map into a dimensionless function ($\delta T/T$) and pull the rest of the Planck function out in front of the integral, discarding the DC term, to which the interferometer is not sensitive.

$$B_\nu(T(\mathbf{x})) = \frac{dB_\nu}{dT}\Big|_{T_{CMB}} T_{CMB}\frac{\delta T}{T_{CMB}} \tag{3.6}$$

The function $\frac{dB_\nu}{dT}$ is as follows (*e.g.* White et al., 1999)

$$\frac{dB_\nu}{dT} = \frac{2k_B}{c^2}\left(\frac{k_BT}{h}\right)^2\frac{x^4e^x}{(e^x-1)^2} \tag{3.7}$$

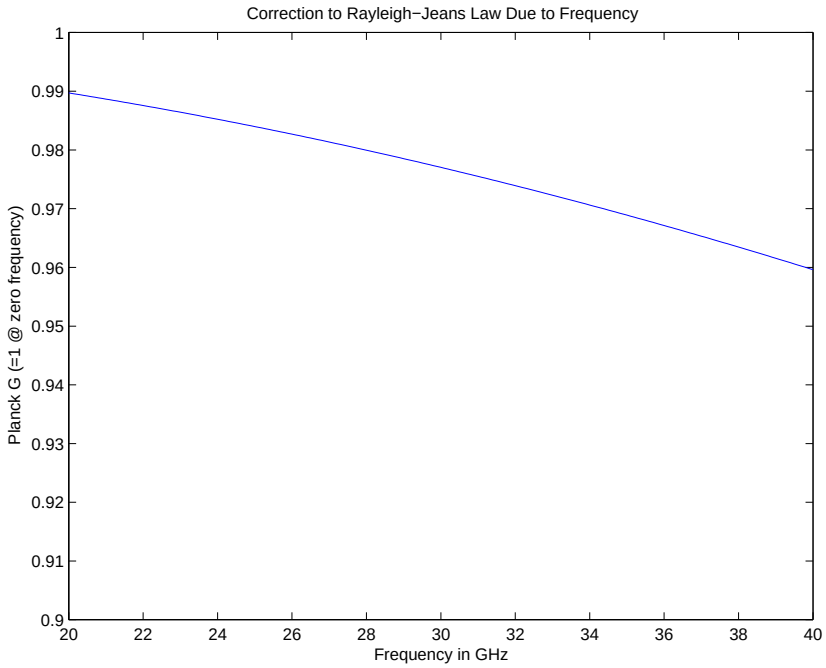


Figure 3.11 Plot showing correction factor multiplied to Rayleigh-Jeans law to get differential Black Body, $\frac{dB_\nu}{dT}$. Note the small scale on the x-axis. Even at moderate frequencies, the true blackbody intensity is very close to that of the Rayleigh-Jeans law.

where $x \equiv h\nu/k_B T_{CMB}$ (unrelated to the vector on the sky $\boldsymbol{x}$ in Equation 3.5). Pull out a factor of x^2, and what is left is a Rayleigh-Jeans law with a correction factor, the Planck g function:

$$\frac{dB_\nu}{dT} = \frac{2k_B}{c^2}\nu^2 g\left(\nu\right) \tag{3.8}$$

where

$$g\left(\nu\right) \equiv \frac{x^2 e^x}{\left(e^x - 1\right)^2} \tag{3.9}$$

The correction to the Rayleigh-Jeans law for the CBI is fairly small. The frequency coverage of the CBI is 26-36 GHz, so for $T_{CMB} = 2.73$K, x ranges between 0.46 and 0.63 and g is between 0.983 and 0.967 (see Figure 3.11). For clarity in writing, we shall adopt the definition of f_T in Myers et al. (2003)

$$f_T\left(\nu\right) \equiv \frac{2k_B T_{CMB}}{c^2} g\left(\nu\right) \tag{3.10}$$

Because the CMB is a Gaussian random field, in the limit of small sky coverage it is a sum of

independent Fourier modes with random phases. Therefore we need to understand the response of the interferometer to a plane wave on the sky, which is most conveniently done by taking the Fourier transform of (3.5). The visibility from the mode with wave vector $\boldsymbol{w}$ is then

$$V = \frac{dB_\nu}{dT} T_{CMB} \frac{\delta \widetilde{T}}{T_{CMB}} (\boldsymbol{w}) \widetilde{A} (\boldsymbol{u} - \boldsymbol{w}) = f_T (\nu) \nu^2 \frac{\delta \widetilde{T}}{T_{CMB}} (\boldsymbol{w}) \widetilde{A} (\boldsymbol{u} - \boldsymbol{w}) \tag{3.11}$$

where $\widetilde{A}$ is the Fourier transform of the primary beam and $\delta \widetilde{T}$ are the temperature fluctuations in Fourier space. And the total response of the interferometer to the sky is this function integrated over $\boldsymbol{w}$:

$$V = f_T (\nu) \nu^2 \int \int \frac{\delta \widetilde{T}}{T_{CMB}} (\boldsymbol{w}) \widetilde{A} (\boldsymbol{u} - \boldsymbol{w}) d^2 \boldsymbol{w} \tag{3.12}$$

To calculate the W_B in Equation 2.21, we need to be able to calculate the correlation between pairs of visibilities. The response of a pair of interferometers to a single Fourier mode on the sky is just their individual responses to a mode integrated over modes:

$$\langle V_1^* V_2 \rangle = f_T (\nu_1) f_T (\nu_2) \nu_1^2 \nu_2^2 T_{CMB}^2 \left(\frac{\delta \widetilde{T}}{T_{CMB}} (\boldsymbol{w}) \right)^2 \widetilde{A}^* (\boldsymbol{u_1} - \boldsymbol{w}) \widetilde{A} (\boldsymbol{u_2} - \boldsymbol{w}) \tag{3.13}$$

We take the complex conjugate to make the product strictly real and independent of the phase of the wave on the sky. This can be integrated over wave space to get the expected response of an interferometer pair to a set of temperature fluctuations.

$$\langle V_1^* V_2 \rangle = f_T (\nu_1) f_T (\nu_2) \nu_1^2 \nu_2^2 T_{CMB}^2 \int \int \left(\frac{\delta \widetilde{T}}{T_{CMB}} (\boldsymbol{w}) \right)^2 \widetilde{A}^* (\boldsymbol{u_1} - \boldsymbol{w}) \widetilde{A} (\boldsymbol{u_2} - \boldsymbol{w}) d^2 \boldsymbol{w} \tag{3.14}$$

Now, the expected value of $\frac{\delta \widetilde{T}}{T}$ is merely the power spectrum of fluctuations.

$$\left\langle \frac{\delta \widetilde{T} (\boldsymbol{w})}{T_{CMB}} \right\rangle^2 = S(w) \tag{3.15}$$

We have replaced $\boldsymbol{w}$ by w since the power spectrum should be independent of angle and only depend on the wavelength of the modes in question. We also need the relation between the Fourier spectrum

$S(w)$ and the angular power spectrum valid for the small-angle approximation, given by White et al. (1999):

$$w^2 S(w) \simeq \frac{\ell(\ell+1)\mathcal{C}_\ell}{(2\pi)^2} \tag{3.16}$$

where $w = \ell/2\pi$. Since the CBI observes at high ℓ, the difference between ℓ and $\ell+1$ is negligible. We can then rewrite

$$S(w) \simeq \mathcal{C}_\ell \mid_{2\pi w=\ell} \tag{3.17}$$

and

$$\langle V_i^* V_j \rangle = f_T(\nu_1) f_T(\nu_2) \nu_1^2 \nu_2^2 T_{CMB}^2 \int\int S(w) \widetilde{A}_1^* (\boldsymbol{u_1} - \boldsymbol{w}) \widetilde{A}_2 (\boldsymbol{u_2} - \boldsymbol{w}) d^2\boldsymbol{w} \tag{3.18}$$

3.3.2 Visibility Window Functions

Since we expect that the CMB fluctuations will be angle-independent, we can do the angular part of the integral separately from the integral in $d|w|$.

$$\langle V_1^* V_2 \rangle = f_T(\nu_1) f_T(\nu_2) \nu_1^2 \nu_2^2 T_{CMB}^2 \int w S(w) dw \int \widetilde{A}_1^* (\boldsymbol{u_1} - \boldsymbol{w}) \widetilde{A}_2 (\boldsymbol{u_2} - \boldsymbol{w}) d\theta \tag{3.19}$$

The angular integral is called the *visibility window function* $\mathrm{W_{ij}(w)}$, or simply the window function (not to be confused with the *band power window functions* of Section 2.5). The window functions contain essentially all of the telescope-specific information. We must now work out the window functions for CBI. Their calculation is greatly simplified if, as was the case for Padin et al. (2001a), the data all have the same pointing center. We also approximated the CBI beam with a 45.1' FWHM Gaussian at 30 GHz (see Figure 3.10 again). We normalize the telescope response to unity in the beam center in physical space, so the beam Fourier transform is

$$\widetilde{A}(\boldsymbol{u}) = \frac{1}{2\pi\sigma_p^2} \exp\left(-\frac{u^2}{2\sigma_p^2}\right) \tag{3.20}$$

where σ_p is the Gaussian σ of the primary beam Fourier transform. After much algebra we can do the window function integral

$$\mathrm{W}_{ij}\left(w\right) = \int \tilde{A}_1^*\left(\boldsymbol{u_1} - \boldsymbol{w}\right) \tilde{A}_2\left(\boldsymbol{u_2} - \boldsymbol{w}\right) d\theta = \frac{1}{4\pi^2 \sigma_{p,1}^2 \sigma_{p,2}^2} \int \exp - \frac{\left(\boldsymbol{u_1} - \boldsymbol{w}\right)^2}{2\sigma_{p,1}^2} \exp - \frac{\left(\boldsymbol{u_2} - \boldsymbol{w}\right)^2}{2\sigma_{p,2}^2} d\theta \tag{3.21}$$

to get

$$\mathrm{W}_{ij}(w) = \frac{1}{4\pi^2 \sigma_{p1}^2 \sigma_{p2}^2} \exp(-Aw^2 - B) 2\pi I_0(Cw) \tag{3.22}$$

The Bessel function I_0 comes about from $\int \exp(cw\cos(\theta)) d\theta = 2\pi I_0(cw)$. The coefficients are

$$A = \frac{1}{2\sigma_{p1}^2} + \frac{1}{2\sigma_{p2}^2} \tag{3.23}$$

$$B = \frac{u_1^2}{2\sigma_{p1}^2} + \frac{u_2^2}{2\sigma_{p2}^2} \tag{3.24}$$

$$C^2 = \frac{u_1^2}{\sigma_{p1}^4} + \frac{u_2^2}{\sigma_{p2}^4} + \frac{2u_1 u_2}{\sigma_{p1}^2 \sigma_{p2}^2} \cos(\theta_{1,2}) \tag{3.25}$$

where $\theta_{1,2}$ is the angle between baseline 1 and baseline 2. An accurate approximation to I_0 valid over the range in which we are interested (baseline lengths of a meter or longer) is

$$\int \exp(a\cos(\theta)) = 2\pi I_0 \simeq \exp(a) \left(\frac{2\pi}{a}\right)^{1/2} \left(1 + \frac{1}{8a}\right) \tag{3.26}$$

So, the final window function for a Gaussian beam and a single-pointing exposure is

$$\mathrm{W}_{ij}(w) = \frac{1}{4\pi^2 \sigma_{p1}^2 \sigma_{p2}^2} \sqrt{\frac{2\pi}{Cw}} \exp\left(-Aw^2 - B + Cw\right) \left(1 + \frac{1}{8Cw}\right) \tag{3.27}$$

It is illustrative to work out the window function for the case of a single baseline compared with itself. In that case, the coefficients are

$$A = \sigma_p^{-2}, B = \left(\frac{u}{\sigma_p}\right)^2, C = \frac{2u}{\sigma_p^2} \tag{3.28}$$

and

$$\mathrm{W}_{ii}\left(w\right)=\frac{1}{4\pi^{2}\sigma_{p}^{3}}\sqrt{\frac{\pi}{uw}}\exp\left(\frac{\left(w-u\right)^{2}}{\sigma_{p}^{2}}\right)\left(1+\frac{\sigma_{p}^{2}}{16uw}\right) \tag{3.29}$$

This is a very reasonable expression—basically, a two-element interferometer is sensitive to modes on the sky that have the same wavelength as the separation of the elements, and the sensitivity from this peak falls off like the primary beam Fourier transform. The factor of two is gone in the denominator because the covariance element is a visibility squared, rather than a simple visibility (square a Gaussian, and the two disappears). The primary beam scaling in the coefficient is at first a bit unusual because we expect the total variance to be proportional to the total area of the beam, which is σ_{p}^{-2}. However, we must integrate W_{ij} across the power spectrum, so we pick up an extra factor of σ_{p}, for a total scaling that is proportional to σ_{p}^{-2}, as expected. We can insert Equation 3.27 into Equation 3.19 to get the total covariance expected for visibilities from a single field:

$$\langle V_{1}^{*}V_{2}\rangle=f_{T}\left(\nu_{1}\right)f_{T}\left(\nu_{2}\right)\nu_{1}^{2}\nu_{2}^{2}T_{CMB}^{2}\int wS(w)dw\mathrm{W}_{1,2}(w,V_{1},V_{2}) \tag{3.30}$$

using the formula for W_{ij} from Equations 3.22 or 3.27. We are now in position to choose a parameterization of the power spectrum, which specifies $S(w)$. Then by integrating Equation 3.30 across the bins in w, we have the window matrices used to find the maximum likelihood power spectrum.

3.4 Complex Visibilities

There is one slight adjustment that needs to be made to go from the complex visibility formulation of the preceeding section to separate real and imaginary estimators (see also Myers et al., 2003). Consider the fundamental definition of the covariance of two visibilities:

$$\mathrm{C}_{ij}=\langle V_{i}^{*}V_{j}\rangle=\langle V_{i,r}V_{j,r}\rangle+\langle V_{i,i}V_{j,i}\rangle+i\langle V_{i,r}V_{j,i}\rangle-i\langle V_{i,i}V_{j,r}\rangle \tag{3.31}$$

We can also rotate one of the visibilities through 180°, which leaves the real part of the visibility unchanged, but flips the sign of the imaginary part. That is, it turns a visibility into its conjugate.

In that case:

$$\mathrm{C}_{i^*j} = \langle V_i V_j \rangle = \langle V_{i,r} V_{j,r} \rangle - \langle V_{i,i} V_{j,i} \rangle + i\langle V_{i,r} V_{j,i} \rangle + i\langle V_{i,i} V_{j,r} \rangle \tag{3.32}$$

This is a set of four relations, since each of the two equations must hold for both its real and imaginary parts. The set of relations can be solved for the covariances between the real and imaginary parts of the visibilities as follows:

$$\langle V_{i,r} V_{j,r} \rangle = \frac{1}{2}(\mathrm{C}_{ij,r} + \mathrm{C}_{i^*j,r}) \tag{3.33}$$

$$\langle V_{i,i} V_{j,i} \rangle = \frac{1}{2}(\mathrm{C}_{ij,r} - \mathrm{C}_{i^*j,r}) \tag{3.34}$$

$$\langle V_{i,r} V_{j,i} \rangle = \frac{1}{2}(\mathrm{C}_{ij,i} + \mathrm{C}_{i^*j,i}) \tag{3.35}$$

$$\langle V_{i,i} V_{j,r} \rangle = \frac{1}{2}(-\mathrm{C}_{ij,i} + \mathrm{C}_{i^*j,i}) \tag{3.36}$$

If baseline i and baseline j are close to each other in UV space, then C_{i^*j} will be small since the conjugate is on the other side of the UV plane. In this case, the *real-real* covariance is the same as the *imaginary-imaginary* covariance, and the real and imaginary parts are equivalent. However, if both C_{i^*j} and C_{ij} are non-zero, then the symmetry is broken and the real part and the imaginary part of the visibility are no longer statistically equivalent, and hence should be treated separately. For that reason, we do treat the problem as one of dimension $2n$ real estimators rather than n complex estimators and use Equations 3.33 through 3.36 to calculate the window matrices.

3.5 Power Spectrum

We measured a power spectrum using the commissioning data described in Section 3.1, the covariances from Section 3.3, point source subtractions from the OVRO 40 meter, and a statistical correction for the signal from sources unmeasured by OVRO calculated by Brian Mason. The analysis was done using a package written by the author. Because the point source formalism we used at this time did not involve projecting out sources of unknown intensity, a substantial source signal

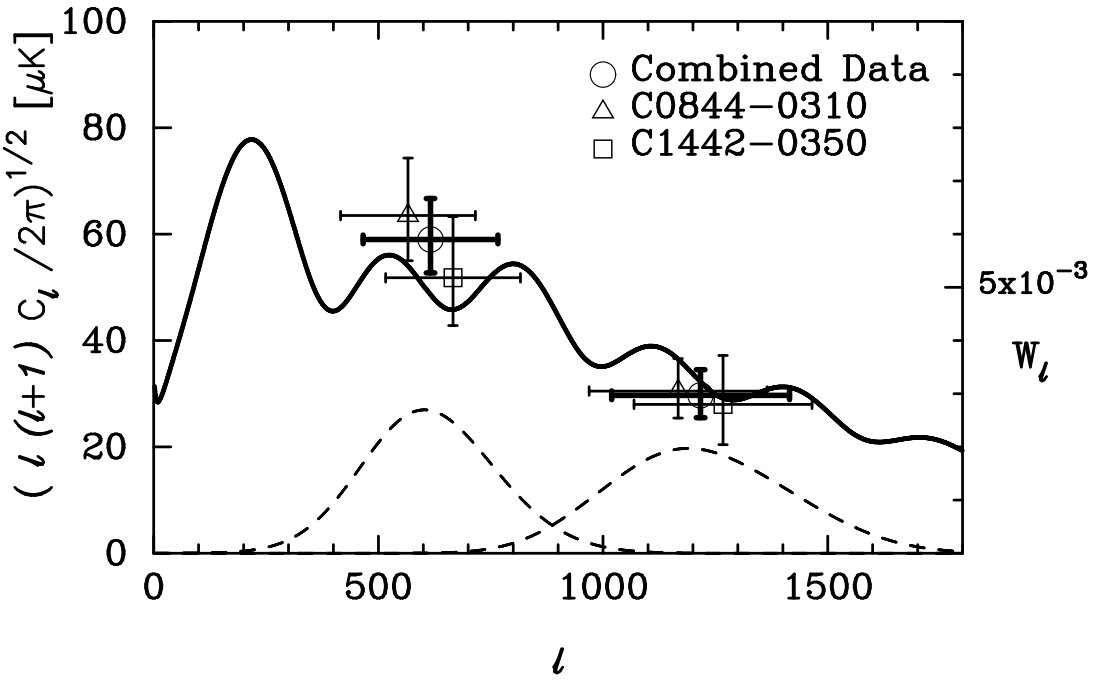


Figure 3.12 Power spectrum plotted in Padin et al. (2001a). The model spectrum is a standard ΛCDM model, with $h = 0.75$, $\Omega_B h^2 = 0.019$, $\Omega_m = 0.2$, and $\Omega_k = 0$. The dashed lines are approximate band power window functions showing the region in ℓ to which the two points are senstive. Unlike other CBI results, the Padin et al. (2001a) results were presented in μK, rather than μK^2. The points have been offset in ℓ for clarity, but actually are sensitive to the same range in ℓ. There is a clear detection of power at $\ell \sim 600$ in the range expected for flat ΛCDM cosmologies, unlike the first BOOMERANG results in de Bernardis et al. (2000) where the power was $< 40\mu K$.

remained in the power spectrum at $\ell \geq 2000$. As a result, the first power spectrum from the CBI consisted of only two points. The amplitude in a bin centered on $\ell = 603$ was $58.7^{+7.7}_{-6.3}$ μK, and the amplitude in a bin centered on $\ell = 1190$ was $29.7^{+4.8}_{-4.2}$ μK. We had not yet switched to using $\mathcal{C}_\ell$, hence quoting the values in μK instead of μK^2, where the bin values are 3445 μK^2 and 882 μK^2. The spectrum is plotted in Figure 3.12, along with approximate band power window functions and a model spectrum from a typical flat ΛCDM cosmology.

3.6 Interpretation and Importance of Spectrum

While the first CBI power spectrum had only two points, they were two very important points. A fundamental prediction of all theories in which the microwave background arises cosmologically at the surface of last scattering is Silk damping (Silk, 1968), the exponential decline in the power

spectrum at large ℓ from photon diffusion. The region of the decline is called the "damping tail" and is unavoidable if the microwave background anisotropies are of primordial origin. The lack of a damping tail would have been a powerful blow against the canonical model of the universe. The two points in Section 3.5 marked the first time the damping tail was measured and were a confirmation of a major prediction of standard cosmology (see White, 2001, for further discussion).

The Padin et al. (2001a) spectrum appeared at an important time, only a few months after the first BOOMERANG (de Bernardis et al., 2000) and MAXIMA (Hanany et al., 2000) spectra were made public. While the principal result of the two experiments was the first precision determination that the universe was geometrically flat, BOOMERANG, and to a lesser extent MAXIMA, had also fueled intense interest because of the apparent lack of signal in the region past the first acoustic peak at $\ell \simeq 600$, where the second peak had been expected. The ratio of the second peak height to the first peak is most sensitive to the physical baryon density in the universe, $\Omega_B h^2$. If real, the most conservative intepretation of the missing second peak would have been that there was a fairly profound misunderstanding of the cosmic baryon content from big bang nucleosynthesis calculations and deuterium line measurements in the Ly-α forest (Tegmark & Zaldarriaga, 2000), and that $\Omega_B h^2$ was about 50% higher than previously believed. The measurement by the CBI at $\ell \sim 600$ was nearly a factor of two higher in $\mathcal{C}_\ell$ than that of BOOMERANG, more in line with the level expected from prior baryon estimates, though a bit high. This was a strong indication that once the CMB experiments converged, the second peak would likely be about at the level expected, which indeed has turned out to be the case. Now, all the major CMB experiments are consistent with each other, and the $\Omega_B h^2$ measured from the CMB (*e.g.* 0.023 ± 0.003 for combined CMB experiments in Sievers et al., 2003) is in good agreement with that measured using other methods, most notably that of Big Bang Nucleosynthesis (Olive et al., 2000; Burles et al., 1999; Tytler et al., 2000). The resolution to the apparent conflict was that the BOOMERANG beam was larger than expected, washing out power on small scales, MAXIMA was consistent with current estimates, and the CBI data happened to have slightly higher than expected power due to cosmic variance and the small sample of only two fields.

The CBI was also able to do some cosmology with the commissioning data although, because of the small area surveyed, it was perforce somewhat limited. The data set was small enough that we were able to do direct likelihood calculations on a grid of models generated using CMBFAST rather than having to do cosmology using the power spectrum. To do this, rather than integrate a flat spectrum model across a band, we integrate $\mathcal{C}_\ell(\ell = 2\pi w)\mathrm{W}_{ij}(w)$ to get the total covariance expected from the CMB. The CBI was able, using only the COBE spectrum as additional information, to rule out intermediate density ($\Omega_{tot} \sim 0.5 - 0.6$) cosmologies at the 90% confidence level. The CBI was able to do this using effectively only two points because of the sharp drop between them. The only places standard power spectra have such large drops is either on the tail end of the first peak, or in the damping tail. If the drop after the first peak is at $\ell \sim 600$, then $\Omega_{tot} \sim 0.3$, while if the drop is due to damping after the third peak, then $\Omega_{tot} \sim 1.0$. With the additional bit of information that there was a first peak at lower ℓ, but without any details as to that peak position or amplitude, the CBI was able to rule out $\Omega_{tot} < 0.7$. Not surprisingly, the CBI also measured a low value for $\Omega_B h^2$ because of its high value at $\ell \sim 600$, with a best fit value of $\Omega_B h^2$=0.009, though the constraint was weak, and the likelihood had only dropped by a factor of 2 at $\Omega_B h^2$=0.019, and a factor of 3 at $\Omega_B h^2$=0.03.

Chapter 4

First-Year Observations and Results

The first-year observations and analysis were a major advance over the commissioning data of Chapter 3. In addition to more data on the first two fields, another deep field was added, as well as three larger-area $\sim 2^\circ \times 2^\circ$ mosaics. The mosaics provide increased ℓ resolution, revealing the shape of the power spectrum in much more detail than is possible with deep fields. The spectrum extraction pipeline was considerably more sophisticated than that of Padin et al. (2001a) as well. The window matrices were calculated using a method based on gridding visibilities written by Steve Myers called CBIGRIDR (Myers et al., 2003). The final spectrum extraction from the window matrices and gridded data was done using MLIKELY, written by Carlo Contaldi, and was based on the slow Equation 2.36, though we have since adopted the fast methods of Chapter 2. My main contribution to the first-year papers was extracting the power spectrum from the mosaics using CBIGRIDR/MLIKELY. This included major work on understanding systematic effects in the mosaic spectra and how to correct for them. This chapter describes my contributions to the first-year data analysis and results. In Section 4.1 I describe my calculation of a statistical correction to the estimated noise. The bias comes about when combining data points whose variances have been estimated by scatter internal to the data points. Uncorrected, the noise bias has a major impact on the high-ℓ power spectrum. In Section 4.2 I discuss improvements to the CBIGRIDR/MLIKELY pipeline that substantially increased the speed. Those speed increases allowed us to push out to higher-ℓ with the mosaic spectrum. In Section 4.3 I describe how we deal with sources in the mosaics,

and some unexpected effects from the sources I discovered in the process of doing the mosaic analysis. In Section 4.5 I describe the data that went into the first-year CBI papers. Finally, in Section 4.6, I describe the final power spectrum from the first-year mosaics and cosmological results from the spectrum.

4.1 Noise Statistics

It is critically important to have a good estimate of the noise in microwave background experiments, especially when the signal is noise-limited and not cosmic variance-limited. Since the noise is in effect subtracted from the data variance, any error in the noise directly biases the power spectrum, and not just the error estimate of the power spectrum. The CBI estimates noise from the scatter of 8.4 second differenced samples during the 8-minute scans. This is an unbiased estimate of the noise in the 8-minute scan. However, if several 8-minute scans are combined, using their measured noises to optimally combine them, the noise estimate becomes biased to an extent that can quite significantly affect the power spectrum at high ℓ if the noise statistics are not correctly treated. I compare the theoretical expectation of the bias to more accurate numerical integrals and Monte Carlo simulations of the data. We use these simulations to determine a final value by which we scale the CBI scatter-based noise estimates in order to get a final, unbiased estimate of the noise. The first-order analytic expression is derived in Appendix A. It is $1 + \frac{4}{\nu}$ if there are ν measurements in each of the 8-minute scans. For the CBI, ν is of order 100 since there are approximately 50 samples per scan, and each sample has both a real and imaginary measurement.

4.1.1 Fast Fourier Transform Integrals

It rapidly becomes exceedingly difficult to get better (higher than first-order) analytic expressions, so numerical methods for evaluating the correction factor under a wide range of circumstances are important (if for no other reason than to check on the analytic expressions). A general brute-force approach to the problem is not very useful because we have many different independent variables (each of the w_i), and so another technique is required if we want to examine the combination of

more than just a few (3-4) 8-minute scans. Fortunately, FFT's are the magic bullet we need. This is because the distribution of the sum of two random variables is the convolution of their individual distribution functions. So, all we have to do is take the FFT of a distribution, and raise it to the power of however many samples we want to combine. The two quantities we need to understand are

$$\left\langle \left(\frac{w_1}{w_1 + \sum_{i>1} w_i} \right)^2 \right\rangle \tag{4.1}$$

and

$$\left\langle \left(\sum w_i \right)^{-2} \right\rangle \tag{4.2}$$

For the first term, we can convolve all of the w_i for $i > 1$ to get a new variable, say q. Then the desired quantity is

$$\left\langle \left(\frac{w_1}{w_1 + q} \right)^2 \right\rangle \tag{4.3}$$

So, we have reduced the problem to a two-dimensional integral, which is quite feasible computationally. The other term becomes even simpler—all the w_i can be combined, to get a one-dimensional integral. The main subtlety is that since the FFT implicitly assumes periodic boundary conditions, the length of region of real-space to be transformed must be large enough so that only one period of the function contributes. Since each w_i is peaked around one, the convolution of n of them will be peaked around n, and the real-space coverage of the distribution that gets transformed must be substantially larger than n. Once one does that, then the answers are quite good. For instance, I checked the expectation value of the first term for $\nu = 50$ and two scans. The theoretical value is $1 + \frac{1}{\nu+1} = 1.019607843$ and the value I get from the FFT integral is 1.019607855.

We expect the first-order calculations to be close for the CBI. The CBI typically has of order 50 points per scan, with both real and imaginary points used in estimating the noise, for a total of 100 degrees of freedom in the PDF. Figure 4.1 shows the correction factor calculated using FFT's to convolve the PDF of single weights. If the correction factor required to scale the variance is expressed as $1 + \frac{x}{\text{d.o.f.}}$, then Figure 4.1 show x for varying numbers of scans, with 100 d.o.f. per scan

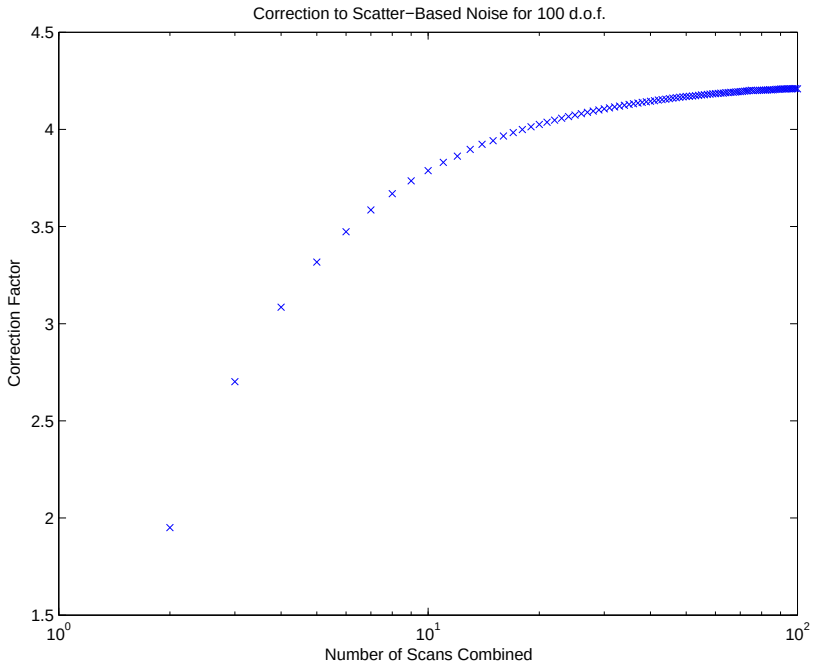


Figure 4.1 Plot of numerical estimates of the correction factor that needs to be applied to scatter-based estimates of the variance. If multiple scans whose noises are estimated from internal scatter are combined with optimal weighting, then there is a systematic underestimate of the true variance of the final, averaged data point. The values plotted are x where the correction factor to be applied to the variance is of the form $1 + \frac{x}{\text{d.o.f.}}$ where d.o.f. is the number of samples in the scan minus the degrees of freedom we may have removed in subtracting off means. First-order calculations predict $x = 2$ for 2 scans, and $x = 4$ for infinitely many scans. The first-order calculations can have substantial corrections to x if there are few d.o.f., but with the CBI's typical value of 100 d.o.f., the first-order prediction is close. Note that few scans are needed to approach the limiting value of the correction. All data points going into the scans have identical variances and are Gaussian distributed.

and each individual point an identically distributed Gaussian. The first-order predictions are 2 for 2 scans and 4 for infinitely many scans. The FFT values are 1.95 for 2 scans, and 4.21 for 100 scans. At 10 scans, the correction factor is 3.8, or about 90% of its limiting value, so the correction factor approaches its limiting value with relatively few scans. Because of roundoff issues in the FFT's, it is difficult to push the numerical integrals to much higher accuracy or to many more scans combined.

4.1.2 Noise Correction Using Monte Carlo

We use Monte Carlo simulations of the noise to estimate the final noise correction factor. There are multiple factors that can break the assumptions in the theoretical calculations that are better

treated by Monte Carlo. Not all scans necessarily have the same number of points, due to outlier tossing or un-matched lead/trail points. Also, the noise on baselines at the same UV point from different receivers will be different, as each receiver has its own system temperature. These effects are difficult to treat theoretically but can be simulated without undue effort. I calculated the final noise correction factor from a set of 50 simulations created using the program MOCKCBI, written by Tim Pearson. MOCKCBI takes a set of visibilities and a map, then replaces the visibilities by the value they would have from the map, and adds Gaussian noise. By forcing MOCKCBI to use the undifferenced estimates of the noise rather than the scatter-based weights of the differenced data, the final combined data points will have the proper noise behavior. The data set can then be combined and χ^2 calculated. To avoid confusion caused by the presence of CMB signals, the maps were simulated with the CMB set to zero. Once simulated, the data were run through the standard pipeline to combine them into scans, and then combine the scans with scatter-based weights into final UV values for each antenna pair. I then calculated the χ^2 values for antenna pairs at identical UV points to estimate the final noise correction. Using the 20 hour deep field as a visibility template, the final noise correction value is 1.057 ± 0.002. The answer has been skewed somewhat by a minor bug in our pipeline program that mis-estimated the degrees of freedom by 1, leading to an error in the noise estimate of about 1%. So, the true value of the noise correction is probably more like 1.047, which is in excellent agreement with the predicted first-order theoretical value of 1.04, and the Fourier integral value of 1.042. The difference is likely due to the fact that some scans have fewer than 100 d.o.f., which will skew the correction to a larger value. The noise correction value that should be used is in actuality probably a bit higher. The reason is that individual UV points are not independent, but rather are correlated because of the primary beam. As such, maximum likelihood is combining, with weights, several different UV points to create independent estimators of the CMB. Those independent estimators will have contributions from many more scans than a single UV point in the final data set, which will have approximately 50 nights' worth of data at each point (since that's how many nights went into the 20 hour deep field). It is for this reason that the result from the Fourier integral calculations that the excess noise converges to its final

value in relatively few scans is critically important. Because of that, the true final value can be only marginally different from the Monte Carlo value for individual UV points.

4.2 GRIDR/MLIKELY Speedups

One would naively think that high-ℓ data wouldn't affect the low-ℓ power spectrum. This would be the case if there were no undesirable radio point sources in our observations. In the presence of sources, though, the high-ℓ data becomes extremely useful, and can be critical if there are very many sources (area per source on the sky comparable to the area of the synthesized beam). One of my major contributions to the CBI papers was optimizing the CBIGRIDR/MLIKELY pipeline to be fast enough to be able to use all the CBI data and to be able to investigate various spectrum properties. This section discusses some of the pipeline improvements. Their utility in testing the spectrum and improving response to sources are discussed in Section 4.3.

The most important speedup was the adoption of a hybrid gridding lattice in CBIGRIDR. The way CBIGRIDR works is to linearly combine ("grid") visibilities to create estimators of the sky intensity at a set of points $\boldsymbol{u}_i$ in the UV plane. Because the underlying sky intensities in the UV plane are uncorrelated (since they are equivalent to estimates of individual $a_{\ell m}$), the variance window functions for the gridded estimators are simple to calculate. During the gridding process, CBIGRIDR keeps track of the noise correlations introduced by the gridding to create the noise correlation matrix for the gridded estimators. There is no *a priori* requirement in CBIGRIDR about where to locate the estimators in the UV plane, but it should be on the scale of the effective beam in UV space. Since the UV beam is set by the sky coverage, the size scale in UV space is the Fourier transform of the half-power point of mosaic map on the sky. The expected behavior is that as the spacing of the estimators shrinks, the spectrum will become more accurate until the spacing reaches a critical level, roughly the Nyquist sampling interval, at which point a further decrease in estimator spacing won't change the spectrum. It is important to get the spacing right, since a too-large spacing loses information, and a too-small spacing increases execution time substantially. If we oversample by a factor of two, it's a factor of four in estimators (two in each dimension of the UV plane), and a factor

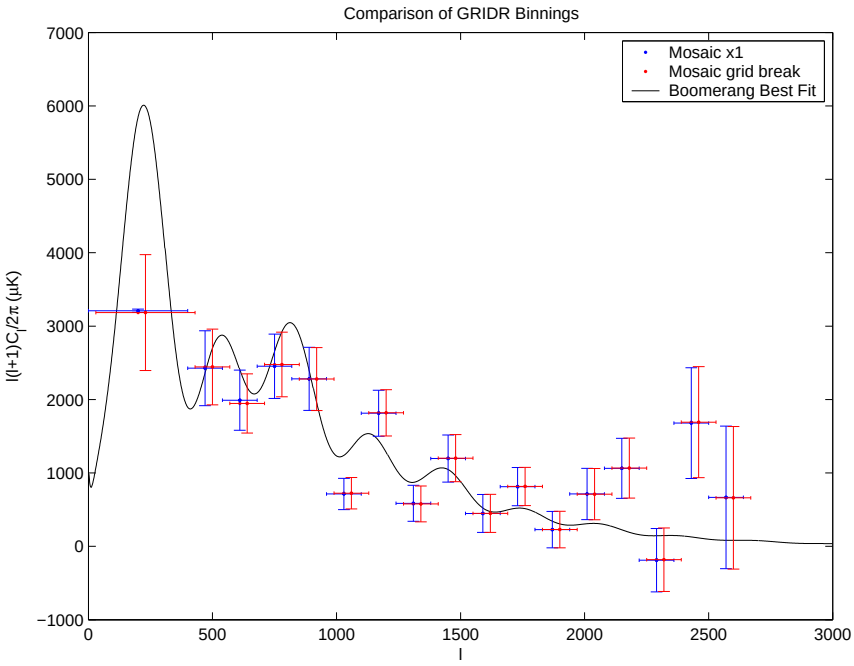


Figure 4.2 Comparison between spectra using a fine mesh in CBIGRIDR and a hybrid mesh with coarser sampling at $\ell > 800$. The two spectra have been offset in ℓ for greater clarity. The artificially low error bar on the first point for the fine mesh spectrum is due to the fact that we initially regularized the first bin, since the CBI rapidly loses sensitivity for ℓ much smaller than about 500. In this case, we regularized to the value from the unregularized spectrum, so the only effect is the small error bar.

of $4^3 = 64$ in run time, so penalty for oversampling is stiff indeed. In investigating the behavior of the output spectrum as the gridding was changed, I found that the high-ℓ spectrum converged at a coarser sampling than the low-ℓ spectrum, by about a factor of two, with the sensitivity change happening at $\ell \simeq 800$. This is presumably because the SNR on low-ℓ estimators is very high, and so the correlations are more important than at high-ℓ, and one needs to trace out the structure in the mosaic FT in more detail. However, we have not investigated the reasons behind the differing sensitivity in detail. Once I uncovered this effect, we changed CBIGRIDR to a hybrid lattice scheme, where estimators were placed on a split mesh, with a fine mesh at $\ell < 800$, and a coarse mesh, sampled half as often, for $\ell > 800$. The spectrum produced from the hybrid grid scheme was virtually identical to that from the uniform, finely-sampled grid. A comparison of the two spectra is shown in Figure 4.2. The speedup from the hybrid mesh happens both in CBIGRIDR, because each visibility is gridded onto fewer estimators, and in the linear algebra part of the pipeline, MLIKELY,

since fewer estimators means smaller matrices. The speedup is a bit less than the canonical 4 in CBIGRIDR ($\frac{1}{4}$ as many estimators) and 64 in MLIKELY ($\frac{1}{4}^3$) because of the fine gridding, but it is close to these values. The number of estimators in a coarse grid to $\ell = 1600$ is the same number as that in a fine grid to $\ell = 800$. So, as long as the upper-ℓ cutoff is noticeably greater than 1600, then the number of estimators is dominated by the coarse mesh, and the speedup is large. Before we used the hybrid mesh, we were only using CBI mosaic data to $\ell = 2600$ because at that point it took over a day on a 32-CPU supercomputer (Compaq GS320 with 733 MHz alpha CPUs with 64GB of RAM) to get a spectrum, and to get to the CBI upper limit of $\ell = 3500$ would have taken a factor of $(3500/2600)^6 = 6$ times as long. In addition to the computational burden, the memory requirements for the larger matrices would have pushed us over the 64 GB available on the computer. While we could perhaps have extracted a single spectrum (though even that was not clear), we would never have been able to test it. In contrast, with the hybrid mesh, it took approximately eight to ten hours to both grid and measure a spectrum to $\ell = 3500$.

I also made a couple of minor modifications to MLIKELY that helped quite a bit, especially when measuring several similar spectra from the same set of gridded estimators. The first was to add an option to start the spectrum fitting with an arbitrary, user-enterable spectrum instead of a constant value. This made the spectrum converge in fewer iterations if one had a good guess (as was the case for the investigation of source parameters in Section 4.3). Also, I found that MLIKELY seemed consistently to underpredict the shift in the spectrum to get to the maximum when iterating by a factor of a bit less than 2. By allowing the user to set a parameter by which MLIKELY scaled its step in the spectrum, I was able to get it to converge in fewer iterations. These two changes meant that MLIKELY converged to 1% of the error bars typically in 2-4 iterations (depending on the quality of the initial spectrum guess), whereas previously, it had been more like 12-14 iterations.

4.3 Source Effects in CBI Data

There is no correlation between low-ℓ and high-ℓ when observing the CMB with an interferometer. The response of a baseline to structure in the UV plane is the autoconvolution of the dish illumination

patterns measured in wavelengths, centered at the UV coordinate of the baseline. As such, a baseline has intrinsically zero response to regions in the UV plane more than twice the dish diameter (in wavelengths) away from the baseline UV position, independent of the shape of the primary beam. So, it is impossible for a 100 cm baseline to observe any CMB in common with a 400 or 500 cm baseline when using the CBI's 90 cm dishes. This suggests that if we are interested in the power spectrum at $\ell \sim 600$, then there is no point including data from $\ell \sim 3000$, since that data cannot contain CMB information in common with any baseline that observed around 600. Consequently, there is no reason in principle to include high-ℓ data where the CMB is not detected, since there is no information contained in the data. In fact, the price paid in running time for keeping high-ℓ data is very large. For a reasonably evenly sampled experiment, if we keep data up to ℓ_{max}, the number of independent patches in the UV plane $n \propto \ell_{max}^2$, and execution time is $\propto n^3$, for a total scaling of ℓ_{max}^6. While not immediately obvious, the presence of sources makes this argument invalid, and consequently it became important to push to as high an ℓ as possible when measuring the first-year power spectrum, even though the power spectrum at the highest ℓ's was thoroughly noise-dominated. In this section, I discuss how radio point sources affect the CBI spectrum and why using all of the CBI data, even that at high-ℓ, improves the low-ℓ power spectrum.

4.3.1 Source Effects on Low-ℓ-Spectrum

Radio point sources are a major contaminant of CMB data, especially at high-ℓ (larger than about 1800 at 30 GHz) where their power can become comparable to or larger than that of the CMB. The best way to deal with them is, of course, to know their fluxes and subtract them off. In practice, there are too many sources to measure them all. There are of order 56 sources per square degree brighter than 2.5 mJy at 1.4 GHz (in NVSS Condon et al., 1998), or about a source every 8 arcminutes. We measure those brighter than 6 mJy at 1.4 GHz with the OVRO 40 meter telescope as in Section 3.1, and subtract those with measured flux greater than 8 mJy at 30 GHz. This leaves substantial uncertainties in the residual flux from the point sources that is difficult to estimate (since the statistics of faint sources at 30 GHz are poorly known) that can add significant amounts of power to

the CBI spectrum at high-ℓ. Because the flux is unknown, and therefore unsubtractable, we instruct the analysis pipeline to set the uncertainty of the source flux to an extremely large number, thereby ignoring any flux it may have during the spectrum extraction. This process is called projecting out sources. To see how to downweight the sources, consider a source of unknown amplitude described by a visiblity vector $\mathbf{\Delta}$. If we then add $q\mathbf{\Delta}^T\mathbf{\Delta}$ to the noise matrix, where q is some large parameter, maximum likelihood sets the noise at the source location to be extremely large, and the spectrum is insensitive to the true flux from the source (Bond et al., 1998). Because there is no way to know what the CMB is doing underneath the source, maximum likelihood loses the information about the CMB at that point as well. To project a source, we need only know its location, and not its flux (since the point of projection is to make flux from the source have no impact on the spectrum). Projection of sources has been successfully in the past by others (*e.g.* Halverson et al., 2002). The parameter q is called the projection amplitude, and is typically a very large number (we currently use 10^5), but not so large as to cause numerical instabilities in the matrix operations.

Fortunately it appears that there is not a population of sources too faint to appear in NVSS (hence with unknown positions) with enough flux to significantly affect the CBI power spectrum, as neither the CBI nor BIMA (Dawson et al., 2002) see any sources at 30 GHz down to a few mJy that aren't present in NVSS. BIMA especially would be sensitive to such a population since they have larger dishes. We would like to project out all of the NVSS sources since we don't know which of them are problematic. If we restrict ourselves to, say, the 100 cm baselines, then the beam size is about 15 arcminutes, and so there are roughly four sources per beam. If the sources are projected out, then almost all the data is lost due to the projection. As we go to higher ℓ, the situation must improve at some level since there are more independent beams in the UV plane, but the total number of modes lost to sources is fixed, since the number of sources is fixed and each one deletes a single mode. The question remains, though, what is maximum likelihood actually doing when it projects out sources, and what are the effects expected in the spectrum?

4.3.2 Two Visibility Experiment

To gain insight into the general behavior of sources in maximum likelihood, let us consider a simple experiment. There is a single source at the center of the observed field, and two interferometer baselines. One baseline is short and observes the microwave background, and the other is sufficiently long so that there is a negligible contribution to it from the CMB. The noise in the two baselines is the same, and they are both equally sensitive to the source. If the assumed source amplitude in the visibilities is defined to be $\sqrt{a}$, then the vector of visibilities is $(\sqrt{a} \quad \sqrt{a})$, and the source matrix is the outer product of the visibility vector. Let us also assume that the expected CMB signal on the short baseline is equal to the noise. Under these assumptions, the noise matrix, the source matrix, and the CMB window matrix are (listing the short visibility first):

$$\mathrm{N} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \mathrm{Source} = \begin{pmatrix} a & a \\ a & a \end{pmatrix} \quad \mathrm{CMB} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{4.4}$$

To project out the source, we let $a \to \infty$. In the simple case we have just discussed, we can analytically examine the behavior of maximum likelihood as we change a. If $a = 0$, then there is no source, and the problem is diagonal. There is only one measurement of the CMB contained in the short visibility, and it has an SNR of one. If a is non-zero, then the effective noise matrix (noise+source) is not diagonal, but we can do a rotation that will make it diagonal. The effective noise matrix is

$$\mathrm{N_{eff}} = \begin{pmatrix} 1+a & a \\ a & 1+a \end{pmatrix} \tag{4.5}$$

which has eigenvectors

$$v_1 = \begin{pmatrix} \sqrt{0.5} \\ \sqrt{0.5} \end{pmatrix} \quad v_2 = \begin{pmatrix} -\sqrt{0.5} \\ \sqrt{0.5} \end{pmatrix} \tag{4.6}$$

and eigenvalues

$$\lambda_1 = 2a + 1 \quad \lambda_2 = 1 \tag{4.7}$$

If we use these to rotate into the space in which the effective noise is diagonal, we have

$$\text{Noise}_{\text{eff,rot}} = \begin{pmatrix} 2a+1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{CMB}_{\text{eff}} = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix} \tag{4.8}$$

We can now take the limit as $a \to \infty$ to see that the effective noise in the first (sum) mode goes to infinity, while the effective noise in the second (difference) mode remains constant at one. But the price paid is that the second mode has an expected power of $\frac{1}{2}$, whereas the short baseline visibility originally had a power of 1. So what maximum likelihood has done is to create an estimator intrinsically free from source contamination though the new estimator is noisier. Because the noise on both baselines has been combined to get the source-free estimator, it is important to measure both visibilities as well as possible. In fact, if one is free to allocate a fixed amount of time between the two visibilities, the optimal SNR is when the time is split evenly (see Figure 4.3).

The source has also coupled visibilities on different scales, which will lead to increased correlations between bins, in much the same way that knocking holes in a map will broaden its Fourier transform. One could add CMB into the long baseline visibility, and then the output source-free mode would have contributions from both the low-ℓ and high-ℓ CMB. In this simple case, it would also be correct to think of maximum likelihood using the long baseline to measure the flux from the source and subtract it. This works because the long baseline is sensitive to only the flux from the source and so is a pure measurement of the source brightness. In the general case, though, there is no such pure measurement, and so there is no estimate of the source flux to subtract. So, it is more correct to think of the process as creating source-free modes rather than subtracting off sources.

4.3.3 Sources in a Single Field

It is also important to study the effects of sources in more realistic situations. Rather than Monte Carlo a set of simulations, it is possible to use window matrices to calculate the expected response. To do this, I created a set of baselines in a single pointing covering a range in ℓ with uniform sampling and noise per area in the UV plane, with 5 point sources projected out, one at the pointing

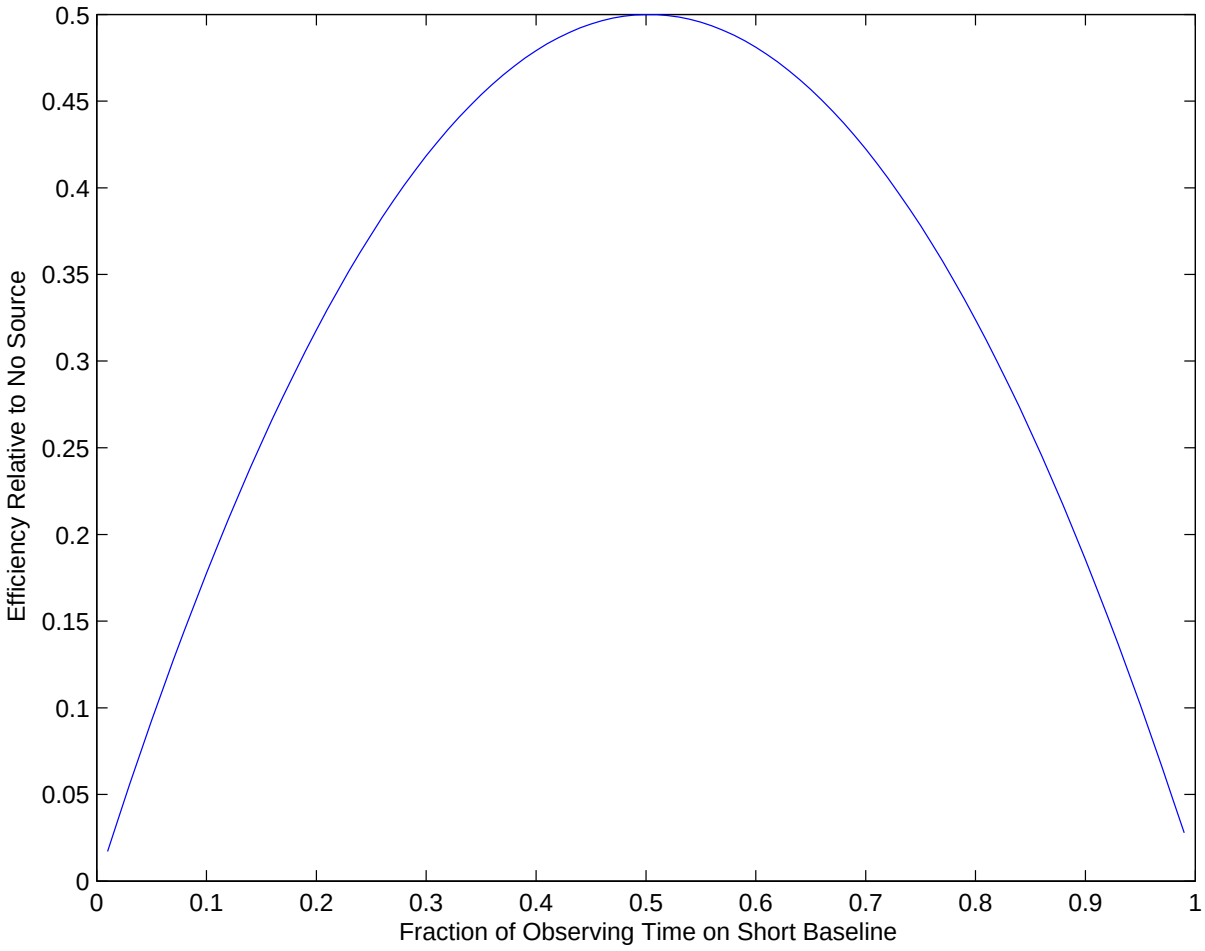


Figure 4.3 Relative efficiency of a two visibility experiment with one long baseline and one short baseline. The short baseline is sensitive to both the CMB and a foreground point source, while the long baseline is sensitive only to the source. The two baselines are equally sensitive to the source. If the source amplitude is unknown, then the optimal distribution of observing time is an even split between the short and long baselines. This is true even though the long baseline contains no information about the CMB.

center and one each at ± 15 arcminutes in RA and Dec. A single window matrix was used, with a flat band power out to $\ell = 780$, in order to investigate the behavior of low-ℓ bands due to sources. There are two numbers of interest: one is the total signal available in the data set, and the other is the fraction lost due to sources. Figure 4.4 show the behavior of these quantities as the ℓ cutoff of the data is changed. The total signal available is just the sum of the eigenvalues in the window matrix after a matrix transformation that takes the noise matrix into the identity matrix. This is equivalent to S/N when cosmic variance is unimportant, as is the case in low-S/N experiments (such as polarization). One can include cosmic variance, but it is more model dependent, depending in detail on the assumed S/N per area in the UV plane, though the general effect is to reduce the fraction of data lost to sources. The blue crosses in Figure 4.4 show how this total available signal varies with ℓ range. As expected, the available signal rapidly converges to its limiting value once the data range gets much past the upper ℓ limit of the window matrix. The same quantity can be calculated in the presence of sources by diagonalizing the noise+source matrix and scaling so that the noise+source elements are all one. The red asterisks show the amount by which the available signal falls short of the no-source available signal. Unlike the no-source case, the available signal continues to rise as the ℓ cutoff is increased since the high-ℓ data continue to help characterize the sources and source-free modes. In this case, a mere 5 sources are sufficient to cost half the data in a single pointing if only the data in the ℓ range of interest are used. In contrast, if the data out to $\ell = 400$ are used, then the price paid because of sources is only 5%. Since there are typically dozens of NVSS sources per field, broad ℓ coverage is critical.

4.4 Source Effects in the First-Year Mosaics

With the speedups in the pipeline from Section 4.2 I was able to extract the power spectrum out to high-ℓ. We had originally planned to use the source projection parameters that had been derived and extensively tested from the deep fields by Brian Mason. The method that he found worked for the deep fields was to measure all sources bright than 6 mJy in NVSS with the OVRO 40m using a 30 GHz, four-channel receiver. Those sources measured brighter than 4σ (about 8 mJy) at 30

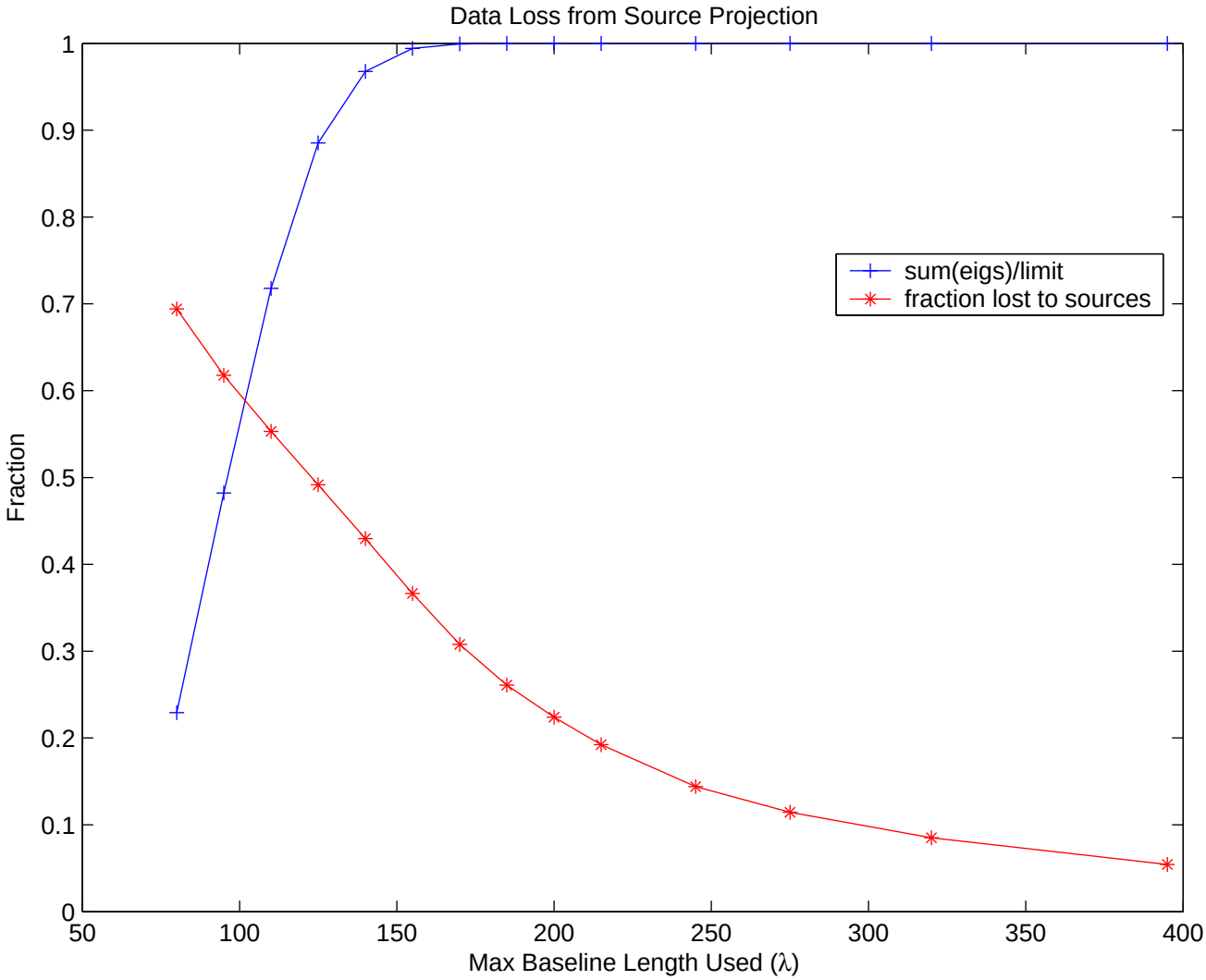


Figure 4.4 Expected behavior of total signal available and signal lost due to sources as the ℓ range of the data is varied.

GHz were subtracted from the data set. The statistics of the NVSS sources detected by OVRO were used to estimate a best spectral index to extrapolate the flux from 1.4 to 30 GHz for the undetected and faint NVSS sources. The source projection matrix we used was the sum of the outer products of the flux from each source gridded onto the estimators, using the extrapolated source brightness for the unmeasured sources. It was this matrix, multiplied by a projection factor, that was added to the noise matrix to remove the sources with known position (there is also a contribution from sources too faint to appear in NVSS, calculated the same way as the source signal in Chapter 3. This contribution is small - see Section 4.6.) For the deep fields, a source projection factor of 100 was sufficient to remove source effects, with the spectrum insensitive to variation in the projection coefficient at values higher than that. Because the mosaic had been much slower to run ($\sim$ 1-2 days on the 32 CPU Dec Alpha machine), we had anticipated setting the mosaic source projection parameters using the deep field source parameters, rather than spend the CPU time to investigate the sources in the mosaics separately. After the improvements to the pipeline, it was fast enough to investigate the effects of different source parameters. In doing so, I found that a substantial source signal remained in the mosaics. We had originally found power at high-ℓ in the mosaics (ℓ above $\sim$1600) of about 1000 μK^2 that would have been very difficult to explain cosmologically, and was about a factor of 2 larger than that in the deep fields in Mason et al. (2003). See Figure 4.5 for the power spectrum. In investigating the mosaic spectrum, I discovered that the spectrum calculated using the deep-field projection level of 100 had not reached the limiting regime at which point sources were truly projected out. While initially surprising (we were after all projecting out similar source populations), the behavior is actually sensible. The reason is that projection works by downweighting the importance of the mode that contains the source information. The weight is, from Chapter 2, $\frac{S}{S+N}$ (again, these are defined in terms of variance and not σ). At high-ℓ, we are thermal-noise limited, which means what the weight is roughly $\frac{S}{N}$. Projecting a source with a fixed amplitude adds a fixed amount to N, dropping the weight of the mode. For the deep fields, the noise per beam at high-ℓ was quite a bit smaller than for the mosaics, typically 1 mJy versus $>$ 4 mJy. A mode with a source at 5 mJy projected out in the deeps will have a weight relative to the weights

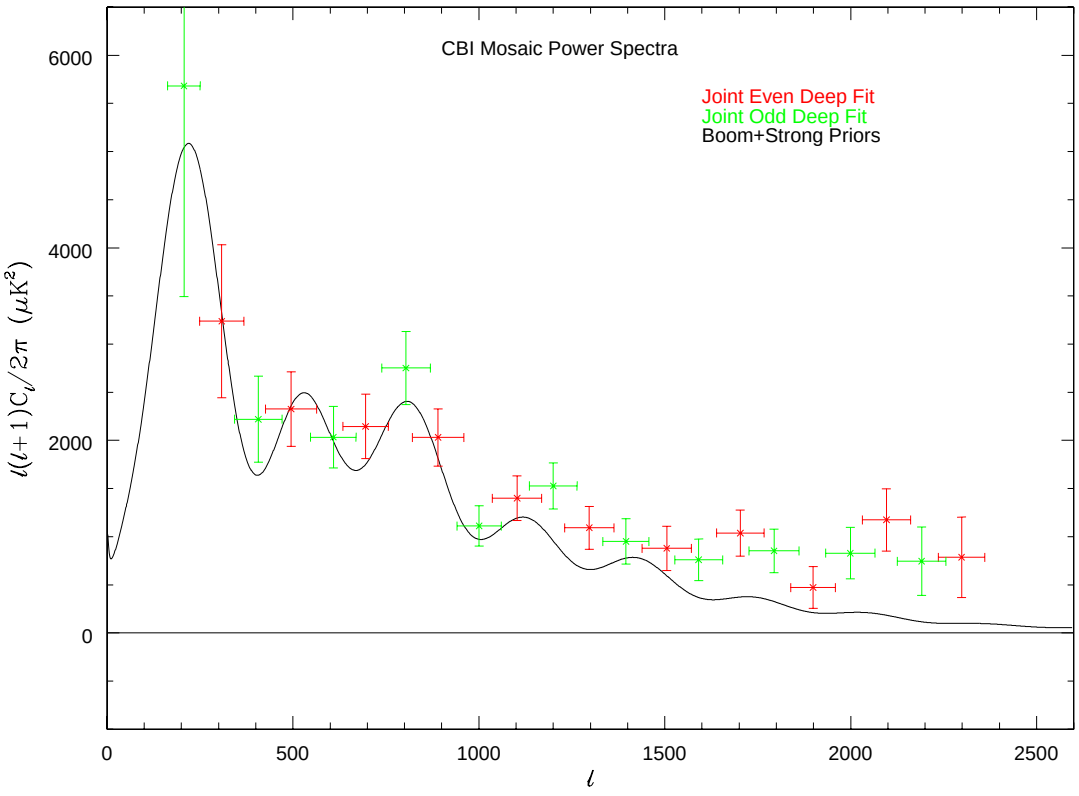


Figure 4.5 Original mosaic power spectrum using deep-field source projection parameters. The high power level at $\ell > 1600$ is due to the inapplicability of deep-field source projection parameters to the mosaic power spectrum.

of other similar modes of $1\text{mJy}^2 + 25\text{mJy}^2$ against 1mJy^2. So, it will be downweighted by a large factor relative to the other data (in this case about 25). The same source projected at the same level in the mosaics will, however, have a relative downweighting of $16 + 25\text{mJy}^2$ against 16mJy^2, or only a bit less than 50%. So, the source will not really have been projected out of the mosaic spectrum even though it is gone from the deep spectrum. See Figure 4.6 to see how the spectrum changes as the source projection level is varied between 4 and 10^4. We finally adopted a projection level of 10^5, the largest value that was comfortably numerically stable. It is, in general, a good idea to use the largest projection value possible. The reason is that modes enter into maximum likelihood like

$$\frac{S}{S+N}\left(\chi^2 - 1\right) \tag{4.9}$$

(from Equation 2.9). As the projection level increases, the weight drops, but so does χ^2, thereby introducing a bias. The projection required is higher for low-ℓ modes since they have a much higher signal, so a high projection level is required to move past their bias regime. It is for this reason that we use a large value for the projection.

To get an idea of the effects discussed in Section 4.3 see Figure 4.7. It is a plot of the spectrum produced with the original, low source level and two data cutoffs, one at $\ell = 2600$ and one at $\ell = 3500$. The cutoff at $\ell = 3500$ contains essentially all the CBI data. The error bars are slightly larger in the low-cutoff spectrum (most easily seen in the bins centered at 900 and 1900), though not substantially so with a source projection amplitude of 100. This is because most faint sources (which constitute most of the sources) are not projected out at low-ℓ when the projection amplitude is 100. The difference between the $\ell = 2600$ and $\ell = 3500$ cutoff errorbars would be substantially larger using a higher projection amplitude. We have never done the direct comparison, though, since by the time we realized the projection level needed to be higher, the high projection, $\ell = 2600$ spectrum would have required a complete re-run of the entire spectrum pipeline. The CPU time was more productively used doing more tests of the $\ell = 3500$ spectrum, so we never produced the high projection, $\ell = 2600$ spectrum.

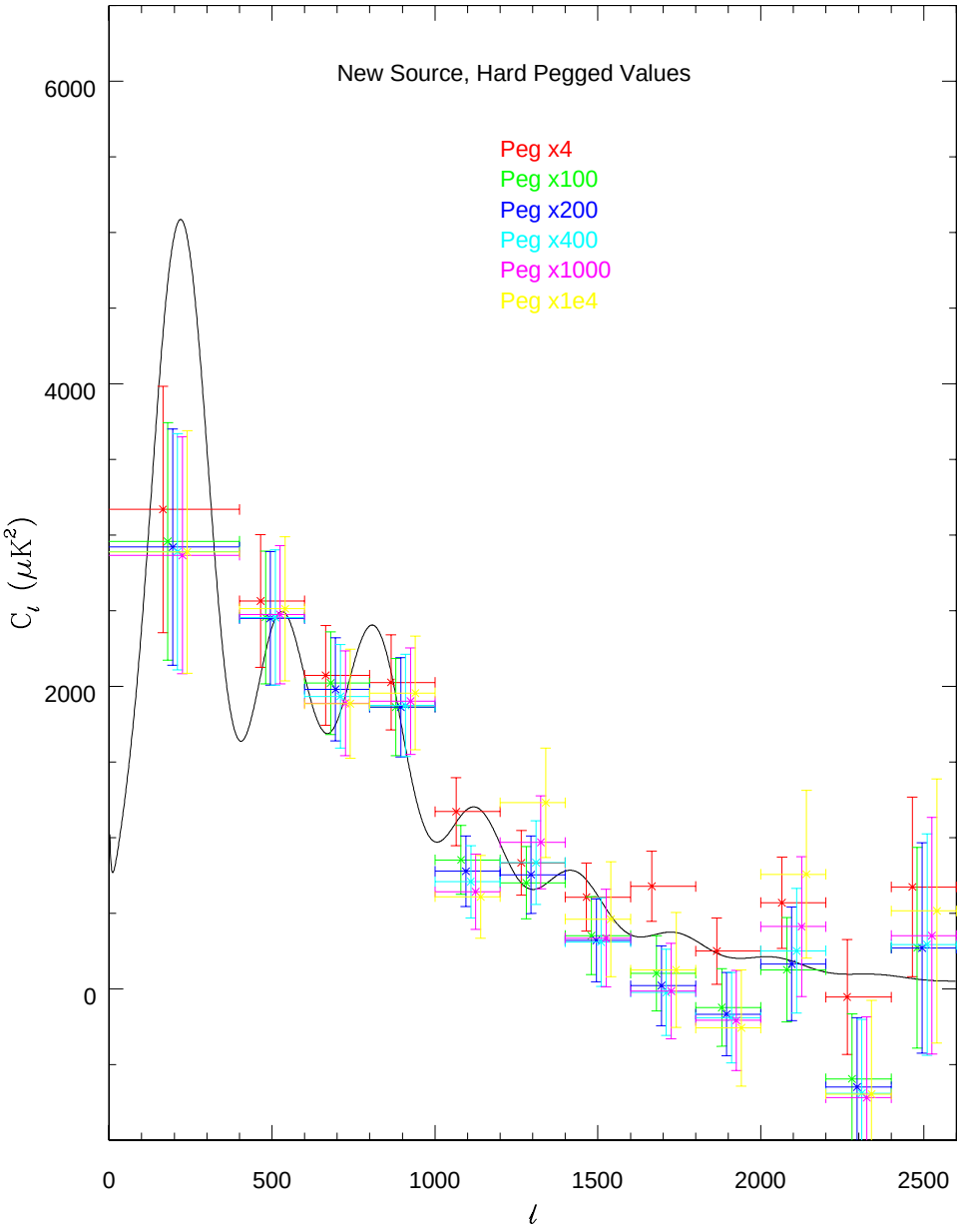


Figure 4.6 Mosaic power spectrum as a function of various source projection levels. Note how the higher projection levels are systematically lower that the projection at 4 times the predicted source amplitude. The lower power level indicates that a substantail fraction of the high-ℓ flux at low levels is due to flux from sources that has not been fully projected out. The dip to low power levels as the projection amplitude is increased followed by a slight rise is typical maximum likelihood behavior, and the reason why as high a projection level as is numerically stable is desired. The final level we used was 10^5.

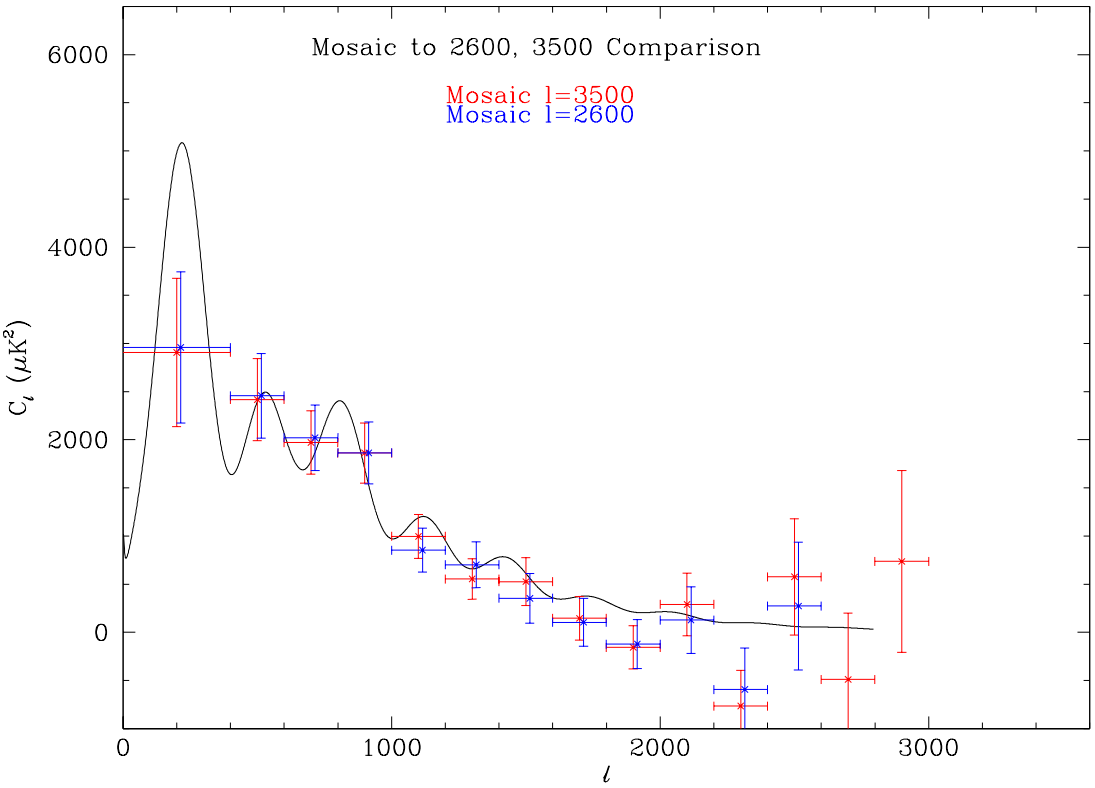


Figure 4.7 Comparison of mosaic power spectra with the data running to $\ell = 2600$ (blue points) and $\ell = 3500$ (red points). The increased error bars with the lower ℓ cutoff can be seen most easily in the bins centered at $\ell = 900$ and $\ell = 1900$. These early runs were done with a projection level of 100, much lower than our final adopted value. The difference between the 2600 and 3500 cutoff would be substantially more striking with the higher projection levels.

4.5 First-Year Data

The first-year data falls into two sets of observations: a set of three deep fields (which includes the deep field data in Padin et al., 2001a), and a set of mosaic data. The mosaic data consist of three mosaics centered at $02^h50 - 03^\circ$, $14^h50 - 03^\circ$, and $20^h50 - 03^\circ$. Each mosaic covers roughly $2^\circ \times 4^\circ$, with the differencing for ground subtraction in the long direction, for an effective coverage of $2^\circ \times 2^\circ$. The individual mosaic pointings are summarized in Pearson et al. (2003). The deep field data are three pairs of differenced fields with the lead fields centered at $08^h44' - 03^\circ10'$, $14^h42' - 03^\circ50'$, and $20^h48' - 03^\circ30'$. The 14 hour and 20 hour deep fields are located inside the 14 hour and 20 hour mosaics, so there is some slight correlation between the mosaic and deep field results. The correlation is not strong, though, since only a couple of nights of the deep data in the 14 hour and 20 hour mosaics was included, and both the 08 hour deep and 02 hour mosaic are entirely independent. The deep data are summarized in Mason et al. (2003). The same observational constraints (night-time, $> 60^\circ$ from the moon, *etc.*), calibration, and differencing schemes discussed in Chapter 3 were used. Source subtraction was again carried out using source measurements from the OVRO 40 meter telescope. Maps of the three mosaics, both source-subtracted and unsubtracted are in Figures 4.8 through 4.10.

4.6 First-Year Results

4.6.1 Power Spectrum

The final first-year power spectrum results are in Table 4.1, and are plotted in Figure 4.11.

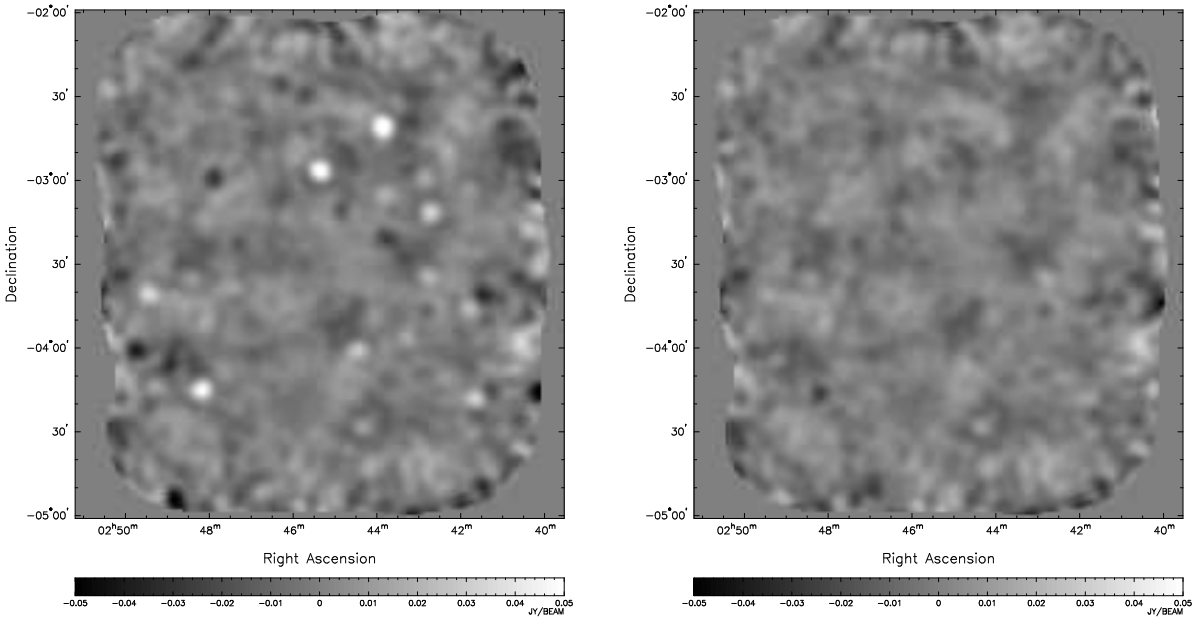


Figure 4.8 Map of the 02 hour mosaic. The left half shows the image before source subtraction, the right half shows the same image with the sources measured by the OVRO 40 meter subtracted. Especially on large scales, the large majority of the structure in the source-subtracted image is CMB and not noise.

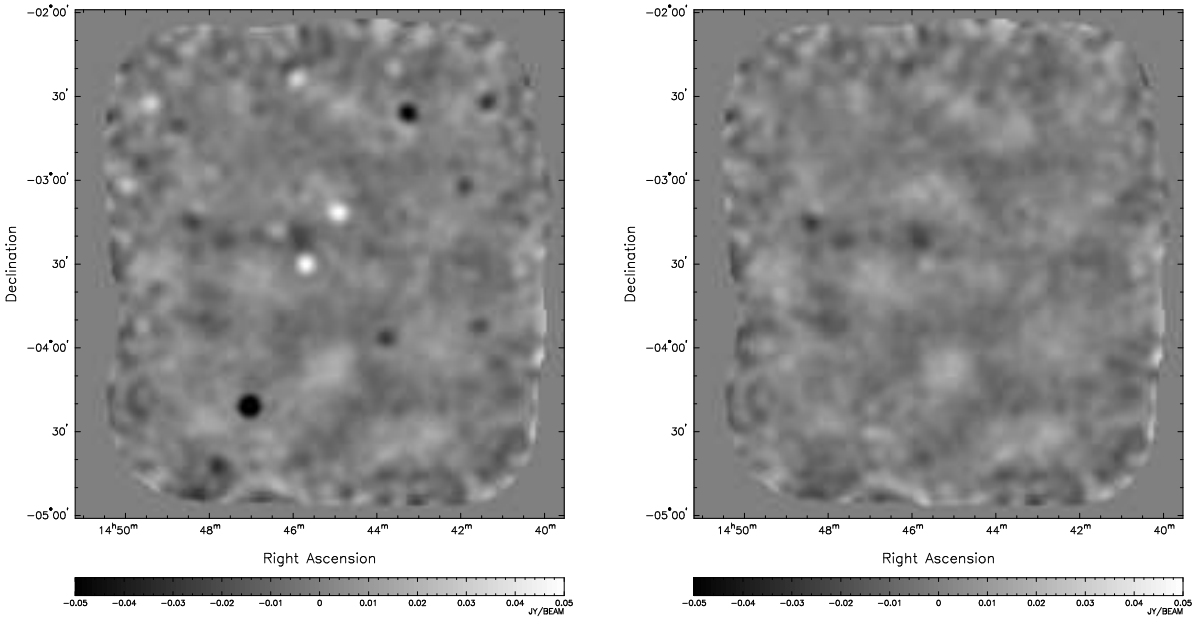


Figure 4.9 Same as Figure 4.8 for the 14 hour mosaic.

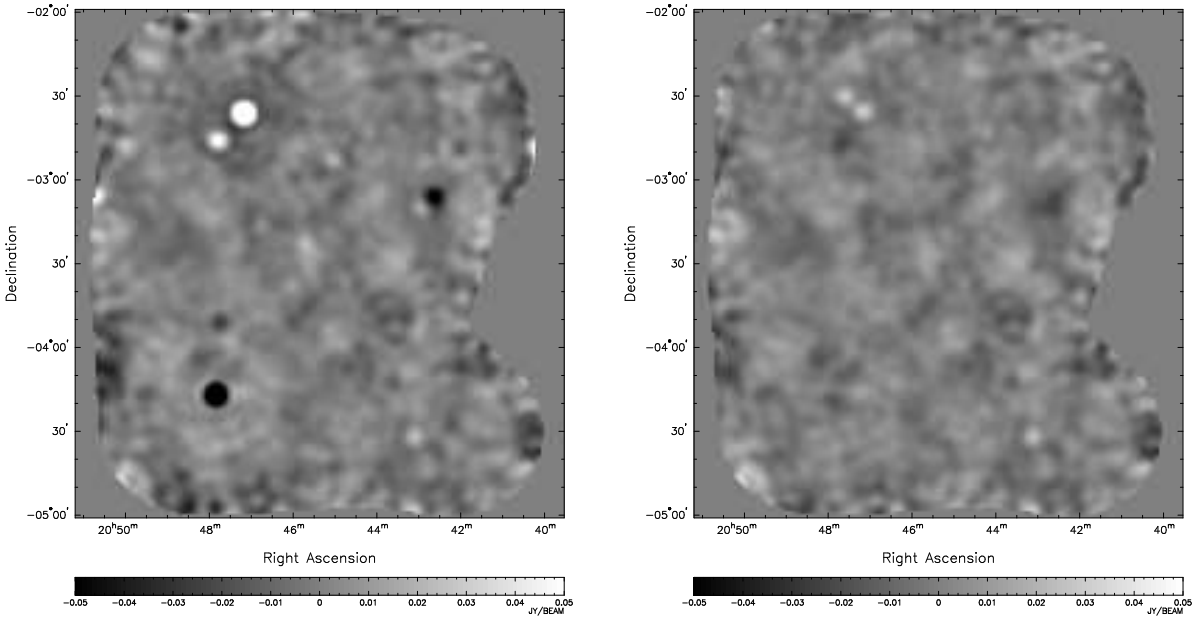


Figure 4.10 Same as Figure 4.8 for the 20 hour mosaic.

Table 4.1. Band Powers and Uncertainties (from Pearson et al. (2003))

ℓ range	$\ell_{\rm eff}$	Band Power $l(l+1)C_l/(2\pi)$ (μK^2)
		Even Binning
0–400	304	2790 ± 771
400–600	496	2437 ± 449
600–800	696	1857 ± 336
800–1000	896	1965 ± 348
1000–1200	1100	1056 ± 266
1200–1400	1300	685 ± 259
1400–1600	1502	893 ± 330
1600–1800	1702	231 ± 288
1800–2000	1899	−250± 270
2000–2200	2099	538 ± 406
2200–2400	2296	−578± 463
2400–2600	2497	1168 ± 747
2600–2800	2697	178 ± 860
2800–3000	2899	1357 ± 1113
		Odd Binning
0–300	200	5243 ± 2171
300–500	407	1998 ± 475
500–700	605	2067 ± 375
700–900	801	2528 ± 396
900–1100	1002	861 ± 242
1100–1300	1197	1256 ± 284
1300–1500	1395	467 ± 265
1500–1700	1597	714 ± 324
1700–1900	1797	40 ± 278
1900–2100	1997	−319± 298
2100–2300	2201	402 ± 462
2300–2500	2401	163 ± 606
2500–2700	2600	520 ± 794

Table 4.1—Continued

ℓ range	$\ell_{\rm eff}$	Band Power $l(l+1)C_l/(2\pi)$ (μK^2)
2700–2900	2800	770 ± 980

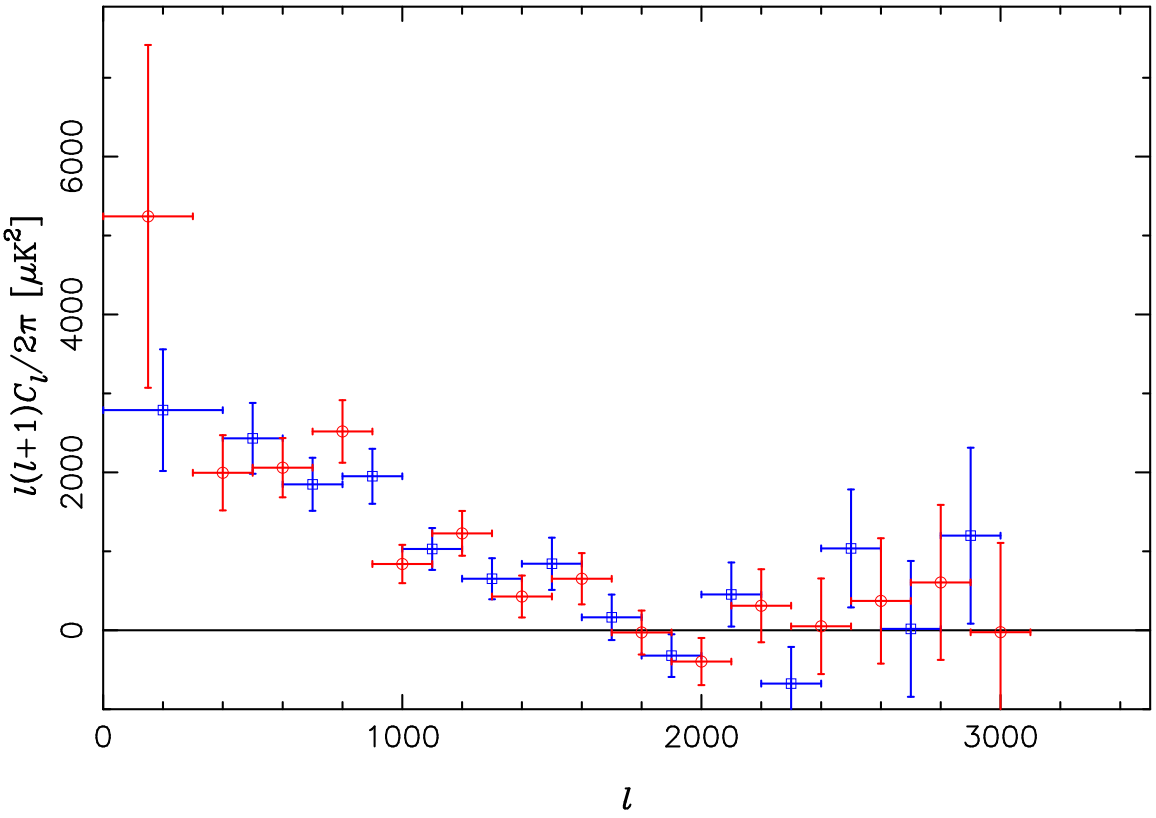


Figure 4.11 Final first-year power spectrum, binning is $\Delta\ell = 200$. Red and blue points are two different binnings for the same data. Adjacent same-colored points are from the same spectrum and are weakly correlated ($\sim 20\%$). Adjacent different-colored points are not independent and we expect their correlations to be very high.

The spectrum was calculated with sources detected by OVRO subtracted, a source projection factor of 10^5, and an isotropic faint source contribution of 0.08 Jy^2 per steradian, or 25 mJy^2 per square degree. There are two completely separate power spectra extracted from the same data using two different binnings, the "even" and "odd" binnings in Table 4.1. On the plot, the "even" binning is the blue points, and the "odd" binning is the red points. Points from within a single binning are basically independent, with correlations $< 20\%$. Adjacent points from different binnings (*e.g.* a red point compared to the nearest blue points) are *not* independent and have unknown correlations, as they were produced in different pipeline runs. Similarly, when using the CBI's power spectra to

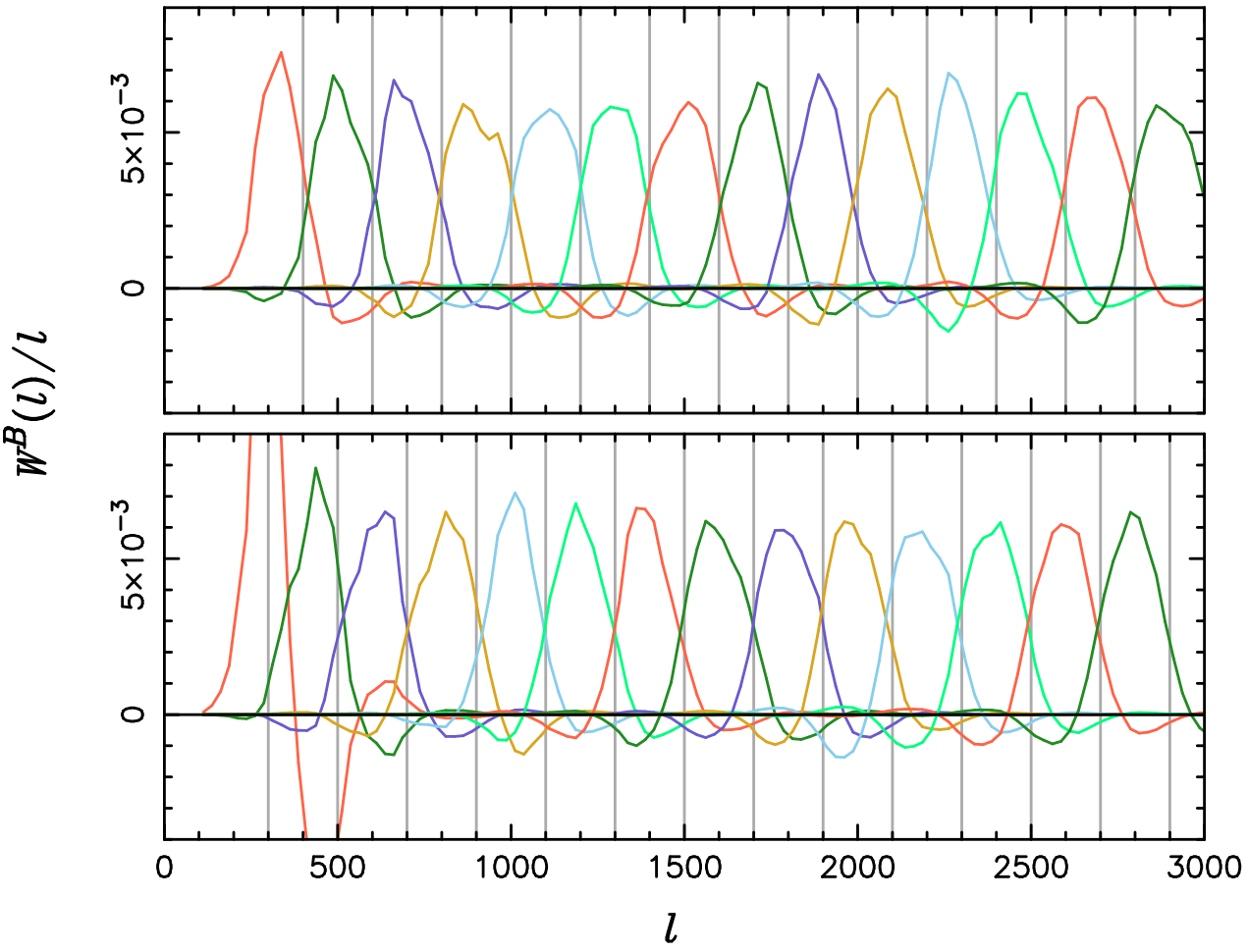


Figure 4.12 The CBI mosaic band power window functions. The upper panel shows the "even" binning and the lower the "odd" binning. The expected value in a CBI bin is $\int \mathcal{C}_\ell W^B(\ell)/\ell$, so the window functions can transform a power spectrum into the experimental space of the CBI power spectrum.

compute cosmological parameters, one should use either the even or the odd binning, but not both. The band power window functions, that describe the sensitivity of the CBI bands to the CMB power at a given ℓ, are in Figure 4.12. They can be used to transform a model $\mathcal{C}_\ell$ spectrum into expected CBI band powers, subject to the caveats of Section 2.5.

The CBI spectrum is in very good agreement with that of other experiments. Figure 4.13 shows the same spectrum along with a reference model from a fit to BOOMERANG data. This model does not depend at all on CBI data, and in fact only depends on BOOMERANG data out to $\ell = 1000$. Figure 4.14 shows the CBI's spectrum plotted along with the actual spectra from DASI,

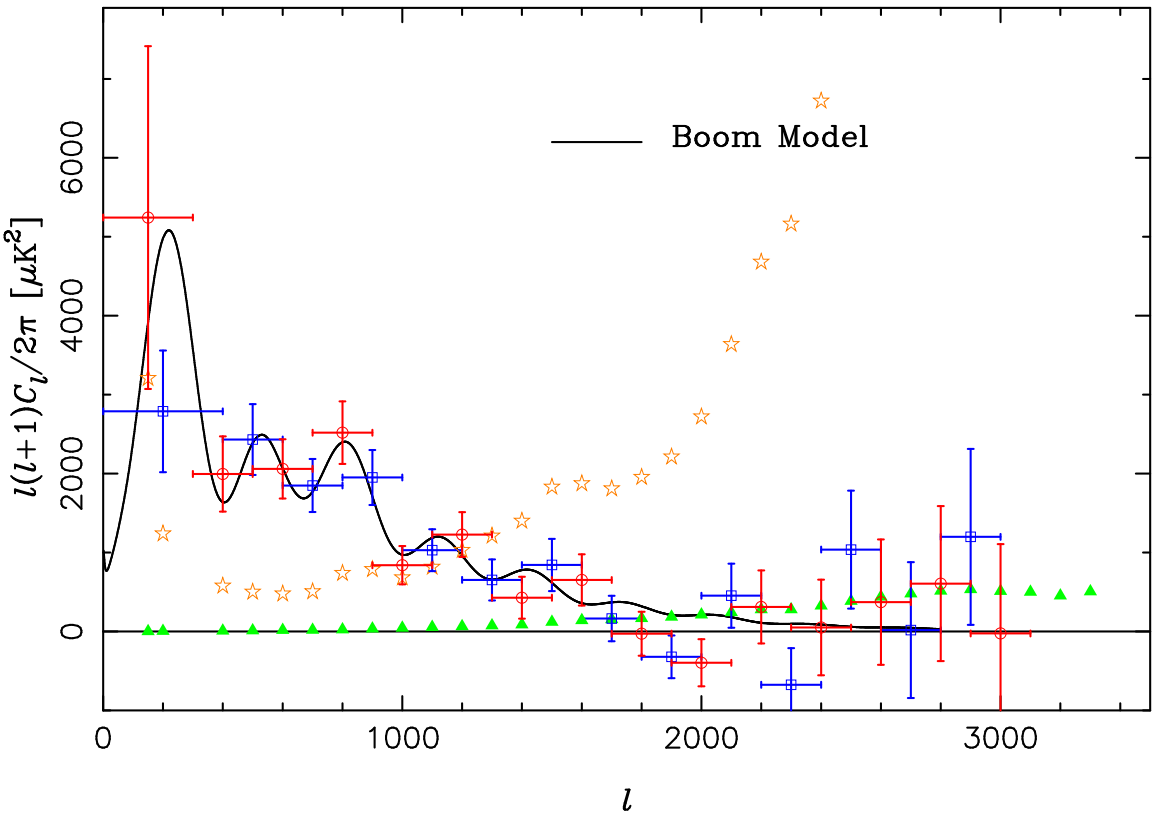


Figure 4.13 Same as Figure 4.11, with a fit to BOOMERANG plotted for reference. The noise spectrum is also plotted, that is the amount of power contributed by noise. If one were to change the estimated noise by a fraction ϵ, the CMB spectrum would shift by ϵ times the noise spectrum. The green triangles are the amount that the isotropic, faint-source correction has shifted the power spectrum. The data follow the curve remarkably well, even though the curve is a fit to an entirely unrelated data set that only extends to $\ell \sim 1000$. The reference model has parameters $\Omega = 1$, $\Omega_{cdm}h^2 = 0.12$, $\Omega_B h^2 = 0.02$, $n_s = 0.975$, and $\tau_c = 0.1$.

BOOMERANG, and MAXIMA. Again, the agreement is excellent between all experiments. The figure also shows by how much the CBI extended the ℓ range over which the CMB power spectrum is measured, as well as the contribution from sources too faint to appear in NVSS.

We also measured the CBI mosaic power spectrum using the same binning as the CBI deep fields, and found that the agreement was good, with $\chi^2 = 5.77$ for 5 degrees of freedom. Of note is the power level at high-ℓ (> 2000) in the deep fields that is higher by $> 3\sigma$ than that predicted by standard cosmologies. We believe this may be the first detection through the CMB power spectrum (rather than pointed observations of clusters) of secondary anisotropy due to the SZ effect (Bond et al., 2002b). Another intriguing suggestion is that of Oh et al. (2003) where the SZ effect due

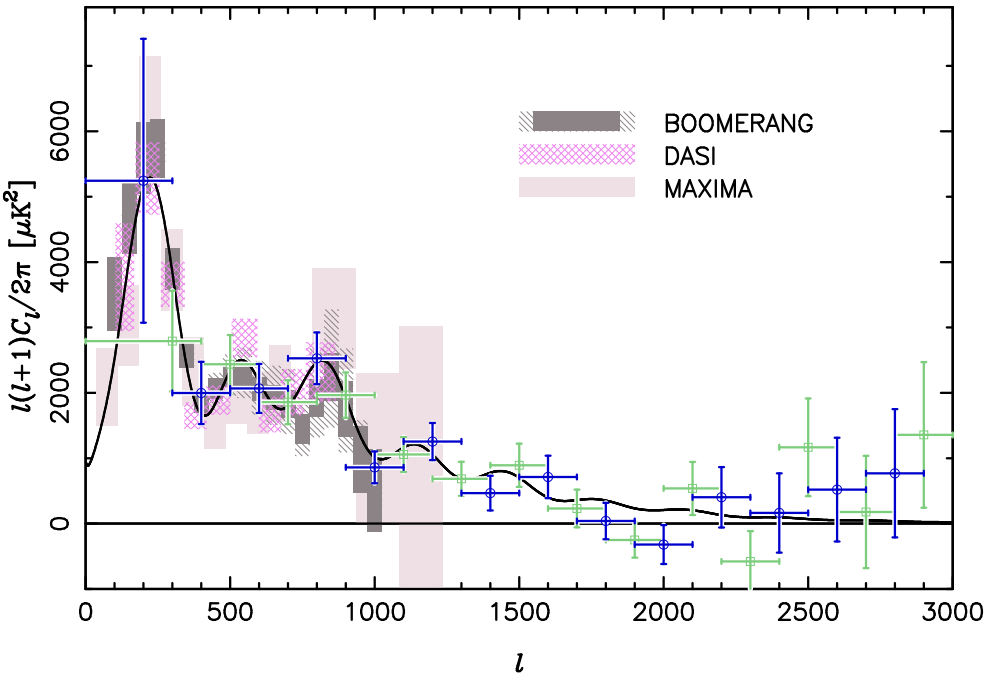


Figure 4.14 CBI spectrum, along with the BOOMERANG, DASI, and MAXIMA spectra. The agreement between all experiments is striking. Also note how much the CBI extends the range over which the CMB power spectrum is measured. From Pearson et al. (2003)

to winds from supernovae in Population III stars is shown to be comparable to the high-ℓ power. The most obvious potential low-z source for this signal is radio point sources, but it is difficult to create even a baroque source population capable of creating such a high power level. The power level is equivalent to a single source of 10 mJy in each field, but since the noise is low ($<$ 1mJy), the flux would have to be split amongst several fainter sources ($<$ 4mJy to be below the confusion limit of the CBI) per field that do not appear in NVSS. However, such a population would appear in Dawson et al. (2002), which consists of higher spatial resolution observations also at 30 GHz with the larger BIMA dishes, as either a population of resolved sources at a few mJy not in NVSS or a large collection of unresolved faint sources. They do not see a new population of resolved sources at a few mJy, and an unresolved population would lead to a much higher power level in their data (at $\ell \sim 7000$) than in the CBI high-ℓ data (at $\ell \sim 2500 - 3000$), rather than the slightly lower value observed. So, the excess power is highly unlikely to be from point sources. The CBI high-ℓ measurement also marked the first time that the CMB had been detected on length scales

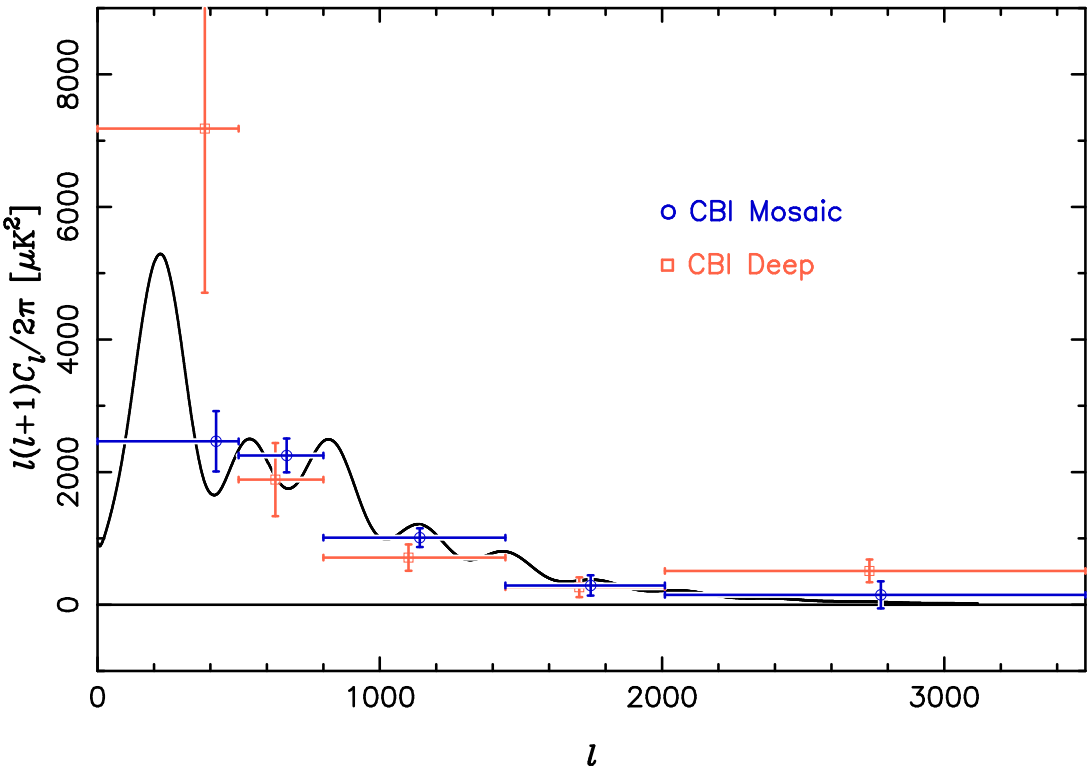


Figure 4.15 Mosaic and deep field spectra, with the mosaic using the same binning as the deep. This makes comparisons between the two sets of spectra straightforward. The agreement between the two is good, with $\chi^2 = 5.57$ for 5 degrees of freedom.

equivalent to masses as low as $10^{14}\,M_\odot$ – the size of virialized clusters in the local universe. The high-ℓ fluctuations are the seeds from which today's galaxy clusters form.

Finally, to see how the CBI spectrum compares to the recently released WMAP (Hinshaw et al., 2003) and ACBAR (Runyan et al., 2003) spectra, see Figure 4.16. This shows, as a teaser, the 2000+2001 mosaic spectrum from the CBI, which represent a substantial improvement over the first-year data. The results from the 2000+2001 have not been released yet, so this work restricts itself to the 2000 data (although the full 2000+2001 data set is used in the spectral index measurements of Chapter 5). Worth mentioning is that because the SZ signal is weaker at the higher frequencies at which ACBAR observes, if the CBI high-ℓ power were due to the SZ effect, one would expected it to be a factor of a few lower in the ACBAR spectrum, consistent with what they observe.

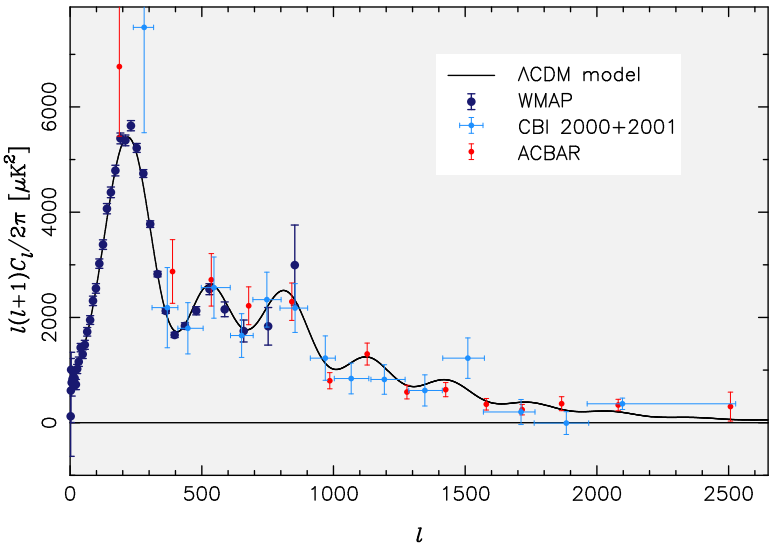


Figure 4.16 Comparison of CBI 2000+2001 data (light blue) with WMAP (dark blue) and ACBAR (red). This is only a single binning of the CBI data. Again, the agreement between different experiments is very good.

4.6.2 Cosmology with the CBI Spectrum

One of the fundamental uses of CMB observations is to measure cosmological parameters both reliably and accurately. We used the CBI spectra to measure parameters both in isolation (using COBE-DMR as a very low-ℓ anchor) and in combination with other experiments. The formalism and results are discussed in detail in Sievers et al. (2003). The basic idea is to approximate the likelihood surface around the peak using an offset lognormal approximation (Bond et al., 2000) to the surface. Predicted bin values can be taken from a model cosmological spectrum $\mathcal{C}_\ell$ and turned into predicted values using the band power window functions. The offset lognormal can then be used to give a likelihood that the model in question would have yielded the observed spectrum. We repeat this procedure for a grid of models to create a likelihood surface for cosmological parameters. The surface can then be projected along various dimensions to give the likelihood of a desired parameter, *e.g.* Ω_k, n_s, etc. The grid of model spectra is described in Table 4.2. In addition, the overall spectrum amplitude $\mathcal{C}_{10}$ is treated as a continuous parameter than can be integrated, rather than requiring a discrete sum on a model grid. We also use various combinations of prior information to try and break some of the parameter degeneracies in the CMB spectrum described in the introduction, such

Table 4.2. Parameter Grid for Likelihood Analysis. From Sievers et al. (2003)

Parameter	Grid:												
Ω_k	0.9	0.7	0.5	0.3	0.2	0.15	0.1	0.05	0	-0.05	-0.1	-0.15	
	-0.2	-0.3	-0.5										
$\omega_{\rm cdm}$	0.03	0.06	0.08	0.10	0.12	0.14	0.17	0.22	0.27	0.33	0.40	0.55	0.8
ω_b	0.003125	0.00625	0.0125	0.0175	0.020	0.0225	0.025	0.030	0.035	0.04	0.05	0.075	
	0.10	0.15	0.2										
Ω_Λ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	
n_s	1.5	1.45	1.4	1.35	1.3	1.25	1.2	1.175	1.15	1.125	1.1		
	1.075	1.05	1.025	1.0	0.975	0.95	0.925	0.9	0.875	0.85	0.825		
	0.8	0.775	0.75	0.725	0.7	0.65	0.6	0.55	0.5				
τ_c	0	0.025	0.05	0.075	0.1	0.15	0.2	0.3	0.4	0.5	0.7		

as the HST H_0 key project, or constraints from large-scale structure measurement. Then the priors can be used to calculate the *a priori* likelihood that a particular model could have given rise to the priors. This likelihood is then multiplied by the likelihood from the power spectrum (in practice their log likelihoods are summed), to give a total likelihood that reflects both the knowledge from the CMB and the knowledge from the priors. The priors used in calculating cosmological parameters using the CBI spectrum are as follows:

1. wk-h - very general constraints designed to be noncontroversial. The Hubble constant is set to $0.45 < h < 0.9$, the age of the universe is restricted to $\mathrm{T}_0 > 10$ Gyr, and $\Omega_m > 0.1$.

2. flat - since CMB data (including the CBI) strongly suggest the universe is close to geometrically flat, a prior with $\Omega_k = 1$ seems reasonable.

3. LSS - a broad constraint on large-scale structure and matter clustering. It takes the form of a constraint on $\sigma_8\Omega_m^{0.56} = 0.47^{+0.02\ +0.11}_{-0.02\ -0.08}$ where the two sets of errors are convolved together, with the first error bar Gaussian and the second uniform. There is also a constraint on the effective shape parameter $\Gamma_{\rm eff} = 0.21^{+0.03\ +0.08}_{-0.03\ -0.08}$. More information on the LSS prior can be found in Bond et al. (2002b).

4. SN - constraint in the $\Omega_m - \Omega_\Lambda$ plane from Type Ia supernovae (see Perlmutter et al., 1999; Riess et al., 1998).

5. HST-h - measurement of the Hubble contant from the HST key project of 72 ± 8, as found in Freedman et al. (2001).

The CBI provided useful cosmological constraints. The cosmological parameters derived from

Table 4.3. Cosmic Parameters for Various Priors Using CBIo140+DMR. From Sievers et al. (2003)

Priors	$\Omega_{\rm tot}$	n_s	$\Omega_b h^2$	$\Omega_{\rm cdm} h^2$	Ω_Λ	Ω_m	Ω_b	h	Age	τ_c
wk-h	$1.00^{0.11}_{0.12}$	$1.08^{0.11}_{0.10}$	$0.023^{0.016}_{0.010}$	$0.16^{0.08}_{0.07}$	$0.43^{0.25}_{0.28}$	$0.59^{0.22}_{0.22}$	$0.083^{0.053}_{0.053}$	$0.58^{0.11}_{0.11}$	$13.9^{2.2}_{2.2}$	< 0.66
wk-h+LSS	$1.05^{0.08}_{0.08}$	$1.07^{0.10}_{0.10}$	$0.029^{0.012}_{0.012}$	$0.10^{0.04}_{0.03}$	$0.67^{0.10}_{0.13}$	$0.39^{0.12}_{0.12}$	$0.095^{0.055}_{0.055}$	$0.60^{0.12}_{0.12}$	$15.4^{2.1}_{2.1}$	< 0.66
wk-h+SN	$1.03^{0.08}_{0.08}$	$1.11^{0.11}_{0.11}$	$0.028^{0.016}_{0.012}$	$0.10^{0.05}_{0.04}$	$0.71^{0.08}_{0.09}$	$0.33^{0.08}_{0.08}$	$0.076^{0.046}_{0.046}$	$0.65^{0.12}_{0.12}$	$14.7^{2.4}_{2.4}$	< 0.67
wk-h+LSS+SN	$1.04^{0.08}_{0.08}$	$1.11^{0.11}_{0.10}$	$0.029^{0.016}_{0.012}$	$0.10^{0.04}_{0.03}$	$0.72^{0.07}_{0.07}$	$0.32^{0.08}_{0.08}$	$0.082^{0.047}_{0.047}$	$0.64^{0.11}_{0.11}$	$15.0^{2.2}_{2.2}$	< 0.67
flat+wk-h	(1.00)	$1.07^{0.11}_{0.10}$	$0.023^{0.010}_{0.008}$	$0.15^{0.06}_{0.04}$	$0.47^{0.25}_{0.27}$	$0.54^{0.24}_{0.24}$	$0.068^{0.028}_{0.028}$	$0.60^{0.12}_{0.12}$	$14.0^{1.4}_{1.4}$	< 0.65
flat+wk-h+LSS	(1.00)	$1.05^{0.15}_{0.09}$	$0.024^{0.011}_{0.009}$	$0.11^{0.02}_{0.02}$	$0.67^{0.10}_{0.13}$	$0.34^{0.12}_{0.12}$	$0.057^{0.020}_{0.020}$	$0.66^{0.11}_{0.11}$	$14.2^{1.3}_{1.3}$	< 0.62
flat+wk-h+SN	(1.00)	$1.10^{0.12}_{0.11}$	$0.025^{0.011}_{0.009}$	$0.11^{0.03}_{0.02}$	$0.70^{0.07}_{0.07}$	$0.30^{0.07}_{0.07}$	$0.052^{0.017}_{0.017}$	$0.70^{0.09}_{0.09}$	$13.8^{1.4}_{1.4}$	< 0.65
flat+wk-h+LSS+SN	(1.00)	$1.08^{0.11}_{0.09}$	$0.025^{0.011}_{0.009}$	$0.11^{0.03}_{0.02}$	$0.71^{0.06}_{0.06}$	$0.29^{0.06}_{0.06}$	$0.053^{0.016}_{0.016}$	$0.69^{0.09}_{0.09}$	$13.9^{1.3}_{1.3}$	< 0.63
flat+HST-h	(1.00)	$1.09^{0.12}_{0.10}$	$0.026^{0.010}_{0.009}$	$0.13^{0.07}_{0.04}$	$0.65^{0.12}_{0.20}$	$0.38^{0.18}_{0.18}$	$0.058^{0.022}_{0.022}$	$0.68^{0.08}_{0.08}$	$13.3^{1.3}_{1.3}$	< 0.65
flat+HST-h+LSS	(1.00)	$1.09^{0.13}_{0.10}$	$0.026^{0.010}_{0.009}$	$0.11^{0.03}_{0.02}$	$0.71^{0.07}_{0.08}$	$0.29^{0.08}_{0.08}$	$0.054^{0.019}_{0.019}$	$0.70^{0.08}_{0.08}$	$13.8^{1.1}_{1.1}$	< 0.64
flat+HST-h+SN	(1.00)	$1.10^{0.12}_{0.11}$	$0.026^{0.010}_{0.009}$	$0.12^{0.03}_{0.03}$	$0.71^{0.06}_{0.06}$	$0.29^{0.07}_{0.07}$	$0.052^{0.017}_{0.017}$	$0.71^{0.07}_{0.07}$	$13.6^{1.2}_{1.2}$	< 0.65
flat+HST-h+LSS+SN	(1.00)	$1.08^{0.11}_{0.09}$	$0.026^{0.010}_{0.009}$	$0.11^{0.02}_{0.02}$	$0.71^{0.06}_{0.05}$	$0.29^{0.06}_{0.06}$	$0.054^{0.017}_{0.017}$	$0.70^{0.07}_{0.07}$	$13.7^{1.1}_{1.1}$	< 0.63

·Estimates of the 6 external cosmological parameters that characterize our fiducial minimal-inflation model set as progressively more restrictive prior probabilities are imposed. (τ_C is put at the end because it is relatively poorly constrained, even with the priors.) Central values and 1σ limits for the 6 parameters are found from the 16%, 50% and 84% integrals of the marginalized likelihood. For the other "derived" parameters listed, the values are means and variances of the variables calculated over the full probability distribution. wk-h requires $0.45 < h < 0.90$, Age > 10 Gyr, and $\Omega_m > 0.1$. The sequence shows what happens when LSS, SN and LSS+SN priors are imposed. While the first four rows allow $\Omega_{\rm tot}$ to be free, the next four have $\Omega_{\rm tot}$ pegged to unity, a number strongly suggested by the CMB data. The final 4 rows show the "strong-h" prior, a Gaussian centered on $h = 0.71$ with dispersion ± 0.076, obtained for the Hubble key project. When the 1σ errors are large it is usual that there is a poor detection, and sometimes there can be multiple peaks in the 1-D projected likelihood.

the CBI+DMR(required anchor in the $\ell = 2 - 40$ range), using a bin size of $\Delta\ell = 140$, are in Table 4.3. We use a finer binning for the cosmology than for plotting spectra to make sure we don't lose any information to overly-large bins. The price is higher correlations, which is correctly treated using the Fisher matrix in the cosmology, but can lead to misleading impressions when the spectrum is looked at visually. The first bin for the $\Delta\ell = 140$ "even" binning has an upper limit of $\ell = 400$, while the first bin for the "odd" binning stops at $\ell = 330$. The CBI is not very sensitive to the spectrum below $\ell \sim 400$, so these cosmological results are basically independent of the first acoustic peak. It is interesting to note that even without the first peak and quite mild restrictions, the CBI measures the universe to be flat to about 10% ($1.00^{+0.11}_{-0.12}$).

The likelihood surface is often more complicated than can be described using simple error bars. Historically, parameters have often had widely separated invervals allowed by the data (such as the Padin et al. (2001a) result that $\Omega_k < 0.4$ or > 0.7), though this is less of a problem now as the data are of higher quality. One dimensional likelihood distributions of cosmological parameters from the CBI spectrum are plotted in Figure 4.17. One might ask how much of the cosmology is prior-driven

rather than CMB-driven. Figure 4.18 shows the parameters from DMR+priors. The parameters in Figure 4.18 are very weakly constrained relative to those with the CBI data in Figure 4.17, which means that the accuracy of the individual parameters is driven by the CBI data and not imposed through the priors used.

Some consistency checks between different binnings of the CBI data, as well as with some other experiments are given in Table 4.4. The CBIo140 and CBIe140 are the "odd" and "even" $\Delta\ell = 140$ CBI binnings. The parameters labelled CBIo140 ($\ell > 610$) are for the CBI "odd" binning, throwing out the spectrum below $\ell = 610$ in order to provide a check on parameters derived from a region of the spectrum with almost no overlap with that of other, lower-ℓ experiments. The CBI $\Delta\ell = 200$ "odd" binning results are under CBIo200, with the deep field results labelled CBIdeep. Finally, some comparisons with the spectra from DASI, BOOMERANG, and All-data (a large collection of experiments that included basically all the CMB results up through summer 2002. Details are given in Sievers et al., 2003). The cosmological parameters maintain a high degree of consistency in all these various checks.

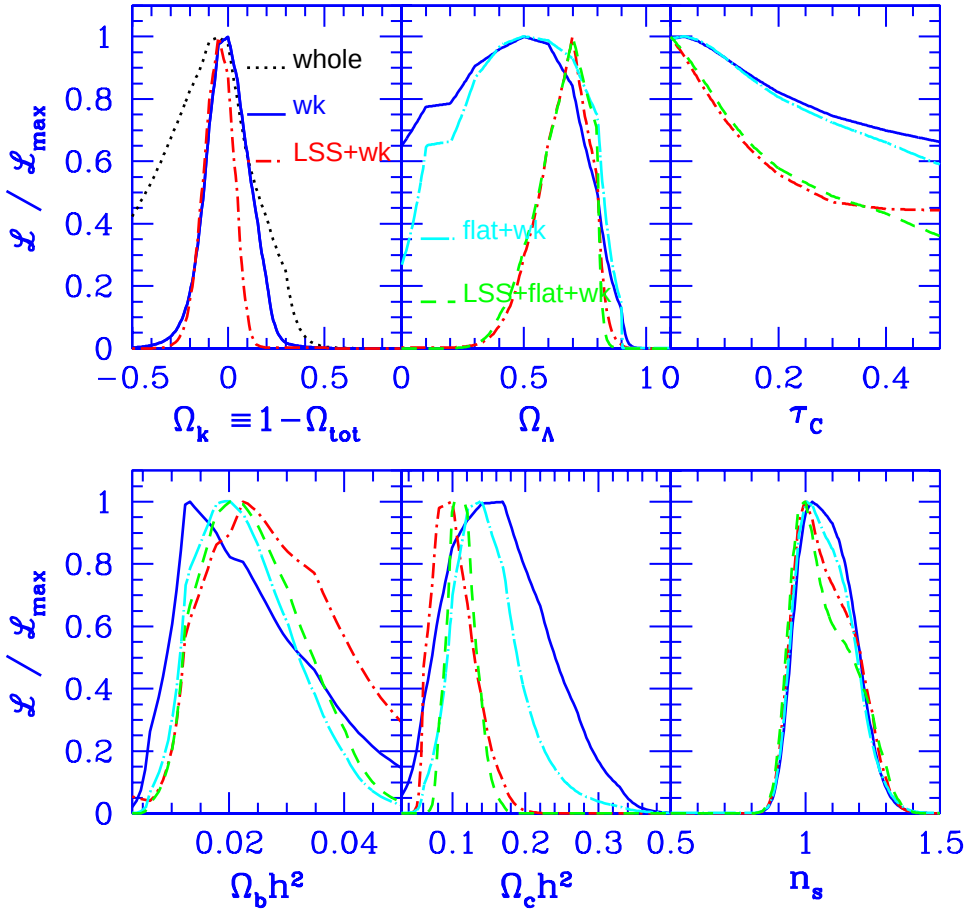


Figure 4.17 1-D projected likelihood functions calculated for the CBIo140+DMR data. All panels include the weak-h (solid dark blue) and LSS+weak-h(short-dash-dotted red) priors. (LSS is the large-scale structure prior.) The Ω_k panel also shows what the whole $\mathcal{C}_\ell$-database gives before the weak-h prior is imposed (black dotted). We note that even in the absence of CMB data there is a bias towards the closed models (Lange et al., 2001). In the other panels, flat+weak-h (long-dashed-dotted light blue) and LSS+flat+weak-h (dashed green) are plotted. Notice how stable the n_s determination is, independent of priors. We see here that, under priors ranging from the weak-h prior to the weak-h+LSS+flat priors, the CBI provides a useful measure of four out of the six fundamental parameters shown. This is independent of the first acoustic peak, where the CBI has low sensitivity, and is also largely independent of the spectrum below $\ell \sim 610$ for all but $\Omega_b h^2$ (see Table 4.4). From Sievers et al. (2003).

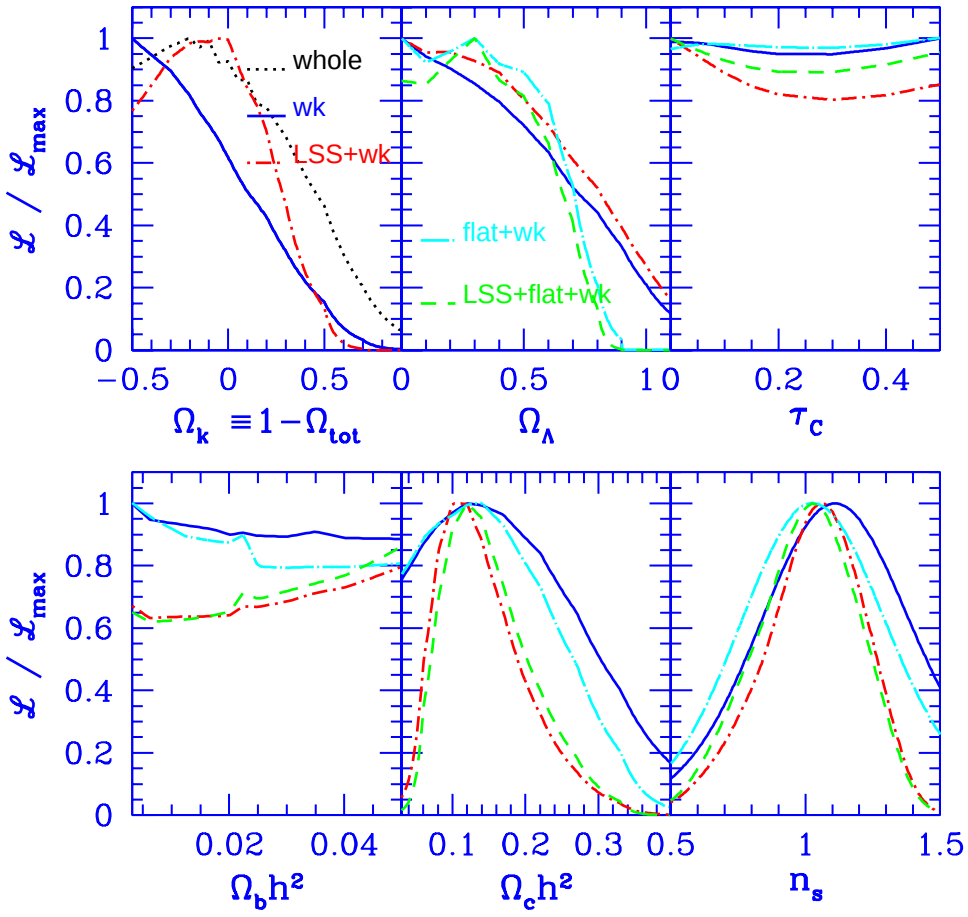


Figure 4.18 Cosmological constraints obtained using DMR alone. This gives an idea of the role of the LSS prior in sharpening up detections for DMR. Note that DMR did reasonably well by itself in first indicating for this class of models that $n_s \sim 1$ (e.g., Bond, 1996). Of course it could not determine ω_b and the structure in Ω_k and Ω_Λ can be traced to $\mathcal{C}_\ell$-database constraints (Lange et al., 2001). Comparison with Fig. 4.17 shows the greatly improved constraints when the CBI data are added. From Sievers et al. (2003).

Table 4.4. CBI Tests and Comparisons. From Sievers et al. (2003)

Priors	$\Omega_{\rm tot}$	n_s	$\Omega_b h^2$	$\Omega_{\rm cdm} h^2$	Ω_Λ	Ω_m	Ω_b	h	Age	τ_c
CBIo140										
wk-h	$1.00^{0.11}_{0.12}$	$1.08^{0.11}_{0.10}$	$0.023^{0.016}_{0.010}$	$0.16^{0.08}_{0.07}$	$0.43^{0.25}_{0.28}$	$0.59^{0.22}_{0.22}$	$0.083^{0.053}_{0.053}$	$0.58^{0.11}_{0.11}$	$13.9^{2.2}_{2.2}$	< 0.66
flat+wk-h	(1.00)	$1.07^{0.11}_{0.10}$	$0.023^{0.010}_{0.008}$	$0.15^{0.06}_{0.04}$	$0.47^{0.25}_{0.27}$	$0.54^{0.24}_{0.24}$	$0.068^{0.028}_{0.028}$	$0.60^{0.12}_{0.12}$	$14.0^{1.4}_{1.4}$	< 0.65
flat+wk-h+LSS	(1.00)	$1.05^{0.15}_{0.09}$	$0.024^{0.011}_{0.009}$	$0.11^{0.02}_{0.02}$	$0.67^{0.10}_{0.13}$	$0.34^{0.12}_{0.12}$	$0.057^{0.020}_{0.020}$	$0.66^{0.11}_{0.11}$	$14.2^{1.3}_{1.3}$	< 0.62
CBIe140										
wk-h	$0.96^{0.14}_{0.13}$	$1.10^{0.11}_{0.11}$	$0.016^{0.013}_{0.007}$	$0.18^{0.08}_{0.07}$	$0.37^{0.28}_{0.26}$	$0.62^{0.23}_{0.23}$	$0.059^{0.044}_{0.044}$	$0.60^{0.11}_{0.11}$	$13.3^{2.1}_{2.1}$	< 0.66
flat+wk-h	(1.00)	$1.08^{0.10}_{0.10}$	$0.018^{0.009}_{0.006}$	$0.15^{0.05}_{0.04}$	$0.44^{0.26}_{0.27}$	$0.56^{0.24}_{0.24}$	$0.059^{0.025}_{0.025}$	$0.58^{0.11}_{0.11}$	$14.1^{1.3}_{1.3}$	< 0.66
flat+wk-h+LSS	(1.00)	$1.06^{0.14}_{0.10}$	$0.020^{0.010}_{0.007}$	$0.11^{0.02}_{0.02}$	$0.68^{0.09}_{0.13}$	$0.32^{0.11}_{0.11}$	$0.049^{0.019}_{0.019}$	$0.67^{0.11}_{0.11}$	$14.3^{1.3}_{1.3}$	< 0.63
CBIo140($\ell > 610$)										
wk-h	$1.06^{0.15}_{0.14}$	$1.14^{0.15}_{0.13}$	$0.068^{0.065}_{0.036}$	$0.15^{0.10}_{0.08}$	$0.44^{0.26}_{0.28}$	$0.71^{0.26}_{0.26}$	$0.264^{0.207}_{0.207}$	$0.59^{0.11}_{0.11}$	$13.1^{1.8}_{1.8}$	< 0.67
flat+wk-h	(1.00)	$1.10^{0.12}_{0.10}$	$0.047^{0.056}_{0.019}$	$0.18^{0.07}_{0.07}$	$0.41^{0.22}_{0.26}$	$0.62^{0.23}_{0.23}$	$0.188^{0.166}_{0.166}$	$0.63^{0.11}_{0.11}$	$12.6^{1.4}_{1.4}$	< 0.66
flat+wk-h+LSS	(1.00)	$1.05^{0.14}_{0.09}$	$0.041^{0.017}_{0.017}$	$0.13^{0.03}_{0.03}$	$0.66^{0.09}_{0.13}$	$0.35^{0.11}_{0.11}$	$0.082^{0.037}_{0.037}$	$0.71^{0.11}_{0.11}$	$13.1^{1.4}_{1.4}$	< 0.62
CBIo200										
wk-h	$1.12^{0.13}_{0.14}$	$1.14^{0.12}_{0.11}$	$0.048^{0.023}_{0.024}$	$0.22^{0.09}_{0.08}$	$0.36^{0.27}_{0.25}$	$0.82^{0.32}_{0.32}$	$0.152^{0.088}_{0.088}$	$0.58^{0.10}_{0.10}$	$12.8^{1.9}_{1.9}$	< 0.67
flat+wk-h	(1.00)	$1.07^{0.11}_{0.09}$	$0.025^{0.015}_{0.010}$	$0.19^{0.09}_{0.07}$	$0.41^{0.25}_{0.26}$	$0.61^{0.24}_{0.24}$	$0.071^{0.033}_{0.033}$	$0.63^{0.11}_{0.11}$	$12.6^{1.7}_{1.7}$	< 0.64
flat+wk-h+LSS	(1.00)	$1.04^{0.14}_{0.08}$	$0.028^{0.014}_{0.011}$	$0.12^{0.03}_{0.02}$	$0.68^{0.09}_{0.13}$	$0.33^{0.11}_{0.11}$	$0.061^{0.025}_{0.025}$	$0.70^{0.11}_{0.11}$	$13.6^{1.4}_{1.4}$	< 0.59
CBIdeep										
wk-h	$1.09^{0.11}_{0.24}$	$1.16^{0.15}_{0.14}$	$0.078^{0.070}_{0.049}$	$0.21^{0.11}_{0.12}$	$0.42^{0.31}_{0.29}$	$0.85^{0.32}_{0.32}$	$0.261^{0.190}_{0.190}$	$0.61^{0.10}_{0.10}$	$12.2^{1.8}_{1.8}$	< 0.67
flat+wk-h	(1.00)	$1.05^{0.12}_{0.11}$	$0.050^{0.072}_{0.034}$	$0.20^{0.09}_{0.14}$	$0.37^{0.26}_{0.25}$	$0.65^{0.24}_{0.24}$	$0.189^{0.188}_{0.188}$	$0.64^{0.11}_{0.11}$	$12.2^{1.9}_{1.9}$	< 0.66
flat+wk-h+LSS	(1.00)	$1.03^{0.13}_{0.11}$	$0.055^{0.064}_{0.035}$	$0.13^{0.04}_{0.04}$	$0.55^{0.16}_{0.28}$	$0.50^{0.23}_{0.23}$	$0.187^{0.179}_{0.179}$	$0.66^{0.13}_{0.13}$	$12.9^{2.0}_{2.0}$	< 0.65
DASI+CBIo140										
wk-h	$1.05^{0.05}_{0.06}$	$1.01^{0.11}_{0.07}$	$0.023^{0.004}_{0.004}$	$0.12^{0.04}_{0.04}$	$0.55^{0.17}_{0.22}$	$0.51^{0.19}_{0.19}$	$0.077^{0.024}_{0.024}$	$0.56^{0.10}_{0.10}$	$15.2^{1.5}_{1.5}$	< 0.63
flat+wk-h	(1.00)	$0.99^{0.08}_{0.05}$	$0.021^{0.004}_{0.003}$	$0.14^{0.03}_{0.03}$	$0.56^{0.18}_{0.26}$	$0.46^{0.21}_{0.21}$	$0.057^{0.017}_{0.017}$	$0.62^{0.11}_{0.11}$	$13.9^{0.8}_{0.8}$	< 0.39

Table 4.4—Continued

Priors	$\Omega_{\rm tot}$	n_s	$\Omega_b h^2$	$\Omega_{\rm cdm} h^2$	Ω_Λ	Ω_m	Ω_b	h	Age	τ_c
flat+wk-h+LSS	(1.00)	$1.00^{0.10}_{0.06}$	$0.022^{0.004}_{0.004}$	$0.12^{0.02}_{0.02}$	$0.66^{0.09}_{0.10}$	$0.33^{0.10}_{0.10}$	$0.048^{0.009}_{0.009}$	$0.68^{0.09}_{0.09}$	$13.8^{0.8}_{0.8}$	< 0.42
DASI+Boom+CBIo140										
wk-h	$1.03^{0.05}_{0.05}$	$0.95^{0.09}_{0.05}$	$0.022^{0.003}_{0.003}$	$0.13^{0.03}_{0.03}$	$0.52^{0.18}_{0.20}$	$0.52^{0.19}_{0.19}$	$0.074^{0.022}_{0.022}$	$0.56^{0.10}_{0.10}$	$15.0^{1.4}_{1.4}$	< 0.52
flat+wk-h	(1.00)	$0.94^{0.06}_{0.04}$	$0.021^{0.002}_{0.002}$	$0.14^{0.03}_{0.03}$	$0.55^{0.18}_{0.28}$	$0.48^{0.21}_{0.21}$	$0.058^{0.016}_{0.016}$	$0.61^{0.11}_{0.11}$	$14.0^{0.5}_{0.5}$	< 0.31
flat+wk-h+LSS	(1.00)	$0.94^{0.08}_{0.04}$	$0.022^{0.002}_{0.002}$	$0.13^{0.02}_{0.02}$	$0.63^{0.09}_{0.11}$	$0.37^{0.10}_{0.10}$	$0.051^{0.008}_{0.008}$	$0.65^{0.07}_{0.07}$	$13.9^{0.5}_{0.5}$	< 0.36
all-data										
wk-h	$1.04^{0.05}_{0.05}$	$0.98^{0.10}_{0.06}$	$0.023^{0.003}_{0.003}$	$0.12^{0.03}_{0.03}$	$0.55^{0.17}_{0.20}$	$0.49^{0.18}_{0.18}$	$0.075^{0.023}_{0.023}$	$0.56^{0.11}_{0.11}$	$15.1^{1.4}_{1.4}$	< 0.57
flat+wk-h	(1.00)	$0.96^{0.09}_{0.05}$	$0.022^{0.003}_{0.002}$	$0.13^{0.03}_{0.03}$	$0.60^{0.15}_{0.26}$	$0.42^{0.20}_{0.20}$	$0.055^{0.015}_{0.015}$	$0.64^{0.11}_{0.11}$	$13.9^{0.5}_{0.5}$	< 0.35
flat+wk-h+LSS	(1.00)	$0.97^{0.09}_{0.05}$	$0.022^{0.003}_{0.002}$	$0.12^{0.02}_{0.02}$	$0.66^{0.09}_{0.12}$	$0.35^{0.10}_{0.10}$	$0.051^{0.008}_{0.008}$	$0.66^{0.08}_{0.08}$	$13.8^{0.5}_{0.5}$	< 0.39

[·] Cosmological parameter estimates as in Table 4.3, except for a variety of data combinations which test and compare results. Only the wk-h, flat+wk-h and flat+wk-h+LSS priors are shown.

One of the most intriguing results is that using just the CBI spectrum at $\ell > 610$ gives parameters consistent from those derived from the spectrum around the first and second peaks from other experiments. It is indeed an impressive consistency check that non-overlapping spectra from different experiments give the same overall properties of the universe! This results gives further confidence that we are indeed seeing a coherent picture of the universe using many different lines of evidence. A final display of the consistency between various experiments can be seen in Figure 4.19. This figure shows the two dimensional 2σ likelihood contours for various parameters with the dark-matter density, $\omega_{\rm cdm}$ for a set of experiments. The fact that all the contours circle the same region in parameter space means that the individual experiments favor similar regions, which is what one hopes for and expects. Again, the degree of consistency among heterogeneous CMB experiments is remarkable.

The final cosmological results using CBI and all data available as of the summer of 2002, including BOOMERANG, DASI, MAXIMA, and VSA, along with a variety of priors, is contained in Table 4.5. This was the most up-to-date parameter set possible at the time. Some of the most interesting results are that the CMB, including the flat, wk-h, LSS and SN priors, but *not* the HST key project h value, gives a Hubble constant of $h = 0.69 \pm 0.05$. The agreement with the HST key project's value of $h = 0.72 \pm 0.08$ (Freedman et al., 2001) is very good, enough so that this author is convinced that we finally indeed know the Hubble constant to better than 10%. The presence of dark energy

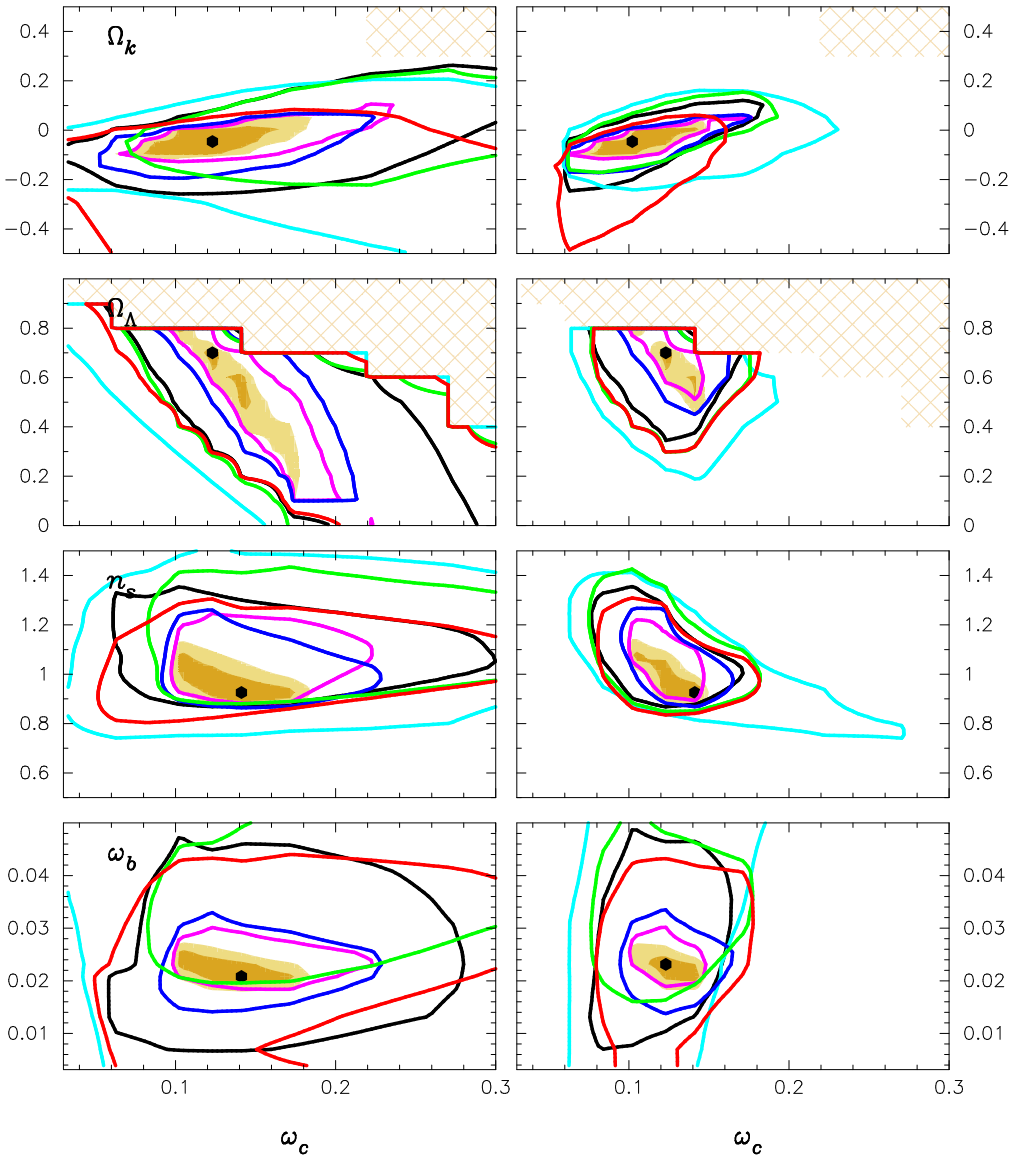


Figure 4.19 Comparison of different experiments. 2-σ likelihood contours for the weak-h prior ($\omega_{\rm cdm}$–Ω_k panel) and flat+weak-h prior for the rest, for the following CMB experiments in combination with DMR: CBIe140 (black), BOOMERANG (magenta), DASI (dark blue), Maxima (green), VSA (red) and "prior-CMB" = BOOMERANG-NA+TOCO+Apr99 data (light blue). Light brown region shows the 2-σ contour when all of the data are taken together, dark brown shows the 1-σ contour. The LSS prior has not been used in deriving the plots on the left, but it has for those on the right. The hatched regions indicate portions excluded by the range of parameters considered (see Table 4.2). This figure shows great consistency as well as providing a current snapshot of the collective CMB data results. Even without the LSS prior (or the HST-h or SN1a priors), localization of the dark matter density is already occurring, but Ω_Λ still has multiple solutions. The inclusion of the SN1a and/or the HST-h priors does not concentrate the bulls-eye determinations much more for the all-data shaded case. Note that the expectation of minimal inflation models is that $\Omega_k \approx 0$, $n_s \approx 1$ (usually a little less). The Big Bang Nucleosynthesis result, $\omega_b = 0.019 \pm 0.002$ also rests comfortably within the bulls-eye. From Sievers et al. (2003).

Table 4.5. Cosmological Parameters from All-Data

Priors	$\Omega_{\rm tot}$	n_s	$\Omega_b h^2$	$\Omega_{\rm cdm} h^2$	Ω_Λ	Ω_m	Ω_b	h	Age	τ_c
wk-h	$1.04^{0.05}_{0.05}$	$0.98^{0.10}_{0.06}$	$0.023^{0.003}_{0.003}$	$0.12^{0.03}_{0.03}$	$0.55^{0.17}_{0.20}$	$0.49^{0.18}_{0.18}$	$0.075^{0.023}_{0.023}$	$0.56^{0.11}_{0.11}$	$15.1^{1.4}_{1.4}$	< 0.57
wk-h+LSS	$1.04^{0.05}_{0.04}$	$1.01^{0.09}_{0.09}$	$0.023^{0.004}_{0.003}$	$0.11^{0.03}_{0.03}$	$0.66^{0.09}_{0.13}$	$0.39^{0.11}_{0.11}$	$0.069^{0.020}_{0.020}$	$0.60^{0.09}_{0.09}$	$15.2^{1.5}_{1.5}$	< 0.60
wk-h+SN	$1.03^{0.05}_{0.04}$	$1.03^{0.08}_{0.08}$	$0.024^{0.003}_{0.003}$	$0.11^{0.02}_{0.02}$	$0.71^{0.06}_{0.07}$	$0.32^{0.08}_{0.08}$	$0.061^{0.020}_{0.020}$	$0.64^{0.09}_{0.09}$	$14.8^{1.6}_{1.6}$	< 0.63
wk-h+LSS+SN	$1.03^{0.05}_{0.04}$	$1.04^{0.08}_{0.08}$	$0.024^{0.004}_{0.003}$	$0.10^{0.02}_{0.03}$	$0.71^{0.06}_{0.06}$	$0.33^{0.06}_{0.06}$	$0.064^{0.020}_{0.020}$	$0.63^{0.09}_{0.09}$	$15.0^{1.6}_{1.6}$	< 0.63
flat+wk-h	(1.00)	$0.96^{0.09}_{0.05}$	$0.022^{0.003}_{0.002}$	$0.13^{0.03}_{0.03}$	$0.60^{0.15}_{0.26}$	$0.42^{0.20}_{0.20}$	$0.055^{0.015}_{0.015}$	$0.64^{0.11}_{0.11}$	$13.9^{0.5}_{0.5}$	< 0.35
flat+wk-h+LSS	(1.00)	$0.97^{0.09}_{0.05}$	$0.022^{0.003}_{0.002}$	$0.12^{0.02}_{0.02}$	$0.66^{0.09}_{0.12}$	$0.35^{0.10}_{0.10}$	$0.051^{0.008}_{0.008}$	$0.66^{0.08}_{0.08}$	$13.8^{0.5}_{0.5}$	< 0.39
flat+wk-h+SN	(1.00)	$0.99^{0.07}_{0.06}$	$0.023^{0.002}_{0.002}$	$0.12^{0.02}_{0.02}$	$0.71^{0.06}_{0.07}$	$0.29^{0.07}_{0.07}$	$0.045^{0.006}_{0.006}$	$0.71^{0.06}_{0.06}$	$13.6^{0.3}_{0.3}$	< 0.37
flat+wk-h+LSS+SN	(1.00)	$0.99^{0.07}_{0.06}$	$0.023^{0.002}_{0.002}$	$0.12^{0.01}_{0.01}$	$0.70^{0.05}_{0.06}$	$0.30^{0.06}_{0.06}$	$0.047^{0.005}_{0.005}$	$0.69^{0.05}_{0.05}$	$13.7^{0.3}_{0.3}$	< 0.39
flat+HST-h	(1.00)	$0.99^{0.07}_{0.06}$	$0.023^{0.002}_{0.002}$	$0.12^{0.02}_{0.02}$	$0.70^{0.08}_{0.11}$	$0.30^{0.10}_{0.10}$	$0.047^{0.009}_{0.009}$	$0.70^{0.08}_{0.08}$	$13.6^{0.4}_{0.4}$	< 0.37
flat+HST-h+LSS	(1.00)	$0.99^{0.08}_{0.06}$	$0.023^{0.002}_{0.002}$	$0.12^{0.02}_{0.02}$	$0.69^{0.06}_{0.08}$	$0.31^{0.08}_{0.08}$	$0.048^{0.006}_{0.006}$	$0.69^{0.06}_{0.06}$	$13.7^{0.3}_{0.3}$	< 0.39
flat+HST-h+SN	(1.00)	$0.99^{0.07}_{0.05}$	$0.023^{0.002}_{0.002}$	$0.12^{0.01}_{0.02}$	$0.71^{0.06}_{0.05}$	$0.28^{0.06}_{0.06}$	$0.045^{0.006}_{0.006}$	$0.71^{0.05}_{0.05}$	$13.6^{0.2}_{0.2}$	< 0.37
flat+HST-h+LSS+SN	(1.00)	$1.00^{0.06}_{0.05}$	$0.023^{0.002}_{0.002}$	$0.12^{0.01}_{0.01}$	$0.70^{0.05}_{0.05}$	$0.30^{0.05}_{0.05}$	$0.047^{0.004}_{0.004}$	$0.69^{0.04}_{0.04}$	$13.7^{0.2}_{0.2}$	< 0.38

· Cosmological parameter estimates as in Table 4.3, but now for all-data. From Sievers et al. (2003)

is also convincingly detected, with a limit on Ω_m using wk-h +LSS+SN (but not flat) of 0.33 ± 0.06, with the limit on $\Omega_{\rm tot}$ of $1.03^{+0.05}_{-0.04}$. There is not really enough information to discriminate between Λ and more generalized forms of non-collapsing energy density such as quintessence (Bond et al., 2002a), but an exotic form of energy is definitely required. The age of the universe is also very well determined, with CMB+flat+HST-h+LSS+SN giving $T_0 = 13.7 \pm 0.2$ Gyr, which is virtually identical to the much-celebrated WMAP result of 13.7 ± 0.2 Gyr (presumably the values and errors would differ given more (in)significant digits).

Chapter 5

A Fast, General Maximum Likelihood Program

In this chapter, I describe a program called CBISPEC that efficiently calculates window matrices and then compresses them. In Section 5.1 I describe how the compression is carried out and some of its advantages. In Section 5.2 I explain how to generalize the techniques of Section 3.3 to mosaicked observations and present a fast algorithm for calculating the window functions for Gaussian primary beams. In Section 5.3 I compare CBISPEC results with other maximum likelihood techniques. Finally, in Section 5.4 I use CBISPEC to constrain potential foreground contributions to the CBI's power spectrum. This is very difficult to do with most traditional maximum likelihood methods because they usually destroy the frequency information necessary for measuring foregrounds.

5.1 Compression

Modern experiments can easily have huge numbers of data points making them computationally intractable if treated naively. As mentioned in Section 2.2, the CBI extended mosaics have approximately 800,000 data points in each. Finding the maximum likelihood spectrum from such a problem would take literally years on a supercomputer. In addition, the memory requirements are enormous - with 20 bins stored as doubles, even if we only keep half of each window matrix (since it is symmetric), we would require over 40 terabytes of memory! The actual independent information contained in the data set is very much smaller. For interferometers, it is on the order of the number

of synthesized beams in the map. The CBI has a $3'$ beam and the extended mosaics cover $\sim 2^\circ \times 4^\circ$, for a total number of beams $\sim 4,000$. While still a large number, this is easily handled using the fast maximization techniques of Chapter 2 even on a desktop machine. It takes a P4 1.4 GHz machine about 2 minutes to invert a 4000 by 4000 symmetric matrix. If we need 10 iterations to converge using three extended mosaics, we could in principle measure the CBI power spectrum from 3 mosaics in $\sim 10 \times 3 \times 2$ minutes, about an hour. Clearly, it is of critical importance to get as close as possible to the theoretical minimum number of estimators that contain all the information in the experiment. Even Nyquist sampling is costly - a factor of 2 in each direction means a factor of 4 in data size and a factor of 64 in execution time! One way of compressing data is optimal sub-space filtering, also called a Karhunen-Loeve transform (see, *e.g.*, White et al., 1999; Tegmark et al., 1998, and references therein). It is conceptually straightforward to carry out a Karhunen-Loeve transform. If necessary, one first rotates into a space in which the noise is identical and uncorrelated for all data points (a so-called "whitening transform"). For the general case of a correlated noise matrix, this requires the Cholesky decomposition of the noise matrix $\mathrm{N} = \mathrm{L}\mathrm{L}^T$ with L a lower triangular matrix. (Of course, there is nothing special about using a lower-triangular factorization: one can just as easily use an upper-triangular matrix.) Once we have L, we use it to rotate the noise matrix, the window matrices, the data vector, and any source matrices. The rotation of a matrix A is:

$$\mathrm{A} \rightarrow \mathrm{L}^{-1}\mathrm{A}\mathrm{L}^{-1\ T}$$

The rotation for a vector (usually the data vector) is

$$\boldsymbol{\Delta} \rightarrow \mathrm{L}^{-1}\boldsymbol{\Delta}$$

This does not leave the likelihood unchanged, but rather shifts the log determinant term by the constant factor $\log|\mathrm{L}|$. It does, however, leave the shape of the likelihood unchanged. Since all we care about is the shape of the likelihood surface, all quantities of interest will remain unchanged, provided we never compare whitened and unwhitened likelihoods.

Once we have whitened the noise, we calculate the signal covariance, and rotate into the space in which the signal covariance is diagonal. Since any rotation of the identity matrix leaves it unchanged, the data in the new basis still have identical, independent noises. Since the modes all have the same noise, and their expected variance is the corresponding eigenvalue, the eigenvalue is then the expected signal to noise ratio of that mode. Furthermore, because the signal part of the covariance is also diagonal, the different modes are truly independent - we have reverted back to the case of uncorrelated data in Section 2.1. If the data are oversampled, as is usually the case, the transformed data set will have a few modes with large SNR, and many with SNR close to zero. Those modes with very small SNR contain information about the noise, but essentially no information about the signal, and the shape of the likelihood surface will be highly insensitive to them. In that case, we might as well throw them away and only use the high signal modes to calculate the power spectrum. So, we can *compress* the data set by cutting the low signal modes. We can do the cutting more efficiently by not using the full eigenvector matrix V and instead using only those eigenvectors corresponding to the eigenvalues we wish to keep. If we denote the $m \times n$ matrix containing the first m eigenvectors by $\mathrm{V}^\dagger$, then the modified rotation is

$$\mathrm{A}^\dagger \longrightarrow \mathrm{V}^\dagger \mathrm{A} \mathrm{V}^{\dagger\ T}$$

The reason this compression method is called *optimal sub-space filtering* is because, for a fixed number of modes m, we have transformed into the $m \times m$ subspace of the original $n \times n$ space that has the highest signal to noise ratio possible.

While this seems at first an attractive solution to the problem of how one compresses the data, in general for CMB data, it is *not good*. There are several major problems. First, it can be prohibitively expensive computationally. We need to do the whitening transform, which for correlated noise requires expensive $\mathcal{O}\left(n^3\right)$ operations, both to calculate L^{-1} and to do the rotation. For interferometery, this is happily not relevant because for a well-functioning system, the receiver noises are uncorrelated. So, rather than having to factor a matrix, we can merely scale each data point and matrix element by the associated visibility noises, so no $\mathcal{O}\left(n^3\right)$ operations are required. More problematic is calculating $\mathrm{V}^\dagger$ in the whitened space. If we expect the compression factor to be large

(m very much smaller than n), then we can calculate only the m eigenvectors with the largest eigenvalues. This is then an $\mathcal{O}\left(mn^2\right)$ operation. If m is only a few times smaller than n, then this step can take as much time as it would have taken to extract the spectrum from the uncompressed data set using the fast technique of Chapter 2! Clearly a faster way of compressing is highly desirable.

Even if the Karhunen-Loeve compression were computationally feasible, it suffers from another drawback. Namely, though it is optimal at maximizing the expected SNR, for typical CMB behaviors, it is bad at retaining the information we want to preserve in the compression. What we desire in a compression method is to do the best job of reproducing the uncompressed spectrum with as few estimators as possible, which in practice is very different from maximizing the SNR. The problem is essentially one of dynamic range. For an interferometer, the response of a baseline to the CMB falls like one over the baseline length squared because long baselines have more fringes, so a long baseline averages over many more independent patches on the sky than a short baseline. On top of this, $\mathcal{C}_\ell$ generally falls rather quickly with increasing ℓ, so the intrinsic signal for a long baseline is much weaker than that for a short baseline. These two factors combined can easily lead to factors of several hundred between the expected variance on a short baseline and the expected variance on a long baseline. To see why this causes optimal subspace filtering to perform very poorly, picture the simple case of an experiment consisting of two pairs of visibilities, one pair at low ℓ, and one pair at high ℓ. The visibilities within the pairs sample almost the same CMB. Clearly, we would like our compression to keep one number for each pair, roughly corresponding to the average value in that pair. If the measurements within a pair are sufficiently similar, then there is essentially no other information contained in them. If they are slightly different, though, the K-L transform will think there is some power in the modes corresponding to the differences. So here is the problem: if the expected power in the difference of the low-ℓ mode is *larger* than the expected power in the average of the high ℓ mode, then the low-ℓ difference will be preferentially kept over the high-ℓ mode. This is problematic for three reasons. First, the K-L transform throws out desirable high-ℓ modes. Second, it keeps undesirable low-ℓ modes that can be problematic in the limit of high SNR, which is frequently the case for CBI low-ℓ data. As the noise drops, the ML tries to push further and further

into the primary beam looking for weaker and weaker signals, so these modes are far more sensitive to errors in the primary beam and, because the expected signal is lower, are more easily corrupted by a fixed error, say due to a point source. The third problem is that because we are keeping too many unwanted modes, the compression is not as efficient as it should be.

The ideal compression algorithm is both fast to run and efficient at keeping only useful data. For the case of interferometers, a modified K-L transform achieves both these goals. One nice feature of interferometers is that closely spaced visibilities in the UV plane are highly correlated (where closely spaced is defined relative to the size of the primary beam FT), while widely spaced visibilities are only weakly correlated, if at all. This is a very general property of interferometric observations of the CMB because interferometers directly sample the FT of the sky, which is just the space in which the CMB is expected to be independent. This does *not* apply to map-making experiments, where pixels widely spaced on the sky are still correlated through long wavelength modes. So, we would expect to be able to break up the entire UV plane into chunks on scales of the primary beam, compress those, and get most of the reduction in size that we expect from a global K-L compression. To fix the optimal filter problem of making poor choices about selecting modes to keep, instead of calculating the compression on the basis of the best-fit spectrum, use a model spectrum during compression that is something like a white-noise spectrum (C_ℓ flat, or $\mathcal{C}_\ell$ rising as ℓ^2). For a white-noise spectrum, the visibilities on long baselines are expected to have the same variance as the short baselines, and so using a white-noise spectrum as the model when forming the covariance matrix used in compression will preserve the desired information while efficiently excising the redundant modes. Strictly speaking, the data combinations kept by this algorithm are no longer normal modes of the covariance matrix, and furthermore, the normal modes change as the power spectrum changes. In practice, we have found that the eigenvectors are highly insensitive to the details of the assumed power spectrum.

I ran a variety of tests on sets of simulated data to examine the sensitivity of the output spectra to the model spectrum used in compression. The simulations were of a typical ΛCDM cosmology, with data from a single deep field. I examined four spectral models for compression, a CMB-like

Table 5.1. Model Spectra Used in Compression Tests

Bin ℓ range	CMB	Flat	Slow-Rise	Rising
$\ell < 900$	3.0×10^{-10}	3.0×10^{-10}	3.0×10^{-10}	1.0×10^{-10}
$900< \ell <1500$	1.0×10^{-10}	3.0×10^{-10}	4.0×10^{-10}	4.0×10^{-10}
$1500< \ell <2100$	3.0×10^{-11}	3.0×10^{-10}	8.0×10^{-10}	9.0×10^{-10}
$2100< \ell <2700$	1.0×10^{-11}	3.0×10^{-10}	1.3×10^{-9}	1.6×10^{-9}
$\ell > 2700$	1.0×10^{-11}	3.0×10^{-10}	2.0×10^{-9}	2.5×10^{-9}

spectrum that falls quickly in ℓ, a flat spectrum with equal power in all bins, a rising spectrum roughly proportional to ℓ^2, and a slowly rising spectrum less steeply increasing than the rising spectrum. The ℓ values for the bins and the corresponding spectral models used during compression are summarized in Table 5.1 The effects of the different compression spectra are shown in Figures 5.1 and 5.2 for the highest- and lowest-ℓ bins, respectively. If one uses the CMB spectrum during compression, one needs more estimators (about 500) to capture all the information in the highest-ℓ bin than either the rising or slow-rise spectra (about 200), with the flat spectrum intermediate (about 300). Conversely, for the first bin, the CMB spectrum performed the best, since its estimators were predominantly sensitive to the the first bin, with good performance down to about 200 estimators. The CMB model may have performed well with even fewer estimators in the first bin, but at that severe a compression level, the high-ℓ bins were so unconstrained that the fits were unable to converge. The rising spectrum began to degrade at about 400 estimators, the slow-rise at around 300, and the flat at 200. Except for a spike between 200 and 300 estimators (presumably due to shot noise in which estimators were kept), the slow-rise spectrum closely matched the flat spectrum in performance in the first bin.

Another way of visualizing the results is to plot, for various compression levels, the scatter in each bin, and connecting bins from the same compression level. The ideal model spectrum used in compression would lead to the scatter in the bins increasing at about the same rate, which would keep the lines horizontal. A model spectrum that too heavily emphasizes one region of ℓ space would lead to a tilt as that region keeps a low scatter while other regions of the spectrum become noisier.

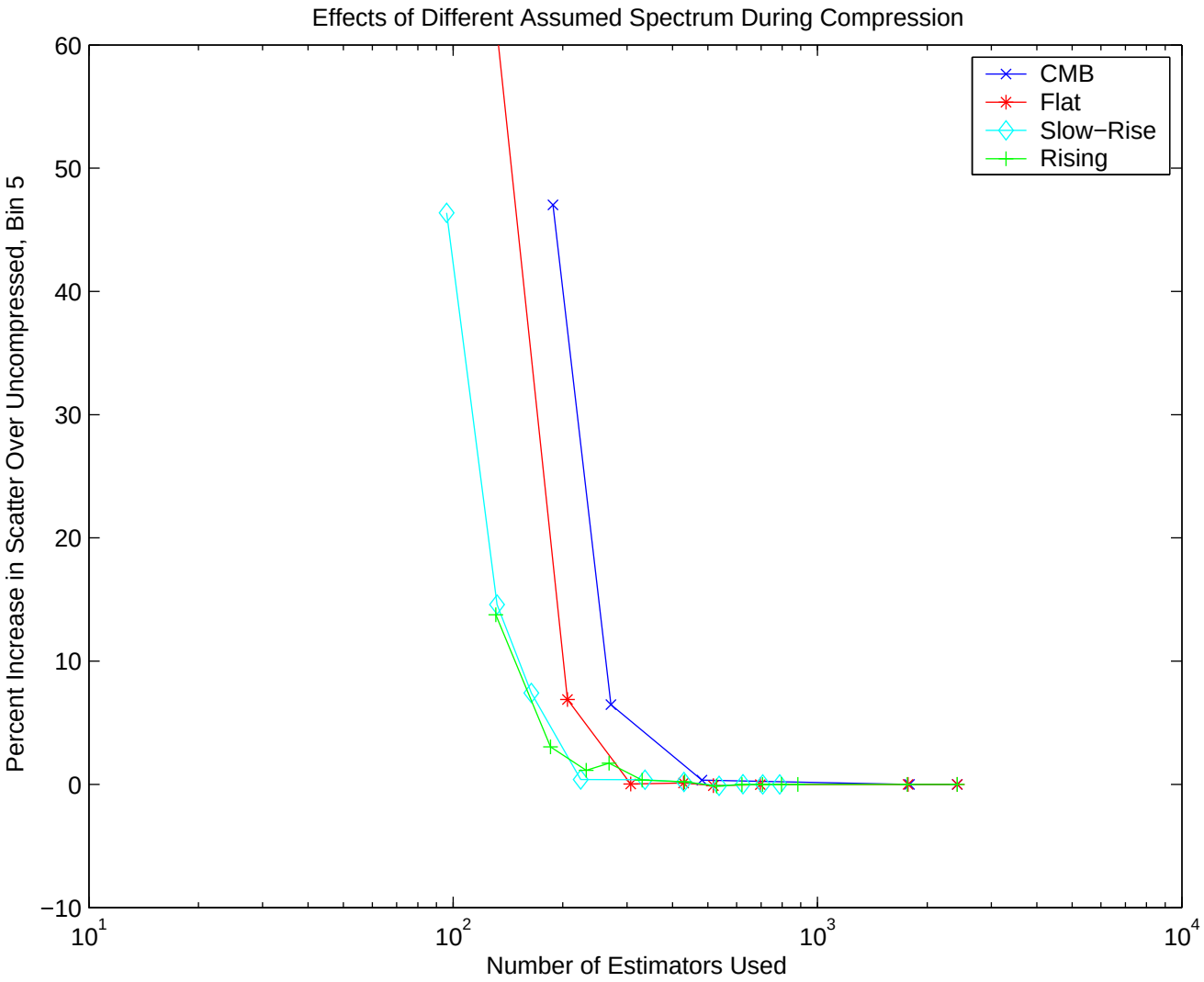


Figure 5.1 Figure showing the effects of different model spectra used during compression. Plotted is the increased scatter from the compression against the number of estimators used (not the compression level!), for bin 5. For a high-ℓ bin, rising spectra should perform best, since they preferentially keep high-ℓ information. This is clearly seen in the plot, as for a fixed number of estimators, the falling CMB spectrum compression performs the worst, followed by the flat spectrum. The rising, and slow-rise spectra both perform well, taking only 200 estimators to have minimally increased scatter, as opposed to 500 for the CMB spectrum.

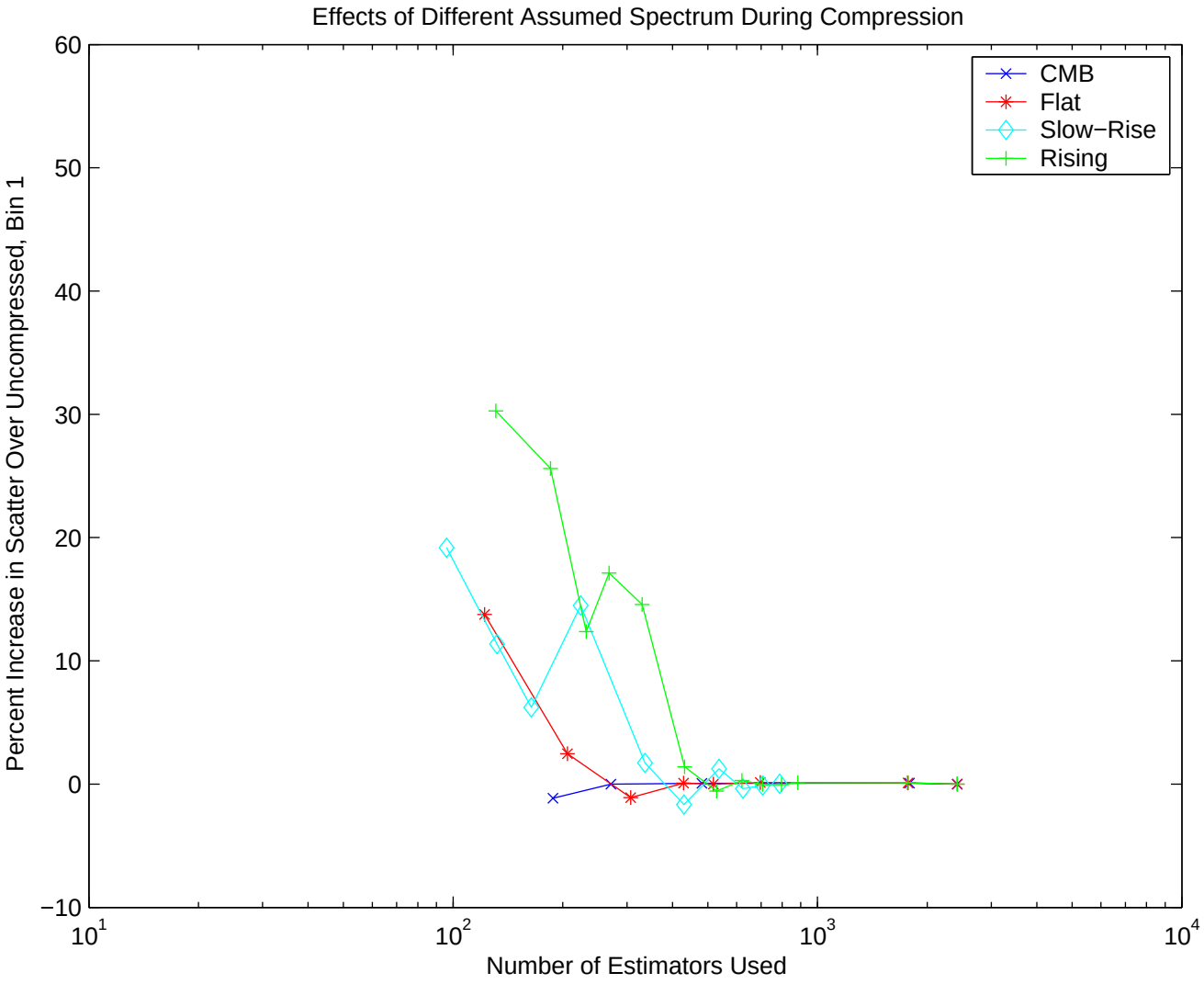


Figure 5.2 Same as Figure 5.1, showing the lowest-ℓ bin. In this case, we expect the falling CMB spectrum to perform better with a fixed number of estimators, since it will preferentially concentrate them at low-ℓ. This is indeed the case, though the penalty associated with using a rising spectrum at low-ℓ isn't as large as that associated with using a falling spectrum at high-ℓ. The price at ~ 300 estimators is 1.7% for the slow-rise spectrum for bin 1, but it is $\sim 5\%$ for the CMB spectrum in bin 5.

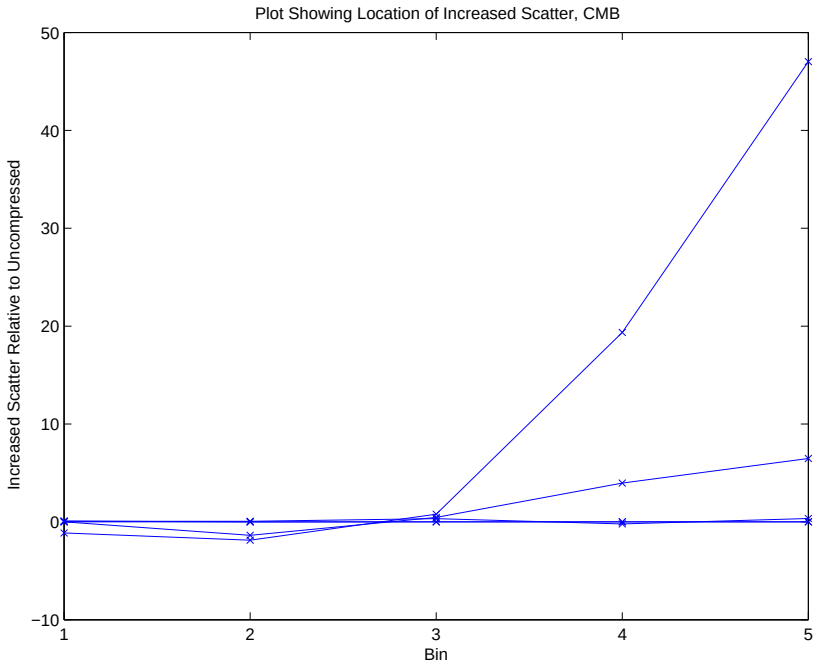


Figure 5.3 Plot showing increase in bin scatters for various compression levels using a CMB spectrum as the model for compression. Each horizontal line connects average bin scatters for a fixed compression level. Clearly, the CMB spectrum underemphasizes the high-ℓ spectrum, since those bins degrade tremendously before the first bins have been affected by the compression. This is the hallmark of a poor choice for the model spectrum.

These plots are shown in Figures 5.3 through 5.6. Using the CMB model spectrum in compression clearly leads to a excessive rise in scatter in the high-ℓ bins which means that too few estimators have been devoted to high-ℓ, with the same true to a lesser extent for the flat spectrum model. The rising model has the opposite problem, with the low-ℓ rising before the high-ℓ. The slow-rise spectrum in Figure 5.5 shows how the scatters should increase with increasing compression, as the bins degrade about equally as fewer and fewer estimators are used.

In practice, the quality of the compression is not terribly sensitive to the parameters, with flat or slowly rising model spectra in the compression performing well at a level of about a few times 10^{-3}. At this level, one keeps about 300 estimators if analyzing a single field, at a cost of $< 1\%$ in increased variance. If we keep 150 complex estimators (for a total of 300) over the UV half-plane out to 560λ, the end of the CBI coverage, then there is a total area of $3280\lambda^2$ per complex estimator, which is a circle in the UV plane of diameter 65λ. This is rather remarkable, since the FWHM of the

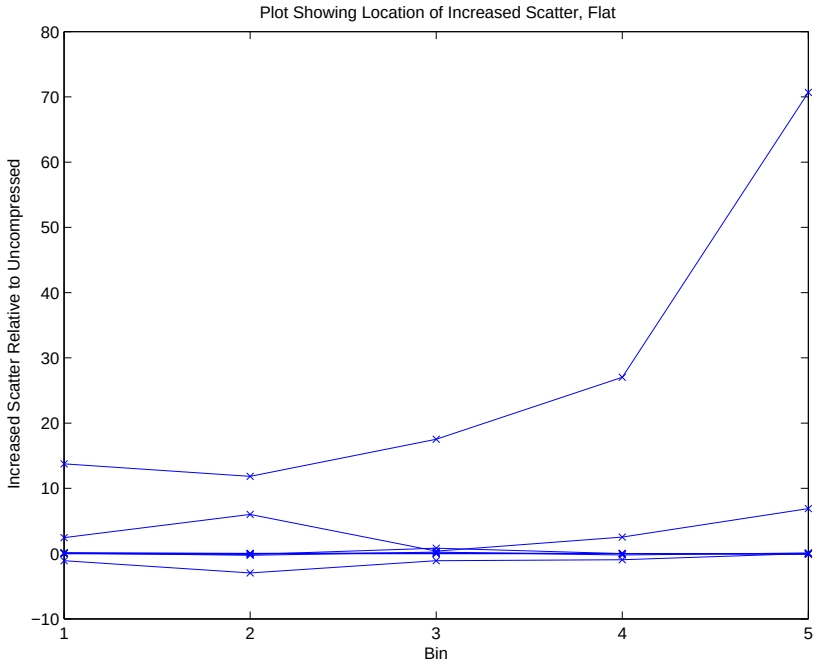


Figure 5.4 Same as 5.3 for a flat spectrum. Better than the CMB model spectrum, the high-ℓ bins are nevertheless overly noisy relative to the low-ℓ bins.

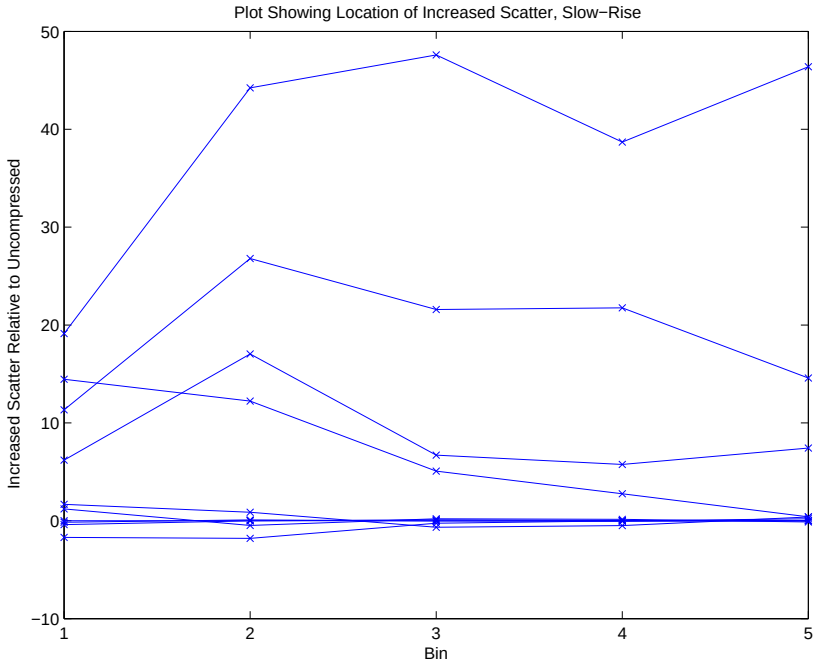


Figure 5.5 Same as 5.3 for a slowly rising spectrum. This is the best overall compression model, with no one region of the spectrum clearly better or worse than the others.

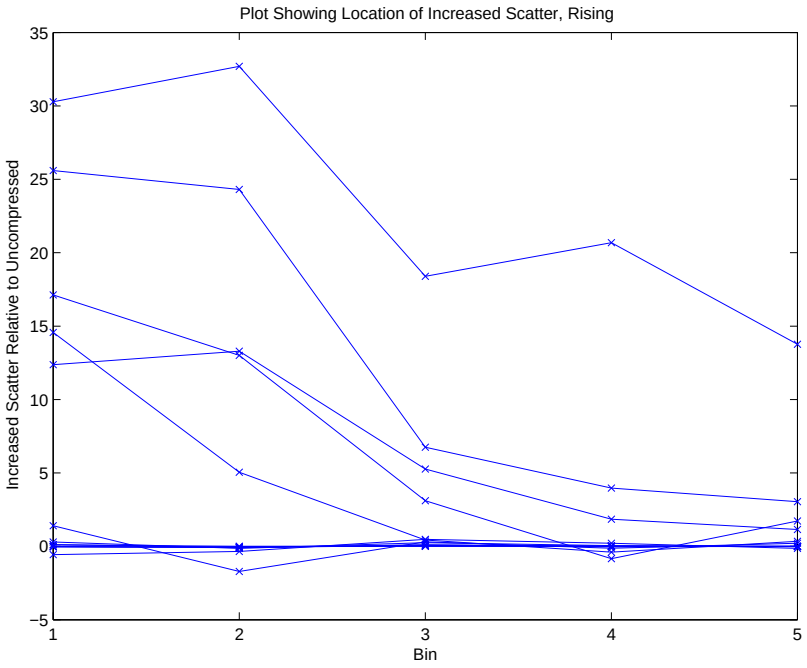


Figure 5.6 Same as 5.3 for a model spectrum rising as ℓ^2. In this case, the low-ℓ region is underemphasized relative to the high-ℓ region of the spectrum. It's better to use a flatter spectrum than ℓ^2 to maintain sensitivity at low-ℓ.

primary beam FT is 67λ, which means that there is almost exactly one estimator per independent patch in the UV plane. We have reached the absolute theoretical minimum amount of information that can do a good job of characterizing the CMB. CBIGRIDR uses 1860 estimators to cover the same region, which corresponds to an estimator footprint diameter of 25λ. This is not bad as it about Nyquist samples the primary beam FT. The n^3 part of the spectrum fitting will go much faster with the highly compressed CBISPEC dataset though, with an expected CBISPEC execution time about $(300/1860)^3$ that of CBIGRIDR, about 0.5%.

This is the basic outline of the compression scheme used by CBISPEC. It is both fast and efficient. To estimate the operation count, we will assume we have an $n \times n$ covariance matrix split into n_{block} blocks with roughly equal number of visibilities, and that we will compress by a factor of f_{comp}, typically about 0.1 for the CBI. To calculate the compression matrix, we need to diagonalize the blocks along the diagonal of the covariance matrix. Each block has roughly n/n_{block} visibilities, so the work required to diagonalize a given block is $\mathcal{O}\left(n/n_{block}\right)^3$. Since we have n_{block} diagonalizations,

the the total effort required to create the compression matrix from the covariance matrix is

$$\frac{n^3}{n_{block}^2} \tag{5.1}$$

So, what are typical values of n_{block}? For the CBI, we have upwards of 25 primary beam patches per individual field. With 80 fields in the extended mosaics (the 2000+2001 data shown in Figure 4.16), we have a total of $\sim$ 2000 blocks. That means the speedup to calculate the compression matrix is $\mathcal{O}\left(10^6\right)$. So, what would have taken a decade can now be done in a few minutes. The other computationally intensive part of the compression is actually carrying out the compression on the large window matrices. We can take advantage of the fact that the compression matrix is a string of isolated blocks to greatly speed up the compression as well. Compressing a matrix takes two steps: first, multiply the uncompressed matrix by the compression matrix on the right. This gives an intermediate matrix of size $n \times f_{comp}n$. Then multiply the intermediate matrix by the transpose of the compression matrix on the left to get the final, compressed matrix. It turns out that, because all relevant matrices are symmetric, we need only calculate half of the intermediate matrix. So, the final number of elements we need to calculate is $\frac{1}{2}n \times f_{comp}n$. Normally, each element would require a set of n multiplications to calculate, but because the compression matrix consists of blocks, we need only use the non-zero elements in the block. Since there are on average n/n_{block} elements, the total operation count to compress is then

$$\frac{1}{2}\frac{f_{comp}}{n_{block}}n^3 \tag{5.2}$$

So, the speedup is a factor of $2n_{block}/f_{comp}$. For the CBI, the compression factor is typically $\sim 10\%$ (although it depends on how oversampled the data are), so the execution time for the extended mosaics is dropped by $\mathcal{O}\left(50,000\right)$. This is not quite as much of a speedup as for calulating the compression matrix, but is subtantial nonetheless, and certainly sufficient to bring the compression into the realm of feasibility. The computational burden for the final compression from the intermediate matrix can be calculated the same way, but the number of elements we need to calculate is

smaller yet by another factor of f_{comp}, so it is always faster than computing the intermediate matrix. The compressed data vector is easy to calculate as it is simply the compression matrix times the uncompressed data vector.

While feasible, even for fairly large problems, the above method can still be substantially sped up. The compression becomes faster as we increase the number of blocks, but at a cost of reducing the compression efficiency. Fortunately, this can be worked around using a multi-stage compression. Notice that the compression matrix compresses blocks of uncompressed visibilities into compressed visibilities without mixing any information between blocks. So, each output, compressed visibility *remains localized* in UV space. So, we can do an initial compression using lots of blocks, then group the blocks into a set of super-blocks and repeat the compression on the newly compressed problem using the larger blocks. So for the case of the CBI, we could split each of our primary-beam sized blocks into 10 (roughly a third in each direction) and do an intial compression that is very fast, but at the cost of f_{comp}. Then we merge those 10 blocks back into a single block, and recompress. Because the partially compressed matrices are already much smaller, the compression using the larger blocks is very fast, and as efficient as if we had done a single, large (but expensive) compression.

This compression method has several useful properties in addition. First, because the compression is based on the modes of a covariance matrix passed to the compression algorithm, the compressed data set will naturally keep high signal modes present in the covariance matrix. So it is easy to create a compressed data set that retains its sensitivity to any desired properties of the data set described by their covariances. This is how CBISPEC can naturally retain sensitivity to the spectral index of the sky signal - by adding a component with α (where α is defined such that a visibility will have signal proportional to frequency raised to the power α) different from 2 to the input compression matrix, CBISPEC will retain not only modes that look like pure CMB, but modes with spectral index α as well. Because modes with intermediate spectral indices can be approximated by a superposition of modes with spectral index α as well as 2, in practice CBISPEC keeps sensitivity to a wide range of spectral indices. Of course, this technique can be used to keep a much wider range of possible data signals in the compressed data set as well. For a demonstration of

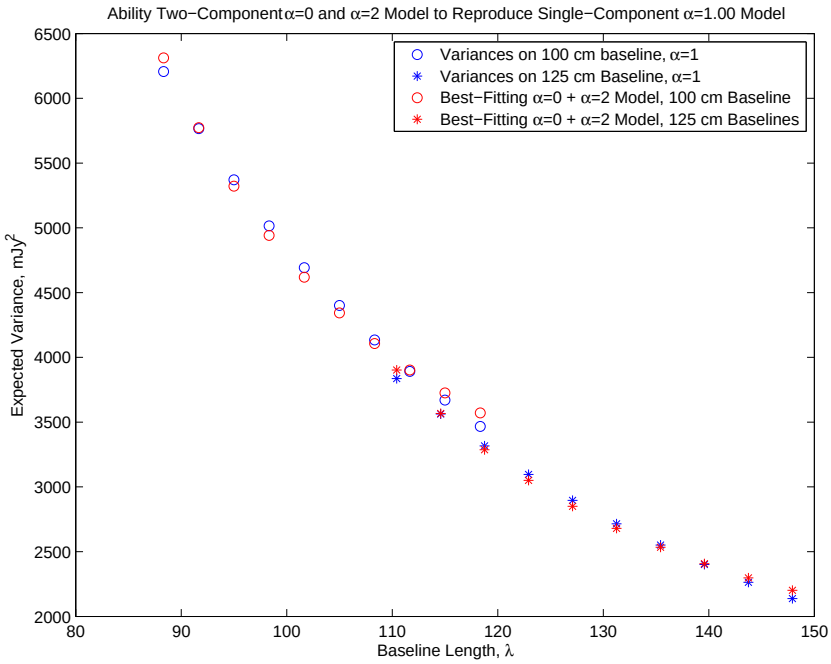


Figure 5.7 Equivalence of single component models with variable spectral index α to two-component spectral index data. The red points are 100 cm and 125 cm baseline expected variances for roughly equal parts of $\alpha = 0$ and $\alpha = 2$. The blue points are 100 cm and 125 cm data at an intermediate spectral index of $\alpha = 1$. The band powers have been adjusted to provide the best overall fit between the two models. The quality of the fit shows that single component models with adjustable spectral index to a good job reproducing multiple-component Gaussian fields with different spectral indices for the components.

how a combination of $\alpha = 0$ and $\alpha = 2$ models can reproduce a single $\alpha = 1$ model, see Figure 5.7. A single band power is applied to each of the $\alpha = 0$ and $\alpha = 2$ models, which were then adjusted to give the best fit to the $\alpha = 1$ model. The two-component model reproduces the $\alpha = 1$ model very faithfully both at different frequencies and different baseline lengths, so the spectral information will indeed be kept during compression.

A second useful property of the compression is that it (usually) only needs to be done once. The compression depends on the *expected* properties of the data, not on the actual values of the data. So, if the data change, through recalibration, new point-source subtraction, etc., only the new data vector needs to be compressed, and the previously calculated compressed window matrices remain unchanged. Compressing the data is at worst an $\mathcal{O}\left(n^2\right)$ operation, and so is a trivial computing burden. This can be tremendously useful when data sets are undergoing incremental changes. The

window matrices must be recompressed if the noise changes, though, since the compression relies on having uniform, independent noises in the data.

A closely related benefit of CBISPEC is that it is extremely efficient at analyzing sets of simulated data. Because only the data themselves change between different realizations and not the statistical properties, the only compression required is again that of the data vector. This makes CBISPEC ideal for Monte Carlo simulations.

5.2 Mosaic Window Functions

In this section we expand the analysis of Section 3.3 to include the calculation of window functions for visibilities with different pointing centers. This is necessary to take advantage of the sharpened ℓ resolution offered by mosaicked observations.

5.2.1 General Mosaic Window Functions

Starting from Equation 3.14, it is straightforward to adjust for the different pointing centers. If I wish to move the primary beam around on the sky, I can equivalently move the mode on the sky, but in the opposite direction. Because the mode is a plane wave, that is equivalent to simply shifting the phase of the mode. If $\boldsymbol{\phi}$ is the vector on the sky by which I have moved, and $\boldsymbol{w}$ is the wavevector of the mode in question, then Equation 3.14 becomes

$$\langle V_1^* V_2 \rangle = f_T(\nu_1) f_T(\nu_2) \int\int \left(\frac{\delta T}{T_{CMB}}(\boldsymbol{w}) \right)^2 \exp(2\pi i \boldsymbol{\phi} \cdot \boldsymbol{w}) \, d^2\boldsymbol{w} \, \widetilde{A}^* (\boldsymbol{u_1} - \boldsymbol{w}) \, \widetilde{A}(\boldsymbol{u_2} - \boldsymbol{w}) \quad (5.3)$$

We can again pull out the angular part of the integral to get the window functions:

$$\mathrm{W}_{ij}(\ell) = \int \widetilde{A}_1^* (\boldsymbol{u_1} - \boldsymbol{w}) \, \widetilde{A}_2 (\boldsymbol{u_2} - \boldsymbol{w}) \exp(2\pi i \boldsymbol{\phi} \cdot \boldsymbol{w}) \, d^2\boldsymbol{w} d\theta$$

5.2.2 Gaussian Beam

The evaluation of the window function proceeds along similar lines to the single pointing window functions. For a Gaussian,

$$\widetilde{A}_i\ (\mathbf{u}) = \frac{1}{2\pi\sigma_i^2} exp(-\frac{\mathbf{u}^2}{2\sigma_i^2})$$

and so the window function is

$$W(w) = \frac{1}{4\pi^2\sigma_1^2\sigma_2^2}\int_0^{2\pi} \exp(-\frac{(\mathbf{u_1}-\mathbf{w})^2}{2\sigma_1^2} - \frac{(\mathbf{u_2}-\mathbf{w})^2}{2\sigma_2^2} + 2\pi i\boldsymbol{\phi}\cdot\mathbf{w})d\theta$$

If we introduce the variables θ_1 and θ_2 for the angles of $\mathbf{u_1}$ and $\mathbf{u_2}$ respectively on the sky, and θ_ϕ for the angle of $\boldsymbol{\phi}$ on the sky, we have

$$W(w) = \frac{1}{4\pi^2\sigma_1^2\sigma_2^2}\int_0^{2\pi} \exp[(-\frac{u_1^2}{2\sigma_1^2}-\frac{u_2^2}{2\sigma_2^2})-w^2(\frac{1}{2\sigma_1^2}+\frac{1}{2\sigma_2^2})+w(\frac{u_1\cos(\theta-\theta_1)}{\sigma_1^2}+\frac{u_2\cos(\theta-\theta_2)}{\sigma_2^2})+2\pi iw\phi\cos(\theta_\phi-\theta)]d\theta$$

The term involving θ_1 and θ_2 can be simplified to $C\cos(\theta-\theta_{eff})$ if we have

$$C^2 = \frac{u_1^2}{\sigma_1^4} + \frac{u_2^2}{\sigma_2^4} + \frac{2u_1u_2}{\sigma_1^2\sigma_2^2}\cos(\theta_1-\theta_2)$$

and

$$\tan(\theta_{eff}) = \frac{\frac{u1}{\sigma_1^2}\sin(\theta_1)+\frac{u2}{\sigma_2^2}\sin(\theta_2)}{\frac{u1}{\sigma_1^2}\cos(\theta_1)+\frac{u2}{\sigma_2^2}\cos(\theta_2)}$$

Let us also define the variables

$$A = \frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2}$$

$$B = \frac{u_1^2}{2\sigma_1^2} + \frac{u_2^2}{2\sigma_2^2}$$

These are the same definitions for A, B, C that we had in the single pointing case. This leaves

$$W(w) = \frac{1}{4\pi^2\sigma_1^2\sigma_2^2}\int_0^{2\pi} \exp(-Aw^2 - B - Cw\cos(\theta-\theta_{eff}) + 2\pi iw\phi\cos(\theta-\theta_\phi))d\theta$$

The terms involving A and B do not have any θ dependence, so they can be pulled out of the integral. If we define $\theta^{'}$ to be $\theta - \theta_{eff}$ and change the limits of integration (which are arbitrary as long as we go exactly once around a circle), we are left with

$$W(w) = \frac{1}{4\pi^2\sigma_1^2\sigma_2^2}\exp(-Aw^2 - B)\int_{-\pi}^{\pi}\exp[-Cw\cos(\theta^{'}) + 2\pi iw\phi\cos(\theta^{'} - \theta_{\phi} + \theta_{eff})]d\theta^{'}$$

We can evaluate an integral of this form

$$\int_{-\pi}^{\pi}\exp[a\cos(\theta) + bi\cos(\theta + \phi)]d\theta$$

quickly by starting with the identities (derivable from, *e.g.* Abramowitz & Stegun, 1965, 9.1.42 and 9.1.43):

$$\exp(ia\cos(\theta)) = J_0(a) + 2\sum_{1}^{\infty} i^n J_n(a)\cos(n\theta)$$

By letting $a \longrightarrow ia$, we also have

$$\exp(a\cos(\theta)) = I_0(a) + 2\sum_{1}^{\infty} I_n(a)\cos(n\theta)$$

Now we have a phase in the complex part, but it is easily dealt with as $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$, so

$$\exp(ia\cos(\theta + \phi)) = J_0(a) + 2\sum_{1}^{\infty} i^n J_n(a)(\cos(n\theta)\cos(n\phi) - \sin(n\theta)\sin(n\phi))$$

If m and n are integers, we have $\int_{-\pi}^{\pi}\cos(n\theta)\cos(m\theta) = \pi\delta_{n,m}$, or 2π if $n = m = 0$. The same result holds for cos $\longrightarrow$ sin, and $\int_{-\pi}^{\pi}\cos(n\theta)\sin(m\theta) = 0$. All $\sin(n\theta)$ terms go away in the integral, as well as all cross terms the product of the two cosin series, so we have the exact result:

$$\int_{-\pi}^{\pi}\exp(a\cos(\theta) + ib\cos(\theta + \phi)) = 2\pi I_0(a)J_0(b) + 4\pi\sum_{1}^{\infty} i^n I_n(a)J_n(b)\cos(n\phi) \qquad (5.4)$$

Now as long as the Bessel functions are quickly calculable, this is a fast way of doing the integral. Fortunately, this is indeed the case. The following recurrence relations hold:

$$J_{n+1} = \frac{2n}{x} J_n(x) - J_{n-1}(x) \qquad I_{n+1} = -\frac{2n}{x} I_n(x) + I_{n-1}(x)$$

They are unstable in the forward direction, but they *are* stable in the downward direction. For calculating high order Bessel functions, *Numerical Recipes* (Press et al., 1992) recommends starting with essentially random starting values for the recurrence relation, and running it downwards. One saves the value at the desired order, and then continues down to order zero, at which point one normalizes things by a call to the zeroth order Bessel function. So, rather than make separate calls to different Bessel functions, we can accumulate the sums of the products of the Bessel functions and normalize at the end, making the whole integral only marginally more work than two calls to high order Bessel routines. It is also important to use the recurrence relations for sines and cosines as well:

$$\cos(n\theta) = \cos((n-1)\theta)\cos(\theta) - \sin((n-1)\theta)\sin(\theta) \qquad \sin(n\theta) = \sin((n-1)\theta)\cos(\theta) + \cos((n-1)\theta)\sin(\theta)$$

In playing around with these, I've found that I should start accumulating the sum at $n_{max} \simeq 2min(|a|, |b|) + 16$, and run someting like 40 iterations beforehand to let the recurrence relations converge. This algorithm runs a few hundred times faster than carrying out numerical integrals to achieve similar accuracy.

5.3 Comparisons with Other Methods

In order for CMB power spectrum measurements to be believable, it is critical that different methods produce very similar spectra from the same data set. The most natural comparison is between CBIGRIDR and CBISPEC. Many other methods are not applicable to interferometer data, such as the one used by WMAP (algorithm described in Oh et al., 1999) and the one used by recent version of

Table 5.2. CBIGRIDR and CBISPEC Comparison

Bin ℓ range	1-R	m	$b(\mu K)$
< 900	4.5e-5	0.988	-2.1
900-1500	7.0e-5	0.986	-2.9
1500-2100	5.1e-5	0.982	0.0
2100-2700	2.4e-5	0.985	-0.1
> 2700	2.3e-5	0.986	0.0

BOOMERANG (described in Hivon et al., 2002), since they require taking a fast spherical harmonic transform of the data in order to calculate the window matrices. Because visibilities are point-like neither in UV space or on the sky (as map pixels are), there is no comparable transform for the CBI dataset.

I report here on a comparison between the power spectra measured by CBIGRIDR and MLIKELY described in Myers et al. (2003) and those measured by CBISPEC and the fast fitting method of Chapter 2 on a set of 100 simulated deep datasets. The agreement between CBISPEC and CBIGRIDR/MLIKELY is excellent, with correlation coefficients $\sim 1-$ a few times 10^{-5}. In addition to the scatter about the best-fit lines being small, the slopes m to the linear fits were in all cases nearly unity, with CBISPEC averaging about 1.5% less than CBIGRIDR, and the offsets from the origin b of a few μK. See Table 5.2 for a summary of the comparison statistics, and Figures 5.8 and 5.9 for CBISPEC and CBIGRIDR fits to the first (high-signal) and last (high-noise) bins, respectively.

5.4 Foreground with CBISPEC

The goal of measuring the primordial anisotropy spectrum is complicated by the presence of astronomical sources in the foreground contributing to the intensity measured at earth. The most prominent foreground signal at 30 GHz is radio point sources, discussed in greater detail in Mason et al. (2003). If the point source positions are known, they can be quite effectively projected out of the data set, making the spectrum insensitive to the actual value of the point source, as seen in

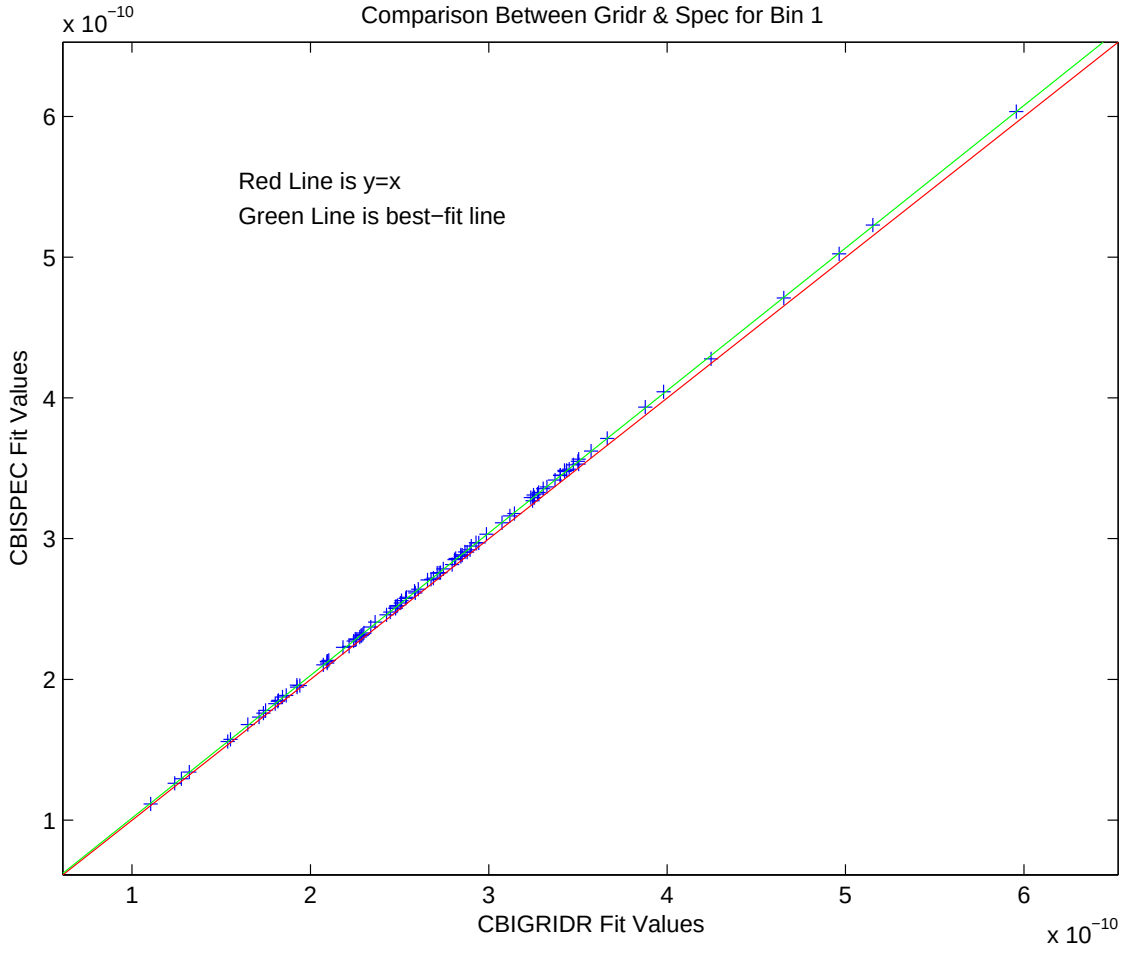


Figure 5.8 Comparison of fit values between CBIGRIDR and CBISPEC, for the first bin (bins defined in Table 5.1). The agreement is very good, with the CBIGRIDR and CBISPEC results almost identical. Statistics of the comparison are in Table 5.2, where m and b are the slope and intercept of best-fit line. The first bin has the highest SNR in the data of any of the bins.

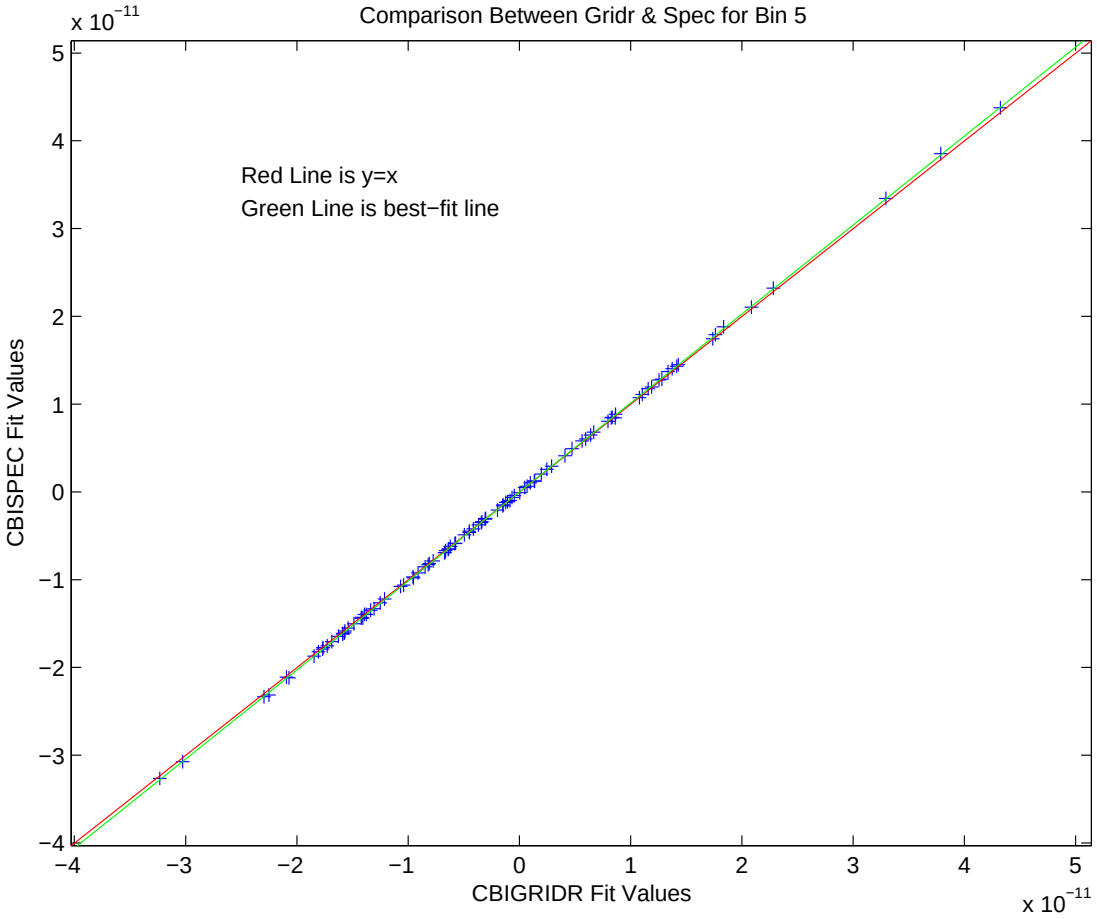


Figure 5.9 Same as Figure 5.8 for the highest-ℓ bin. This bin has the lowest SNR of any of the bins. Again, the agreement between CBIGRIDR and CBISPEC is excellent.

Chapter 4. This technique has been effectively used by many CMB experiments (e.g. Mason et al., 2003; Pearson et al., 2003; Halverson et al., 2002). In addition, the power from faint radio sources can be estimated, with reasonable accuracy, from the number counts of brighter sources (Mason et al., 2003). So, the effects of point sources are calculable and removable. The uncertainties in the spectrum due to point sources are negligible for all but the very smallest scales for the CBI, and even on those scales, the uncertainty is much smaller than the measured spectrum.

More problematic is the signal due to diffuse galactic foregrounds, such as synchotron radiation or bremsstrahlung. The major difficulty is they are rather poorly understood on the small scales and high frequencies at which the CBI operates. Consequently, we wish to constrain the limits on diffuse foreground emission from the CBI dataset itself. Unlike the point sources, where there is information about their expected level, their expected spectral shape, and their expected angular power spectrum, the only information we have to work with on the diffuse foregrounds is that their spectral indices will likely be substantially lower than that of the CMB ($\alpha \sim -0.7 - 0$ vs. $\alpha \sim +2$). The ideal thing to do is to make otherwise identical maps covering a wide range of frequencies with high signal to noise, and then measure the component only with the spectral shape of a 2.73 degree blackbody. For the CBI, which works in the Rayleigh-Jeans regime, we have to use the CBI fractional bandwidth of ~ 0.3 to distinguish between the CMB and foregrounds. A major application of CBISPEC is placing limits on potential foreground signals using the CBI's spectral discrimination. CBIGRIDR is unsuitable for this task since it assumes all data has a single frequency behavior during gridding, destroying frequency information in the process.

We place limits on the potential contribution of foreground sources through a two-part procedure. The first part is to measure a single best-fitting spectral index to the low-ℓ data and its uncertainty. We then use that spectral index to limit what fraction of the total signal could have come from a foreground.

5.4.1 Measuring the Spectral Index

To measure the overall spectral index α of a data set, we assume a single spectral index, calculate the window functions using that spectral index, fit an angular power spectrum, and record the likelihood of that spectrum, and repeat for a new assumed spectral index. Gradually, this builds up the curve that describes the likelihood as a function of α. The peak of the curve is the best-fitting spectral index, with the uncertainty in α given by the width of the likelihood curve. We have to re-fit the power spectrum at each likelihood (rather than simply change α and evaluate the likelihood) because, in general, the power spectrum is degenerate with α, and if we don't re-fit, the constraint will be artificially tight.

It is straightforward and fast to re-calculate the uncompressed window functions when varying the spectral index. By looking at the general form for the window function, Equation 3.18, one can see that the sensitivity to frequency is contained only in the coefficient, $f_T(\nu_1) f_T(\nu_2)$, defined in Equation 3.10. So, the window function for a given spectral index is simply the original CMB window function (which we have already calculated) divided by $f_T(\nu_1) f_T(\nu_2) \nu_1^2 \nu_2^2$ and multiplied by $\nu_1^\alpha \nu_2^\alpha$. This must be done before compressing, as the compression mixes together visibilities of different frequencies. If one wishes to compress, it is also essential to use the same compression matrix at all spectral indices, or else the likelihoods will not be directly comparable. Otherwise, once the window matrices have had α applied to them, the compression and fitting procedures proceed exactly as in the pure CMB case.

When measuring foregrounds with the CBI, it is important to know what the expected best-fit spectral index is. While we expect it to be in the vicinity of the Rayleigh-Jeans value of 2, there is potentially a strong degeneracy between α and the shape of the underlying CMB power spectrum. In fact, for an interferometer with baselines of a single length, the degeneracy is almost perfect. Figure 5.10 shows the degeneracy between spectral index and slope of the power spectrum. Plotted are the expected variances for the 10 CBI channels on a 100 cm baseline. The blue points are the canonical, flat CMB spectrum expected variances. The red points are the expected variances for the same $\mathcal{C}_\ell$ spectrum, with a frequency spectral index of ν^α applied, for $\alpha = 0$. The green points

show the degeneracy between the flat, ν^0 spectrum with a spectrum $\propto \ell^{-6.4}$ and a CMB frequency dependence. For clarity in the plot, the difference between the red and green points has been magnified by a factor of 10, as they otherwise lie on top of each other. Clearly, no single baseline length can discriminate between a CMB spectrum sloping in ℓ and a foreground spectrum flat in ℓ. If the CMB spectrum is falling (as the trend is in the ℓ range covered by the CBI), then the best-fitting value of α can be substantially less than 2, even if there are no foregrounds present. Once we add baselines sampling the same ℓ region at different frequencies, though, the degeneracy is broken. In Figure 5.11 same data in as in Figure 5.10 are plotted, along with the identical models evaluated for a 125 cm baseline. The blue, green, and red stars on the right are the 125 cm models, with the crosses the 100 cm data. The degeneracy is best broken in the overlap region where 100 cm and 125 cm baselines sample the same ℓ range at different frequencies. This is a more difficult measurement than that of the power spectrum since the best handle on α comes from the difference between a few channels rather than the average of all channels. Consequently, to measure foregrounds well, we need groups of baselines of similar, but slightly different, lengths, preferably with high SNR.

Consequently, we reconfigured the CBI in July 2000 to have 3 125 cm baselines in addition to 7 100 cm baselines. Since the SNR is important, we use only the 100/125 cm baselines in measuring α, as the sensitivity drops quickly at higher ℓ and contamination from radio point sources becomes relatively more important. To determine the expected value of α as well as one measure of its uncertainty, I used MOCKCBI to create simulations as close as is feasible to the CBI, using a purely CMB sky. I also included realistic point-source populations (using the number counts in Mason et al., 2003) and subtracted off simulated OVRO 40 meter fluxes with errors in order to get the point-source population as close to reality. Because we use only the low-ℓ, short baseline data for determining foregrounds, the effective beam on the sky is very large. This makes projection out individual point sources unpracticable since each source in effect removes a patch the size of the synthesized beam on the sky. When we use high-ℓ data with its small synthesized beam, there is lots of sky left after removing the sources, but that is not the case if we use only the low-ℓ data. Fortunately, the expected signal from point sources unmeasured by OVRO and the residuals caused

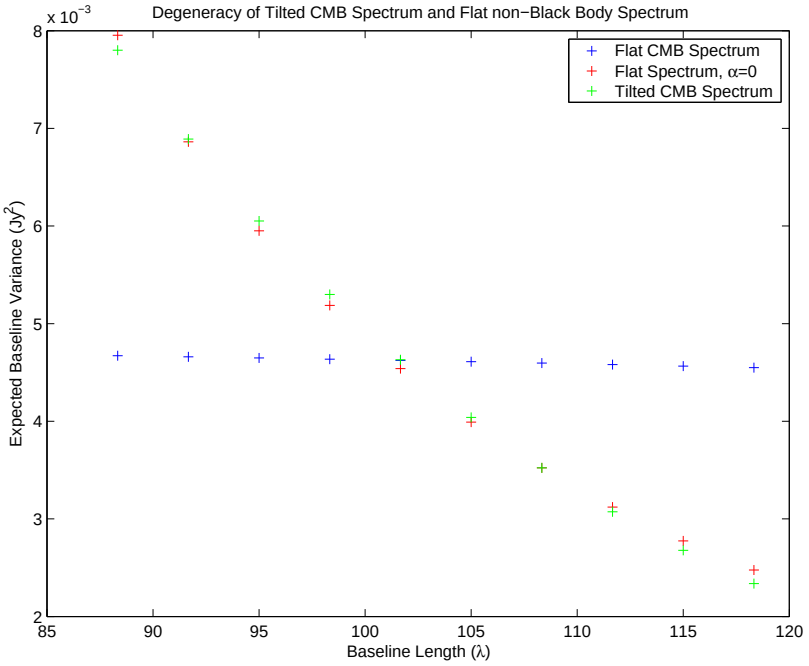


Figure 5.10 Figure showing the degeneracy for a single baseline between a tilt in the power spectrum ($\mathcal{C}_\ell \propto \ell^\gamma$) and a flat power spectrum with a non-Black Body spectrum. Blue points are the expected variances on the 10 CBI channels for a 100 cm baseline, assuming a flat CMB spectrum. Red points are the expected variances for the same flat spectrum in ℓ, with a frequency spectral index of ν^α applied with $\alpha = 0$. Green points are the expected variances for a non-flat CMB spectrum with a power law applied that best matches the $\alpha = 0$ points. The best-fit law for the green points is $\mathcal{C}_\ell \propto \ell^{-6.4}$. For ease of viewing, the difference between the red and green points has been amplified by a factor of 10.

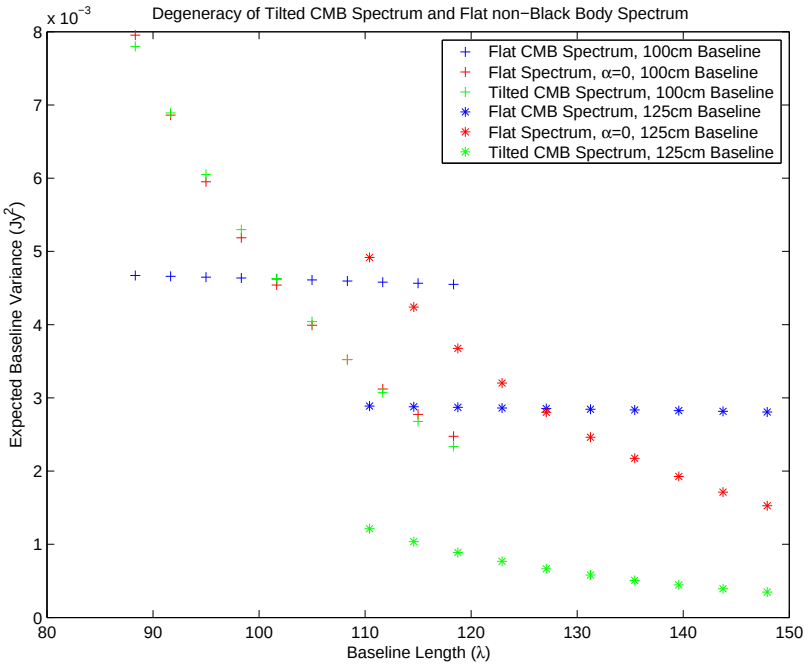


Figure 5.11 Same as Figure 5.10, this time with a 125 cm baseline added. Color scheme is the same as Figure 5.10. The crosses are the 100 cm baseline, and the asterisks are the 125 cm baseline. Again, the difference between the red and green crosses has been amplified by a factor of 10 for clarity, but there is no scaling on the 125 cm points. The addition of the 125 cm baseline has broken the degeneracy between the flat, ν^0 spectrum and the $\ell^{-6.4}$, Planck spectrum. For these parameters, in the region in which the two baselines overlap at $110 - 120\lambda$, the predicted values for the 125 cm visibilities differ by a factor of 4 when the 100 cm visibilities are degenerate.

by errors in the OVRO subtraction are quite small at $\ell \sim 600$ (the point source spectrum is rising as ℓ^2 while the CMB is falling off, so while care is required in the treatment of point sources at $\ell \sim 2500$, their effect at $\ell \sim 600$ is negligible). Finally, I used the measured CBI primary beam in the simulations, rather than the Gaussian approximation used by CBISPEC, in order to account for any potential bias caused by using a Gaussian beam.

After creating a set of 90 simulations based on the 02 hour mosaic, I analyzed them for the single best-fitting spectral index. I used the data below $\ell = 770$, and for computational efficiency grouped the fields into 3×3 mosaic blocks. The sample mean was $\alpha = 2.0528$ with the scatter about the mean of 0.24. The sample scatter agrees well with the uncertainties derived by the curvature of the likelihood around the peaks, which was 0.27. Because of the good agreement between the sample variance and the uncertainty measured by the likelihood curvature, we can again adopt the likelihood curvature errors for spectral index as we did for total power. Also, since the simulations seem consistent with the expected value of 2, I adopt that as the target value for the real data. See Figure 5.12 for the histogram of the best fitting values of α for the individual simulations.

5.4.2 The Spectral Index Measured by CBI

With the simulation results from the previous section in hand, we are now in a position to intepret the spectral index measured from the actual data. The pipeline used to process the results is identical to that used for the simulations, save for the noise correction factor from Section 4.1 required when using real data. The fields are again divided into 3×3 patches. The individual field results are in Table 5.3. We know that there is an extremely bright (~ 1 Jy) source in the northern extension of the 02 hour mosaic that leaves significant artifacts in the maps. Not surprisingly, this source also has a significant impact on the spectral index of the 02 hour field. With the 4 patches around the source left in, the 02 hour field has a best-fit α of 1.47 ± 0.22, 2.46σ away from 2. When these patches are removed, the best-fit α rises to 1.72 ± 0.25, only 1.1σ from 2. Also, these four patches have the highest power levels amongst all 31 individual patches that comprise our mosaic data set, with the lowest of these four about 1000 μK^2 higher than the next-highest patch. See Figure 5.13 for the

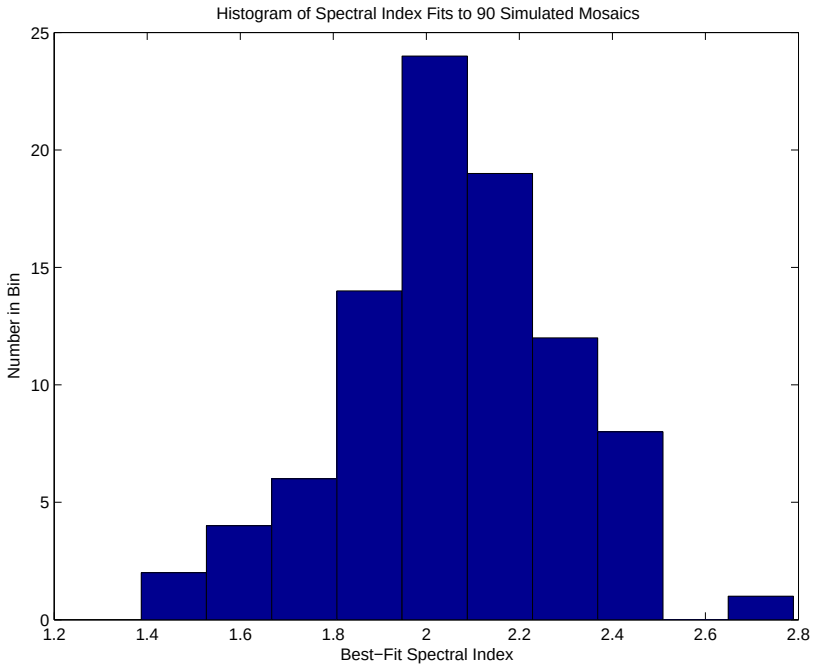


Figure 5.12 Histogram of spectral index fits to a flat band power CMB model, made using simulations based on the 02 hour mosaic. The expected value of α is 2, and the simulations do indeed cluster around it. The mean of the distribution is $\alpha = 2.05$ with 1-σ scatter of 0.24.

power/α plot with the four anomalous patches marked in blue. The odds of these four being the highest is 1 in 31,465. For comparison, that is only about twice as likely as getting dealt a straight flush in poker, without ever drawing. Since these patches are clearly corrupted, we remove them in the joint fit. The best-fit α for the entire mosaic set is then 1.76 ± 0.13, a difference of 1.85σ from 2. This is consistent with pure CMB, though perhaps a mildly suggestive of the presence of a weak foreground. Unfortunately, it will be challenging to place much tighter constraints on potential forground contamination. By looking at what fraction of the total signal is required to come from a foreground at given α in order to make the visibility window function of CMB + foreground agree with the best-fit α, we can estimate the upper limits on possible foreground signals. In our case, if the entire difference from 2 is ascribed to the presence of a free-free foreground ($\alpha \simeq -0.1$), then the free-free signal makes up only $\simeq$12% of the total power at the center of CBI's band. If instead the foreground source were synchotron, then it could contribute only $\simeq$ 8.5%. So, to have a reliable estimate of the foreground contribution from an outside source, the outside source would

Table 5.3. Spectral Indices of CBI Mosaics

Field	α	$\sigma(\alpha)$	like(peak)-like(2)
C14 hour	1.63	0.22	1.38
C20 hour	1.90	0.21	0.11
C02 hour	1.47	0.22	3.02
C02 hour notop	1.72	0.25	0.62
Joint notop	1.76	0.13	1.70

need to have a signal-to-noise per pixel of the CMB at 30 GHz and $\ell \sim 600$ substantially larger than 10. While WMAP indeed has an all-sky foreground map at 30 GHz (Bennett et al., 2003), their signal-to-noise is poor on a per-pixel basis at these scales. As follow-up to this work, I do plan to try to estimate the foreground contribution from the WMAP maps, but the sensitivity will almost certainly be much poorer than the CBI internal sensitivity to foregrounds.

As a final note on foregrounds, if foregrounds were indeed the cause of the shift in α away from 2, then we would expect an anti-correlation between the spectral indices of the individual patches and their power levels. Since the CMB is statistically identical for all the patches, a stronger foreground should mean a higher power level in addition to a lower spectral index. The plot of α versus band power for all 31 patches is Figure 5.13. The blue crosses are the four contaminated patches at the north end of the 02 hour mosaic. The remaining 27 patches are marked with red asterisks. If we exclude the four patches, then there is actually a positive correlation between α and band power—the opposite of what one would expect from foregrounds. The correlation is extremely weak ($r = 0.12$) and highly insignificant ($prob(r > 0.12) = 48\%$ for Gaussian data). While a statistically very uniform foreground would not introduce an anti-correlation, it would indeed be baroque if our three mosaics, separated by 90° and at different galactic latitudes had very similar foregrounds.

5.4.3 Future Improvements

The are a number of relatively straightforward improvements than will be made to the current version of CBISPEC, greatly improving performance.

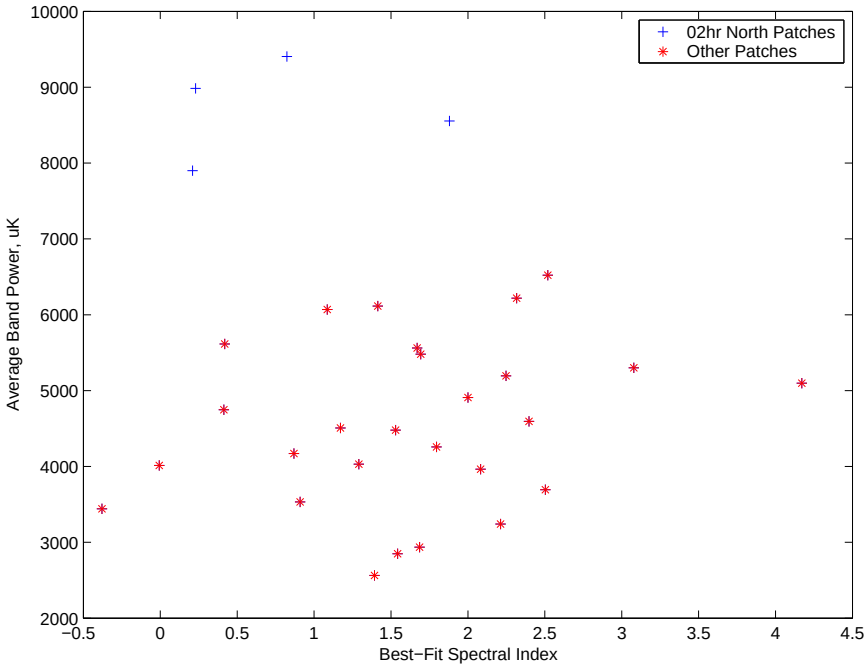


Figure 5.13 Figure showing the distribution of spectral indices of the individual 3 by 3 chunks of the CBI data, plotted against their low-ℓ power levels. The four blue fields with low values of α and high power are adjacent patches at the north end of the 02 hour mosaic contaminated by a very bright point source. For this reason, they have not been used in measuring the spectral index of CBI data.

The most important change will be the indexing of pre-calculated window function. Currently, it takes 45 minutes on a 4 CPU es45 HP server (1.0 GHz alpha CPU's) to calculate the raw, uncompressed window matrices between two fields (in this case the 14 hour deep field). Tweaking the Bessel function sums of Section 5.2 can cut the operation count by a factor of two, but calculating the mosaic window functions will never be a very cheap operation (in contrast, CBIGRIDR takes just over a minute). For a given UV coverage and a given set of pointings, however, they only need to be calculated once, ever. Furthermore, if the UV coverage of all fields is identical, which is the case for the CBI between reconfigurations, then pairs of fields with the same angle between the fields will have identical window functions. So, we expect to be able to use the window functions, calculated for a single pair of fields, many times when working with an entire mosaic. To estimate how many separate fields we will need to calculate, picture the pointings on an evenly spaced grid in RA and dec. This is how the CBI has observed. Then, ignoring $\cos(\delta)$, the vector between any pair of fields is also the vector that connects one of (any) two corners and another field. So, if we

calculate all the window functions between only the two corner fields and the rest of the fields in the mosaic, we will have all necessary window functions. If we have a total of m pointings, then we need to calculate and store $2m$ sets of window functions. This means that the window function calculations will scale quite well for CBISPEC, with the computational burden scaling linearly with mosaic area. The CBIGRIDR scaling is different. If I double the size in each direction of a mosaic, then I have four times as much area, and four times as many visibilities that need gridding. I also need to grid onto estimators of only half the size because my sampling of the UV plane is now finer, for a total of four times as many estimators. So, a factor of 4 in area leads to a factor of 16 in work, which is a scaling of area^2. So, CBISPEC should behave better for larger areas than CBIGRIDR. The approximation used, ignoring the $\cos(\delta)$ is a good one for the CBI, since we have restricted ourselves to regions within 5° of the equator. The cosine of 5° is 0.996, which means that if we calculate the window functions assuming that $\cos(\delta) = 0.998$, we will never make an error of more than 0.2%. The effect will be to smear the spectrum in ℓ by 0.2%, which is negligible. Even at $\ell = 3500$, the highest value to which the CBI is sensitive, that is an error of only $\delta\ell = 7$.

When the reuse of precalculated window functions is in place, I intend to revisit the foreground analysis of the mosaic data. Rather than break up the mosaics into three field by three field chunks, it will be simple to treat the mosaic in its entirety. This should tighten the foreground constraints somewhat because there is appreciable overlap between the chunks. Not taking advantage of that overlap leads to a penalty in SNR, since some redundant information is being handled separately. Treating the mosaic as a whole will give the best possible foreground constraints from the CBI data.

Because foregrounds will be more of a concern for polarization observations than CMB total-intensity observations, a major future task for CBISPEC will be constraining the polarization spectral index. This will require updating CBISPEC to do polarization. To do this, we will need to calculate mosaic polarization window functions, which are similar to the standard mosaic window functions, modulo an extra sine in the integral.

None of these changes should be difficult, and I hope to implement them soon, especially so that we can measure the polarized foregrounds in the upcoming CMB polarization results. Martin

Shepherd, who has done the lion's share of the actual coding using the algorithms I have developed here, is currently occupied working on another project. Once that is finished, which should be in a few months, we will update CBISPEC, making it a far more powerful tool.

Chapter 6

Conclusion

In this thesis, we have discussed observations of the Cosmic Microwave Backgroud with the Cosmic Background Imager. The CBI is a highly sensitive interferometer working at 30 GHz optimized for observations of the CMB in the multipole region $500 < \ell < 3500$. It was the first instrument to measure the damping tail of the CMB spectrum, the falloff in power at small scales due to photon diffusion before recombination, originally in Padin et al. (2001a), and then with more detail in Mason et al. (2003) and Pearson et al. (2003). The CBI also measured an unexpectedly large amount of power at large-$\ell(> 2000)$ in Mason et al. (2003), which is possibly the first time that secondary anisotropy due to the Sunyaev-Zeldovich effect has been measured statistically either from clusters (Bond et al., 2002b) or the first generation of stars (Oh et al., 2003) rather than in pointed observations of known galaxy clusters. The CBI also measured the CMB on scales of present-day galaxy clusters for the first time. We have also used the angular power spectrum measured by the CBI to constrain cosmological parameters (Sievers et al., 2003) both alone (using COBE-DMR as an anchor at low-ℓ) and in concert with other experiments. The parameters derived from the combination of experiments are some of the most precise ever determined, with our best determination (using data from CMB, large-scale structure, and Type Ia supernovae observations) of the flatness of the universe to be $\Omega_{tot} = 1.03^{+0.05}_{-0.04}$. Since the universe appears to be flat to high accuracy, as predicted by inflation, we adopt a flat universe prior in further parameter estimates. Our best parameter values are calculated using the previous data, the flat prior, and also the Hubble Space Telescope key project result for the value of the Hubble constant. The parameter limits

are: primordial fluctuation spectrum $n_s = 1.00^{+0.06}_{-0.05}$, physical baryon density $\Omega_B h^2 = 0.023^{+0.002}_{-0.002}$, physical cold dark matter density $\Omega_{cdm} h^2 = 0.12^{+0.01}_{-0.01}$, and cosmological constant $\Omega_\Lambda = 0.70^{+0.05}_{-0.05}$. We do not derive a useful constraint on the optical depth to reionization, $\tau_c < 0.38$. We also place limits on parameters of interest derived from these fundamental model parameters: the density of matter relative to critical is $\Omega_m = 0.30^{+0.05}_{-0.05}$, the density of baryons relative to critical is $\Omega_B = 0.047^{+0.004}_{-0.004}$, the Hubble constant is $h = 0.69^{+0.04}_{-0.04}$, and the age of the universe is $13.7^{+0.2}_{-0.2}$Gyr.

The author has participated in many phases of CBI acitivty. Initially, I helped in the construction of the CBI, including assembling and testing the CBI receivers. One of my key contributions was developing the analysis pipeline used to measure the power spectrum contained in Padin et al. (2001a) as well as the derivation of some weak cosmological constraints from that dataset. Another was to participate in the development of the pipeline described in Myers et al. (2003) and use it to measure the spectrum in Pearson et al. (2003). This included the calculation of a statistical correction to our noise estimate required to make it unbiased, numerous speedups in the pipeline that allowed us to measure the spectrum from CBI mosaics to higher ℓ, and a fuller understanding of the effects of radio point sources on the CBI spectrum, which took advantage of the high-ℓ data in the CBI mosaics to reduce the impact of the sources. I also describe a major improvement to the algorithm used to find the maximum likelihood spectrum that will be described in Sievers (2004, *in prep*) that we have adopted into our current pipeline. Finally, I have developed a flexible algorithm that efficiently compresses CBI datasets while maintaining considerable freedom in the choice of information retained. Martin Shepherd and I have coded these algorithms into a program called CBISPEC that I have used to constrain possible diffuse galactic foregrounds present in the CBI data. I find that diffuse foregrounds contribute no more than about 12% of the CBI signal at $\ell \sim 600$ for a bremsstrahlung-like spectral index of $\alpha = -0.1$, with the data consistent with no foregrounds at all at a level of 1.85σ. Finally, Patricia Udomprasert and I have developed and tested optimal methods for treating the noise introduced by the CMB into observations with the CBI of the Sunyaev-Zeldovich effect in clusters of galaxies. By properly weighting the data, we achieve a significant reduction in the uncertainty in H_0 measured using our cluster data, with the potential

of even greater reductions if we survey larger regions around the clusters.

Appendix A

First-Order Expectation of Noise Correction Factor

This appendix derives the theoretical expectation of the noise correction factor. We have many identical measurements (scans) we want to combine. Each scan is made up of many data points, and the estimated error of the scan comes from the scatter of those internal data points. Our final estimator is the weighted average of the scans, using the scatter-based errors for the weighting of each individual scan. While the noise estimate for a single scan is unbiased, there is a bias introduced when we combine many scans. In the limit of combining many statistically identical scans, with each scan made up of ν independent data points, the bias is, to first order, $1 + \frac{4}{\nu}$. We expect to scale our estimates of the noise by something close to this quantity in order to get an unbiased noise estimate.

A.1 Statistical Basics

We will need several basic statistical results in order to work out the epxectation of the noise. The required results are presented here. Throughout this appendix, the variance of a variable x will be written Var(x).

A.1.1 Variance of a Product

We need to know the variance of the product of two independent random variables (not necessarily identically distributed). It is easily shown to be (*e.g.* Mood et al., 1974):

$$\mathrm{Var(xy)} = \langle \mathrm{x}^2 \rangle \langle \mathrm{y}^2 \rangle - \langle \mathrm{x} \rangle^2 \langle \mathrm{y} \rangle^2 \tag{A.1}$$

for independent variables x and y. One can easily verify the following general result, again dependent only upon x and y being independent:

$$\mathrm{Var(xy)} = \mathrm{Var(x)Var(y)} + \langle \mathrm{x} \rangle^2 \mathrm{Var(y)} + \langle \mathrm{y} \rangle^2 \mathrm{Var(x)} \tag{A.2}$$

It is also worth noting explicitly that, if the expectation values of the variables are zero, the variance of the product is the product of the variances: $\mathrm{Var(xy)} = \mathrm{Var(x)Var(y)}$. Also, if only one of the variables has an expectation of zero, then we have the following result (say for $\langle y \rangle = 0$):

$$\mathrm{Var(xy)} = \mathrm{Var(x)Var(y)} + \langle \mathrm{x} \rangle^2 \mathrm{Var(y)} = \mathrm{Var(y)} \left(\mathrm{Var(x)} + \langle \mathrm{x} \rangle^2 \right) = \mathrm{Var(y)} \langle \mathrm{x}^2 \rangle \tag{A.3}$$

This form will get used often below.

A.1.2 Expectation of $f(x)$

We also need to understand how to calculate the expectation value of functions of variables. Say we have random variable x whose distribution $p(x)$ is relatively well-localized (by which we mean that p has finite moments). If we desire the expectation of some function $\langle f(x) \rangle$, then we can Taylor-expand the function and express the expectation in terms of derivatives of f and moments of p.

$$\langle f(x) \rangle = \langle f(x_0) + (x - x_0) \frac{df}{dx}\Big|_{x=x_0} + (x - x_0)^2 \frac{1}{2!} \frac{df^2}{d^2x}\Big|_{x=x_0} + ... \rangle \tag{A.4}$$

We can break up the expectation into different terms since the expectation of the sum is the sum of the expectations. Furthermore, since the derivatives are constant, they can be pulled out of the expectation, yielding:

$$\langle f(x)\rangle = \langle f(x_0)\rangle + \langle (x-x_0)\rangle \frac{df}{dx}\Big|_{x=x_0} + \langle (x-x_0)^2\rangle \frac{1}{2!}\frac{df^2}{d^2x}\Big|_{x=x_0} + ... \tag{A.5}$$

If we set the reference value x_0 to be the expectation of x, then the second term goes to zero, since $\langle x - x_0\rangle = \langle x\rangle - x_0 = 0$ if $x_0 = \langle x\rangle$. So, we have

$$\langle f(x)\rangle = f(\bar{x}) + \langle (x- \bar{x})^2\rangle \frac{1}{2!}\frac{d^2 f}{dx^2}\Big|_{x=\bar{x}} + \sum_{n=3}^{\infty} \langle (x- \bar{x})^n\rangle \frac{1}{n!}\frac{d^n f}{dx^2}\Big|_{x=\bar{x}} \tag{A.6}$$

Since the expectation value in the second term is simply the variance of the distribution, the expectation value to second order is:

$$\langle f(x)\rangle = f(\bar{x}) + \frac{1}{2}\mathrm{Var(x)}\frac{\mathrm{d^2 f}}{\mathrm{dx^2}}\Big|_{\mathrm{x=\bar{x}}} + \mathcal{O}\left(\mathrm{x} - \langle \mathrm{x}\rangle\right)^3 \tag{A.7}$$

Let us use this formula to work out the specific case of $\frac{1}{x}$. The second derivative of $\frac{1}{x}$ is $\frac{2}{x^3}$, so we can plug that in to get:

$$\left\langle \frac{1}{x}\right\rangle = \frac{1}{\bar{x}} + \frac{1}{2}\mathrm{Var(x)}\frac{2}{\bar{\mathrm{x}}^3} = \frac{1}{\bar{\mathrm{x}}}\left(1 + \frac{\mathrm{Var(x)}}{\bar{\mathrm{x}}^2}\right) + ... \tag{A.8}$$

A.1.3 Some Relevant Distributions

Several different probability distributions are relevant to our noise issues. Even though the individual 8.4 second samples are (assumed) Gaussian-distributed, we encounter more distributions than just the Gaussian because we estimate the variances from the scatter of data points rather than knowing the underlying variance. The incomplete Γ and χ^2 distributions are taken from Press et al. (1992). The single variance estimators are the sums of Gaussians random variables squared, hence they are distributed like χ^2 random variables. These can be derived from the incomplete Γ distribution,

which is

$$P\left(a,x\right)=\frac{1}{\Gamma(a)}\int_{0}^{x}e^{-t}t^{a-1}dt \tag{A.9}$$

where $\Gamma(a+1)$ is a factorial. A χ^2 with ν degrees of freedom is

$$\text{Prob}\left(\chi^{2}\langle\chi^{*2}\right)=\text{P}\left(\frac{\nu}{2},\frac{\chi^{*2}}{2}\right) \tag{A.10}$$

If we have n independent members drawn from a Gaussian population with an underlying variance 1 (our data), denote the variance of our particular data set by v, and the degrees of freedom by d ($n-1$ if we're estimating the mean from the data, n if we're not). We can then combine Equations A.9 and A.10, then rescale to get the cumulative distribution function (CDF) of v. Differentiating the CDF yields the probaility distribution function (PDF), which is:

$$\text{PDF}(v)=\frac{d\left(\frac{dv}{2}\right)^{\frac{d}{2}-1}e^{-\frac{dv}{2}}}{2\Gamma(\frac{d}{2})} \tag{A.11}$$

It's fairly easy to show that the first few expectations are

$$\langle v\rangle=1 \tag{A.12}$$

$$\langle v^{2}\rangle=1+\frac{2}{d} \tag{A.13}$$

$$\text{Var}(v)=\langle v^{2}\rangle-\langle v\rangle^{2}=\frac{2}{d} \tag{A.14}$$

The general expectation relation is done by integrating by parts and comparing the resulting integral to the expectation value of the order below it. Here is the answer:

$$\langle v^{n}\rangle=\left(1+\frac{2(n-1)}{d}\right)\langle v^{n-1}\rangle \tag{A.15}$$

We can use this to calculate negative moments as well as positive moments:

$$\langle v^0 \rangle = \left(1 + \frac{-2}{d}\right) \langle v^{-1} \rangle = 1 \tag{A.16}$$

$$\langle v^{-1} \rangle = \frac{1}{1 - \frac{2}{d}} \simeq 1 + \frac{2}{d} \tag{A.17}$$

The next order, and the variance of $\frac{1}{v}$ are

$$\langle v^{-2} \rangle = \frac{1}{1 - \frac{2}{d}} \frac{1}{1 - \frac{4}{d}} \simeq 1 + \frac{6}{d} \tag{A.18}$$

$$\mathrm{Var}\left(\frac{1}{\mathrm{v}}\right) = \left\langle \frac{1}{\mathrm{v}^2} \right\rangle - \left\langle \frac{1}{\mathrm{v}} \right\rangle^2 = \frac{\frac{2}{\mathrm{d}}}{\left(1 - \frac{4}{\mathrm{d}}\right)\left(1 - \frac{2}{\mathrm{d}}\right)^2} \tag{A.19}$$

We also need the variance of v^{-2}:

$$\langle v^{-4} \rangle = \left[\left(1 - \frac{2}{d}\right)\left(1 - \frac{4}{d}\right)\left(1 - \frac{6}{d}\right)\left(1 - \frac{8}{d}\right)\right]^{-1} \simeq 1 + \frac{20}{d} \tag{A.20}$$

$$\mathrm{Var}\left(\mathrm{v}^{-2}\right) \simeq 1 + \frac{20}{\mathrm{d}} - \left(1 + \frac{6}{\mathrm{d}}\right)^2 \simeq \frac{8}{\mathrm{d}} \tag{A.21}$$

Another important distribution is the F distribution. It is the distribution of the ratio of two empirically determined variance estimates if they are drawn from samples with the same intrinsic variance. It is based on the incomplete Beta function,

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1}(1 - t)^{b-1} dt \tag{A.22}$$

with $B(a, b)$ the complete beta function, $B(a, b) = \int_0^x t^{a-1}(1 - t)^{b-1} dt$ (Press et al. (1992)). I_x is unrelated to the modified Bessel function I_n. The CDF of an F is

$$\mathrm{Prob}\left(F_{obs} \rangle F(\nu_1, \nu_2)\right) = I_{\frac{\nu_2}{\nu_2 + \nu_1 F}}\left(\frac{\nu_2}{2}, \frac{\nu_1}{2}\right) \tag{A.23}$$

where F is the ratio of sample variance 1 to sample variance 2 with ν_1 and ν_2 degrees of freedom,

respectively. After a change of variables in the expression for I, a differentiation, and some algebra, one form of the PDF is

$$PDF(F) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} F^{\frac{\nu_1}{2}-1} \left(F + \frac{\nu_2}{\nu_1}\right)^{-\frac{\nu_1+\nu_2}{2}} \left(\frac{\nu_1}{\nu_2}\right)^{-\frac{\nu_2}{2}} \tag{A.24}$$

To take moments of F, we need to integrate $F^n PDF(F)$. The general integral is of the form

$$\int_0^\infty x^n (x+a)^{-m} dx \tag{A.25}$$

One can integrate by parts n times to get (omitting lengthy algebra):

$$\int_0^\infty x^n (x+a)^{-m} dx = \frac{\Gamma(n+1)\Gamma(m-n-1)}{\Gamma(m)} a^{-(m-n-1)} \tag{A.26}$$

This integral only converges if $m \rangle n+1$. We can quickly check the normalization of the F distribution using this. We have $n = \frac{\nu_1}{2} - 1$, $m = \frac{\nu_1+\nu_2}{2}$, and $a = \frac{\nu_2}{\nu_1}$. The integral is then:

$$\frac{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}{\Gamma(\frac{\nu_1+\nu_2}{2})} \left(\frac{\nu_2}{\nu_1}\right)^{-\frac{\nu_2}{2}} \tag{A.27}$$

Also, let us work out moments of the F distribution:

$$\langle F^p \rangle = \frac{\Gamma(\frac{\nu_1}{2}+p)\Gamma(\frac{\nu_2}{2}-p)}{\Gamma(\frac{\nu_1+\nu_2}{2})} \left(\frac{\nu_2}{\nu_1}\right)^{-\frac{\nu_2}{2}+p} \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_2}{2}} \tag{A.28}$$

$$= \frac{\Gamma(\frac{\nu_1}{2}+p)}{\Gamma(\frac{\nu_1}{2})} \frac{\Gamma(\frac{\nu_2}{2}-p)}{\Gamma(\frac{\nu_2}{2})} \left(\frac{\nu_2}{\nu_1}\right)^{p} \tag{A.29}$$

The expectation of F is ($p = 1$):

$$\frac{\nu_1}{2} \frac{1}{\frac{\nu_2}{2}-1} \frac{\nu_2}{\nu_1} = \frac{\nu_2}{\nu_2 - 2} \tag{A.30}$$

The second moment is

$$\left(\frac{\nu_1}{2}+1\right)\frac{\nu_1}{2}\frac{1}{\frac{\nu_2}{2}-1}\frac{1}{\frac{\nu_2}{2}-2}\left(\frac{\nu_2}{\nu_1}\right)^2 = \frac{1+\frac{2}{\nu_1}}{(1-\frac{2}{\nu_2})(1-\frac{4}{\nu_2})} \tag{A.31}$$

And the variance (calculated from the first two moments) is

$$\frac{1+\frac{2}{\nu_1}-\frac{6}{\nu_2}-\frac{4}{\nu_1\nu_2}}{\left(1-\frac{4}{\nu_2}\right)\left(1-\frac{2}{\nu_2}\right)^2} \tag{A.32}$$

To first order in the degrees of freedom (since second order correction to the CBI will be down by a factor of order $100^2 = 0.01\%$), the variance is

$$\frac{1+\frac{2}{\nu_1}-\frac{6}{\nu_2}}{1-\frac{4}{\nu_2}-\frac{4}{\nu_1}} \tag{A.33}$$

A.2 Combining Two Identical Data Points

The simplest case we can consider is that of two visibilities made up of several 8 second observations (y_1 and y_2) with scatter weights w_1 and w_2. The scatter weights are merely the reciprocal of the estimated variance on y_1 and y_2 calculated using the observed variance of their constituent 8 second observations. Let the true mean have been subtracted off, so $\langle y_i \rangle = 0$. Further, let the underlying variances be the same, and the number of degrees of freedom be the same, denoted by ν. The output visibility V is then

$$V = \frac{w_1 y_1 + w_2 y_2}{w_1 + w_2} = \frac{y_1}{1+\frac{w_2}{w_1}} + \frac{y_2}{1+\frac{w_1}{w_2}} \tag{A.34}$$

The the variance of V is just the sum of the variances of the terms since the y_i are uncorrelated with expectation value of zero. If we use the formula for the variance of a product where one expectation is zero (Equation A.3), we have

$$\mathrm{Var(V)} = \mathrm{Var(y_1)}\left\langle \frac{1}{(1+\frac{\mathrm{w_2}}{\mathrm{w_1}})^2} \right\rangle + \mathrm{Var(y_2)}\left\langle \frac{1}{(1+\frac{\mathrm{w_1}}{\mathrm{w_2}})^2} \right\rangle \tag{A.35}$$

Now it is a property of Gaussians that there is no correlation between the variance of y_i and its weight w_i. In other words, if I tell you that the individual data points that comprise a single y_i all happened to lie very close together, that does *not* imply that y_i is likely to lie closer to its expectation value. This is not generally the case: Consider a distribution that has a very sharp central peak with extremely broad tails with small area. If I draw a set of data points that all come from the central peak, then their mean will be close to the true mean, and the measured scatter will be small. If, however, some points from the tails are included in the data set, then the variance of the mean will be significantly increased, as will the scatter of the data points. The case of a boxcar distribution is even odder: if my points all come from the same small region, there is no reason to think that that small region is the center of the distribution. However, if the points are very widely spaced out, they must actually give a *better* estimate of the mean. The highest possible scatter variance is when the points are evenly placed on the two edges of the distribution - in which case the estimate of the mean is almost perfect. So, for a data set drawn from a given boxcar distribution, the worse the estimated error on the mean is, the better the estimate of the mean actually is, and the better the estimated error is, the worse the estimate of the mean actually is! The Gaussian is a distribution that precisely balances these two things so that the variance of the estimate is uncorrelated with the estimate of the variance. It is straightforward to demonstrate this empirically through Monte Carlo simulations.

Now the quantity $\frac{w_2}{w_1}$ is distributed precisely as an F distribution with degrees of freedom ν_1 and ν_2. In this case they're the same, so it's an $F_{\nu,\nu}$. The desired quantity by which we need to scale the variance is then of $\langle(1+F_{\nu,\nu})^{-2}\rangle$. Fortunately, for the case of equal degrees of freedom, we know exactly how to calculate this! Not only can we calculate the moments of F, we can also easily calculate powers of $\left(F+\frac{\nu_2}{\nu_1}\right)$. So, if $\nu_1=\nu_2$, we have

$$\langle(1+F_{\nu,\nu})^{-2}\rangle = \frac{1}{2}\left(\frac{1+\frac{\nu}{2}}{1+\nu}\right) = \frac{1}{4}\frac{\nu+2}{\nu+1} = \frac{1}{4}\left(1+\frac{1}{\nu+1}\right) \tag{A.36}$$

So, the total variance is (the factor of two $\frac{1}{4} \rightarrow \frac{1}{2}$ comes from the two identical terms):

$$\mathrm{Var(V)} = \mathrm{Var(y_i)}\frac{1}{2}\left(1 + \frac{1}{\nu+1}\right) \tag{A.37}$$

And the expectation value of the estimated final variance is

$$\left\langle \frac{\mathrm{Var(y_i)}}{w_1+w_2} \right\rangle \simeq \frac{\mathrm{Var(y_i)}}{2\langle w\rangle}\left(1 + \frac{2\mathrm{Var(w)}}{(2\langle w\rangle)^2}\right) \simeq \frac{\mathrm{Var(y_i)}}{2}(1-\frac{2}{\nu})(1+2\frac{2}{\nu}/4) \simeq \frac{\mathrm{Var(y_i)}}{2}(1-\frac{1}{\nu}) \tag{A.38}$$

where we have used the expansion for $\langle \frac{1}{x} \rangle$, and kept things to first order in ν. Empirically, I find that a better value is

$$\frac{\mathrm{Var(y_i)}}{2}\left(1 - \frac{1}{\nu+1}\right) \tag{A.39}$$

but, as Numerical Recipes says, if this makes a big difference, you are probably up to no good anyways. So, if we want the expectation of the variance of V to equal the expectation of our estimate, we need to scale by a factor of

$$\frac{1/2(1+1/(\nu+1))}{1/2(1-1/\nu+1)} \simeq 1 + \frac{2}{\nu} \tag{A.40}$$

This approximation works quite well for even fairly small degrees of freedom. For 20,000 pairs of averaged data points, I find that for 4 degrees of freedom, the predicted factor is 1.5, empirical 1.502; for 9 degrees of freedom, the predicted factor is 1.222, empirical 1.217, and for 29 degrees of freedom, the prediced factor is 1.069, and the empirical is 1.069.

A.3 Combining Many Identical Data Points

Let us now take the limit in which we combine many identically distributed data points with scatter weights. We will again keep terms only to first order in ν, and neglect terms down by n from the leading term, where n is the number of scans we combine. First, find the true variance of the

estimator:

$$\mathrm{Var(V)} = \mathrm{Var}\left(\frac{\sum \mathrm{w_i y_i}}{\sum \mathrm{w_i}}\right) \tag{A.41}$$

Again, the y_i are uncorrelated, so this expression becomes

$$\mathrm{Var(V)} = \mathrm{nVar}\left(\frac{\mathrm{w_1 y_1}}{\sum \mathrm{w_i}}\right) = \mathrm{nVar\,(y_1)}\left\langle\left(\frac{\mathrm{w_1}}{\sum \mathrm{w_i}}\right)^2\right\rangle \tag{A.42}$$

Let us now work only on the expectation term. If the number of data points is large enough, the correlation between the numerator and denominator becomes negligible, and the expectation of the product becomes the product of the expectations:

$$\left\langle\left(\frac{w_1}{\sum w_i}\right)^2\right\rangle = \langle w_1^2\rangle\left\langle\left(\sum w_i\right)^{-2}\right\rangle \tag{A.43}$$

We have already calculated the first term - $\langle w^2\rangle \simeq 1 + \frac{6}{\nu}$. We can calculate the second term using the power series expansion for expectations:

$$\left\langle\left(\sum w_i\right)^{-2}\right\rangle = \left\langle\sum w_i\right\rangle^{-2} + \frac{6}{2}\mathrm{Var}\left(\sum \mathrm{w_i}\right)\left\langle\sum \mathrm{w_i}\right\rangle^{-4} + ... \tag{A.44}$$

Now the variance of the sum is the sum of the variances, so it depends on n^1. Also, the expectation of the sum is the sum of the expectations, so it also depends on n^1. This leaves an n dependency of the first term of n^{-2}, and of the second term of n^{-3}, so the second term becomes negligible as n becomes large. Now let us actually calculate the expectation:

$$\left\langle\sum w_i\right\rangle = n\langle w\rangle \simeq n(1 + \frac{2}{\nu}) \tag{A.45}$$

$$\left\langle\sum w_i\right\rangle^{-2} \simeq n^{-2}(1 + \frac{2}{\nu})^{-2} \simeq n^{-2}\left(1 - \frac{4}{\nu}\right) \tag{A.46}$$

And the final variance of the estimator is then

$$\mathrm{Var(V)} \simeq \mathrm{nVar(y)} \left(1 + \frac{6}{\nu}\right) \mathrm{n}^{-2} \left(1 - \frac{4}{\nu}\right) \tag{A.47}$$

The expectation of our estimate of the variance is

$$\mathrm{Var(y)} \left\langle \frac{1}{\sum \mathrm{w_i}} \right\rangle \mathrm{Var(y)} \simeq \mathrm{n}^{-1} \langle \mathrm{w} \rangle^{-1} \simeq \mathrm{Var(y)n}^{-1} \left(1 - \frac{2}{\nu}\right) \tag{A.48}$$

And the factor by which we have misestimated is the ratio of the two estimates:

$$\simeq n^{-1}\mathrm{Var(y)} \left(1 + \frac{2}{\nu}\right) / \mathrm{Var(y)n}^{-1} \left(1 - \frac{2}{\nu}\right) \simeq 1 + \frac{4}{\nu} \tag{A.49}$$

It is this first-order factor of $1 + \frac{4}{\nu}$ to which we expect the noise bias to converge for many scans.

Appendix B

CMB Weighting in SZ Cluster Observations

In addition to observations of the CMB, a major campaign of the CBI has been a survey of clusters at $z < 0.1$ with the goal of measuring H_0 to an accuracy of 10% independently of other, more traditional methods. This appendix describes simulations carried out by Patricia Udomprasert using algorithms developed by the author that minimize the impact of the CMB on the cluster observations. For our nearby clusters, the CMB is a major source of noise. The algorithm and results are more fully discussed in Udomprasert (2003).

The Sunyaev-Zeldovich (SZ) effect is the heating of CMB photons as they scatter off of hot gas in galaxy clusters. It is a rich source of information about galaxy clusters (see, *e.g.*, Rephaeli, 2002, for a recent review), especially when combined with other sources of information about the hot gas, such as X-Ray observations. Observations of the SZ effect in nearby ($z \sim 0.1 - 0.2$) clusters are especially useful since the X-Ray data are of better quality, and a fixed angular resolution leads to better physical resolution in closer clusters. However, the CMB is a major contaminant in observations of these nearby clusters. It is, in fact, the single biggest contaminant for the sample of clusters at $z < 0.1$ observed by the CBI, with typical CMB signals of 55μK compared to cluster signals of a few hundred μK. The CMB is much less of a problem for use in more distant (and hence smaller in angular size) clusters as the power in the CMB falls rapidly on decreasing scales. The best way to separate the CMB signal from the SZ signal is to have multi-frequency observations spanning the SZ null at 217 GHz. Then one can use the fact that the SZ effect appears as a hole in the CMB at

frequencies below the null and a bright spot in the CMB at frequencies above it. This is clearly an observationally expensive proposition, and since no high-frequency observations of our clusters are available, some other way must be found of treating the CMB. The spectral coverage of the CBI is of limited use here since the difference in frequency behavior between the SZ effect and the CMB are of no consequence in the CBI bandpass.

To measure the SZ effect, one usually generates a model on the sky (typically an isothermal β model, though this discussion applies equally well to other parameterizations of cluster structure), predicts the values that visibilities have under the assumed model, and compares those predictions to the actual data. The best-fit model is the one with the minimum value of χ^2. Unlike measuring the CMB power spectrum, models generate predicted values for the visibilities rather than predicted variances, so mis-estimates of the noise lead to incorrect error bars rather than biased models. There are some simplistic ways of treating the CMB when fitting clusters. The easiest is to simply ignore the CMB, since the cluster parameters will be unbiased. A better way is to estimate the noise on each visibility from the CMB and add it in quadrature to the thermal noise in the visibility before calculating χ^2. This gives better results than ignoring the CMB, but is not optimal because it does not correctly take into account the fact that nearby visibilities have correlated CMB values. This means that uncertainties will be larger than they need to be, and error estimates will still be incorrect.

The correct way to treat the CMB is to transform the visibilities into a set of estimators where both the thermal noise and CMB signal are uncorrelated. Once we have done this, the CMB and thermal noise can be combined into an effective noise, and since each point is independent, a χ^2 fit is simple to carry out. Furthermore, the χ^2 values reflect the true goodness-of-fit, and so errors can be accurately estimated. To do this requires several steps. The first is to calculate the covariance matrix of the data given a known (from outside sources) CMB spectrum. Then divide each visibility by its noise, applying the same scaling to the covariance matrix and to the model visibilities (a whitening transform of Section 5.1). Once we have done this, the noise matrix becomes the identity matrix. This is important because a rotation of the identity matrix remains the identity matrix.

Were the noises unequal, then if we rotated the noise matrix, it would no longer be diagonal. This means that the thermal noise would no longer be uncorrelated between estimators. Next, one finds the eigenvalues and eigenvectors of the whitened CMB covariance matrix. Finally, one uses the eigenvector matrix to rotate both the covariance matrix, the noise matrix, the data, and the model data. The signal from the CMB is now uncorrelated between estimators, and the thermal noise remains so. We can now directly calculate χ^2 for the model:

$$\chi^2 = \sum \frac{(x_i - m_i)^2}{1 + \lambda_i} \tag{B.1}$$

where x_i is the value of the i^{th} rotated estimator, m_i is the value for the i^{th} estimator predicted by the model, and $1 + \lambda_i$ is the variance of the i^{th} estimator (because we have whitened the thermal noise, the thermal noise in each estimator is unity). This method is optimal since we have used all the information in the CMB covariance matrix, which fully describes the properties of the CMB (assuming Gaussianity). The results of the simulations are described in Table B.1 which compares the uncertainty in $h^{-1/2}$ (which is proportional to the central temperature decrement) when measured ignoring the CMB signal to the uncertainty when using optimal weighting. These errors in these simulations are representative of the data already taken by the CBI on the clusters named in Table B.1. The net effect is to reduce the ensemble uncertainty in $h^{-1/2}$ from 0.178 to 0.130, a reduction of 27%, which should lead to a reduction in uncertainty on H_0 of 47%. We used a total of 1000 simulations for each cluster, with a standard ΛCDM model for the CMB with $h = 0.7$.

To visualize how the weighting scheme works, picture the behavior of both the model and the CMB in the UV plane. If the cluster were a point source, it would have equal amplitude in all baselines. In reality, clusters have finite size, so the response of visibilities to the cluster will be uniform and large for baselines much shorter than the inverse of the cluster size in radians. Baselines much longer than the inverse cluster size resolve the cluster, leading to a reduction in signal. The detailed behavior of the falloff in signal with baseline length is dependent on the detailed shape of the cluster. As the size of the cluster shrinks, longer and longer baselines are expected to retain

Table B.1 Comparison of Predicted Errors in $h^{-1/2}$for no Weighting and Eigenmode Weighting

Cluster	β-FWHM	$\sigma_{\rm nowt}$	$\sigma_{\rm eigwt}$
A85	8.80	0.373	0.292
A399	12.54	0.423	0.379
A401	8.58	0.272	0.210
A478	3.77	0.251	0.183
A754	16.96	0.291	0.264
A1651	6.68	0.437	0.324
A2597	1.92	0.902	0.589
CMB error in $h^{-1/2}$for sample		0.178	0.130
H_0 for sample with uncertainty due to CMB		67^{+25}_{-16}	67^{+17}_{-12}

Results of simulations showing increase in accuracy in fit parameters when using our transformed estimators compared to ignoring the CMB. The only free parameter is the cluster central temperature decrement, with the location and shape of the cluster determined externally (such as from X-Ray data). The first column is the cluster simulated, the second is the cluster FWHM in arcmin, the third is the scatter in $h^{-1/2}$ with no CMB weighting, the fourth is the scatter in $h^{-1/2}$ using our weighting. The central decrement is proportional to $h^{-1/2}$, so we quote the results in terms of $h^{-1/2}$ as it directly relates to our cosmological constraints. The net effect of the weighting scheme is to reduce the uncertainty in $h^{-1/2}$ from 0.178 to 0.130. Reprinted from Udomprasert (2003).

good sensitivity to the cluster. The CMB is a set of independent patches in the UV plane with size set by the Fourier transform of the primary beam, and amplitude set by the power spectrum $\mathcal{C}_\ell$ at the distance from the origin of the patch in question. This means that the CMB noise in each patch usually falls quickly with increasing baseline length, so the most useful visibilities are those from long enough baselines to have low CMB response but short enough not to resolve the cluster. Small clusters have more of these visibilities than large clusters, where the SZ signal can fall off almost as quickly at the CMB. In addition to the small clusters being less corrupted by the CMB, we expect the weighting scheme to improve the fits to the small clusters more than those to the large clusters since the small clusters have high signal visibilities relatively unaffected by the CMB that can be preferentially used, whereas the large clusters have no such visibilities. This behavior is seen in Table B.1, where the improvement in small clusters is indeed more than that in the large clusters.

It is worth noting that we can think of the weighting scheme as using a single noisy estimate of the cluster signal in each independent patch. So the uncertainty in the cluster decrement is approximately the SNR in each patch divided by the square root of the number of independent patches. If the size of the independent patches were shrunk we would have more of them in a given

region of the UV plane, and therefore a better determination of the cluster properties. What is the size of the patches? It is the Fourier transform of the observed area. For a single pointing, this is the size of the Fourier transform of the primary beam, but if we mapped out a larger area, it would be the Fourier transform of the entire survey region. A larger map means a smaller Fourier transform, which leads to the counterintuitive result that our measurement of the cluster becomes more precise as we observe more blank sky around it! Essentially, surveying a larger region allows us to better characterize the behavior of the CMB underneath the cluster. This is a potentially powerful (and perhaps the only) way of increasing the accuracy with which sensitive instruments working in narrow frequency ranges can measure cluster structure.

Bibliography

Abramowitz, M. & Stegun, I. A. 1965, Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables (Dover Publications)

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., & Sorensen, D. 1999, LAPACK Users' Guide, 3rd edn. (Philadelphia, PA: Society for Industrial and Applied Mathematics)

Bennett, C. L., Hill, R. S., Hinshaw, G., Nolta, M. R., Odegard, N., Page, L., Spergel, D. N., Weiland, J. L., Wright, E. L., Halpern, M., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., & Wollack, E. 2003, ApJS, 148, 97

Benoît, A., Ade, P., Amblard, A., Ansari, R., Aubourg, É., Bargot, S., Bartlett, J. G., Bernard, J.-P., Bhatia, R. S., Blanchard, A., Bock, J. J., Boscaleri, A., Bouchet, F. R., Bourrachot, A., Camus, P., Couchot, F., de Bernardis, P., Delabrouille, J., Désert, F.-X., Doré, O., Douspis, M., Dumoulin, L., Dupac, X., Filliatre, P., Fosalba, P., Ganga, K., Gannaway, F., Gautier, B., Giard, M., Giraud-Héraud, Y., Gispert, R., Guglielmi, L., Hamilton, J.-C., Hanany, S., Henrot-Versillé, S., Kaplan, J., Lagache, G., Lamarre, J.-M., Lange, A. E., Macías-Pérez, J. F., Madet, K., Maffei, B., Magneville, C., Marrone, D. P., Masi, S., Mayet, F., Murphy, A., Naraghi, F., Nati, F., Patanchon, G., Perrin, G., Piat, M., Ponthieu, N., Prunet, S., Puget, J.-L., Renault, C., Rosset, C., Santos, D., Starobinsky, A., Strukov, I., Sudiwala, R. V., Teyssier, R., Tristram, M., Tucker, C., Vanel, J.-C., Vibert, D., Wakui, E., & Yvon, D. 2003, A&A, 399, L19

Bond, J. R. 1996, in Les Houches Lectures, 469–674

Bond, J. R., Contaldi, C., Pogosyan, D., Mason, B., Myers, S., Pearson, T., Pen, U., Prunet, S., Readhead, T., & Sievers, J. 2002a, in American Institute of Physics Conference Series, 15–33

Bond, J. R., Contaldi, C. R., Pen, U. L., Pogosyan, D., Prunet, S., Ruetalo, M. I., Wadsley, J. W., Zhang, P., Mason, B. S., Myers, S. T., Pearson, T. J., Readhead, A. C., Sievers, J. L., & Udomprasert, P. S. 2002b, astro-ph/0205386

Bond, J. R. & Efstathiou, G. 1984, ApJ, 285, L45

—. 1987, MNRAS, 226, 655

Bond, J. R., Jaffe, A. H., & Knox, L. 1998, Phys. Rev. D, 57, 2117

—. 2000, ApJ, 533, 19

Borrill, J. 1999, astro-ph/9911389

Burles, S., Nollett, K. M., Truran, J. W., & Turner, M. S. 1999, Physical Review Letters, 82, 4176

Cartwright, J. K. 2002, PhD thesis, California Institute of Technology

Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., & Broderick, J. J. 1998, AJ, 115, 1693

Dawson, K. S., Holzapfel, W. L., Carlstrom, J. E., Joy, M., LaRoque, S. J., Miller, A. D., & Nagai, D. 2002, ApJ, 581, 86

de Bernardis, P., Ade, P. A. R., Bock, J. J., Bond, J. R., Borrill, J., Boscaleri, A., Coble, K., Crill, B. P., De Gasperis, G., Farese, P. C., Ferreira, P. G., Ganga, K., Giacometti, M., Hivon, E., Hristov, V. V., Iacoangeli, A., Jaffe, A. H., Lange, A. E., Martinis, L., Masi, S., Mason, P. V., Mauskopf, P. D., Melchiorri, A., Miglio, L., Montroy, T., Netterfield, C. B., Pascale, E., Piacentini, F., Pogosyan, D., Prunet, S., Rao, S., Romeo, G., Ruhl, J. E., Scaramuzzi, F., Sforna, D., & Vittorio, N. 2000, Nature, 404, 955

Dicke, R. H., Peebles, P. J. E., Roll, P. G., & Wilkinson, D. T. 1965, ApJ, 142, 414

Fixsen, D. J., Cheng, E. S., Cottingham, D. A., Eplee, R. E., Isaacman, R. B., Mather, J. C., Meyer, S. S., Noerdlinger, P. D., Shafer, R. A., Weiss, R., Wright, E. L., Bennett, C. L., Boggess, N. W., Kelsall, T., Moseley, S. H., Silverberg, R. F., Smoot, G. F., & Wilkinson, D. T. 1994, ApJ, 420, 445

Freedman, W. L., Madore, B. F., Gibson, B. K., Ferrarese, L., Kelson, D. D., Sakai, S., Mould, J. R., Kennicutt, R. C., Ford, H. C., Graham, J. A., Huchra, J. P., Hughes, S. M. G., Illingworth, G. D., Macri, L. M., & Stetson, P. B. 2001, ApJ, 553, 47

Fukugita, M., Sugiyama, N., & Umemura, M. 1990, ApJ, 358, 28

Grainge, K., Carreira, P., Cleary, K., Davies, R. D., Davis, R. J., Dickinson, C., Genova-Santos, R., Gutiérrez, C. M., Hafez, Y. A., Hobson, M. P., Jones, M. E., Kneissl, R., Lancaster, K., Lasenby, A., Leahy, J. P., Maisinger, K., Pooley, G. G., Rebolo, R., Rubiño-Martin, J. A., Sosa Molina, P. J., Ödman, C., Rusholme, B., Saunders, R. D. E., Savage, R., Scott, P. F., Slosar, A., Taylor, A. C., Titterington, D., Waldram, E., Watson, R. A., & Wilkinson, A. 2003, MNRAS, 341, L23

Halverson, N. W., Leitch, E. M., Pryke, C., Kovac, J., Carlstrom, J. E., Holzapfel, W. L., Dragovan, M., Cartwright, J. K., Mason, B. S., Padin, S., Pearson, T. J., Readhead, A. C. S., & Shepherd, M. C. 2002, ApJ, 568, 38

Hanany, S., Ade, P., Balbi, A., Bock, J., Borrill, J., Boscaleri, A., de Bernardis, P., Ferreira, P. G., Hristov, V. V., Jaffe, A. H., Lange, A. E., Lee, A. T., Mauskopf, P. D., Netterfield, C. B., Oh, S., Pascale, E., Rabii, B., Richards, P. L., Smoot, G. F., Stompor, R., Winant, C. D., & Wu, J. H. P. 2000, ApJ, 545, L5

Haslam, C. G. T., Klein, U., Salter, C. J., Stoffel, H., Wilson, W. E., Cleary, M. N., Cooke, D. J., & Thomasson, P. 1981, A&A, 100, 209

Haslam, C. G. T., Stoffel, H., Salter, C. J., & Wilson, W. E. 1982, A&AS, 47, 1

Hinshaw, G., Spergel, D. N., Verde, L., Hill, R. S., Meyer, S. S., Barnes, C., Bennett, C. L.,

Halpern, M., Jarosik, N., Kogut, A., Komatsu, E., Limon, M., Page, L., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L. 2003, ApJS, 148, 135

Hivon, E., Górski, K. M., Netterfield, C. B., Crill, B. P., Prunet, S., & Hansen, F. 2002, ApJ, 567, 2

Hu, W., Scott, D., Sugiyama, N., & White, M. 1995, Phys. Rev. D, 52, 5498

Knox, L. 1999, Phys. Rev. D, 60, 103516

Kogut, A., Spergel, D. N., Barnes, C., Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Limon, M., Meyer, S. S., Page, L., Tucker, G. S., Wollack, E., & Wright, E. L. 2003, ApJS, 148, 161

Komatsu, E. & Seljak, U. 2002, MNRAS, 336, 1256

Kovac, J. M., Leitch, E. M., Pryke, C., Carlstrom, J. E., Halverson, N. W., & Holzapfel, W. L. 2002, Nature, 420, 772

Lange, A. E., Ade, P. A., Bock, J. J., Bond, J. R., Borrill, J., Boscaleri, A., Coble, K., Crill, B. P., de Bernardis, P., Farese, P., Ferreira, P., Ganga, K., Giacometti, M., Hivon, E., Hristov, V. V., Iacoangeli, A., Jaffe, A. H., Martinis, L., Masi, S., Mauskopf, P. D., Melchiorri, A., Montroy, T., Netterfield, C. B., Pascale, E., Piacentini, F., Pogosyan, D., Prunet, S., Rao, S., Romeo, G., Ruhl, J. E., Scaramuzzi, F., & Sforna, D. 2001, Phys. Rev. D, 63, 42001

Lee, A. T., Ade, P., Balbi, A., Bock, J., Borrill, J., Boscaleri, A., de Bernardis, P., Ferreira, P. G., Hanany, S., Hristov, V. V., Jaffe, A. H., Mauskopf, P. D., Netterfield, C. B., Pascale, E., Rabii, B., Richards, P. L., Smoot, G. F., Stompor, R., Winant, C. D., & Wu, J. H. P. 2001, ApJ, 561, L1

Leitch, E. M., Readhead, A. C. S., Pearson, T. J., & Myers, S. T. 1997, ApJ, 486, L23+

Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473

Lineweaver, C. H. 1997, in Microwave Background Anistropies, 69–

Lyth, D. H. & Riotto, A. A. 1999, Phys. Rep., 314, 1

Mason, B. S., Pearson, T. J., Readhead, A. C. S., Shepherd, M. C., Sievers, J., Udomprasert, P. S., Cartwright, J. K., Farmer, A. J., Padin, S., Myers, S. T., Bond, J. R., Contaldi, C. R., Pen, U., Prunet, S., Pogosyan, D., Carlstrom, J. E., Kovac, J., Leitch, E. M., Pryke, C., Halverson, N. W., Holzapfel, W. L., Altamirano, P., Bronfman, L., Casassus, S., May, J., & Joy, M. 2003, ApJ, 591, 540

Mather, J. C., Cheng, E. S., Cottingham, D. A., Eplee, R. E., Fixsen, D. J., Hewagama, T., Isaacman, R. B., Jensen, K. A., Meyer, S. S., Noerdlinger, P. D., Read, S. M., Rosen, L. P., Shafer, R. A., Wright, E. L., Bennett, C. L., Boggess, N. W., Hauser, M. G., Kelsall, T., Moseley, S. H., Silverberg, R. F., Smoot, G. F., Weiss, R., & Wilkinson, D. T. 1994, ApJ, 420, 439

Miller, A. D., Caldwell, R., Devlin, M. J., Dorwart, W. B., Herbig, T., Nolta, M. R., Page, L. A., Puchalla, J., Torbet, E., & Tran, H. T. 1999, ApJ, 524, L1

Mood, A. M., Graybill, F. A., & Boes, D. C. 1974, Introduction to the Theory of Statistics, 3rd Edition (McGraw-Hill)

Myers, S. T., Contaldi, C. R., Bond, J. R., Pen, U.-L., Pogosyan, D., Prunet, S., Sievers, J. L., Mason, B. S., Pearson, T. J., Readhead, A. C. S., & Shepherd, M. C. 2003, ApJ, 591, 575

Netterfield, C. B., Ade, P. A. R., Bock, J. J., Bond, J. R., Borrill, J., Boscaleri, A., Coble, K., Contaldi, C. R., Crill, B. P., de Bernardis, P., Farese, P., Ganga, K., Giacometti, M., Hivon, E., Hristov, V. V., Iacoangeli, A., Jaffe, A. H., Jones, W. C., Lange, A. E., Martinis, L., Masi, S., Mason, P., Mauskopf, P. D., Melchiorri, A., Montroy, T., Pascale, E., Piacentini, F., Pogosyan, D., Pongetti, F., Prunet, S., Romeo, G., Ruhl, J. E., & Scaramuzzi, F. 2002, ApJ, 571, 604

Oh, S. P., Cooray, A., & Kamionkowski, M. 2003, MNRAS, 342, L20

Oh, S. P., Spergel, D. N., & Hinshaw, G. 1999, ApJ, 510, 551

Olive, K. A., Steigman, G., & Walker, T. P. 2000, Phys. Rep., 333, 389

Padin, S., Cartwright, J. K., Joy, M., & Meitzler, J. C. 2000, IEEE Trans. Antennas Propagat., 48, 836

Padin, S., Cartwright, J. K., Mason, B. S., Pearson, T. J., Readhead, A. C. S., Shepherd, M. C., Sievers, J., Udomprasert, P. S., Holzapfel, W. L., Myers, S. T., Carlstrom, J. E., Leitch, E. M., Joy, M., Bronfman, L., & May, J. 2001a, ApJ, 549, L1

Padin, S., Cartwright, J. K., Shepherd, M. C., Yamasaki, J. K., & Holzapfel, W. L. 2001b, IEEE Trans. Instrum. Meas., 50, 1234

Padin, S., Shepherd, M. C., Cartwright, J. K., Keeney, R. G., Mason, B. S., Pearson, T. J., Readhead, A. C. S., Schaal, W. A., Sievers, J., Udomprasert, P. S., Yamasaki, J. K., Holzapfel, W. L., Carlstrom, J. E., Joy, M., Myers, S. T., & Otarola, A. 2002, PASP, 114, 83

Peacock, J. A. 1999, Cosmological Physics (Cambridge University Press)

Pearson, T. J., Mason, B. S., Readhead, A. C. S., Shepherd, M. C., Sievers, J. L., Udomprasert, P. S., Cartwright, J. K., Farmer, A. J., Padin, S., Myers, S. T., Bond, J. R., Contaldi, C. R., Pen, U.-L., Prunet, S., Pogosyan, D., Carlstrom, J. E., Kovac, J., Leitch, E. M., Pryke, C., Halverson, N. W., Holzapfel, W. L., Altamirano, P., Bronfman, L., Casassus, S., May, J., & Joy, M. 2003, ApJ, 591, 556

Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., & The Supernova Cosmology Project. 1999, ApJ, 517, 565

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in C: The Art of Scientific Computing (Cambridge University Press)

Rephaeli, Y. 2002, Space Science Reviews, 100, 61

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., & Tonry, J. 1998, AJ, 116, 1009

Ruhl, J. E., Ade, P. A. R., Bock, J. J., Bond, J. R., Borrill, J., Boscaleri, A., Contaldi, C. R., Crill, B. P., de Bernardis, P., De Troia, G., Ganga, K., Giacometti, M., Hivon, E., Hristov, V. V., Iacoangeli, A., Jaffe, A. H., Jones, W. C., Lange, A. E., Masi, S., Mason, P., Mauskopf, P. D., Melchiorri, A., Montroy, T., Netterfield, C. B., Pascale, E., Piacentini, F., Pogosyan, D., Polenta, G., Prunet, S., & Romeo, G. 2002, ArXiv Astrophysics e-prints

Runyan, M. C., Ade, P. A. R., Bock, J. J., Bond, J. R., Cantalupo, C., Contaldi, C. R., Daub, M. D., Goldstein, J. H., Gomez, P. L., Holzapfel, W. L., Kuo, C. L., Lange, A. E., Lueker, M., Newcomb, M., Peterson, J. B., Pogosyan, D., Romer, A. K., Ruhl, J., Torbet, E., & Woolsey, D. 2003, astro-ph/0305553

Sachs, R. K. & Wolfe, A. M. 1967, ApJ, 147, 73

Scott, P. F., Carreira, P., Cleary, K., Davies, R. D., Davis, R. J., Dickinson, C., Grainge, K., Gutiérrez, C. M., Hobson, M. P., Jones, M. E., Kneissl, R., Lasenby, A., Maisinger, K., Pooley, G. G., Rebolo, R., Rubiño-Martin, J. A., Sosa Molina, P. J., Rusholme, B., Saunders, R. D. E., Savage, R., Slosar, A., Taylor, A. C., Titterington, D., Waldram, E., Watson, R. A., & Wilkinson, A. 2003, MNRAS, 341, 1076

Seljak, U. & Zaldarriaga, M. 1996, ApJ, 469, 437

Sievers, J. 2004, in prep

Sievers, J. L., Bond, J. R., Cartwright, J. K., Contaldi, C. R., Mason, B. S., Myers, S. T., Padin, S., Pearson, T. J., Pen, U.-L., Pogosyan, D., Prunet, S., Readhead, A. C. S., Shepherd, M. C., Udomprasert, P. S., Bronfman, L., Holzapfel, W. L., & May, J. 2003, ApJ, 591, 599

Silk, J. 1968, ApJ, 151, 459

Smoot, G. F., Bennett, C. L., Kogut, A., Wright, E. L., Aymon, J., Boggess, N. W., Cheng, E. S., de Amici, G., Gulkis, S., Hauser, M. G., Hinshaw, G., Jackson, P. D., Janssen, M., Kaita, E., Kelsall, T., Keegstra, P., Lineweaver, C., Loewenstein, K., Lubin, P., Mather, J. andMeyer, S. S., Moseley, S. H., Murdock, T., Rokke, L., Silverberg, R. F., Tenorio, L., Weiss, R., & Wilkinson, D. T. 1992, ApJ, 396, L1

Spergel, D. N., Verde, L., Peiris, H. V., Komatsu, E., Nolta, M. R., Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Page, L., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L. 2003, ApJS, 148, 175

Tegmark, M., Hamilton, A. J. S., Strauss, M. A., Vogeley, M. S., & Szalay, A. S. 1998, ApJ, 499, 555

Tegmark, M. & Zaldarriaga, M. 2000, Physical Review Letters, 85, 2240

Tytler, D., O'Meara, J. M., Suzuki, N., & Lubin, D. 2000, Physica Scripta Volume T, 85, 12

Udomprasert, P. S. 2003, PhD thesis, California Institute of Technology

Vittorio, N. & Silk, J. 1984, ApJ, 285, L39

—. 1992, ApJ, 385, L9

White, M. 2001, ApJ, 555, 88

White, M., Carlstrom, J. E., Dragovan, M., & Holzapfel, W. L. 1999, ApJ, 514, 12

White, M., Scott, D., & Silk, J. 1994, ARA&A, 32, 319