Analysis of Cosmic Microwave Background Polarisation

A thesis submitted to The University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2011

Michael Preece
School of Physics and Astronomy
Contents

List of Figures ........................................................................... 7

List of Tables ............................................................................ 13

Abstract ...................................................................................... 15
Declaration .................................................................................. 17
Copyright Statement ..................................................................... 18

1 Introduction ............................................................................. 19

1.1 Definition of CMB temperature power spectrum ..................... 21
1.2 Polarisation in the CMB .......................................................... 22
1.2.1 Definition of polarisation ..................................................... 22
1.2.2 Polarisation position angles ................................................. 23
1.2.3 Spin spherical harmonics and analysis of CMB polarisation .. 23
1.2.4 Properties of E and B modes ............................................... 25
1.3 Partial-sky observations and foreground masking .................... 28
1.4 Errors in CMB observations ..................................................... 30
1.4.1 Basics of real CMB observations ......................................... 30
1.4.2 Cosmic variance ................................................................. 31
1.4.3 Effects of noise and beams .................................................. 32
1.5 Inflationary theory and CMB anisotropies ............................... 34
1.5.1 The theory of inflation ......................................................... 34
1.5.2 Inflationary power spectra and types of model ................. 36
## CONTENTS

1.5.3 Fluctuations in the CMB ........................................... 38
1.5.4 Generation of polarisation in the CMB .......................... 42
1.5.5 Anisotropies from tensor perturbations ......................... 43
1.5.6 B-mode power spectrum ......................................... 44

2 Simulations of the detection of B-modes using the Pure-$C_\ell$ method 47
2.1 Monte-Carlo simulations of the standard pseudo-$C_\ell$ method .... 49
2.1.1 Generation of masked CMB maps ............................... 50
2.1.2 Deconvolution of masked CMB maps ............................ 52
2.1.3 Computation of statistics for output power spectrum .......... 57
2.1.4 Addition of noise into simulations .............................. 57
2.2 Issues with the pseudo-$C_\ell$ method ............................ 59
2.3 Removing problems with E-B mixing ............................... 64
2.3.1 Definition of pure-$C_\ell$ .................................... 64
2.3.2 Implementation of pure-$C_\ell$ algorithm ......................... 70
2.3.3 Apodisation of pure-$C_\ell$ masks ............................... 75
2.4 Simulations of pure and standard pseudo-$C_\ell$ ..................... 81
2.5 Conclusions .................................................. 85

3 Survey strategies and future instruments 89
3.1 Theoretical errors for masked CMB measurements ................. 90
3.1.1 Theory behind optimal errors .................................. 90
3.1.2 Realistic noise for CMB experiments .......................... 92
3.2 Effects of survey size on pure-$C_\ell$ errors ......................... 98
3.3 Calculation of variance for pseudo-$C_\ell$ methods ................. 100
3.3.1 E-B mixing in the standard pseudo-$C_\ell$ approach .......... 100
3.3.2 Calculation of error for pure-$C_\ell$ ........................... 103
3.3.3 Analytic expression for errors ................................ 107
3.4 Conclusions .................................................. 112

ANALYSIS OF CMB POLARISATION
## 4 Point source removal

4.1 Effect of point sources on B-mode polarisation measurements
   - 4.1.1 Generation of point-source maps
   - 4.1.2 Simple study of point-source effects

4.2 Large-scale effects of point-source removal

4.3 Effect of point-sources on B-mode detection

4.4 Analytic Calculation of point-source errors
   - 4.4.1 Comparison of analytic formula to simulations

4.5 Conclusions

## 5 Null test of CMB polarisation data using position angles

5.1 Basic principles of the method

5.2 Properties of the probability distribution
   - 5.2.1 Distribution of Q and U using a multivariate Gaussian
   - 5.2.2 Uncorrelated pixels
   - 5.2.3 Correlated pixels

5.3 Effect of systematic errors
   - 5.3.1 Definition and properties of systematic errors
   - 5.3.2 Detection of systematic errors in experiments
   - 5.3.3 Simulations of uncorrelated noise
   - 5.3.4 Signal-dominated regime

5.4 Application to WMAP 5
   - 5.4.1 Properties of WMAP polarisation measurements
   - 5.4.2 Simulations of WMAP noise
   - 5.4.3 WMAP without foreground removal
   - 5.4.4 WMAP with foreground removal

5.5 Effect of systematic errors on measurements of B-modes

5.6 Chi-squared for correlated random variables
   - 5.6.1 Ordinary $\chi^2$ distribution
## CONTENTS

5.6.2 Case of equal eigenvalues ................................. 179  
5.6.3 Solution in general case ................................. 181  
5.6.4 Alternate method for deriving formulae ................. 185  
5.6.5 Extension of method to non-central chi-squared distribution . 185  
5.7 Mean of the sum of Bernoulli distributed variables ............... 190  
5.8 Variance of number of pixels in histogram bins ................................. 193  
  5.8.1 Formulae for variance in uncorrelated and correlated case ....... 193  
  5.8.2 Uncorrelated case ........................................ 194  
  5.8.3 Extension to correlated pixels ................................ 195  
5.9 Determination of probability distribution of $\alpha$ ......................... 198  
  5.9.1 Calculation of covariance matrix for a Gaussian CMB ........... 199  
5.10 Probability distribution for systematic errors ......................... 203  
5.11 Conclusions ................................................. 206  

6 Conclusions and future work ........................................ 209  

References ......................................................... 213
# List of Figures

1.1 Pure E and B mode patterns for m=0 centred on the North Pole. . . . 26

1.2 EE (left) and BB (right) power spectra for \( r = 0.1 \), calculated using CAMB with the parameters given at the beginning of Chapter 2. On the BB spectrum, the black line is the primordial power spectrum, and the dashed red line is the gravitational lensing component. . . . . . . 46

2.1 QUIJOTE mask in Galactic co-ordinates, using a Mollweide projection. Red areas are unmasked, blue areas are masked. . . . . . . . . . 60

2.2 \( \Delta C^{TT}_\ell / C^{TT}_\ell \) (left) and \( \Delta C^{BB}_\ell / C^{BB}_\ell \) (right) without binning (top) and with a bin size of \( \Delta \ell = 30 \) (bottom) for the QUIJOTE mask (red line), an azimuthal mask with the same \( f_{\text{sky}} \) (blue line) and the theoretical errors as given by Equation (3.2) (black line) . . . . . . . . . . . . . . . . . . . . . . 61

2.3 Mask showing the 20% of the sky with the lowest foreground contamination, in Galactic coordinates with a Mollweide projection. . . . . . 63

2.4 Comparison between errors on BB spectrum for a mask based on Carretti et al. (2006) (right) and an azimuthal mask with the same \( f_{\text{sky}} \) (left), for all values of \( \ell \) up to \( \ell = 20 \) (top) and for multiples of 10 up to \( \ell = 256 \) (bottom) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 63
LIST OF FIGURES

2.5 Detectability of B-modes given measurements up to a maximum $l$ of 100 (left) and 500 (right) assuming no noise or lensing for a 50 square-degree (top), 1000 square-degree (middle) and half-sky (bottom) spherical cap mask as a function of the apodisation length, $r_c$, for a Cosine (blue solid) and Gaussian (black dashed) apodisation function . . . . 78

2.6 Simulated pure-$C_\ell$ (left) and pseudo-$C_\ell$ (right) spectra with (top) and without (bottom) E-modes for a 10000 square-degree spherical cap mask. 82

2.7 Simulated pure-$C_\ell$ (left) and pseudo-$C_\ell$ (right) spectra with (top) and without (bottom) E-modes for a 1000 square-degree spherical cap mask. 83

2.8 Simulated pure-$C_\ell$ (left) and pseudo-$C_\ell$ (right) spectra with (top) and without (bottom) E-modes for a point source mask with one source per 100 square-degrees. . . . . . . . . . . . . . . . . . . . . . . . . . . . 83

2.9 Simulated pure-$C_\ell$ (left) and pseudo-$C_\ell$ (right) spectra with (top) and without (bottom) E-modes for a point source mask with one source per 10 square-degrees. . . . . . . . . . . . . . . . . . . . . . . . . . . . 84

3.1 QUIJOTE optimal errors: the top panel is the S/N and the bottom panel is the optimal $f_{\text{sky}}$. The solid red line is for $f_{\text{sky}} = 0.15$ (QUIJOTE $f_{\text{sky}}$) and a 1 degree beam, the dashed black line is for the calculated optimal $f_{\text{sky}}$ with a 1 degree beam, the dot-dashed green line is for the QUIJOTE $f_{\text{sky}}$, 5 degree beam and the dotted blue line is for the calculated optimal $f_{\text{sky}}$ with a 5 degree beam. . . . . . . . . . . . . 97

3.2 Signal to noise ratio $\Delta C_\ell/C_\ell$ (left) and error relative to the optimal errors $\Delta C_\ell/\Delta C_\ell(\text{optimal})$ (right) given by Equation (3.2) for azimuthal masks with a half sky cut (black line) and 10000 deg$^2$ (green dotted), 5000 deg$^2$ (red dashed), 1000 deg$^2$ (blue long dashed) and 500 deg$^2$ (magenta dot-dashed) spherical cap cuts . . . . . . . . . . . . . 99
3.3 Comparison of errors from Monte-Carlo simulations for unapodised (black) and apodised (blue) mask with unapodised optimal errors (green) computed using Equation (3.2), apodised optimal errors (red) computed using Equation (3.17) and errors due to E-B mixing (magenta) computed using Equation (3.18) for $r = 0.1$ (left) and $r = 0.05$ (right), using $\Delta l = 10$ (top) and $\Delta l = 50$ (bottom)  

4.1 Noise-to-signal ratio of simulations of the effect of masking out point-sources for a full-sky mask (black), 10000 square degree spherical cap (red) and 1000 square degree spherical cap (blue) compared with theoretical estimates (dashed line) and simulations without sources (dotted line) in the case of one source per 100 deg$^2$ (left) and one source per 10 deg$^2$ (right).  

4.2 Point-source-only simulations at three different values of $N_{\text{side}}$ with 6000 sources masked out. Optimal errors are shown by solid black lines, $N_{\text{side}} = 128$ by dotted green lines, $N_{\text{side}} = 256$ by dashed red lines and $N_{\text{side}} = 512$ by solid blue lines.  

4.3 Noise-to-signal ratio of simulations of the effect of masking out point-sources as a function of $\ell$ for masks containing no sources (black solid line) and one source per 10000 (magenta dot-dashed), 1000 (cyan dashed), 25 (red dashed), 10 (blue dotted) and 2.5 (green solid) square degrees for a full sky survey (top left) and spherical cap masks of 1000 square degrees (top right) and 10000 square degrees (bottom).
LIST OF FIGURES

4.4 Comparison of simulations involving a point-source mask applied to a full sky survey (top), 10000 deg\(^2\) spherical cap (middle) and 1000 deg\(^2\) cap (bottom) with those from simulations without point-sources for one source per 1000 (magenta long-dashed), 100 (black solid), 25 (red dashed) and 10 (blue dotted) square degrees. On the left the ratio of the standard deviations of the power spectrum with and without sources is plotted, and on the right the results are scaled by an \(N_s\) factor as explained in the text. Note the close agreement between the results in the right-hand plots. 123

4.5 Comparison of the simulated noise-to-signal ratio (black solid line), without deconvolution applied, to the theoretical value as given by Equation (4.37) (red dotted line) for one source per 100 deg\(^2\) (left) and one source per 10 deg\(^2\) (right). 142

5.1 Histogram of the \(\hat{n}\) in each bin for white noise (\(C_\ell = \text{const.}\)) (top left) and three CMB-like power spectra, one with \(r = 1\) (top right), one with \(C_{BB}^\ell = 0\) (bottom left) and the final one with \(C_{EE}^\ell = 0\) (and \(r = 1\)) (bottom right), with the solid line being the mean of 1000 simulations and the error bars being the expected variance for one realisation, calculated from the simulations. Note that the scales are not the same. It can be seen that, whilst the white noise case follows the standard \(\sqrt{n}\) error formula, the CMB-like spectra have errors which are considerably larger, vary between bins and also depend on the particular power spectrum provided. 152

5.2 Fractional error (\(\sigma_n/n\)) plotted against \(N_{\text{pix}}\) for pure-E mode (left) and pure-B mode (right) CMB runs. The dashed lines are \(\frac{1}{\sqrt{n}}\), the error for uncorrelated pixels, the dotted lines show the fractional error in the bin with the lowest error, and the solid lines show the error in the bin where the error is largest. 155
5.3 Probability \( P(\chi^2 > x) \) against \( \lambda \) for values of \( x \) corresponding to (from lowest to highest) 50\%, 75\%, 90\%, 95\%, 99\%, 99.9\% and 99.99\% confidence levels. ........................................... 158

5.4 Histogram of number of pixels in each bin, plus error bars showing the standard deviation of the number of pixels in the bin, for different values of \( \sigma_{QQ} \) and \( \sigma_{UU} \), with \( \sigma_{QU} = 0 \), with theoretical probability distributions also included. ....................... 160

5.5 Histogram of number of pixels in each bin, plus error bars showing the standard deviation of the number of pixels in the bin, for different values of \( \sigma_{QU} \), with \( \sigma_{QQ} = \sigma_{UU} = 1 \), with theoretical probability distributions also included. ....................... 161

5.6 Histogram of \( \hat{n} \) against \( \alpha \) including systematic effects for different values of \( \epsilon_1 \): (top left) \( \epsilon_1 = 0.1 \); (top right) \( \epsilon_1 = 0.05 \); (bottom left) \( \epsilon_1 = 0.01 \); (bottom right) \( \epsilon_1 = 0.005 \). I estimate that it should be possible to detect values of \( \epsilon_1 \) of around 0.01. ......................... 163

5.7 Histogram of \( \hat{n} \) against \( \alpha \) including systematic effects for different values of \( \epsilon_2 \): (top left) \( \epsilon_2 = 0.1 \); (top right) \( \epsilon_2 = 0.05 \); (bottom left) \( \epsilon_2 = 0.01 \); (bottom right) \( \epsilon_2 = 0.005 \). I estimate that it should be possible to detect values of \( \epsilon_2 \) of around 0.01. ......................... 164

5.8 \( \hat{n} \) vs. \( \alpha \) for simulated WMAP single-frequency maps. ................... 168

5.9 Comparison of \( \hat{n} \) vs. \( \alpha \) for the simulated and observed WMAP single-frequency maps. Note the peaks at \( \alpha = -20^\circ \) and \( 70^\circ \) in the masked maps, which appear to exist at all frequencies. ......................... 170

5.10 Comparison of \( \hat{n} \) vs. \( \alpha \) for the simulated and observed WMAP single-frequency maps with foreground reduction applied. ......................... 172

5.11 RMS difference in \( \alpha \) vs. \( r \) (left) and \( \epsilon_{QQ} \) (right). The red lines are the simulated values, and the green lines (not visible on the left hand plot due to absolute agreement with the simulations) are the formulae given in Equations (5.28) and (5.29), respectively. ......................... 174
LIST OF FIGURES

5.12 Effect on E and B mode spectra of addition of systematic errors of $\epsilon_1 = 0.1$ (left) and $\epsilon_2 = 0.1$ (right). The red line is the input spectrum and the black points are the output spectrum, with the error bars derived from the simulations. Note how the EE spectrum is unchanged and the output BB spectrum looks like a scaled-down version of the input EE spectrum.
List of Tables

5.1 Table of reduced $\chi^2$ values for the difference between the five-year WMAP foreground-reduced maps and simulations of the noise and CMB signal expected in these maps . . . . . . . . . . . . . . . 173
LIST OF TABLES
ABSTRACT OF THESIS submitted by Michael Preece
for the Degree of Doctor of Philosophy and entitled
Analysis of Cosmic Microwave Background Polarisation. March 2011

Observations of the Cosmic Microwave Background (CMB) radiation are an extremely important tool for understanding the Universe. The next generation of CMB experiments will attempt to measure the polarisation signal. In particular, the detection of B-mode polarisation, which is mainly generated by gravitational waves from the very early Universe, would provide a strong indicator for the energy level of inflation. However, due to the relative weakness of this signal, and the fact that there exists a much stronger E-mode signal, detecting B-mode polarisation poses several technical challenges. In particular, the standard method of CMB analysis, the pseudo-$C_\ell$ method, is insufficient for the purposes of analysing B-mode polarisation, because of the difficulties of separating E and B mode polarisation on a cut sky. However, an alternate method, the pure-$C_\ell$ method, has recently been outlined, which removes the effect of E-B mixing by subtracting out ambiguous modes.

In order to test this method, I have written code to implement my own version of the algorithm, as well as the ordinary pseudo-$C_\ell$ method. I have then used this code to compare the two methods both with and without E-modes, thereby demonstrating that the pure-$C_\ell$ method does not suffer from issues with E-B mixing. However, I also find that, due to the effects of subtracting out ambiguous modes, the performance of the pure-$C_\ell$ method is degraded on large scales, by a factor which is dependent on the length of the mask boundary. I have, therefore, investigated this effect. I find that, in general, the issues with this method are limited to only the very largest scales, and they are unlikely to cause significant problems for most masks. However, due to the nature of point source masks, these are particularly susceptible to such effects. Therefore, I
have considered this case more carefully.

Here, I find that there is indeed a significant effect up to multipoles of \( \ell \sim 50 \), and I have discovered that there is a simple relationship between the additional error caused by a point source mask and the number of sources. Due to the rapidly rising (in \( \ell^2 C_\ell \)) nature of the point-source power-spectrum, this issue can probably be avoided by only using the point-source mask for high-\( \ell \) measurements and, thus, the effect of point sources on the pure-\( C_\ell \) method will probably be limited. The exception to this, however, is the case where an exceptionally large number of sources must be masked out. In this case, the requirement to apodise the mask in order to implement the method will result in the analysis breaking down.

Additionally, I have also looked at several other aspects of B-mode detection. I used various optimal error formulae, in conjunction with my pure-\( C_\ell \) code, to attempt to determine the optimal scan strategy for a given set of parameters, with a particular focus on the QUIJOTE experiment. Ultimately, it was found that this experiment in its original form would only be able to detect values of \( r \) of around 0.2, although increasing the number of beams would improve this slightly. Finally, I have outlined a novel null test that is designed to detect systematic errors in CMB experiments using polarisation position angles. I have showed that this method will be able to detect shear-like systematic errors at a level of less than 1\%.
Declaration

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

MICHAEL PREECE
Copyright Statement

(i) Copyright in text of this thesis rests with the Author. Copies (by any process) either in full, or of extracts, may be made only in accordance with instructions given by the Author and lodged in the John Rylands University Library of Manchester. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the Author.

(ii) The ownership of any intellectual property rights which may be described in this thesis is vested in The University of Manchester, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement.

(iii) Further information on the conditions under which disclosures and exploitation may take place is available from the Head of School of Physics and Astronomy.
Introduction

The Cosmic Microwave Background (CMB) radiation is highly red-shifted electromagnetic radiation that was formed in the early Universe. It is predicted by the Big Bang theory and was first discovered by Arno Penzias and Robert Wilson in 1964 (Penzias and Wilson 1965). The early Universe was a mixture of electrons, ionised nuclei and photons. Thomson scattering of photons led to them being tightly coupled to the baryons, forming a photon-baryon fluid. By a temperature of about 3000 K, recombination of the electrons and nuclei is almost complete. After this time, the mean free path of the photons is effectively infinite, and they free stream towards us. Thus, we see the surface of last scattering of the primordial photons which, due to the expansion of the universe, has been red-shifted to a temperature of 2.7 K.

The initial discovery of the CMB was an extremely important piece of evidence for the Big Bang theory (Dicke et al. 1965). However, it is now important as a probe of the early stages of the Universe. The early Universe was very uniform (which in itself is a problem, since regions which could not seemingly have been in causal contact appear to have the same temperature to an accuracy of about 1 part in 10⁵, see Section 1.5), but must have had some level of density perturbations, or else the structure that we see today could not have formed. In fact, it is believed that these fluctuations are provided by quantum fluctuations in the very early Universe. There are several problems with the Hot Big Bang theory and so a process called inflation is postulated to have occurred

MICHAEL PREECE
in the very early Universe, in order to solve these.

Inflation is an extremely rapid accelerating expansion of the Universe that is believed to have occurred around $10^{-35}$ seconds after the Big Bang (see Section 1.5 for further details). Quantum fluctuations in the inflaton field are blown up and lead to density perturbations, seen in the current structure of the Universe, and temperature anisotropies, which can be measured in the CMB radiation. Various experiments have been carried out to measure these fluctuations. The first major one was the COBE satellite (see e.g. Smoot et al. 1991), which was first conceived of in 1975, and launched in 1989. This was followed by the WMAP satellite (see e.g. Bennett et al. 2003a), which measured the power at smaller angular scales, and is nearing the end of its observational life. The most recent full-sky satellite mission is Planck (see e.g. Mandolesi et al. 2002), which is due to release its first result set at the end of 2012. This is measuring in nine different frequency bands (in order to aid in the removal of non-primordial contributions, which do not follow the black-body spectrum of the CMB), to an angular resolution of five arc minutes (compared to 20 arc minutes for WMAP and seven degrees for COBE).

The WMAP5 data has constrained the six parameter $\Lambda$CDM model to high precision (Komatsu et al. 2009, Dunkley et al. 2009), and the Planck satellite will improve on this still further. However, the present frontier of CMB research is the measurement of polarisation (Readhead et al. 2004, Leitch et al. 2005, Montroy et al. 2006, Page et al. 2007, Wu et al. 2007, Bischoff et al. 2008, Brown et al. 2009). Not only is this significantly weaker, but it is spin-2 leading to the dichotomy between E and B modes (Zaldarriaga and Seljak 1997, Kamionkowski et al. 1997). E-modes are created by all inhomogeneities, but B-modes can only be due to vorticity and gravitational waves at linearised order. Since primordial vorticity is not generated during inflation, detection of B-modes could be seen as an indirect detection of gravitational waves, something which would constrain the energy-scale at which inflation took place.
1.1 Definition of CMB temperature power spectrum

Since the CMB is generated by random fluctuations, we cannot predict its exact temperature realisation. Instead, we can only predict its statistical properties as a function of angular size. In order to do this, it is necessary to write the fluctuations in terms of spherical harmonics

\[ T(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\theta, \phi), \]

(1.1)

where \( \ell \) indicates the angular scale of the mode, \( m \) indicates the direction of the mode and the \( Y_{\ell m} \) functions form a complete basis set for a spherical surface. Further, the power spectrum of the CMB can now be defined as the set of values

\[ \hat{C}_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} a_{\ell m} a_{\ell m}^*. \]

(1.2)

According to most theories of the early Universe, the CMB should be statistically isotropic, with perturbations that are approximately Gaussian, although low-levels of non-Gaussianity are expected to exist. If this is the case, then all cosmological information is contained within these \( C_\ell \) coefficients, with \( \langle a_{\ell m} a_{\ell' m'}^* \rangle = C_\ell \delta_{\ell \ell'} \delta_{mm'} \) and, thus, CMB observations are usually reduced to the power spectrum for analysis. The \( \ell \) value is directly related to the angular scale of the fluctuations represented by it. As a rough relation, the angle, \( \theta \), covered on the sky by a fluctuation that produces a peak in the power spectrum at \( \ell \) is \( \theta \approx \left( \frac{180}{\ell} \right)^\circ \).

The \( \ell = 0 \) term corresponds to the average CMB temperature (2.7K), and this is subtracted out before analysis (and, indeed, is often not measured, as a result of the experimental strategy). The \( \ell = 1 \) terms correspond to the Doppler shift caused by the motion of the Earth relative to the CMB, and is also removed before analysis. Consequently, the \( \ell = 2 \) term is the first one used in analysis of the CMB.

To determine the \( a_{\ell m} \) (and therefore \( C_\ell \)) coefficients, the properties of the \( Y_{\ell m} \) functions need to be known. They are defined as the azimuthal part of the solution to Laplace’s equation in spherical polar co-ordinates, and can be shown to form an or-
1: INTRODUCTION

thonormal basis set on the surface of the unit sphere, and so

$$\int Y_{\ell m}(\theta, \phi)Y^{*}_{\ell' m'}(\theta, \phi)d\Omega = \delta_{\ell \ell'} \delta_{mm'} . \quad (1.3)$$

Further, the $m < 0$ terms can be related to the $m > 0$ ones by

$$Y_{\ell(-m)} = (-1)^m Y_{\ell m}^* , \quad (1.4)$$

which also implies that $Y_{\ell0}$ is real.

Equation (1.3) allows us to write an expression for the $a_{\ell m}$ coefficients in terms of $T$.

Generally, we define them using

$$a_{\ell m} = \int \Delta T(\theta, \phi)Y_{\ell m}^*(\theta, \phi)d\Omega , \quad (1.5)$$

where $\Delta T(\theta, \phi) = T(\theta, \phi) − \langle T(\theta, \phi) \rangle$. Also, using Equation (1.4) it can be shown that, if $T$ is a real function,

$$a_{\ell(-m)} = (-1)^m a_{\ell m}^* . \quad (1.6)$$

1.2 Polarisation in the CMB

1.2.1 Definition of polarisation

The CMB radiation can also be polarised, and measurement of this polarisation is an important part of current CMB research. This polarisation is generated by Thomson scattering of anisotropic radiation, where the anisotropies are mainly produced by velocity gradients in the pre-recombination photon-baryon fluid (see Sections 1.5.4 and Zaldarriaga (2004) for more details). This mechanism produces a fractional polarisation of about 10% of the anisotropies.

For a quasi-monochromatic electromagnetic wave propagating in the direction of $\hat{n}$, with arbitrary polarisation, we can define (see Kamionkowski et al. 1997)

$$E_i = a_i(t) \cos(\omega_0 t - \theta_i(t)) , \quad (1.7)$$
where $E_i$ is the electric field in the direction of the unit vector $\hat{e}_i$, with $i = 1$ or $i = 2$ and the $\hat{e}_i$ vectors together with $\hat{n}$ form an orthogonal basis set. Then, the Stokes' parameters can be defined as

\[
I = \langle |a_1|^2 \rangle + \langle |a_2|^2 \rangle ,
\]

\[
Q = \langle |a_1|^2 \rangle - \langle |a_2|^2 \rangle ,
\]

\[
U = \langle a_1 a_2 \cos(\theta_1 - \theta_2) \rangle ,
\]

\[
V = \langle a_1 a_2 \sin(\theta_1 - \theta_2) \rangle .
\]

The parameter $I$ describes the absolute intensity (i.e. in this case, $T$), and so is not relevant to polarisation. $V$ measures circular polarisation and, since Thomson scattering produces only linear polarisation, is expected to be zero for the CMB, and so is also unimportant. However, $Q$ and $U$ measure linear polarisation, and are used to parametrise the CMB polarisation.

### 1.2.2 Polarisation position angles

In certain cases, it is useful to parametrise polarisation in a different manner. In particular, one useful quantity is the polarisation position angle, $\alpha$, which is defined in terms of the Stokes parameters $Q$ and $U$ by

\[
\alpha = \frac{1}{2} \tan^{-1}\left(\frac{U}{Q}\right) ,
\]

and can take values between $-90^\circ$ and $90^\circ$. Since this is a coordinate-dependent quantity, it is not often discussed in the context of the CMB. However, it is commonly used in the context of other astrophysical sources, and will be used in Chapter 5 as a novel null test for systematics. From the definition of $\alpha$, it can easily be seen that $Q = P \cos 2\alpha$ and $U = P \sin 2\alpha$, where $P = \sqrt{Q^2 + U^2}$ is the total polarised intensity.

### 1.2.3 Spin spherical harmonics and analysis of CMB polarisation

To analyse CMB polarisation, an analog of the spherical harmonics used in Section 1.1 is used. Since the values of $Q$ and $U$ are dependent on the chosen set of axes, the
power spectrum cannot simply be defined based on the $Q$ and $U$ values (the results would be meaningless). Instead, we proceed by defining $P(\Omega) = Q(\Omega) + iU(\Omega)$. Rotating by an angle $\phi$ about the direction of observation, gives $P(\Omega) \rightarrow e^{-2i\phi}P(\Omega)$ and thus $P$ is a spin 2 variable (see Newman and Penrose (1966) for more details). The ordinary spherical harmonics are not spin 2 functions, and so linear combinations of them cannot describe the value of $P$ for all chosen bases. However, generalisations of the spherical harmonics, known as spin-weighted spherical harmonics can be defined, and these have the required spin (Newman and Penrose 1966). A spin raising operator $\delta$ can then be defined, and its complex conjugate, $\bar{\delta}$ is a spin lowering operator. The spin-$s$ spherical harmonics, $sY_{\ell m}$ are then defined as

$$sY_{\ell m} = \sqrt{\frac{(\ell - s)!}{(l + s)!}} Y_{\ell m} \quad 0 \leq s \leq l \quad (1.13)$$

$$sY_{\ell m} = \frac{(l - s)!}{(l + s)!} (-1)^s \delta Y_{\ell m} \quad -l \leq s \leq 0 \quad (1.14)$$

$$sY_{\ell m} = 0 \quad l < |s| \quad (1.15)$$

More details can be found in Newman and Penrose (1966) and Goldberg and Macfarlane (1967).

These spin-weighted spherical harmonics have properties that are analogous to those of the original spherical harmonics. Most importantly, they form an orthonormal basis set for spin-$s$ quantities. Also, in analogy to Equation (1.4), it can be shown that

$$sY_{\ell m} = (-1)^{m+s} sY_{\ell(-m)} \quad (1.16)$$

Since $P$ is spin 2, it can be written as

$$P(\Omega) = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} 2a_{\ell m} Y_{\ell m} \quad , \quad (1.17)$$

and, for the complex conjugate,

$$P^*(\Omega) = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} -2a_{\ell m} Y_{\ell m} \quad . \quad (1.18)$$
Linear combinations of these coefficients, \(a_{\ell m}^E = (2a_{\ell m} + 2a_{\ell m})/2\) and \(a_{\ell m}^B = i(-2a_{\ell m} - 2a_{\ell m})/2\) can then be defined, and using the definitions of \(2a_{\ell m}\) and \(-2a_{\ell m}\), expressions for these in terms of the polarisation field can be found:

\[
\begin{align*}
da_{\ell m}^E &= \frac{1}{2} \int (P(\Omega) Y_{\ell m}^* + P^*(\Omega) Y_{\ell m}) d\Omega, \\
da_{\ell m}^B &= \frac{i}{2} \int (P^*(\Omega) Y_{\ell m}^* - P(\Omega) Y_{\ell m}) d\Omega.
\end{align*}
\]

Using this definition, and Equation (1.16), an analogue of Equation (1.6) is obtained

\[
da_{\ell(-m)}^P = (-1)^m (a_{\ell m}^P)^*,
\]

where \(P\) represents either \(E\) or \(B\). The importance of these particular variable combinations will be shown in Section 1.2.4.

Using the above, we define a set of six power spectra, \(C_{\ell}^{TT}\), \(C_{\ell}^{TE}\), \(C_{\ell}^{TB}\), \(C_{\ell}^{EE}\), \(C_{\ell}^{EB}\) and \(C_{\ell}^{BB}\), which are defined (generically) as

\[
\hat{C}_{\ell}^{XY} = \langle a_{\ell m}^X (a_{\ell m}^Y)^* \rangle,
\]

where \(a_{\ell m}^T\) is defined in Equation (1.5), and the others are defined in Equations (1.19) and (1.20) respectively. Once again, if the CMB is statistically isotropic with perturbations that are Gaussian, then all cosmological information is contained within the power spectra. However, unlike the \(C_{\ell}^{XX}\) terms, the cross terms (\(C_{\ell}^{XY}\) where \(X \neq Y\)) can be either positive or negative. There are also other ways to calculate the power spectra in the presence of polarisation, one of which is to use tensor spherical harmonics as outlined in, for example, [Kamionkowski et al.] (1997).

### 1.2.4 Properties of E and B modes

The division of the polarisation \(a_{\ell m}\) coefficients into \(E\) and \(B\) has so far appeared to be arbitrary. However, it is in fact extremely useful. If we define real space, spin 0 functions from these coefficients (\(E_l(\Omega) = \sum_m a_{\ell m}^E Y_{\ell m}(\Omega)\), and similarly for \(B\)), it is found that the function \(E\) has even parity, whereas the function \(B\) has odd parity (see...
Figure 1.1 for a diagram of E vs. B modes. This is analogous to electric and magnetic fields, hence the notation \( E \) and \( B \). However, this split is extremely important. Due to the nature of the fluctuations they produce, density perturbations, which are expected to be the major source of polarisation (and, indeed, anisotropies in general) in the CMB, can only generate polarisation modes with even parity, meaning that they can only generate E mode polarisation (Zaldarriaga 2004).

Figure 1.1: Pure E and B mode patterns for \( m=0 \) centred on the North Pole.

On the other hand, gravitational waves can generate both E and B mode polarisation. This is important because gravitational waves are expected to be formed during
inflation with a magnitude dependent on the energy scale at which inflation took place (see Section 1.5.6), and so a measurement of the primordial B-mode spectrum will be strong evidence that inflation did indeed occur, as well as giving a strong constraint on the energy scale. The important parameter for determining the strength of B-mode polarisation is $r$, the tensor-to-scalar ratio. This is (usually) defined as the ratio of the primordial power in gravitational waves to that in curvature perturbations. However, E mode polarisation can also be converted into B mode polarisation by gravitational lensing (Zaldarriaga 2004). This provides a significant constraint on our ability to detect primordial gravitational waves, but could also provide information about the mass distribution in the Universe, giving indirect information about density fluctuations at relatively large $\ell$ values.

Of the six ways of combining the $a_{\ell m}$ values to make $C_\ell$ values, only four are expected to be non-zero. This is because $T$ and $E$ have even parity, whereas $B$ has odd parity, so, assuming that all interactions involved in forming the CMB photons conserve parity, there can be no mechanism for coupling $B$ with either $T$ or $E$. As a result, it is expected that $C_\ell^{EB} = C_\ell^{TB} = 0$. If this is not found to be true, then it either means that there are systematic errors in the data, or else parity is not conserved in the processes that formed the CMB (a remarkable finding). Of the four non-zero terms, $C_\ell^{TT}$ (the temperature moment) is by far the largest, followed generally by $C_\ell^{TE}$, which is about 30 times smaller, with $C_\ell^{EE}$, which is about 5 times smaller than this. The smallest term is $C_\ell^{BB}$, which has yet to be detected, but the tensor to scalar ratio, $r$, (which is roughly the ratio of $C_\ell^{BB}$ to $C_\ell^{EE}$ in the absence of lensing) has been determined to be $r < 0.55$ (Page et al. 2007), although currently the best constraints on this quantity come from measurements of the temperature and E-mode polarisation spectra, rather than from upper limits on the B-mode spectrum.
1.3 Partial-sky observations and foreground masking

The method outlined in the previous two sections can only be applied if the whole sky is being considered, since the $Y_{\ell m}$ functions will not be orthonormal otherwise. However, most CMB experiments cannot provide accurate measurements of the CMB temperature and polarisation over the whole sky. For a ground-based telescope, only part of the sky can usually be observed. However, even if the entire sky can be measured (such as for a satellite-based experiment), some of the data will be contaminated with foreground emissions, especially in the galactic plane. As a result, the analysis of the CMB can generally only be done on a certain fraction of the sky. In the case of e.g. WMAP, the fraction of the sky used ($f_{\text{sky}}$) is reasonably large (about 0.7), cutting out the galactic plane and bright point sources. However, for many ground based telescopes, the fraction of the sky observed is small. The basic maths behind each is the same, although the computational requirements may be different.

Since the formalism outlined previously cannot be used, some other method for analysing the temperature anisotropies is required in this case. There are several ways of doing this. The most commonly used method, which is the main topic of my thesis, is the pseudo-$C_\ell$ method (initially derived in [Peebles 1973]) and variants thereof. There are other possible methods for doing this. In particular, the optimal method is the Maximum-Likelihood method, which uses the entire data set in order to determine the most likely estimate for the power spectrum and the errors. However, this method is $O(l^4)$ in memory requirements and $O(l^6)$ in operations, and hence is not feasible for the resolutions used for modern experiments.

Since the spherical harmonics do not form an orthonormal basis on the cut sphere, the full sphere must be considered in order to determine the $C_\ell$ coefficients. Thus, cut-sky observations need to be converted into full-sky observations. This can be done trivially by multiplying the observed data by a mask function, which is zero for parts of the sky which are not observed or excluded due to foreground or other contamination.
A new temperature function, \( \tilde{T} \) can then be defined as

\[
\tilde{T}(\Omega) = M^T(\Omega)T(\Omega). \tag{1.23}
\]

The mask function, \( M^T \), can take any arbitrary value (not only 0 or 1) for a given \( \Omega \), so we can, for example e.g. weight the contribution to \( \tilde{T} \) by the amount of noise. Obviously it is necessary that \( M = 0 \) if \( T \) has not been measured. I can now define the pseudo-\( C_\ell \)s in the same way that I defined an estimator for the ordinary \( C_\ell \)s earlier

\[
\tilde{C}_\ell = \frac{1}{2\ell + 1} \left( \sum_{m=-\ell}^{\ell} |\tilde{a}_{\ell m}|^2 \right), \tag{1.24}
\]

where

\[
\tilde{a}_{\ell m} = \int \Delta T(\Omega) Y^*_{\ell m}(\Omega) d\Omega. \tag{1.25}
\]

This can also be extended to polarisation by defining \( \tilde{P}(\Omega) = P(\Omega)M^P(\Omega) \), where I have assumed (reasonably) that the mask is the same for \( Q \) and \( U \), but not necessarily the same for temperature and polarisation. Doing this, a set of six pseudo-\( C_\ell \) coefficients, similar to those defined in Section 1.2 for the unmasked case, are obtained.

However, these pseudo-\( C_\ell \)s are biased estimators for the true \( C_\ell \). Therefore, the pseudo-\( C_\ell \)s need to be debiased in order to recover the true \( C_\ell \). Making the assumption that the mask spherical harmonics have the property that the ensemble average \( \langle a_{\ell m}^* a_{\ell' m'} \rangle = C_\ell \delta_{\ell \ell'} \delta_{mm'} \), it can be seen that the pseudo-\( C_\ell \) values must have the same property, and can therefore only depend on \( C_\ell \). In this case,

\[
\langle \tilde{C}^{XY}_\ell \rangle = \sum_{A,B} \sum_{\ell'} M^{XY,AB}_{\ell \ell'} \langle C^{AB}_{\ell'} \rangle. \tag{1.26}
\]

Since \( T \) and \( P \) are separate functions, masking them does not cause any mixing. However, \( E \) and \( B \) are both defined from \( P \), so it is possible for mixing between them to occur. It is found that (see appendix A in Brown et al. (2005) for details) only eight of the matrix elements are non-zero, and there are only five unique elements. These are \( M^{TT,TT} \), \( M^{TE,TE} = M^{TB,TT} \), \( M^{EE,EE} = M^{BB,EE} \), \( M^{EE,BB} = M^{BB,EE} \) and \( M^{EB,EB} \). This implies that, of the six power spectra, only the EE and BB power spectra undergo any

MICHAEL PREECE

29
mixing when the pseudo-$C_\ell$s are calculated. Note that, by convention (and because it is approximately flat), we generally measure and plot $\langle T_\ell \rangle = l(l+1)C_\ell/2\pi$ rather than $C_\ell$ itself and, thus, this is the quantity I will usually be working with in this thesis.

The mask matrix $M$ can be calculated (see Brown et al. (2005) and Section 2.1.2), and can thus be inverted (in principle) in order to obtain an unbiased estimator for the real $C_\ell$ coefficients from the measured pseudo-$C_\ell$s. This process is known as deconvolution, by analogy with Fourier theory. Note, however, that using the pseudo-$C_\ell$ method causes the variance of the $C_\ell$ estimate to be larger than the cosmic variance (see Equation (1.27)). This can be viewed as a result of the loss of information caused by cutting out part of the sky. The error in B-mode measurements also includes an additional term, due to the E-B mixing. Since $C_{EE}\gg C_{BB}$, this mixing has little effect on E-mode measurements, however for large sky cuts it has a substantial effect on the detection of B-mode signals. This effect is discussed quantitatively in Section 2.2.

1.4 Errors in CMB observations

1.4.1 Basics of real CMB observations

The previous sections assume that the measurements are continuous (and not pixelised), so the functions $T$ and $P$ are continuous. In practice, this cannot be the case, since there is no way to measure them with infinite precision, and nor could the data be stored. Hence, the sky must be divided into pixels for the purpose of measurement. Doing this on a sphere in a way that allows fast calculations and has equal area pixels (important for noise addition/removal) is non-trivial. The standard pixelisation method (used on WMAP) is called HEALPix, which is explained in detail in Górski et al. (2005). This is the pixelisation method that I have used throughout my thesis. Once a pixelisation is chosen, most of the previous formulae can be trivially written in a discrete format, and there are standard HEALPix libraries that carry out most of the calculations in the previous sections, converting a $(T, Q, U)$ map automatically into a
1.4: ERRORS IN CMB OBSERVATIONS

set of $a_{\ell m}$ values (for $T$, $E$ and $B$), up to some maximum $\ell$ value. Therefore, I have used the HEALPix libraries for most of the calculations in my thesis.

Furthermore, the previous analysis has also ignored errors, both random errors and systematic errors (including the effects of foregrounds). In this thesis, I have mainly concentrated on random errors (which are simpler to handle), and I will discuss these here. There are two types of random errors affecting CMB observations. These are intrinsic random errors due to the stochastic nature of the $a_{\ell m}$ values (plus additional errors from the measurement method in the case of cut skies), and also errors due to noise and beam effects from the instrument. In this section, I will only consider the case of full-sky observations.

1.4.2 Cosmic variance

The most fundamental source of error in the measurement of the CMB power spectra is cosmic variance. Since the fluctuations in the CMB are caused by quantum fluctuations during inflation (see Section 1.5.3), the fluctuations in each $a_{\ell m}$ are random, with a variance that depends on the theoretical value of $C_\ell$. Therefore, even if a perfect measurement of the sky could be made, with no external sources of error, there would still be a fundamental level of uncertainty in our estimate of the theoretical $C_\ell$. Equations (1.6) (for $T$) and (1.21) (for $E$ and $B$) imply that the $a_{\ell m}$s for a given $\ell$ can be specified using only $2\ell + 1$ independent real variables, all normally distributed. Since $\langle |a_{\ell m}|^2 \rangle$ is independent of $m$ (by isotropy), and using the definition of $C_\ell$, it is found that the intrinsic standard deviation on $C_\ell^{XX}$ is

$$\sigma(C_\ell^{XX}) = \sqrt{\frac{2}{2\ell + 1}} C_\ell^{XX}, \quad (1.27)$$

which applies to all of the auto-power spectra (TT, EE, BB). Whilst cosmic variance is negligible for large $\ell$, it provides a very strong limit on the ability to determine the $C_\ell$ values for low $\ell$. The case where the cross-power spectra (TE, TB, EB) are being considered is more complicated, since $\sigma(C_\ell^{XY})$ depends not only on $C_\ell^{XY}$, but also on

MICHAEL PREECE
1: INTRODUCTION

$C^{XX}_{\ell}$ and $C^{YY}_{\ell}$. The formula that must be used can be determined from the properties of the $a_{\ell m}$ coefficients. It is found that

$$\sigma(C^{XY}_{\ell}) = \sqrt{\frac{1}{2\ell + 1} \left( \langle C^{XY}_{\ell} \rangle^2 + \langle C^{XX}_{\ell} \rangle \langle C^{YY}_{\ell} \rangle \right)}.$$  

(1.28)

1.4.3 Effects of noise and beams

To analyse real experimental data, noise from instrumental effects must also be considered. For this section, I will ignore the effects of foregrounds and I will assume that the whole sky has been used in the analysis. Initially, I will consider the temperature only ($C^{TT}_{\ell}$), and I will then extend this to include polarisation and cross spectra. First, we define the measured harmonic temperature coefficients, $d_{\ell m}^{T}$, as (Kamionkowski et al. 1997)

$$d_{\ell m}^{T} = \int T^\text{map}(\Omega) Y^*_\ell(\Omega) d\Omega \approx \sum_{j=1}^{N\text{pix}} \frac{4\pi}{N_{\text{pix}}} T^\text{map}_j Y_{\ell m}(\theta_j, \phi_j),$$  

(1.29)

where $T^\text{map}_j$ is the temperature of the $j$th pixel, and $(\theta_j, \phi_j)$ is its location on the sky. Note that $d_{\ell m}$ are the measured coefficients, and include noise and beam effects. The observed map is the sum of the cosmological signal and the noise, and, therefore, $d_{\ell m} = a_{\ell m} + a_{\ell m}^{\text{noise}}$. Assuming no correlation between noise and the cosmological signal, then (Kamionkowski et al. 1997)

$$\langle d_{\ell m}^{T} d_{\ell' m'}^{T*} \rangle = B^2_{\ell} C^{TT}_{\ell} \delta_{\ell \ell'} \delta_{mm'} + \langle a_{\ell m}^{T,\text{noise}} (a_{\ell m}^{T,\text{noise}})^* \rangle,$$

(1.30)

where $B_{\ell}$ is the beam window function and accounts for smearing due to the finite beam width of any observing instrument. For a beam that is Gaussian in shape, $B_{\ell} = \exp(-\ell(\ell + 1)\sigma_b^2)$, where $\sigma_b = \theta_{\text{FWHM}} / \sqrt{8 \ln 2}$, and $\theta_{\text{FWHM}}$ is the full-width at half-maximum of the beam. Under the assumption that the noise is an uncorrelated Gaussian random variable, I can write (Kamionkowski et al. 1997)

$$\langle a_{\ell m}^{T,\text{noise}} (a_{\ell' m'}^{T,\text{noise}})^* \rangle = N_{TT}^\ell \delta_{\ell \ell'} \delta_{mm'},$$

(1.31)
where $N_{\ell}^{TT}$ is the power spectrum of the noise. Assuming the noise has constant variance, with $\langle T_i^{\text{noise}} T_j^{\text{noise}} \rangle = (\sigma_{\text{pix}}^T)^2 \delta_{ij}$, it is found that

$$N_{\ell}^{TT} = \frac{4\pi (\sigma_{\text{pix}}^T)^2}{N_{\text{pix}}} ,$$

(1.32)

where $N_{\text{pix}}$ is the total number of pixels. This formula implies that the noise has a flat power spectrum ($N_{\ell} = \text{const}$). Using Equation (1.30), I can find a formula to determine an unbiased estimate of the cosmological $C_{\ell}$s in terms of the observed $C_{\ell}$ values ($\tilde{C}_{\ell}$), defined by $\tilde{C}_{\ell} \delta_{\ell'\ell} \delta_{mm'} = \langle d_{\ell m}^{T} d_{\ell' m'}^{T} \rangle$ and the noise spectrum $N_{\ell}$ (defined above)

$$\hat{C}_{\ell}^{TT} = \frac{1}{B_{\ell}^2} (\tilde{C}_{\ell}^{TT} - N_{\ell}^{TT}) .$$

(1.33)

This process can then be extended to polarisation, again making the same approximations. If I now define $a_{\ell m}^{X_{\text{noise}}}$ as above, but extending $X$ to include $E$ and $B$, and make the same assumptions as before, I obtain

$$\langle a_{\ell m}^{X_{\text{noise}}} a_{\ell' m'}^{X_{\text{noise}}} \rangle = N_{\ell}^{XY} \delta_{\ell'\ell} \delta_{mm'} .$$

(1.34)

Under the assumption that the noise has constant variance, with zero cross-correlation (so $N_{\ell}^{XY} = 0$ for $X \neq Y$), an analog to Equation (1.32) can be found [Kamionkowski et al. 1997]

$$N_{\ell}^{XX} = \frac{4\pi (\sigma_{\text{pix}}^X)^2}{N_{\text{pix}}} ,$$

(1.35)

where $\sigma_{\text{pix}}^E = \sigma_{\text{pix}}^B = \sigma_{\text{pix}}^P$. Also, Equation (1.33) can be extended to become

$$\hat{C}_{\ell}^{XY} = \frac{1}{B_{\ell}^2} (C_{\ell}^{XY} - N_{\ell}^{XY}) ,$$

(1.36)

where the beam is assumed to be the same for both $T$ and $P$ measurements. These formulae enable us to calculate an unbiased estimate for the real $C_{\ell}$ values given a map with a known r.m.s. noise and a known beam width (both of which must be known for an experiment to be useful), and thus are very useful in analysis of noisy CMB maps.

MICHAEL PREECE
1.5 Inflationary theory and CMB anisotropies

1.5.1 The theory of inflation

The existence of the CMB very strongly suggests that the Universe had at one point been very hot and dense, as predicted by the Big Bang theory. However, there are several big problems with the standard Big Bang model. The first is the flatness problem. In an ordinary, decelerating FRW universe, a flat ($\Omega = 1$) model is an unstable equilibrium, and the Universe would be expected to move rapidly away from flatness. However, it is observed today to be almost flat, which means that it must have been extremely flat in its early history.

Another problem is the horizon problem. In a normally-expanding universe, the horizon size increases with time, which means that most of the observable Universe today would have never been in causal contact when the primordial fluctuations that generate the CMB were formed. However, despite this, the CMB is still isotropic to an accuracy of $10^{-5}$, which is a problem since there can be no mechanism to bring them into thermal equilibrium without causal contact.

To solve these problems, the idea of inflation, an extremely rapid expansion of the Universe during its very early history, was introduced. During inflation, the Universe expands at a rapidly accelerating rate (as opposed to the usual decelerating expansion). As a result of this rapid acceleration, $\Omega$ tends very rapidly towards one, providing a natural explanation of the flatness of the Universe. Furthermore, the superluminal expansion that occurs during inflation means that the entire observable Universe can come from a single causally-connected region (see Liddle and Lyth (2000) for more details).

During inflation, the observable Universe increases in size by many orders of magnitude. Therefore $N$, the number of e-folds of growth is defined as

$$N = \ln \left( \frac{a_{\text{end}}}{a_{\text{start}}} \right).$$

If inflation lasts for a sufficiently large number of e-folds (at least $N=60$), then we will...
end up with a universe which still looks flat today, regardless of the initial curvature.

In general, we can describe inflation by considering a scalar field, $\phi(x,t)$, which represents a hypothetical spin 0 particle, called the Inflaton, with interaction potential $V(\phi)$. The properties of this particle are unknown since it only exists at around the Grand Unification scale, which is of the order of $10^{15}$ GeV.

To study inflation, we need to consider a scalar field in a curved space-time. The Lagrangian density for a particle in GR is given by

$$L = \sqrt{-g} \left( \frac{1}{2} \partial_\mu \phi \partial^\mu \phi - V(\phi) \right), \quad (1.38)$$

where $\sqrt{-g} = \sqrt{-\det(g_{\mu\nu})} = a^3$. Using this, we can find (Liddle and Lyth 2000)

$$\rho_\phi = \frac{1}{2} \dot{\phi}^2 + \frac{1}{2} |\nabla \phi|^2 + V(\phi), \quad (1.39)$$

$$P_\phi = \frac{1}{2} \dot{\phi}^2 - \frac{1}{6} |\nabla \phi|^2 - V(\phi). \quad (1.40)$$

This gives the Equation of motion for the Inflaton to be

$$\ddot{\phi} + 3H \dot{\phi} - \nabla^2 \phi a^2 + \frac{dV}{d\phi} = 0. \quad (1.41)$$

To solve this, we assume that the scalar field is spatially homogeneous and that the slow-roll approximation (SRA) is valid, meaning that the $\ddot{\phi}$, $\dot{\phi}^2$ and $\nabla^2 \phi$ terms can be neglected. We then obtain the slow-roll equations

$$H^2 \approx \frac{8\pi G}{3} V, \quad (1.42)$$

$$3H \dot{\phi} = -\frac{dV}{d\phi}. \quad (1.43)$$

The slow-roll parameters can also be defined as (Liddle and Lyth 2000)

$$\epsilon(\phi) \equiv \frac{1}{16\pi G} \left( \frac{V'}{V} \right)^2, \quad (1.44)$$

$$\eta(\phi) \equiv \frac{1}{8\pi G} \left( \frac{V''}{V} \right). \quad (1.45)$$

For the SRA to hold, it is required that $\epsilon \ll 1$ and $|\eta| \ll 1$, and inflation is said to have ended when these parameters become one. These parameters characterise many
of the properties of inflation. For example, the total number of e-foldings, \(N\), that the
Universe grows by during inflation can be found (using the SRA) to be
\[
N = -8\pi G \int_{\phi_i}^{\phi_f} \frac{V}{V'} d\phi,
\]
(1.46)
and thus the total amount of inflation which occurs is dependant on \(\epsilon(\phi)\) only. There
are also other properties associated with inflation, such as the tensor and scalar spectral
indices and the tensor/scalar ratio (see Section 1.5.2), which can also be written in
terms of the slow-roll parameters.

1.5.2 Inflationary power spectra and types of model

To create the large-scale structure that we see in the local Universe and in the CMB,
there must be fluctuations in the early Universe. These fluctuations are generated,
according to most models, by quantum fluctuations in the Inflaton field. For a simple
derivation, I can write (Liddle and Lyth 2000)
\[
\delta\phi(x, t) = \sum_k \delta\phi_k(t) e^{ikx},
\]
(1.47)
where \(r = a(t)\mathbf{x}\), with \(r\) being the physical position, and therefore the physical wavenum-
ber is \(k/a\).

These fluctuations in the Inflaton field lead to fluctuations in the energy density
during inflation, and therefore fluctuations in the matter density following inflation. By
homogeneity, these fluctuations only depend on \(k = |k|\), and the scalar power spectrum
can be defined as
\[
P_S(k) = \left(\frac{\delta\rho}{\rho}\right)^2 \approx \frac{H^4}{\dot{\phi}^2}.
\]
(1.48)
In general,
\[
P_S(k) = A \left(\frac{k}{k_{\text{pivot}}}\right)^{n_s - 1},
\]
(1.49)
where \(A\) is the amplitude of the spectrum, \(n_s\) (\(= n_s(k)\) in general) is the scalar spectral
index and \(k_{\text{pivot}}\) is the pivot scale. There are also tensor fluctuations (gravitational
waves), which have a spectrum given by

\[ P_T(k) = A r \left( \frac{k}{k_{\text{pivot}}} \right)^{n_T}, \]

(1.50)

where \( A \) is the amplitude of the scalar spectrum, \( r \) is the tensor/scalar ratio, \( n_T \) is the tensor spectral index and \( k_{\text{pivot}} \) is the pivot scale. The most important quantities for inflationary theory are the spectral indices and the tensor/scalar ratio, and these can be written in terms of the slow-roll parameters as (Liddle and Lyth 2000)

\[ n - 1 = 2\eta - 6\epsilon, \]

(1.51)

\[ n_T = -2\epsilon, \]

(1.52)

\[ r = \frac{P_T(k_{\text{pivot}})}{P_S(k_{\text{pivot}})} = 16\epsilon. \]

(1.53)

Assuming that the SRA approximation holds exactly, and that the power spectra are exact power laws, \( n_T = -r/8 \). In practice, however, it is possible that running of the spectral index can occur (i.e. \( n_S \neq \text{const.} \)).

As can be seen, there are a very large number of possible inflationary models, corresponding to different choices for \( V \). Generally, these models can be roughly categorised according to the values of \( \epsilon \) and \( \eta \) (as measured at around the time at which the fluctuations coming back into the horizon today first left the horizon during inflation) into three categories (see Kinney and Riotto 2006), and whilst the descriptions given here may not represent all possible models, they are illustrative of those most commonly studied. These are as follows.

1. Large field models (\( -\epsilon < \eta \leq \epsilon \)): These are models for which \( V'' \) is small relative to \( V' \). These models invoke Chaotic inflation - they start with the field randomly displaced from the origin by an amount of order the Planck mass, and the minimum in the potential is at the origin. The Inflaton field then rolls from the initial (random) value towards zero. Examples of models like this are power law models, \( V \propto \phi^n \) and exponential models, \( V \propto e^{nb} \). These models have moderately large tensor/scalar ratios, since \( \epsilon \) is large.
2. Small field models ($\eta < -\epsilon$): These models have $V'' < 0$, and so have negative $\eta$. They start with $\phi = 0$, which is a maximum for the potential, and the field then rolls down the potential gradient towards a local minimum. These models can occur naturally as a result of spontaneous symmetry breaking (which occurs when the minimum for a field changes to a maximum, and the field then falls towards a new minimum), giving a good potential explanation for what causes inflation to begin. An example of such a model is $V \propto (1 - (\frac{\phi}{\mu})^n)$. Since $\epsilon$ is small here, the tensor/scalar ratio is very low, as is $n_T$.

3. Hybrid models ($0 < \epsilon < \eta$): These models have two fields involved in inflation. The Inflaton field is the field which causes the actual inflation, and it starts, once again, with $\phi \neq 0$, and moves towards the minimum at the origin. However in this case, it interacts with another field. When the Inflaton field falls to a certain value, it causes the other field to undergo a phase transition, which leads to the end of inflation as the system falls into the new minimum. In this case, the end of inflation does not simply occur as a result of violating the SRA, as in the other two models.

1.5.3 Fluctuations in the CMB

In order to understand the CMB, it is necessary to understand how it is generated from the initial inflationary perturbations. The perturbation in an observable, $g$, can be written as (Liddle and Lyth 2000)

$$g_k = T_g(t, k)R_k,$$  \hspace{1cm} (1.54)

where $R_k$ is the primordial curvature perturbation and $T_g$ is the transfer function. The curvature perturbation can be written in terms of the Inflation field (Liddle and Lyth 2000)

$$R_k = -\left(\frac{H}{\dot{\phi}} \delta \phi_k\right)_{t=t_0},$$  \hspace{1cm} (1.55)
where \( t^* \) is the time when the curvature perturbation freezes-in, a few e-folds after the scale leaves the horizon.

Considering the observable, \( g \), to be the CMB radiation and expanding \( g \) in terms of spherical harmonic coefficients, the CMB power spectrum can be obtained. Extending Equation (1.54), and assuming the rotational invariance of the Universe, gives (Liddle and Lyth 2000)

\[
A_{\ell m} = \frac{4\pi}{2\pi^2} \int_0^{\infty} T_\Theta(k, l) R_{\ell m}(k) k dk,
\]

(1.56)

where \( T_\Theta \) is the temperature transfer function, which is independent of \( m \) by rotational invariance, and \( R_{\ell m} \) are the spherical harmonic coefficients of the curvature perturbation. From this, it is found that

\[
C_\ell = 4\pi \int_0^{\infty} T_\Theta^2(k, l) P_R(k) \frac{dk}{k},
\]

(1.57)

where \( P_R \) is the spectrum of the curvature function, and, thus, the \( C_\ell \) coefficients can be written in terms of the transfer function and the inflationary parameters given earlier.

The transfer function is determined mostly by the physics of the photon-baryon fluid, as well as cosmic history following recombination. There are several different contributions to the transfer function which give us most of the observed structure in the CMB. At very large scales, above the horizon scale at the time of last scattering, causal processes which acted on the photon-baryon fluid can have no effect. In this case, the only contributions to the transfer function comes from the Sachs-Wolfe effect and the (smaller) integrated Sachs-Wolfe (ISW) effect (Sachs and Wolfe 1967), which are both caused by the gravitational red-shift of the CMB photons as they travel towards us. These give

\[
\Delta T \left|_k \right. = \frac{1}{3} \Phi(k, \eta_{\text{dec}}) + 2 \int_{\eta_{\text{dec}}}^{\eta_0} \Phi \, d\eta,
\]

(1.58)

where \( \Phi \) is the gravitational potential, \( \eta_{\text{dec}} \) is the conformal time at recombination and \( \eta_0 \) is the conformal time today. Here, the first term corresponds to the potential at the last scattering surface (\( \Phi(\eta_0) = 0 \)), and the second term is due to the photons being...
red-shifted as they move through the time-evolving potential towards us. Assuming matter domination, this gives a transfer function of the form (derived in Liddle and Lyth 2000)

\[ T_\Theta(k, \ell) = -\frac{1}{5} j_\ell \left( \frac{2k}{H_0} \right), \]  

(1.59)

where \( j_\ell \) is a Bessel function of the first kind and, hence,

\[ \ell(\ell + 1)C_\ell \approx \frac{25\pi}{8} P_R \left( \frac{\ell H_0}{2} \right). \]  

(1.60)

For \( \ell \ll l_{LS} \) (where \( l_{LS} \) is the angular scale corresponding to recombination), the first term in Equation (1.58) dominates and the spectrum is expected to be flat (if \( n_S = 1 \)). The ISW effect is zero for a static potential (such as in a matter-dominated universe), but is non-zero in a \( \Lambda \)-dominated universe, and thus a measurement of the ISW effect would give us more information about the nature of dark energy. Unfortunately, it only has an effect at very large scales (\( l < 10 \)), which means that its usefulness is limited by cosmic variance (Equation (1.27)).

For higher \( \ell \), the calculation is far more complicated, and cannot be done exactly analytically. However, it is possible to obtain approximate formulae which give us an idea of the processes involved. The first, and most important, effect is acoustic oscillations. The physics behind this is discussed in Zaldarriaga (2004), and I will describe the results here. Prior to recombination, the photons and baryons can be considered as a single fluid, with pressure due to the photons and inertia due to the massive baryons. This fluid can undergo acoustic oscillations, generating standing waves in the photon-baryon fluid. The contribution to the temperature anisotropy from each \( k \) mode is dependent on both the density perturbation, and the velocity, due to Doppler shifting. These perturbations are \( \pi/2 \) out of phase with each other, but the density induced anisotropies are slightly larger, and thus there are peaks in the spectrum. The density induced anisotropies can be written in terms of the density perturbation as

\[ \frac{\Delta T_k}{T} = \frac{1}{4} \delta_r(k, \eta_{dec}). \]  

(1.61)
It is possible to find an expression for $\delta$ in terms of $k$, for both sub-horizon and super-horizon scales, and plugging this expression into the above equation gives

$$\left\langle \frac{T^2_k}{T^2_{CMB}} \right\rangle = \frac{8\pi^2}{9} k^3 |A(k)|^2 \left\{ \begin{array}{ll} 1 & k < k_{\text{dec}} \\ \cos^2 \left( \frac{2\pi}{\sqrt{3}} \left( \frac{k}{k_{\text{dec}}} - 1 \right) \right) & k > k_{\text{dec}} \end{array} \right., \quad (1.62)$$

where $A(k)$ is a constant that depends on the scalar and tensor power spectra, with $k^3 |A(k)|^2$ constant for $n_S = 1$, i.e. a scale-invariant, or Harrison-Zel’dovich, spectrum (see Harrison 1970 and Zeldovich 1970).

From the above equation, it can be seen that there are peaks at $k = k_{\text{dec}} \left( 1 + n \sqrt{3} \right)$. In a similar way, the velocity power spectrum can be determined. Using the assumption that fluctuations of an angular scale $\ell$ are produced mainly from wavevectors with $k \approx k^* = \ell/\eta_0$, we get a spectrum for the acoustic peaks (for $\ell > \ell_{\text{dec}}$) of

$$\ell(\ell + 1)C_{\ell} = \frac{8\pi^2}{3} \left( \frac{\ell}{\eta_0} \right)^3 |A\left( \frac{\ell}{\eta_0} \right)|^2 \left( \frac{1}{3} \left( \cos^2 \left( \frac{2\pi}{\sqrt{3}} \left( \frac{\ell}{\eta_0} \eta_{\text{dec}} \right) - 1 \right) \right) + c_s^2 \sin^2 \left( \frac{2\pi}{\sqrt{3}} \left( \frac{\ell}{\eta_0} \eta_{\text{dec}} \right) - 1 \right) \right), \quad (1.63)$$

Where $\eta_{\text{dec}}$ is the conformal time at decoupling and $\eta_0$ is the conformal time now. Baryon inertia of the fluid leads to $c_s^2 < 1/3$, and so the first term is dominant, as explained previously, leading to acoustic peaks in the CMB at

$$\ell = \frac{\eta_0}{\eta_{\text{dec}}} \left( 1 + n \sqrt{3} \right), \quad (1.64)$$

where $n \in \mathbb{N}_0$. The position of the acoustic peaks depends on the sound horizon size at recombination, and the distance to recombination, in turn, gives information about the matter content and curvature of the Universe.

There are also other effects which can modify this basic peak structure. For large $\ell$, Silk damping (Silk 1967) is the major effect. The photon-baryon fluid is not a perfect fluid, since the photons have a finite mean free path. This results in damping on very small scales, which can be modelled by multiplying the transfer function by $\exp(-k^2/k_s^2)$, with $k_s \propto \eta_{\text{dec}}^{1/2} (\Omega_b h^2)^{-1/2}$, where $\eta_{\text{dec}}$ is the time at which recombination occurs. This effect gives us information about the baryon density of the Universe and the details of recombination. There are also other effects, summarised in Zaldarriaga (2004).
Further, there is also some additional contribution to the temperature power spectrum from tensor perturbations. Since (as shown in Section 1.5.5) the tensor spectrum is largest at very low-$\ell$, the effect of this contribution is to increase the observed power spectrum at values of $\ell$ below $\sim 100$. This can be used to constrain the value of $r$ to a certain extent, and WMAP have used this method to determine that $r < 0.3$. However, due to uncertainties in the expected scalar power spectrum and the effects of cosmic variance, this limit cannot be improved on significantly by measurements of the temperature power spectrum and, thus, to obtain more competitive limits on the value of $r$ measurements of the primordial B-mode power spectrum (which is produced only by tensor perturbations) will be necessary.

1.5.4 Generation of polarisation in the CMB

Polarisation in the CMB is produced by Thomson scattering of anisotropic radiation. For this to happen, the incoming radiation intensity must have a quadrupole moment. There are several different ways for this to be produced (see Zaldarriaga 2004), but only two of these are particularly important. The main contribution (as described in Zaldarriaga 2004) is from the velocity gradients in the photon-baryon fluid that exists prior to recombination. However, for this to occur, the photons need to have a mean free path that is large enough that the difference between the fluid velocity at successive scatterings is significant. Polarisation can therefore only be generated when the mean free path is moderately large. However, to generate polarisation, Thomson scattering must occur, and this requires a reasonably small mean free path (by definition). As a result, polarisation can only be generated in a relatively narrow region of time around recombination, and thus the fractional polarisation of the CMB is only about 10% of the temperature anisotropies.

There are equations analogous to Equation (1.56), for each set of $a_{lm}$ coefficients (Liddle and Lyth 2000). Using these, I get an equation for $C_{\ell}^{XY}$

$$C_{\ell}^{XY} = 4\pi \int_0^{\infty} T_X(k, l)T_Y(k, l)P_{XY}(k) \frac{dk}{k},$$

(1.65)
where \( P_{XY} = P_R + P_T \) for \( X, Y = T, E \), \( P_T \) for \( X = Y = B \) and 0 otherwise. Therefore, we need to find the transfer function. Firstly I will consider E-mode polarisation. Since the power spectrum \( P_{XY} \) is the same as for the temperature case, the only difference between TT, TE and EE modes is the transfer function. Once again, Zaldarriaga (2004) gives the theory behind the oscillations. For polarisation, most of the power is generated by velocity perturbations, since polarisation requires a quadrupole moment in the density, which (to first order in \( k\delta\tau_R \)) is only generated by velocity perturbations.

Therefore

\[
\ell(\ell + 1)C_{\ell}^{EE} \approx \frac{8\pi^2}{9} \left( \frac{\ell}{\eta_0} \right)^3 \mathcal{A} \left( \frac{\ell}{\eta_0} \right)^2 \kappa^2 \left( \frac{\ell\eta_{dec}}{\eta_0} \delta\tau_R \right)^2 \sin^2 \left( \frac{2\pi}{\sqrt{3}} \left( \frac{\ell\eta_{dec}}{\eta_0} - 1 \right) \right),
\]  

(1.66)

where \( \kappa \) is a numerical constant that depends on the visibility function of the photon-baryon fluid and \( \delta\tau_R \) is the length of recombination. This equation has an extra factor of \( \ell^2 \) compared to the equation for TT, due to the requirement for anisotropic radiation to generate polarisation. Wavelengths much smaller than the mean free path are once again damped by Silk damping, resulting in the spectrum peaking on the scale of the mean free path of the photons. Since density perturbations produce very little polarisation (Zaldarriaga 2004), due to the requirement for a quadrupole moment, the oscillations are much sharper in the polarisation case than in the temperature case.

The TE power spectrum can be understood by combining the transfer functions for the temperature and polarisation cases. This gives a spectrum

\[
C_{\ell}^{TE} \sim \sin \left( \frac{2\pi}{\sqrt{3}} \left( \frac{\ell\eta_{dec}}{\eta_0} - 1 \right) \right) \cos \left( \frac{2\pi}{\sqrt{3}} \left( \frac{\ell\eta_{dec}}{\eta_0} - 1 \right) \right),
\]  

(1.67)

and thus TE has its minima and maxima halfway between the TT minima/maxima. The resultant spectrum is shown in Figure 1.2.

### 1.5.5 Anisotropies from tensor perturbations

In addition to the anisotropies generated from scalar perturbations that have been discussed in the previous sections, there are also anisotropies generated by tensor perturbations, with a power spectrum given by Equation (1.50). As a result, the tensor
power spectrum is lower than the scalar power spectrum by a factor of roughly $r$, the tensor-to-scalar ratio. However, because tensor perturbations are generated by gravitational waves which were produced in the inflationary epoch, they are negligible at small scales (due to the decay of the amplitude once the waves are inside the horizon) and, thus, the tensor power spectrum is highest at very large scales, and falls off significantly by $\ell = 100$. Further, unlike scalar perturbations, tensor modes are not restricted to producing anisotropies with only electric-type parity. For temperature measurements, as shown in section 1.5.3, this is not significant, but in the case of polarisation it is extremely important, since the restriction on the parity of scalar modes means that they can only generate E-mode polarisation and not B-mode polarisation, and this, in turn, gives us a strong potential constraint on the intensity of gravitational waves generated by inflation and, thus, of the energy level at which inflation occurred.

### 1.5.6 B-mode power spectrum

Finally, I will consider the B-mode spectrum. This is split into two parts, the primordial part, and the part due to gravitational lensing. The primordial part is, as stated above, generated by gravitational waves from the inflationary epoch. As a result, the power spectrum of primordial B-modes depends on the tensor power spectrum rather than the scalar power spectrum and, thus, the constant $A$ and the spectral index also depend on that of the tensor power spectrum, as given in Equation (1.50). Further, as shown in section 1.5.5, there is little power in the tensor spectrum at high-$\ell$ and, thus, the B-mode spectrum peaks at around $\ell = 80$. However, the acoustic oscillations in the transfer function still follow a similar pattern to the scalar case, and thus the primordial B-mode spectrum still has its peaks and troughs in roughly the same place as the EE spectra.

The most important result for B-mode polarisation is that we can relate, through the gravitational waves, the amplitude of the peak in the B-mode spectrum to the energy
1.5: INFLATIONARY THEORY AND CMB ANISOTROPIES

scale of inflation \( (Zaldarriaga 2004) \)

\[
\sqrt{\frac{\ell(\ell + 1)C_{\ell}^{BB}}{2\pi}} = 0.024 \left( \frac{E_{\text{inf}}}{10^{16} \text{GeV}} \right)^2 \mu K.
\] (1.68)

The gravitational lensing component is caused by gravity acting to slightly change the direction of CMB photons, which allows them to pick up a B-mode component. However, this component is not expected to follow Gaussian statistics, since it is a product of two different (Gaussian) fields, and so it may be possible to use the non-Gaussianity to clean the B-mode maps and recover the primordial B-mode spectrum, at least partially. The gravitational lensing effect is dominant on small scales, due to the fact that it is generated from lensing of E-modes, the spectrum of which peaks at much higher-\( \ell \) (~ 1000) than the primordial B-mode spectrum. As a result, the gravitational lensing component swamps the primordial B-modes at high-\( \ell \). The BB spectrum with \( r = 0.1 \) (along with the EE spectrum for comparison) is shown in Figure 1.2.

MICHAEL PREECE

45
Figure 1.2: EE (left) and BB (right) power spectra for $r = 0.1$, calculated using CAMB with the parameters given at the beginning of Chapter 2. On the BB spectrum, the black line is the primordial power spectrum, and the dashed red line is the gravitational lensing component.
Simulations of the detection of B-modes using the Pure-$C_\ell$ method

As shown in Section 1.3, in order to measure CMB polarisation in the case of a realistic experiment (where the CMB can only reasonably be measured over part of the sky), a method is required to estimate the true power spectrum of the CMB from the masked, noisy CMB map that comes from the experiments. For such a method to be useful, the statistics of the resultant power spectrum also need to be calculated, either using theoretical formulae or simulations, in order that the errors in the observed spectrum can be estimated. In reality, whilst theoretical calculations can be useful for initial studies into the likely usefulness of an experiment, and for attempting to determine the optimal scan strategy or parameters of the experiment (see Chapter 3 for examples), real experiments are always far too complex to calculate accurate theoretical errors, and thus in order to determine the true properties of such experiments, Monte-Carlo simulations are required.

There are many different methods of computing the CMB power spectrum (or, more generally, cosmological parameters) from a masked CMB map. However, the method that I have mainly concentrated on for the purpose of my thesis is the pseudo-$C_\ell$ method, which is outlined in Section 1.3 and variants thereof. In particular, I have
looked at the levels to which B-mode polarisation could potentially be measured using this method.

The pseudo-$C_\ell$ method is a very fast and simple to implement method of calculating CMB power spectra from the power spectrum of a masked map, and for temperature maps provides an estimate with an error that is close to optimal. However, since the transformation from $Q$ and $U$ Stokes’ parameters to E and B modes is non-local, cutting out part of the sky introduces mixing between them. This is not a problem for measuring E-modes, since these are considerably stronger. However, for B-mode detection, this mixing introduces an additional error term into the pseudo-$C_\ell$ formulation, resulting in a substantial degradation in the ability to detect B-mode polarisation (Challinor and Chon 2005). This is, potentially, a significant issue for the detection of B-modes, because this mixing results in a mask-dependant minimum value of $r$ which can be detected, with $r = 0.05$ being the minimum detectable for a survey with $f_{\text{sky}} = 0.01$, even in the absence of noise or other external sources of error. Fortunately, this can be corrected using the pure-$C_\ell$ formulation (Smith (2006), Smith and Zaldarriaga (2007)), which removes E-B mixing by cutting out ambiguous modes. However, whilst this removes E-B mixing, it also adds an additional error, proportional to the boundary size, which affects the low-$\ell$ performance of the method (although usually not to the extent that E-B mixing affects ordinary pseudo-$C_\ell$), as well as favouring masks with shorter boundaries.

In this chapter, I will begin by describing the code that I have developed to carry out Monte-Carlo simulations of polarised, masked CMB realisations and determine the mean and standard deviation of the power spectrum that is estimated using the pseudo-$C_\ell$ method, and the methods I used in doing so. Then, I will demonstrate that the pseudo-$C_\ell$ method is not adequate for many future experiments and, following on from this, I will describe an alternative method, called the pure-$C_\ell$ method (Smith 2006), which which is the main topic of my thesis. I will outline the differences in my implementation relative to the pseudo-$C_\ell$ method and qualitatively explain how it fixes the issues with the pseudo-$C_\ell$ method. Then, I will compare the two methods and
2.1: MONTE-CARLO SIMULATIONS OF THE STANDARD PSEUDO-$C_\ell$ METHOD

demonstrate the effectiveness of the pure-$C_\ell$ formulation.

Unless otherwise stated, throughout this thesis I have been using CMB power spectra generated by the CAMB package (Lewis et al. 2000), using a standard CMB model with $\Omega_b = 0.022$, $\Omega_{CDM} = 0.12$, $\Omega_k = 0$, $H_0 = 70$ km s$^{-1}$Mpc$^{-1}$, $n_S = 1$ and $n_T = 0$. The relative strength of the tensor modes is set using the parameter $r$, which is defined as the ratio of the tensor and scalar amplitudes, $r = A_T/A_S$, and a range of values have been used in this thesis.

2.1 Monte-Carlo simulations of the standard pseudo-$C_\ell$ method

There are several potential methods for analysing CMB maps. The optimal (but not necessarily unbiased) method is the Maximum Likelihood method, which produces the estimator with the lowest possible error. However, this method scales as $N_{\text{pix}}^3$, and thus cannot feasibly be applied to large maps, although it can potentially be used at very large scales. The standard method for analysing CMB maps in order to extract power spectra is the pseudo-$C_\ell$ method (Peebles 1973), sometimes in combination with the use of the maximum-likelihood method for low-$\ell$ (Efstathiou 2006). However, as with any method of analysis, there are issues with this method. In this case, the issue is that E and B mode polarisation cannot be perfectly separated on a partial sky (Challinor and Chon 2005), and thus mixing occurs between them. In order to test this effect, I must perform simulations of experiments using the method.

In this section, I will outline the method used to calculate the statistics for a given mask using Monte-Carlo simulations. In order to do this, I need to be able to simulate, in a simplified manner, the pipeline for a real experiment, with the addition of the ability to generate random CMB maps with the correct statistics, to repeat this simulation for a sufficiently large number of times to obtain good estimates of the mean and variance of the pseudo-$C_\ell$ estimator (unless otherwise specified, I have used 1000 real-
isations, since this gives an error in the mean of 3% of the standard deviation, which is sufficient for my purposes, but does not take too long to run) and to have a method of calculating the statistics of the CMB from these realisations. As is standard for CMB analysis (see Section 1.3), I have defined \( \langle T_\ell \rangle = \ell (\ell + 1) C_\ell / 2 \pi \), and this quantity, rather than the original power spectrum, \( C_\ell \), is ultimately output by my program and used in all analysis.

### 2.1.1 Generation of masked CMB maps

#### Creation of realisations

The first requirement for my program is to generate CMB realisations from given power spectra. In general, it is expected that the CMB will be approximately Gaussian, and due to their opposite parity, there is expected to be no cross-correlation between B and either T or E. It is expected, therefore, that all the information in the CMB can be described by the TT, EE, TE and BB power spectra. Given these four spectra, CMB realisations can be constructed by assuming that the \( a_{\ell m} \) values are Gaussian with mean zero and variance \( C_\ell \), which is accurate to the levels required for my simulations. To take account of the TE cross-correlation, I define

\[
a_{\ell m}^T = \zeta_1 \sqrt{C_\ell^{TT}}, \quad a_{\ell m}^E = \frac{C_\ell^{TE}}{\sqrt{C_\ell^{TT}}} + \zeta_2 \sqrt{C_\ell^{EE} - \left( \frac{C_\ell^{TE}}{C_\ell^{TT}} \right)^2},
\]

\[
a_{\ell m}^B = \zeta_3 \sqrt{C_\ell^{BB}},
\]

where \( \zeta_1, \zeta_2, \) and \( \zeta_3 \) are complex Gaussian random variables with unit variance, except for \( m = 0 \), where the symmetry properties of the \( a_{\ell m} \)s (Equation (1.6)) imply that they are purely real. Note that Equations (1.6) and (1.21) imply that the \( a_{\ell m} \) values for \( m < 0 \) can be calculated in terms of those for \( m > 0 \), and thus only the \( a_{\ell m} \) values for \( m \geq 0 \).
need be stored, with the empirical power spectrum then given by
\[ C_{\ell}^{XY} = \frac{1}{2\ell + 1} \left( a_{\ell0}^X a_{\ell0}^Y + 2 \sum_{m=1}^{\ell} \Re(a_{\ell m}^X (a_{\ell m}^Y)^*) \right), \quad (2.4) \]

where, as stated above, \( a_{\ell 0} \) must be real.

Once I have calculated these \( a_{\ell m} \) values, I can then convert them into a CMB map using the HEALPix alm2map routine (Górski et al. 2005), resulting in a CMB map which will have the same statistical properties as a real CMB map with the same initial power spectrum.

**Masking and calculation of pseudo-\( C_\ell \) values**

However, merely computing a CMB map is not sufficient to model any realistic experiment since, due to foreground contamination and partial-sky coverage, no real experiment can use data taken over the whole sky. Thus, it is necessary to consider the effects of partial sky observations on the levels of error. Therefore, for each realisation I apply a mask to the calculated CMB map, to get a masked map that represents the data that would be used by a real experiment to determine the power spectrum. The observed CMB map is computed from the input CMB map using the following definition
\[
\tilde{T}(\Omega) = T(\Omega) W_T(\Omega), \quad (2.5)
\]
\[
\tilde{Q}(\Omega) = Q(\Omega) W_P(\Omega), \quad (2.6)
\]
\[
\tilde{U}(\Omega) = U(\Omega) W_P(\Omega), \quad (2.7)
\]

where the weight functions \( W_T \) (for temperature) and \( W_P \) (for polarisation) describe the mask (in almost all cases I have set \( W_T = W_P \)). In regions which are not observed, or where foreground contamination is too large, the weight function is set to zero, so that observations from that region are not included in the power spectrum computation. In observed, uncontaminated regions, the weight function can either be set uniformly to one, or some form of apodisation (a process where the mask is smoothed such that it changes continuously from 1 in the centre to 0 in the masked region) or statistical...
weighting can be used. For temperature and E-mode measurements, the best option is generally to set the weight function uniformly to one. However, it can be shown (Challinor and Chon 2005) that, because of issues with E-B mixing, B-mode measurements are most effective when the weight functions are smooth, and thus to detect B-modes to the highest possible level of accuracy, apodisation must be applied to the masks.

Once the mask has been applied, the pseudo-$C_{\ell}$ power spectrum of the resultant convolved map can be computed. However, as shown in Section 1.3, this spectrum is not an unbiased estimator of the original CMB power spectrum. Therefore, a method to retrieve an estimate for the original power spectrum from the pseudo-$C_{\ell}$ spectrum is required.

### 2.1.2 Deconvolution of masked CMB maps

As discussed in Section 1.3 and above, on a cut sky, it is necessary to use the pseudo-$C_{\ell}$ method to calculate the CMB power spectrum in the case where only part of the sky is included in the analysis. However, once the pseudo-$C_{\ell}$ values have been calculated, they must be deconvolved so as to remove the effects of the mask and obtain the correct power spectrum, and this section outlines the method used to do this.

In Section 1.3, it was shown that, in order to compute the real CMB power spectrum from the pseudo-$C_{\ell}$ power spectrum, the mask needs to be taken into account. From the definition of the the pseudo-$a_{\ell m}$s, it can be shown (Section 1.3) that

$$\langle \tilde{C}_{\ell}^{XY} \rangle = \sum_{A,B} \sum_{\ell'} M_{\ell \ell'}^{XYAB} \hat{C}_{\ell}^{AB}. \quad (2.8)$$

Where $\hat{C}_{\ell}$ is an unbiased estimator for the true power spectrum. Therefore, in order to compute the original power spectrum, I must to compute and invert the mask matrix in order to deconvolve the effect of the mask and, thus, determine an unbiased estimate of the original input CMB. There are two steps to this. Firstly, the mask matrix must be computed, and secondly Equation (2.8) needs to be inverted to solve for the estimated power spectrum.
Calculation of mask matrix

Because the mask matrix depends only on the mask and not on the input power spectrum, it only needs to be calculated once for each individual mask. Thus, rather than calculating it as part of the main program, I instead calculate it separately, store the matrix itself in a data file and then read the data from the file each time I run the main program. To do this, I require an explicit formula for the mask matrix. This can be computed from the definition of the pseudo-\(C_\ell\) values, coupled with various properties of the spherical harmonics.

The calculation of the mask matrix is carried out in Brown et al. (2005), with the result for the temperature component being

\[
M^{TT,TT}_{\ell\ell'} = \frac{(2\ell + 1)}{4\pi} \sum_{\ell''} W^{TT}_{\ell\ell'} \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{array} \right)^2, \tag{2.9}
\]

where the coefficients \(W^{TT}_{\ell\ell'}\) are the \(C_\ell\)s for the temperature mask

\[
W^{TT}_{\ell\ell'} = \frac{1}{2\ell + 1} \left( \sum_{m=-\ell}^{\ell} |w^{TT}_{\ell m}|^2 \right), \tag{2.10}
\]

where

\[
w^{TT}_{\ell m} = \int W_T(\Omega) Y^{*}_{\ell m}(\Omega) d\Omega,
\]

and

\[
\left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ m & m' & m'' \end{array} \right)
\]

is the Wigner 3-j symbol. Note that for \(m = m' = m'' = 0\), the Wigner 3-j symbol is zero for \(\ell + \ell' + \ell''\) odd.

The other terms (\(M^{TT,TT}, M^{TE,TE} = M^{TB,TT}, M^{EE,EE} = M^{BB,BB}, M^{EE,BB} = M^{BB,EE}\) and \(M^{EB,EB}\)) are also given in Brown et al. (2005). In particular, the terms involving EE and BB are given by

\[
M^{PP,PP'}_{\ell\ell'} = \frac{(2\ell + 1)}{8\pi} \sum_{\ell''} W^{PP}_{\ell\ell'} \left( (-1)^{(\ell+\ell'+\ell'')} \pm 1 \right) \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{array} \right)^2, \tag{2.12}
\]
where $W_{\ell}^{pp}$ is the power spectrum of the polarisation mask and the $\pm$ sign is a plus for $P = P'$ and a minus for $P \neq P'$ ($P = E$ or $B$). Note that for the cross term, $M^{EE, BB}$, all terms with $\ell + \ell' + \ell''$ even are zero and, consequentially, the $\ell = \ell'$ term is also zero, hence this term is generally much smaller than the $M_{pp, pp}$ term. As a result, it is generally negligible when considering the EE power spectrum, which has much more power than the BB spectrum. However, it is not negligible when calculating the BB power spectrum, since the small size of the multiplying factor is cancelled out by the fact that the EE power spectrum is much larger.

Given Equation (2.12), and the input mask, it is then possible to compute the mask matrix. In order to do this, I first compute the power spectrum of the mask, using HEALPix. Then, I loop over all possible values of $\ell''$ (up to twice the input $C_{\ell}$ value), using the function DRC3JM\(^1\) to compute the values of the 3j function for each set of $\ell$ values, and calculate the mask matrix from this.

### Deconvolution of pseudo-$C_{\ell}$ values

Once I have the mask matrix and masked pseudo-$C_{\ell}$ power spectrum, I must then solve Equation (2.8) in order to determine the estimated CMB power spectrum for a given realisation. In order to do this, I require a routine to solve efficiently the matrix equation given. For the case of the TT power spectrum, and indeed for all spectra other than the EE and BB ones, there are no cross-terms, and thus this is simply a standard two-dimensional matrix equation. However, because of E-B mixing, the case of the EE and BB power spectra is more complicated. I will first deal with the simpler case, which applies to the temperature power spectrum and all the cross power-spectra.

In the temperature (and cross-spectra) case, I must solve a simple matrix equation of the form $\tilde{C}_{\ell} = M_{\ell' \ell} \hat{C}_{\ell'}$. In order to do this, I use two functions, taken from Press et al. (1992). The first one, called ludcmp, performs a UD decomposition on the matrix, finding matrices $U$ and $D$ such that $M = UD$, where $U$ is an upper diagonal matrix and $D$ a lower diagonal one. Given this, it is then trivial to solve equations of the form $M \tilde{C}_{\ell} = C_{\ell}$.

\(^1\)SLATEC mathematical library - http://www.netlib.org/slatec/index.html
the form $y = U Dx$ by solving $y = U z$ followed by $z = Dx$, both of which can quickly be solved using simple substitution, performed using the routine lubksb.

The advantage of this method is that the (slow) factorisation process need only be applied to the matrix $M$ once, and then any equation of the form $y = M x$ can be efficiently solved. In order to take full advantage of this fact, rather than running the deconvolution program after each iteration of the Monte-Carlo process, I instead store the calculated pseudo-$C_\ell$ values for each iteration in an array, and then deconvolve them all at the end. This enables me to compute an estimate for the CMB power spectrum for each realisation (of which there are usually 1000) whilst only being required to run the factorisation algorithm once for each matrix.

However, in order to calculate the EE and BB power spectra, I must take into account the E-B mixing term. In this case, the following equations are obtained

$$\tilde{C}^{EE}_\ell = M^{EEE}_{\ell\ell'} \hat{C}^{EE}_{\ell'} + M^{EEBB}_{\ell\ell'} \hat{C}^{BB}_{\ell'}, \quad (2.13)$$

$$\tilde{C}^{BB}_\ell = M^{BBEE}_{\ell\ell'} \hat{C}^{EE}_{\ell'} + M^{BBBB}_{\ell\ell'} \hat{C}^{BB}_{\ell'}. \quad (2.14)$$

As a result of this mixing, solving for the EE and BB power spectra is more complicated than in the other cases. In order to obtain the solution, these equations must first be rearranged to give $\hat{C}^{EE}_\ell$ in terms of $\tilde{C}^{EE}_\ell$, $\hat{C}^{BB}_\ell$ and the inverse of the mask matrix. Then, these values can be plugged into the expression for $\hat{C}^{BB}_\ell$, which can then be solved for. This requires the equation solving routine to be used more often, to compute the inverse of the mask matrix. However, because the slow step is the factorisation, which need only be done once, this calculation is not substantially slower than the temperature case.

Once this has been done, the resultant output should, in theory, be an array containing an ensemble of unbiased estimates of the input CMB power spectrum. However, in general, whilst this method is, in principle, capable of working for arbitrarily small cuts, in practice the matrices become almost singular when $f_{\text{sky}}$ is small, due to large amounts of mixing between different $C_\ell$ values and, in this case, it is necessary to combine several of the values together to produce a single binned value.
Small $f_{\text{sky}}$ cuts and binning

For a given value of $f_{\text{sky}}$, the pseudo-$C_\ell$ values are correlated over a range of roughly $\Delta l = 1/\sqrt{f_{\text{sky}}}$, since the mask power spectrum, which determines the mask matrix and, thus, the level of correlation between different $\ell$ values, is generally band limited to $\ell$ values within $\Delta l$ of each other. For small $f_{\text{sky}}$ values, this correlation causes very large errors to appear in the individual $C_\ell$ values, resulting in estimates with very large error bars and, consequentially, large fluctuations in the mean values. This leads to data which is totally unusable. This problem can be eliminated by combining the pseudo-$C_\ell$ values that are output by each realisation into bins, of size approximately $\Delta l$ (Brown et al. 2005). Provided that the bin size is less than $\sim \Delta l$, then little actual information is lost, since the correlation between the pseudo-$C_\ell$ values means that little useful information remains in the power spectrum at these levels.

The $C_\ell$ values need to be combined into bins. The general way of doing this is by defining a binned $C_\ell$

$$C'_b = \sum_\ell P_{b\ell} C_\ell .$$

(2.15)

The most commonly used definition of the binned $C_\ell$s is

$$C'_b = \sum_{\ell = (b-1)\Delta l+2}^{b\Delta l+1} \frac{\ell(\ell + 1)}{2\pi \Delta l} C_\ell ,$$

(2.16)

where the $\ell(\ell + 1)$ term ensures that the term summed over is (approximately) flat throughout the bin (necessary to allow the deconvolution process to work correctly), and the usage of $+1$ in the upper limit is so as to exclude $\ell = 0$ and $\ell = 1$ from the bins. The mask matrix and pseudo-$C_\ell$ must also be binned. The binned pseudo-$C_\ell$s can be defined in any way we like, so I will define them such that the mask matrix can simply be written as

$$M'_{bb'} = \sum_{\ell = (b'-1)\Delta l + 2}^{b'\Delta l + 1} \sum_{\ell' = (b-1)\Delta l + 2}^{b\Delta l + 1} \frac{\ell(\ell + 1)M_{\ell\ell'}}{\ell'(\ell' + 1)\Delta l} ,$$

(2.17)

which corresponds to defining the binned pseudo-$C_\ell$s such that $\tilde{C}_b = \sum \frac{\ell(\ell + 1)}{2\pi} \tilde{C}_\ell$. Using
these methods, it is possible to deconvolve the pseudo-$C_\ell$ values and thus recover an unbiased estimate of the true power spectrum.

2.1.3 Computation of statistics for output power spectrum

The deconvolution routine of my program produces an array of power spectrum estimates, one for each realisation. Using this data, I can compute the mean power spectrum, the standard deviation in the estimate of the spectrum, the covariance matrix of the power spectra and the likelihood function of the power spectrum. The mean merely acts as a consistency check. Since I know the power spectrum that was initially input into the routine, provided the algorithm is working correctly (and provided the pseudo-$C_\ell$ method itself produces an unbiased estimator of the power spectrum), the output mean should be the same as the input spectrum, to within the standard $\sqrt{n}$ error margin.

However, of far more importance is the standard deviation. Since each individual realisation is a simulation of a potential real experiment (with the only difference being the input CMB, which is randomly generated each time from the given power spectrum), the standard deviation of the simulations represents the intrinsic random error (due to masking, issues with the pseudo-$C_\ell$ method and the effects of any applied noise or beam smoothing) in the CMB power spectrum as measured by the experiment. Thus, the level of accuracy to which the real CMB power spectrum can be determined by an experiment with the given mask and noise properties. Thus, this algorithm makes it possible to determine the accuracy of any experiment and, thus, to determine the level of B-modes which an experiment could potentially detect if the data were analysed using the pseudo-$C_\ell$ method.

2.1.4 Addition of noise into simulations

So far, I have considered simulations without the effects of noise or telescope beams. However, in order to simulate realistic experimental data, the ability to add noise to
maps and remove it from the power spectrum is required. Addition of noise and beam effects to maps is generally reasonably trivial, at least in the case of uncorrelated noise. However, since considering the effects of realistic noise is not a major part of my thesis, my program is only capable of adding uniform, uncorrelated noise to a map, along with a circular beam. The methodology for doing this differs depending on the type of noise. For uniform noise (which does not necessarily have to be the same for different power spectra, so e.g. the TT noise could be different from the EE noise), the simplest way of doing this is to treat the noise power spectrum as a separate entity, generate a set of $a_{lm}$ values for it and then add them to the $a_{lm}s$ already generated for the CMB. Beams are added by simply multiplying the input $C_\ell$ values by the beam window function. Non-uniform noise, however, would have to be added pixel by pixel (since it cannot be completely described by a power spectrum in the manner that white noise can be), although this is not a significant issue since maps must be produced in any case. Also, currently, my program is only capable of adding uniform noise to TT, TE, EE and BB (due to the method used).

For the case without masking, noise removal is trivial (although the fact that only the mean value of the noise spectrum is known provides an additional source of error), and the expected errors can be exactly calculated. This case is discussed in Section 1.4.3. Removing the noise in the case of a masked map is slightly more complicated, since the noise and beam effects are mixed with the CMB power spectrum to generate the pseudo-$C_\ell$ values. However, this can easily be worked around by simply deconvolving the pseudo-$C_\ell$ values, and then removing the noise and beam effects.

However, the case where the $C_\ell$ values need to be binned to obtain a sensible result is considerably more difficult to handle, since the noise and beam functions are $\ell$-dependent, and so cannot be removed from the binned $C_\ell$ values after deconvolution without causing a potential bias in the estimator. In this case, therefore, the noise must be removed during the deconvolution process itself. However, since how the noise affects the $C_\ell$ values is known, this is easily achievable. First, I will consider the unbinned case. Using Equation (1.36), it is possible to write an expression for the...
output of the deconvolution routine (which will include the noise) in terms of $C_\ell/2\pi$ and the noise power spectrum

$$\langle \tilde{C}^{XY}_\ell \rangle = \sum_{A,B} \sum_{\ell} M^{XYAB}_{\ell\ell'} (B_{\ell'})^2 C^{AB}_{\ell} + N^{AB}_{\ell},$$

(2.18)

where $N_\ell$ and $B_\ell$ are the noise and beam power spectra, as defined in Section 1.4.3.

Using this, it is found that

$$\tilde{C}^{XY}_\ell = \sum_{A,B} \sum_{\ell} R^{XYAB}_{\ell\ell'} C^{AB}_{\ell},$$

(2.19)

where I have defined

$$\tilde{C}^{XY}_\ell = \tilde{C}^{XY}_\ell - \sum_{A,B} \sum_{\ell} M^{XYAB}_{\ell\ell'} N^{AB}_{\ell},$$

(2.20)

and

$$R^{XYAB}_{\ell\ell'} = M^{XYAB}_{\ell\ell'} (B_{\ell'})^2.$$  

(2.21)

Thus, I have defined an effective pseudo-$C_\ell$ and mask matrix which allows us to determine the real $C_\ell$ values directly from the pseudo-$C_\ell$s. This is not particularly useful in the unbinned case, but can trivially be extended to binning using Equation (2.17), and, unlike post-deconvolution removal, this method will provide an unbiased estimate of the power spectrum.

### 2.2 Issues with the pseudo-$C_\ell$ method

In order to demonstrate the issues with the pseudo-$C_\ell$ method, I have carried out simulations comparing the level of error in the B-modes with that in the temperature power spectrum (the error in the E-mode spectrum is similar in nature to the error in the temperature, and thus is not included). For this purpose, I have used a simulation of the QUIJOTE experiment (Rubino-Martin et al. 2008) as an example of the potential issues that exist with the pseudo-$C_\ell$ method. I have plotted the signal to noise ratio for both the expected QUIJOTE mask with the WMAP polarisation mask applied (see Figure 2.1) and an spherical cap azimuthal mask with the same value of $f_{\text{sky}}$ (0.15) for
2: SIMULATIONS OF THE DETECTION OF B-MODES USING THE PURE-C_\ell METHOD

both the TT and BB spectra (with $r = 0.1$) both without binning and with a bin size of $\Delta l = 30$, along with the theoretical errors for each case, calculated using the naive optimal error formula shown later in Section 3.1.1. The results are shown in Figure 2.2

Figure 2.1: QUIJOTE mask in Galactic co-ordinates, using a Mollweide projection. Red areas are unmasked, blue areas are masked.

For the temperature case, the theoretical errors seem to work well in the binned case, although they are less accurate for the unbinned case, and even using the genuine mask (as opposed to the azimuthal equivalent) only worsens the results by a factor of 2 or so. For the B-modes, however, it is obvious that the optimal error formula is totally inadequate. At best, it gives an estimate for the error that is too low by a factor of 10, and even for the binned case with the azimuthal mask, there is a factor of 200 difference at high $\ell$ values. In fact, it appears that binning in this case does not significantly improve the difference between the optimal and simulated errors (although it has some effect for the azimuthal case at low $\ell$). This is because the binning fixes the problem with mixing between different $\ell$ values, whereas the source of the discrepancy here is in fact the E-B mixing term that arises in the deconvolution process. Whilst the
2.2: ISSUES WITH THE PSEUDO-$C_\ell$ METHOD

Figure 2.2: $\Delta C^{TT}_\ell / C^{TT}_\ell$ (left) and $\Delta C^{BB}_\ell / C^{BB}_\ell$ (right) without binning (top) and with a bin size of $\Delta l = 30$ (bottom) for the QUIJOTE mask (red line), an azimuthal mask with the same $f_{\text{sky}}$ (blue line) and the theoretical errors as given by Equation (3.2) (black line)
estimator for $C_{\ell}^{BB}$ is unbiased by construction, it is not true that, for a given realisation, $C_{\ell}^{BB} = 0 \Rightarrow \hat{C}_{\ell}^{BB} = 0$. As a result, the E-B mixing acts effectively as an additional source of noise. This term is largely independent of the bin size (and indeed the value of $C_{\ell}^{BB}$), and as a result, binning does not make much difference. Thus, it is clear that the standard pseudo-$C_{\ell}$ method is not adequate for the purpose of detecting B-mode polarisation. Further, it is obvious that, as expected, the azimuthal mask performs substantially better than the non-azimuthal one, even in the binned case, due to the much higher boundary length and, consequentially, larger amount of E-B mixing.

To demonstrate this effect further, I have considered the idea discussed in Carretti et al. (2006), which suggests searching for polarisation in the WMAP data by selecting the 20% of the sky with the lowest foreground contamination. To do this, I have produced graphs of the B-mode polarisation for a mask based on this idea (see Figure 2.3), together with an equatorial strip azimuthal mask of the same $f_{\text{sky}}$ (Figure 2.4). From these graphs, it can be seen that this idea does not work anywhere near as well as an argument using optimal errors would predict. In fact, it can be seen that except at very low $\ell$ values, the azimuthal mask is about 10 times better than the non-azimuthal mask. This seems to fit with the result from the QUIJOTE simulations, and indicates that this method of excluding foreground contamination in order to measure the CMB polarisation is unlikely to be exceptionally successful, at least using this power spectrum estimation methodology.

From these results, it is clear that, due to E-B mixing the standard pseudo-$C_{\ell}$ method is not sufficiently accurate to detect B-mode polarisation to the level that will be required for many future experiments. Indeed, Challinor and Chon (2005) have shown that, for a survey with an $f_{\text{sky}}$ of 0.01, the pseudo-$C_{\ell}$ method could only detect $r = 0.05$. Thus, a method of analysing the data which is not susceptible to the effects of E-B mixing is required. Smith (2006) have presented such a method, and I have produced an independent implementation of it, which will be outlined in the next section.
2.2: ISSUES WITH THE PSEUDO-C$_T$ METHOD

Figure 2.3: Mask showing the 20% of the sky with the lowest foreground contamination, in Galactic coordinates with a Mollweide projection.

Figure 2.4: Comparison between errors on BB spectrum for a mask based on Carretti et al. (2006) (right) and an azimuthal mask with the same $f_{\text{sky}}$ (left), for all values of $\ell$ up to $\ell = 20$ (top) and for multiples of 10 up to $\ell = 256$ (bottom)
2.3 Removing problems with E-B mixing

In the previous section, it has been shown that as a result of E-B mixing, the standard pseudo-$C_\ell$ method is highly non-optimal for the purpose of detecting B-mode polarisation. Whilst the power spectrum could, in theory, be computed using the Maximum Likelihood (ML) method which will, by definition, give you the minimum possible error for a given spectrum, this method is extremely slow, meaning that it is only viable for calculations at low $\ell$.

Since the problem is mixing between E and B mode polarisation, if it were possible to prevent this from occurring, or else to remove the modes that had undergone mixing, then this problem would no longer exist. This is the principle behind the Pure-$C_\ell$ method, which is described in Smith (2006) and expanded on in Smith and Zaldarriaga (2007). This method defines an estimator for $C^{BB}_\ell$ that has strictly zero contribution from E-modes (as opposed to the statistically zero contribution that E-modes have in the standard pseudo-$C_\ell$ mechanism), and thus eliminates the problems with E-B mixing. In this section, I will explain how the pure-$C_\ell$ power spectrum is defined and how I have implemented it in my code.

2.3.1 Definition of pure-$C_\ell$

There are two different (but mathematically equivalent) methods of defining the pure-$C_\ell$ power spectrum. These two methods are given below.

Counter-term formalism

The first, described in Smith (2006), is the counter-term formalism. First, define the E and B modes using the E and B operators $E_{ab}$ and $B_{ab}$, which are defined in Smith (2006) to be

\begin{align}
E_{ab} &= -\nabla_a \nabla_b + \frac{1}{2} g_{ab} \nabla^2 \\
B_{ab} &= \frac{1}{2} \epsilon_{ac} \nabla^c \nabla_b + \frac{1}{2} \epsilon_{bc} \nabla^c \nabla_a
\end{align}

(2.22) (2.23)
where \( g_{ab} \) is the all-sky metric and \( \epsilon_{ab} \) is the antisymmetric tensor, and which have even and odd parity, respectively. Then E and B spherical harmonic functions can be defined as

\[
Y^E_{(\ell m)ab} = \frac{1}{\sqrt{(\ell - 1)(\ell + 1)(\ell + 2)}} \epsilon_{ab} Y^{(2.24)}_{\ell m},
\]

\[
Y^B_{(\ell m)ab} = \frac{1}{\sqrt{(\ell - 1)(\ell + 1)(\ell + 2)}} B_{ab} Y^{(2.25)}_{\ell m},
\]

which allows the E and B multipoles to be defined as

\[
a^P_{\ell m} = \int \Pi^a_{ab}(\Omega) Y^P_{(\ell m)ab}(\Omega) d\Omega,
\]

where \( \Pi^a_{ab} \) is the polarisation tensor and \( P = E, B \).

Now suppose part of the map is masked. This is equivalent to replacing \( \Pi^a_{ab} \) with \( W(\Omega) \Pi^a_{ab} \). However, by writing the pseudo-\( a_{\ell m} \)s in terms of \( B_{ab} \), I find that

\[
\tilde{a}^B_{\ell m} = \frac{1}{\sqrt{\ell - 1} \ell (\ell + 1) (\ell + 2)} \int \Pi^{ab}(\Omega) W(\Omega) B_{ab} Y^*_m(\Omega) d\Omega.
\]

Since multiplication in position space causes mixing of E and B modes, this quantity is a mixture of E and B modes, rather than being pure-B, as the original \( a_{\ell m} \)s are. This problem is the cause of E-B mixing, and hence the extreme sub-optimality of the pseudo-\( C_\ell \) method when applied to detecting B-modes. However, modifying the definition slightly to

\[
\tilde{a}^{B,\text{pure}}_{\ell m} = \frac{1}{\sqrt{\ell - 1} \ell (\ell + 1) (\ell + 2)} \int \Pi^{ab}(\Omega) B_{ab}(W(\Omega) Y^*_m(\Omega)) d\Omega,
\]

allows it to be shown \textbf{[Smith 2006]} that, provided the function \( W \) and its derivative vanish at the boundary (so again, apodisation is required) for a noiseless CMB realisation containing only E-modes, \( \tilde{a}^{B,\text{pure}}_{\ell m} \) will always be zero. As a result, there is no leakage from E into B, and thus the errors in B are considerably lower. In effect, this is removing ‘ambiguous’ modes (i.e. those which could come from either E or B) from the estimator for the B-mode power. Note that whilst an equivalent definition can easily be applied to the E-mode spectrum, in this case it would actually increase the...
error, since the effect of E-B mixing on E-mode measurements is negligible. Thus, the ordinary definition of the E-mode power spectrum is still used.

By expanding the above expression, I can write

\[ \tilde{a}^B_{\ell m, \text{pure}} = \tilde{a}^B_{\ell m} + \tilde{a}^C_{\ell m} + \tilde{a}^S_{\ell m}, \tag{2.29} \]

where \( \tilde{a}^B_{\ell m} \) are the ordinary pseudo-\( C_\ell \) \( a_{\ell m} \)s, given by

\[ \tilde{a}^B_{\ell m} = \frac{1}{2i} \int P(\Omega) Y^*_{\ell m}(\Omega) d\Omega - \int P^*(\Omega) Y_{\ell m}(\Omega) d\Omega, \tag{2.30} \]

and \( \tilde{a}^C_{\ell m} \) and \( \tilde{a}^S_{\ell m} \) are, respectively, spin-1 and spin-0 counter-terms (the \( C \) stands for “curl”, and this is the spin-1 equivalent of \( B \) (the spin-1 equivalent of \( E \), labelled \( G \) for “gradient” will also be used later), and the \( S \) stands for “scalar”), given by

\[ \tilde{a}^C_{\ell m} = \frac{N_\ell}{2i} \int [1 \tilde{P}(\Omega) Y^*_{\ell m}(\Omega) d\Omega + \int [1 \tilde{P}^*(\Omega) Y_{\ell m}(\Omega) d\Omega, \tag{2.31} \]

\[ \tilde{a}^S_{\ell m} = \frac{N'_\ell}{2i} \int (2 \tilde{P}(\Omega) - 2 \tilde{P}^*(\Omega)) Y^*_{\ell m}(\Omega) d\Omega, \tag{2.32} \]

where I have defined

\[ \tilde{P}(\Omega) = \mathcal{W}(\Omega)(Q(\Omega) + iU(\Omega)), \tag{2.33} \]

with \( \mathcal{W}(\Omega) \) being the spin-\( s \) counter-term

\[ \mathcal{W}(\Omega) = \begin{cases} \delta^s\mathcal{W}(\Omega) & s \geq 0 \\ \delta^{-s}\mathcal{W}(\Omega) & s < 0 \end{cases}, \tag{2.34} \]

and the normalisation factors being given by

\[ N_\ell = \frac{1}{\sqrt{(\ell - 1)\ell(\ell + 1)(\ell + 2)}}, \tag{2.35} \]

\[ N'_\ell = \frac{2}{\sqrt{(\ell - 1)(\ell + 2)}}, \tag{2.36} \]

Now, using the properties of the spin-weighted spherical harmonics, the pure-\( C_\ell \) \( a_{\ell m} \)s can be written in terms of the mask \( a_{\ell m} \)s and \( 3j \) functions. It is then possible to find ex-
expressions for the three terms (the original pseudo-$C_\ell$ term and the two counter-terms):

\[
\tilde{a}^B_{\ell m} = \frac{(-1)^m}{2} \sum_{\ell',\ell'',m',m''} \left( (1 + (-1)^{\ell + \ell' + \ell''}) a^B_{\ell'm'} + i(1 - (-1)^{\ell + \ell' + \ell''}) a^E_{\ell'm'} \right) \times w^P_{\ell',\ell'',m',m''} \begin{pmatrix} -2 \\ 2 \\ 0 \end{pmatrix}, \tag{2.37}
\]

\[
\tilde{a}^C_{\ell m} = \frac{(-1)^m N'_\ell}{2} \sum_{\ell',\ell'',m',m''} \left( (1 + (-1)^{\ell + \ell' + \ell''}) (a^B_{\ell'm'} w^G_{\ell'} - a^E_{\ell'm'} w^C_{\ell'}) \right) B^{-m',m''}_{\ell',\ell''} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}, \tag{2.38}
\]

\[
\tilde{a}^S_{\ell m} = \frac{(-1)^m N'_{\ell}}{2} \sum_{\ell',\ell'',m',m''} \left( (1 + (-1)^{\ell + \ell' + \ell''}) (a^B_{\ell'm'} w^E_{\ell'} - a^E_{\ell'm'} w^B_{\ell'}) \right) B^{-m',m''}_{\ell',\ell''} \begin{pmatrix} 0 \\ 2 \\ -2 \end{pmatrix}, \tag{2.39}
\]

where

\[
w^P_{\ell m} = 0 w_{\ell m} \tag{2.40}
\]

\[
w^G_{\ell m} = \frac{1}{2} w_{\ell m} - \frac{1}{2} w_{\ell m} \tag{2.41}
\]

\[
w^C_{\ell m} = \frac{1}{2i} w_{\ell m} + \frac{1}{2i} w_{\ell m} \tag{2.42}
\]

\[
w^E_{\ell m} = \frac{2}{2i} w_{\ell m} + \frac{2}{2i} w_{\ell m} \tag{2.43}
\]

\[
w^B_{\ell m} = \frac{2}{2i} w_{\ell m} - \frac{2}{2i} w_{\ell m} \tag{2.44}
\]

with $s w_{\ell m}$ being the $a_{\ell m}$s of the spin-s counter-term (see Section 2.3.2 for an explanation.
of the spin-1 terms), and

\[
\begin{pmatrix}
 s_1 \\
 s_2 \\
 s_3 \\
\end{pmatrix}
[
 \begin{align*}
 B_{\ell_1,\ell_2,\ell_3}^{m_1,m_2,m_3} &= \int s_1 Y_{\ell_1 m_1}(\Omega)s_2 Y_{\ell_2 m_2}(\Omega)s_3 Y_{\ell_3 m_3}(\Omega)d\Omega \\
 &= \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \left( \begin{array}{ccc}
 \ell_1 & \ell_2 & \ell_3 \\
 -s_1 & -s_2 & -s_3 \\
 m_1 & m_2 & m_3 \\
\end{array} \right) .
\end{align*}
\]

(2.45)

Note that these expressions still involve both E and B, and thus it does not appear that the resultant pure-\(C_\ell\) pseudo B-mode power spectrum will be independent of the E-modes, as required. However, I have shown (using the \(3j\) recursion relations) that, provided the counter-terms are defined as shown earlier, the E-mode terms cancel out when the three expressions are added together and, thus, the pure-\(C_\ell\) B-mode power spectrum is indeed independent of E-modes.

Given these definitions, the pure-\(C_\ell\) pseudo power spectrum can be trivially defined, and the mask matrix can also be computed.

\(\chi_B\) formalism

There is another, more intuitive, method by which the pure-\(C_\ell\) formalism can be derived, as shown in Smith and Zaldarriaga (2007). Suppose a scalar field, \(\chi_B\), is defined using

\[
\chi_B \equiv \frac{\delta \delta(Q + iU) - \delta \delta(Q - iU)}{2i}.
\]

Then, the field defined by \(\chi_B\) can be shown to be a pure-B field (Smith and Zaldarriaga 2007), meaning that it has no contribution from E-modes and, thus, will be zero in the absence of any B-modes in the input power spectrum. Further, since it is a scalar quantity, it can be handled using ordinary scalar spherical harmonics, as opposed to the spin-weighted spherical harmonics that are usually required when dealing with polarisation, which makes the maths easier. However, since \(\chi_B\) depends on the derivative of
the polarisation field, it is non-trivial to calculate for an observed CMB map (requiring the use of a finite differencing method on a noisy CMB map), and hence the counter-term formalism is easier to work with in the case of simulations. Now, I can define an equivalent to the pseudo-$a_{\ell m}$s for this field as

$$\tilde{a}_{\ell m} = \frac{1}{\sqrt{(\ell - 1)\ell(\ell + 1)(\ell + 2)}} \int \chi_B(\Omega) W(\Omega) Y_{\ell m}(\Omega) d\Omega. \quad (2.47)$$

Note that this is a scalar quantity, and it depends only on the mask matrix (rather than derivatives thereof, as in the case of the counter-term formalism), which makes this formula considerably easier to handle mathematically.

Now, on expanding this formula in terms of the mask, as was done in the last section, it is found that

$$\tilde{a}_{\ell m} = \sqrt{\frac{(\ell - 1)\ell(\ell + 1)(\ell + 2)}{(\ell - 1)\ell(\ell + 1)(\ell + 2)}} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \quad (2.48)$$

At first glance, this looks totally different from the formula given by Equation (2.29) and Equations (2.37), (2.38) and (2.39). It explicitly does not depend on the E-mode spectrum, and it is considerably simpler. However, I am able to prove, from the $3j$ function recursion relations, that the two formulae are, in fact, identical, provided that $W(\Omega)$ and its derivative are zero at the boundary, and the equivalence of the two methods can also be shown using integration by parts (Smith and Zaldarriaga 2007).

**Properties of Pure-$C_\ell$**

Due to the presence of derivatives in the formula for computing the counter-terms, the mask and its derivative are required to be smooth for the method to work. As a result, masks used in the pure-$C_\ell$ formalism must be apodised, or else the method will be extremely inefficient. Furthermore, they need to be apodised in a manner which makes the first derivative of the mask continuous, which in the case of a realistic mask is not necessarily trivial.
2: SIMULATIONS OF THE DETECTION OF B-MODES USING THE PURE-\(C_\ell\) METHOD

Smith and Zaldarriaga (2007) have shown that the two methods given above are, in fact, equivalent, and, thus, they can be used interchangeably, and results derived from one can be used in relation to the other. Thus, in the rest of this thesis I will switch between the two methods according to their relative usefulness for a specific calculation or implementation. In order to do this, it is useful to consider how the two methods differ, and their relative advantages and disadvantages.

Each of the two mathematical descriptions has advantages over the other in specific circumstances. The counter-term formalism is more flexible, since it allows for situations where you might wish to use the pure-\(C_\ell\) method with counter-terms that do not exactly cancel the E-B mixing, and it is considerably simpler to use computationally, since it only requires taking the derivative of the mask (which can be done in harmonic space and only needs to be done once), rather than that of the CMB maps. However, the \(\chi_B\) formalism is more intuitive, and is substantially easier to handle mathematically. Because of these properties, I have generally used the counter-term formalism for computational work and the \(\chi_B\) formalism for theoretical calculations.

2.3.2 Implementation of pure-\(C_\ell\) algorithm

The steps required to run simulations using the pure-\(C_\ell\) method are essentially the same as those required to simulate the ordinary pseudo-\(C_\ell\) method. However, the requirement to include counter-terms complicates certain parts of the calculation, in particular the calculation of the mask matrix and the mask itself. There are three fundamental modifications to consider. The addition of counter-terms to the mask (and their calculation), the computation of the mask matrix from the mask and the requirement to be able to deal with the spin-1 counter-terms, which necessitated writing a spin-1 equivalent of the HEALPix map2alm and alm2map routines to convert between harmonic and real space for spin-1 maps.
Conversion between harmonic and real space for spin-1 maps

In order to carry out CMB simulations, it is necessary to have a mechanism to convert a map from real space to harmonic space and back again. Since CMB temperature maps are spin-0, and polarisation maps are spin-2, HEALPix routines to convert maps with these spin-weights between real and harmonic space have existed for a long time. However, until recently, there was no need to make such a conversion for other spin-weights, and thus HEALPix routines to handle these cases did not exist. Whilst there are now HEALPix libraries which can handle general spin weights, at the time when I wrote my pure-$C_\ell$ code no such libraries existed, and thus I was required to write the code to handle spin-1 maps myself.

In order to do this, I took the existing HEALPix polarisation routines, and modified them to take account of the differences between spin-1 and spin-2 maps. For this purpose, it is useful to look at how the more standard spin-2 maps are handled, in order to generalise this method to maps with other spin-weights.

Now, as shown in Section (1.2.3), for a spin-2 map, linear combinations of the standard spin $\pm 2$ harmonic coefficients which have the useful property that $a^p_\ell(-m) = (-1)^m(a^p_{\ell m})^*$ (Equation (1.21)) can be defined. These are, in fact, the familiar E and B modes, as defined in Equations (1.19) and (1.20). However, by looking back at the definition of the spin-2 polarisation field $P$ in terms of the Stokes’ parameters, it is possible to consider this process in reverse. Rather than considering the polarisation field as being constructed from the Stokes’ parameters, the Stokes’ parameters themselves can be defined by taking the pure spin-2 polarisation field and splitting it into its real and imaginary parts, by defining $Q = (P + P^*)/2$ and $U = (P - P^*)/2i$.

Since both my code and the HEALPix code is designed to handle polarised maps stored as separate $Q$ and $U$ maps, and to handle coefficients with the property $a^p_\ell(-m) = (-1)^m(a^p_{\ell m})^*$ (since these are parity eigenstates, and are, thus, the most natural way to represent the data), it is useful to define analogous quantities for the spin-1 case. For the maps themselves, this is trivial. Just as a spin-2 coefficients can be split into
their real and imaginary parts, I can similarly define spin-1 equivalents to the Stokes’ parameters as \( Q' = (P + P^*)/2 \) and \( U' = (P - P^*)/2i \), where \( P \) is now a field with spin-weight 1. However, the case of the harmonic coefficients is slightly more complicated. This is because the standard spherical harmonic coefficients have the property that

\[
sY_{\ell m} = (-1)^{m+s} sY_{\ell(-m)},
\]

(2.49)

and, thus, the coefficients work differently in cases with odd and even spin. In fact, it can be seen that the linear combinations \( G \) and \( C \), defined by,

\[
aG_{\ell m} = \frac{1}{2} a_{\ell m} - \frac{-1}{2} a^{*}_{\ell m},
\]

(2.50)

\[
aC_{\ell m} = \frac{1}{2} a_{\ell m} + \frac{-1}{2i} a^{*}_{\ell m},
\]

(2.51)

have the property that \( a^{\ell (-m)} = (-1)^m(a^P_{\ell m})^* \), as required. Thus, these two combinations of the spin \( \pm 1 \) harmonic coefficients are direct analogies of the E and B mode coefficients in the case of spin-1, and thus can be used in the same way.

It is then possible to modify the HEALPix routines to deal with spin-1 maps. In order to do this, recursion relations for the spin-1 spherical harmonics, analogous to those for spin-2 harmonics that are used by the HEALPix polarisation routine, are required, and these are given in Smith (2006). By replacing the recursion relations in the spin-2 HEALPix polarisation code with these spin-1 analogues, I was able to produce a routine which converts between the harmonic and real space versions of spin-1 maps.

**Calculation of mask matrix**

As for the standard pseudo-\( C_\ell \) method, in order to use the pure-\( C_\ell \) method, the mask matrix must be calculated, so that the convolved pure-\( C_\ell \) power spectrum can be de-convolved to get an unbiased estimate for the true power spectrum. In principle, the method is the same as for the standard pseudo-\( C_\ell \) case. Because of the counter-terms, the calculation is considerably more complicated.
2.3: REMOVING PROBLEMS WITH E-B MIXING

However, the mask matrix can also be calculated more simply using the $\chi_B$ formalism. In this case, it is simply necessary to take Equation (2.48) and use it to compute the power spectrum of $\chi_B$. Using this, I obtain

$$M_{\ell \ell'}^{\chi\chi_{BB}} = \frac{(2\ell + 1)(2\ell' + 2)(\ell' - 1)\ell' (\ell' + 1)(\ell' + 2)}{4\pi(\ell - 1)\ell(\ell + 1)(\ell + 2)} \sum_{\ell''} W_{\ell \ell'}^{PP} \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{array} \right)^2. \quad (2.52)$$

However, although this is a much simpler formula than the counter-term version (which can easily be computed (using a similar method to the one used here) from Equation (2.29) and Equations (2.37), (2.38) and (2.39), but which has a very large number of terms and is thus not shown here), it also assumes that the counter-terms are exact, and thus is less accurate than the full counter-term form in the case where the counter-term method is used to compute the masked CMB. As a result, it cannot be used in my simulation routine, although it is useful for theoretical calculations.

**Calculation of counter-terms**

Unlike in the case of the ordinary pseudo-$C_\ell$ method, it is not sufficient merely to apply the mask to the CMB map, due to the requirement to subtract off the effects of E-B mixing using counter-terms. Instead, the mask has three components. The ordinary mask, which is determined in the same way as for the pseudo-$C_\ell$ case, and the spin-1 and spin-2 counter-term components, which must be calculated using the definitions given in Section (2.3.1).

The obvious way to compute these components is to use their definition in terms of the spin raising and lowering operators to differentiate the mask, either numerically or, for simple cases, analytically. However, whilst this will indeed work, it is not particularly convenient, since it will likely require the use of a finite-differencing scheme, which is non-trivial to implement for a HEALPix pixelisation grid. However, fortunately, due to the properties of spin-weighted spherical harmonics, there is a short-cut which allows the calculation to be performed in harmonic space.

Now, using the definition of the counter-terms (Equation (2.34)) and the standard
definition of $a_{\ell m}$ coefficients, it can be shown that

$$s W_{\ell m} = \int \delta^s(W(\Omega)) Y_{\ell m}^*(\Omega) d\Omega , \quad (2.53)$$

and, similarly, that

$$-s W_{\ell m} = \int \tilde{\delta}^s(W(\Omega)) -s Y_{\ell m}^*(\Omega) d\Omega . \quad (2.54)$$

Using integration by parts, the first equation can be re-written as

$$s W_{\ell m} = (-1)^s \int W(\Omega) \delta^s(s Y_{\ell m}^*(\Omega)) d\Omega \quad (2.55)$$

$$= (-1)^s \int W(\Omega) (\tilde{\delta}^s(s Y_{\ell m}^*(\Omega)))^* d\Omega . \quad (2.56)$$

Then using the definition of the spin-weighted spherical harmonics (Equation (1.13)), it can be shown that

$$s W_{\ell m} = (-1)^{2s} \sqrt{\frac{(\ell + s)!}{(\ell - s)!}} W_{\ell m} , \quad (2.57)$$

which, in turn, gives us that

$$s W_{\ell m} = \sqrt{\frac{(\ell + s)!}{(\ell - s)!}} W_{\ell m} , \quad (2.58)$$

with a similar calculation giving that

$$-s W_{\ell m} = (-1)^s \sqrt{\frac{(\ell + s)!}{(\ell - s)!}} W_{\ell m} . \quad (2.59)$$

Now, as shown in the previous section, useful linear combinations of these variables can be defined. In the case of spin-2 harmonics, there are the familiar E and B mode harmonics, and in the case of spin-1 I have called them G and C harmonics. Under the definition given in the last section, I find

$$w_{\ell m}^G = \sqrt{\ell(\ell + 1)} W_{\ell m} , \quad (2.60)$$

$$w_{\ell m}^C = 0 , \quad (2.61)$$

$$w_{\ell m}^E = \sqrt{(\ell - 1)(\ell + 1)(\ell + 2)} W_{\ell m} , \quad (2.62)$$

$$w_{\ell m}^B = 0 . \quad (2.63)$$
Once these have been defined, it is a simple task to compute the mask counter-terms, by using the HEALPix alm2map routine (as well as my own version, for the spin-1 case) to convert them back into real-space maps. This method is considerably easier and quicker, for general masks, than attempting to use finite differencing, and thus I have used it to calculate the counter-terms throughout my work.

### 2.3.3 Apodisation of pure-$C_\ell$ masks

As discussed in Section 2.3, for the pure-$C_\ell$ method to work properly, the counter-terms must be finite. Since they are formally defined using the spin raising and lowering operators, in order for this to be true (in the limit of infinite resolution), the mask and its first derivative must be continuous. Therefore, one of the major issues for pure-$C_\ell$ is finding a suitable method for apodising the masks. Smith and Zaldarriaga (2007) suggest an algorithm for computing the optimal mask apodisation, including counter-terms. However, in the case where there is little or no noise, this method does not converge, and thus it is not useful in the case of my simulations, which have been excluding noise for the purposes of better understanding the underlying method.

The simplest way to apply apodisation is to use a nearest-pixel approach. For each pixel in the unmasked (observed) region, I calculate the distance to the nearest point on the edge of the unmasked region (hereafter denoted by $r$). Then, I apply an apodisation function $f(r, r_c)$ (where $r_c$ is the apodisation length) to the mask. The apodisation function is required to have the following properties.

1. $f$ and $\frac{df}{dr}$ are continuous, and are both zero for $r = 0$ (i.e. at the edge of the mask).
2. $f = 0$ for $r \leq 0$.
3. $f = 1$ for $r \gg r_c$.

The first property is a necessary (but not sufficient) condition for the mask and its derivative to be continuous. The second is necessary to ensure that masked areas are...
zeroed and the final one is necessary to ensure that, far away from the boundary, the mask does not modify the observed CMB data. Note, however, that whilst these conditions are necessary to produce a suitable mask, they do not guarantee that the derivative of the mask will be continuous, since the differential of \( r \) does not, in general, vary smoothly across the mask. This can result in the spin-0 counter-term being very large, which results in the method breaking down. However, in the special case of an azimuthal mask, then this method does indeed produce a mask which is continuous and differentiable, as required for the pure-\( C_\ell \) method. In the more general case, however, some smoothing algorithm is required to ensure that the mask is differentiable.

There are many different possible apodisation schemes, but for my thesis I have mainly used two. The first is a cosine apodisation, with the mask given by \( M = 1 - \cos^2(r/r_c) \) (with \( M = 1 \) for \( r > r_c \)), and the second is a Gaussian apodisation, where the mask is given by \( 1 - \exp(-(r/2\sigma)^2) \) (where \( \sigma = r_c \sqrt{8 \log 2} \)), with the mask smoothly tapering to 1 for \( r > kr_c \) (where \( k \) is usually around 5), in order to avoid computing very small corrections to the mask at arbitrarily large values of \( r \). In this section, I investigate different mask apodisation algorithms, in order to determine the best method of apodising the masks, and also to determine how sensitive the method is to alterations in the apodisation method used. First, I have considered the simple case where the mask is azimuthal, which allows for an easy study of the effects of different apodisation schemes. Then, I have extended my analysis to non-azimuthal masks, considering the potential pitfalls which arise in this situation and the possible ways around them.

**Apodisation with azimuthal masks**

In order to implement the pure-\( C_\ell \) method effectively, it is necessary to consider the best apodisation scheme. In addition, it is also necessary to consider how sensitive the method is to variations in the apodisation scheme, since if the effect of having a non-optimal apodisation is relatively small, less effort need be put into determining the optimal apodisation function. If, on the other hand, the results turn out to be highly
sensitive to the apodisation scheme, then determining the optimal function will be of paramount importance to the implementation of the method. Therefore, I have conducted simulations at \( N_{\text{side}} = 256 \), using Gaussian and cosine apodisation functions with apodisation lengths of 0.1° (which is essentially equivalent to applying no apodisation), 0.5°, 1°, 2.5°, 5° and 10°. This allows me to compare the performance of the two different apodisation schemes I have suggested (which will, in turn, give an indication of the likelihood that an alternate scheme will be substantially better than either) and also to both find the optimal apodisation length and study how sensitive the pure-\( C_\ell \) method is to changes in it.

In order to determine the effectiveness of each scheme, I have computed the fractional error in the value of \( r \) that would be measured by the experiment (in the idealised case where there was no noise and no lensing signal) by summing the squares of the \( S/N \) ratio in each bin up to some maximum \( \ell \) value, as explained later in Section (3.1.2). This gives us an indication of the relative usefulness of each apodisation scheme, with a larger fractional error indicating that the scheme is less effective. In order to test the apodisation schemes over as wide a range of situations as possible, I have plotted the results for simulations with three different azimuthal masks, a half-sky mask, to demonstrate what happens in the case of a survey which covers a substantial proportion of the sky, a circular region of total area 1000 deg\(^2\) centred on the north pole of the map (masks of this form will be referred to as spherical cap masks throughout the rest of this thesis), which represents a reasonable-sized CMB survey, and a 50 deg\(^2\) one, which is probably too small for use in any realistic experiment but which could provide a reasonable model of what might happen in the case of less smooth surveys. In order to distinguish between the high-\( \ell \) and low-\( \ell \) effects, I have used two different maximum values of \( \ell \). The plots with \( \ell_{\text{max}} = 100 \) demonstrate the effects of different apodisation schemes on large scales, and the plots with \( \ell_{\text{max}} = 500 \) demonstrate the effects on smaller scales. These results are plotted in Figure 2.5.

These results show clearly that varying the apodisation function and length makes relatively little difference, unless an exceptionally short apodisation length is chosen.
2: SIMULATIONS OF THE DETECTION OF B-MODES USING THE PURE-$C_\ell$ METHOD

Figure 2.5: Detectability of B-modes given measurements up to a maximum $l$ of 100 (left) and 500 (right) assuming no noise or lensing for a 50 square-degree (top), 1000 square-degree (middle) and half-sky (bottom) spherical cap mask as a function of the apodisation length, $r_c$, for a Cosine (blue solid) and Gaussian (black dashed) apodisation function

ANALYSIS OF CMB POLARISATION
The optimal value (of those studied) seems in all cases to be 2.5°, with little difference between the cosine and Gaussian apodisation functions. However, any value above this seems to work equally as well, except for the 50 square-degree mask, where there is a slight degradation in the S/N ratio. However, it is clear from these results that, as expected, very small apodisation lengths do not work. Overall, I can conclude that the pure-\(C_\ell\) method is not particularly sensitive to the apodisation method chosen, and thus I have not investigated this any further.

### Issues with non-azimuthal masks

However, whilst the algorithms used above work perfectly fine for azimuthal masks, in the case of non-azimuthal masks, there is a problem. The apodisation method defined previously works by computing the distance between the pixel in question and the nearest masked pixel, and apodising accordingly. This guarantees that the apodised mask will be smooth, but for the majority of masks it will not be differentiable, since it takes account of only the nearest point, which can suddenly change between two adjacent pixels. Therefore, there needs to be some method for smoothing the mask, to ensure that it is both continuous and differentiable.

A simple way to do this would be to smooth the mask by applying a beam to it. This will smooth out any sharp features, and thus guarantees that the mask will both be continuous and differentiable, resulting in small counter-terms. However, this alone will not work since pixels outside of the observed areas must be fully masked (since there is no usable data there), and thus the smoothed mask will have a discontinuity at the edge of the observed region. Therefore, some combination of the smoothed mask and the original mask is required, to remove the sharp features in the mask whilst also allowing it to go smoothly to zero at the mask boundary. Since the smoothing algorithm will remove power from the mask, it is preferable to avoid applying smoothing whenever possible. Therefore, I have defined a weighting scheme to calculate the final
2: SIMULATIONS OF THE DETECTION OF B-MODES USING THE PURE-\(C_\ell\) METHOD

\[ W_{\text{tot}} (\Omega) = k (\Omega) W_{\text{smooth}} (\Omega) + (1 - k (\Omega)) W_{\text{orig}} (\Omega), \quad (2.64) \]

where \(W_{\text{orig}}\) is the original (unsmoothed) mask, \(W_{\text{smooth}}\) is \(W_{\text{orig}}\) smoothed using a beam smoothing function with beam width of \(\theta\) (a free parameter) and \(k(\Omega)\) is a function which is defined such that it is zero when \(W_{\text{orig}}\) is zero, and that it asymptotes to 1 as the counter-terms tend to infinity. The particular expression that I have used is as follows

\[ k (\Omega) = \tanh \left( \frac{W_0 (\Omega)}{c} \left( \frac{|W_1 (\Omega)|}{\int |W_1 (\Omega')| d\Omega'} + \frac{|W_2 (\Omega)|}{\int |W_2 (\Omega')| d\Omega'} \right) \right). \quad (2.65) \]

Here \(W_0\) is the unsmoothed mask, \(W_1\) and \(W_2\) are the unsmoothed counter-terms, \(\int X (\Omega') d\Omega'\) is the integral of the function \(X\) taken over the apodised regions of the mask and \(n\) and \(c\) are arbitrary constants which will be discussed later. This expression is designed such that, when either counter-term is considerably larger than its average value over the apodised region of the mask, the term inside the tanh function will become large and, thus, \(k\) will tend to 1, and, conversely, as \(W_0\) tends to 0 (i.e. at the edge of the mask), the inner term and, thus, \(k\) itself will tend to 0. This will, of course, mean that the parts of the mask with the largest counter-terms (which, since the counter-terms are calculated from derivatives of the mask, will correspond to those parts where the mask is either discontinuous or is not differentiable) will be smoothed out, reducing the counter-terms in those regions and, consequentially, improving the effectiveness of the pure-\(C_\ell\) method.

This model has three arbitrary parameters. These are the two arbitrary constants, \(c\) and \(n\), and the beam length used to compute the smoothed mask. The constant \(c\) controls how unusually large the counter-terms have to be before the smoothed mask comes into effect. A larger value of \(c\) will mean that the smoothed mask is used less often. However, if it is set too large, then the method may not smooth out the discontinuities in the mask as it is supposed to. The parameter \(n\), on the other hand, controls the effect of \(W_0\) on the smoothing process. If \(n\) is large, then only very small values of
$W_0$ will reduce the smoothing to any significant extent, which will mean that smoothing will only be disfavoured towards the mask boundary, as required. However, a very large value of $n$ will mean that the effect of a small $W_0$ on the smoothing process will be negligible and, as a result, the mask could become discontinuous at the boundaries. The smoothed mask is generated by using the standard beam-smoothing routine (in harmonic space), with a beam width $\theta$ that is an arbitrary free parameter. It appears, from testing the algorithm, that values of $\theta$ that are around twice the diameter of a pixel are appropriate. This makes sense, since the smoothing should only apply to pixels that are very close to the site of the discontinuity.

In order to determine the values to use for these parameters in my thesis, I carried out simulations using various point-source masks. I found that the values which gave the best overall performance for the method were $c = 2$, $\theta = 0.5^\circ$ and $n = 3$ and, thus, these are the values that I have chosen to use whenever the smoothing process is applied in this thesis.

### 2.4 Simulations of pure and standard pseudo-$C_\ell$

As previously discussed, the pure-$C_\ell$ method is designed to eliminate problems caused by E-B mixing that arise from the pseudo-$C_\ell$ method. However, I have as yet not demonstrated the veracity of my implementation. Further, there are also issues caused by the removal of ambiguous modes from the computation of the B-mode power spectrum in the pure-$C_\ell$ method which result in an increase in the error relative to the expected error that you would get from the ordinary pseudo-$C_\ell$ method without E-B mixing. In order to prove the effectiveness of my code, and to characterise the performance of the pure-$C_\ell$ method, I have carried out simulations comparing the pure and pseudo-$C_\ell$ methods in the case of a normal CMB power spectrum (with $r = 0.01$) and, to demonstrate the relative effects of E-B mixing on the two methods, I have also carried out simulations where the E-mode polarisation was artificially set to zero. In order to investigate the effectiveness of pure-$C_\ell$ in a variety of cases, I have considered...
four different masks, two spherical cap masks, of size 10000 deg$^2$ (Figure 2.6) and 1000 deg$^2$ (Figure 2.7), with a cosine apodisation function applied with an apodisation length of 4.5$^\circ$ and two point source masks, one with one source per 100 square-degrees (Figure 2.8) and one with one source per 10 square-degrees (Figure 2.9), using a cosine apodisation function with an apodisation length of 1.5$^\circ$ and applying the smoothing algorithm outlined in the previous section. Each simulation is done using $N_{\text{side}} = 512$ and a bin size of $\Delta \ell = 10$, with a CMB power spectrum with $r = 0.01$.

![Figure 2.6: Simulated pure-$C_\ell$ (left) and pseudo-$C_\ell$ (right) spectra with (top) and without (bottom) E-modes for a 10000 square-degree spherical cap mask.](image)

Comparing the results for the pure-$C_\ell$ method with and without E-mode polarisation, it is immediately obvious that there is no difference and, thus, that my implementation of the pure-$C_\ell$ method does indeed remove the effects of E-B mixing entirely. Conversely, it is equally obvious just how much of an effect E-B mixing is having on the standard pseudo-$C_\ell$ method. Excepting for the point source masks at low-$\ell$ (which work reasonably well due to the relatively low proportion of the sky masked out), the effect of E-B mixing on the pseudo-$C_\ell$ method is catastrophic, increasing the error on each bin by almost an order of magnitude in even the best of cases. Thus, these simulations demonstrate conclusively that the pseudo-$C_\ell$ method is not adequate for the...
2.4: SIMULATIONS OF PURE AND STANDARD PSEUDO-$C_\ell$

Figure 2.7: Simulated pure-$C_\ell$ (left) and pseudo-$C_\ell$ (right) spectra with (top) and without (bottom) E-modes for a 1000 square-degree spherical cap mask.

Figure 2.8: Simulated pure-$C_\ell$ (left) and pseudo-$C_\ell$ (right) spectra with (top) and without (bottom) E-modes for a point source mask with one source per 100 square-degrees.
2: SIMULATIONS OF THE DETECTION OF B-MODES USING THE PURE-$C_\ell$ METHOD

Figure 2.9: Simulated pure-$C_\ell$ (left) and pseudo-$C_\ell$ (right) spectra with (top) and without (bottom) E-modes for a point source mask with one source per 10 square-degrees.

The purpose of detecting relatively small values of $r$, as expected. However, of more interest is the comparison between the pure-$C_\ell$ method and the pseudo-$C_\ell$ method in the case where there are no E-modes included in the simulation, since this gives some indication of the effect which removing the ambiguous modes has on the pure-$C_\ell$ method.

Looking at the results, it can be seen that, in general, the pure-$C_\ell$ method performs slightly worse at low $\ell$ than the ordinary pseudo-$C_\ell$ method without the effects of E-B mixing. However, for the spherical cap masks, this effect is relatively small. For the 10000 square-degree spherical cap mask, there is only a significant increase in the error in the lowest bin, and even there it only increases by a factor of four. On smaller scales, the increase is negligible, with even the second bin ($\ell = 16$) having an increase in the error of only 30% or so, and above $\ell = 30$ the error increases by less than 3%. The 1000 square-degree mask is slightly worse, with the lowest bin now entirely unobservable and an increase in error of over 60% for the second bin, but even so the fractional increase in the error due to the pure-$C_\ell$ method still falls to around 20% by the $\ell = 36$ bin and to just over 5% by $\ell = 50$. Overall, the pure-$C_\ell$ method compares well to the no E-B mixing ordinary pseudo-$C_\ell$ results, except at very large scales. This
is not overly surprising since the extra error in the pure-$C_\ell$ method is expected to come from the boundary terms, and an azimuthal mask has a low boundary length.

In the case of the point-source masks, the picture is similar, although slightly worse than before. For the case where there is one source per 100 square degrees (which gives about 400 sources to mask out), the first bin is, again, unobservable (which will prevent detection of the reionisation peak using this method), and the error in the second bin increases by over 4 times, but beyond this point the fractional increase in error falls off rapidly, and by $\ell = 50$ it has fallen to below 5%. For the case of one source per 10 square degrees, however, the error is increased up to $\ell$ of around 50 (for the $\ell = 46$, there is a 50% increase, and for $\ell = 36$ the error more than doubles), and for the first two bins there is an increase of more than an order of magnitude on the first two bins, resulting in them becoming entirely unobservable using this method. Therefore, in this case, the ability to observe the B-mode spectrum at low-$\ell$ will be substantially compromised.

Overall, these results show that the pure-$C_\ell$ method does indeed eliminate the problems of E-B mixing and that, whilst there is an effect from boundary terms at low values of $\ell$, it is not that big a problem, except perhaps in the case of the masks with longer boundary layers, such as point-source masks. However, it is likely that some other method, such as perhaps a Maximum-Likelihood (ML) analysis, will be needed in order to measure the power spectrum at very low $\ell$ values, since the pure-$C_\ell$ method seems to break down there. This will be important for the measurement of the recombination peak at low-$\ell$, since this resides entirely in the lowest $\ell$-bin used here, and thus would not be detectable using the pure-$C_\ell$ method alone.

### 2.5 Conclusions

In this chapter, I have outlined my implementation of the ordinary pseudo-$C_\ell$ method and the pure-$C_\ell$ method. I have explained how I wrote code to generate CMB maps, apply a mask to them and then remove the effects of this mask using the two methods.
I have explained the technical challenges that this posed, and have shown how I overcame them. I have also demonstrated the differences between the two methods. Using these implementations, I then performed several preliminary simulations, to demonstrate the correctness of my code and to show the potential issues with the methods.

There are several conclusions which can be made from the preliminary simulations that I have made. Firstly, it is clear from the ordinary pseudo-$C_\ell$ simulations that this method is not sufficient to analyse B-mode polarisation data. Whilst it will likely be able to detect the levels of B-modes that are being searched for by the present generation of experiments, it will substantially degrade their accuracy, and many future experiments will simply not work at all. The reason behind this is the leakage of power from the E-mode power spectrum into the B-mode spectrum due to the incomplete separation of E and B modes that occurs on a partial sky. This is shown by the simulations in Section 2.4, which demonstrate that, if the E-modes are artificially suppressed, then the pseudo-$C_\ell$ method works well.

For the pure-$C_\ell$ method, there are also several important points to take forward. Because the method uses counter-terms, which are defined using derivatives of the mask, the mask is required to be both smooth and differentiable. As a result, apodisation needs to be applied to the mask. This poses a potential issue, since the ideal apodisation function is not well-known. However, it turns out that, for an azimuthal mask at least, the effectiveness of the method is not significantly affected by modifications to the apodisation function, provided that the apodisation length is sufficiently large (2.5° or larger). Whilst there are potential issues with non-azimuthal masks (since the most obvious apodisation method does not guarantee that such a mask will be differentiable, as is required), I have designed a smoothing algorithm which appears to solve this problem and allows the pure-$C_\ell$ method to work effectively on all types of mask.

Finally, there are several conclusions that can be made about the pure-$C_\ell$ method itself. Whilst the method is substantially better than the ordinary pseudo-$C_\ell$ method in the majority of cases, if the E-modes are artificially suppressed, then the pure-$C_\ell$ method is actually slightly worse. This is due to the removal of ambiguous modes from
2.5: CONCLUSIONS

the calculation of the power spectrum in the pure-$C_\ell$ method. However, whilst this issue seems to make the power spectrum almost impossible to measure in the lowest bin, the effect is strongly localised to the largest scales, with even the most pathological of masks (such as a point source mask, which has an extremely long boundary relative to the total area of the mask) only having a noticeable increase in the error in the pure-$C_\ell$ method (relative to the pseudo-$C_\ell$ method without E-B mixing) for values of $\ell$ below around 50. Overall, it is found that the pure-$C_\ell$ method is an adequate replacement for the ordinary pseudo-$C_\ell$ method for the purposes of estimating the B-mode power spectrum, although it may be necessary to use an alternative method (such as the maximum-likelihood method) in order to calculate the power spectrum on the very largest scales.

Whilst the main purpose of this chapter is to demonstrate my implementation of the pure-$C_\ell$ method shown in Smith (2006) and Smith and Zaldarriaga (2007) and explain the reasoning behind using it, I do nevertheless produce several important results, which will be expanded on in the following two chapters. In particular, I have demonstrated the error properties of the method in a much more systematic way than is shown in the original papers, and in doing so have demonstrated both the superiority of the method over the ordinary pseudo-$C_\ell$ method and also the inadequacy of the method for low-$\ell$ measurements.
2: SIMULATIONS OF THE DETECTION OF B-MODES USING THE PURE-$C_\ell$ METHOD

ANALYSIS OF CMB POLARISATION
3

Survey strategies and future instruments

In order to improve the chances of detecting B-modes, it seems sensible to carefully consider the optimal design for future experiments. In the case of temperature and E-mode polarisation experiments, this can usually be done reasonably effectively by simply using optimal error formulae (see, for example, Kamionkowski et al. 1997), and determining the optimal survey design accordingly. When considering the ordinary pseudo-$C_\ell$ method, this is simply a question of what fraction of sky should be observed in order to minimise the loss of information due to measurement on a small fraction of the sky whilst also maximising the amount of time spent observing each pixel, in order to minimise the noise.

However, in the case of the pure-$C_\ell$ method, there is an additional effect to consider. As shown in Section 2.3, whilst the pure-$C_\ell$ method does an excellent job of removing the E-B mixing which massively degrades the performance of ordinary pseudo-$C_\ell$ estimators, it also adds an additional error, proportional to the boundary size, which affects the low-$\ell$ performance of the method. Effectively, this favours masks with shorter boundaries over those with complicated boundary structures. Therefore, unlike in the case of temperature or E-mode measurements, not only the survey size but also the
3: SURVEY STRATEGIES AND FUTURE INSTRUMENTS

survey shape must be carefully considered in order to ensure the best possible chance of B-mode detection.

In this chapter, I will begin by using the optimal error formula given in [Kamionkowski et al., 1997] to outline a method for determining the ideal scan strategy for any given experiment. I will then use my pure-$C_\ell$ code to attempt to understand how the sub-optimality of the pure-$C_\ell$ method will affect these results. Finally, I will attempt to extend the calculation carried out for the pseudo-$C_\ell$ method in [Challinor and Chon, 2005] to the pure-$C_\ell$ case.

3.1 Theoretical errors for masked CMB measurements

3.1.1 Theory behind optimal errors

To set the scene, I will assume that the errors for B-mode experiments are approximately optimal, and use this assumption to compute a formula for the optimal survey size (and the level of B-mode polarisation that it could detect) as a function of the noise level. In [Scott et al., 1994], it is shown that a good approximation for the error in the power spectrum estimate can be obtained by dividing the cosmic variance by a factor of $f_{\text{sky}}$, the fraction of the sky left uncovered by the mask

$$ \frac{\sigma(C_\ell)}{C_\ell} = \frac{2}{(2\ell + 1)f_{\text{sky}}} \cdot$$ (3.1)

This approximation can be understood to be a result of information loss from masking a fraction $(1 - f_{\text{sky}})$ of the sky. The result for cosmic variance in the full-sky case comes from the fact that for a given $C_\ell$, there are $2\ell + 1$ degrees of freedom at each $\ell$ (the $2\ell + 1$ $a_{\ell m}s$). In the masked case, the number of degrees of freedom is effectively reduced to $(2\ell + 1)f_{\text{sky}}$, hence the formula. However, this formula ignores the effects of mixing the different $C_\ell$ values, and so is only applicable if the matrix is approximately diagonal (or block diagonal in the case of binning).

However, this formula by itself is of limited use, since it only applies to tempera-
3.1: THEORETICAL ERRORS FOR MASKED CMB MEASUREMENTS

ture, and also assumes there are no noise or beam effects (see Section 1.4.3). Therefore, it needs to be extended to include them and also to deal with the fact that there is more than one kind of spectrum in the case of polarisation. This can, in fact, be done quite simply if I once again assume that there is no mixing between different spectra. If this is the case, then the same reasoning as used for the temperature case will apply to the polarisation case, and I can determine an approximate formula for the variance in this case by dividing the formula for cosmic variance in the polarised case (Equation (1.27)). This gives

\[
\sigma(C_{\ell}^{XX}) = \sqrt{\frac{2}{2(2\ell + 1)f_{\text{sky}}}C_{\ell}^{XX}},
\]

(3.2)

and

\[
\sigma(C_{\ell}^{XY}) = \sqrt{\frac{\left(\left(C_{\ell}^{XY}\right)^2 + C_{\ell}^{XX}C_{\ell}^{YY}\right)}{2(2\ell + 1)f_{\text{sky}}}}
\]

(3.3)

However, these formulae still do not take account of the effects of noise. Once again ignoring the mixing between different \(\ell\) values, the variance of the \(C_{\ell}\) values can be calculated in the case where there is noise and a beam but no masking, and then simply divide by \(f_{\text{sky}}\) to add the effect of the mask. To do this calculation, I must first define an unbiased estimator of the cosmological \(C_{\ell}\). Using Equation (1.36), I can define

\[
\hat{C}_{\ell}^{XY} = \frac{1}{B_{\ell}^2} \left(\tilde{C}_{\ell}^{XY} - \langle N_{\ell}^{XY}\rangle\right)
\]

(3.4)

where \(\tilde{C}_{\ell}^{XY}\) is the measured pseudo-\(C_{\ell}\) and \(\langle N_{\ell}^{XY}\rangle\) is the mean value of \(N_{\ell}^{XY}\). The standard deviation of this statistic can then be calculated, which gives the error in our estimate of each term in the power spectrum. This calculation is done in Kamionkowski et al. (1997), and it is found that the error in the case of noise can be determined by replacing \(C_{\ell}^{XY}\) in the formula for optimal errors (Equation (1.27)) with \(\tilde{C}_{\ell}^{XY}\), where

\[
\tilde{C}_{\ell}^{XY} = C_{\ell}^{XY} + \frac{N_{\ell}^{XY}}{B_{\ell}^2}.
\]

(3.5)
so, for example, the equation for BB becomes
\[
\sigma(C_{BB}^\ell) = \sqrt{\frac{2}{(2\ell + 1)f_{\text{sky}}} \left( C_{BB}^\ell + N_{BB}^\ell B_\ell^2 \right)},
\]
(3.6)
where
\[
B_\ell = e^{-\frac{1}{2}\sigma^2(\ell+1)},
\]
(3.7)
is the beam power spectrum, with
\[
\sigma = \frac{\theta_{\text{FWHM}}}{\sqrt{8 \ln 2}},
\]
(3.8)
where \( \theta_{\text{FWHM}} \) is the full-width at half-maximum of the Gaussian beam profile. These equations can then be used to determine the minimum possible error that could be achieved by a particular CMB experiment.

### 3.1.2 Realistic noise for CMB experiments

An important variable to consider when designing an experiment is what proportion of the sky to survey. In the case of zero noise, it is trivially obvious from Section \[ \text{3.1.1} \] that the survey should cover the largest proportion of sky that is possible given the constraints of the location and the scanning properties of the telescope. However, in the case of a noisy survey, this is not necessarily the case, since there is then a trade-off between the proportion of sky observed and the amount of time spent observing each pixel. Using the known properties of white noise, the noise per pixel can be determined as a function of the amount of time spent observing it, which allows, in conjunction with Equation \[ \text{3.6} \], the computation of the optimal proportion of the sky to observe with any given experiment.

The noise level of the instrument is generally parametrised by the Noise Equivalent Temperature (NET), which I will write as \( T_\sigma \). Defining the number of simultaneous beams (defined such that each beam measures both \( Q \) and \( U \) polarisation) to be \( n_B \), the noise per pixel, \( \sigma_{\text{pix}} \), is given by
\[
\frac{\sigma_{\text{pix}}}{\sqrt{4\pi f_{\text{sky}}}} = \frac{T_\sigma}{\sqrt{n_B \Omega_B t}},
\]
(3.9)
where $\Omega_B = 4\pi/N_{\text{pix}}$ is the pixel area and $t$ is the total time for which the experiment produces data on the CMB polarisation (excluding observations which are masked out). Using Equation (1.35), the noise spectrum is given by

$$N_\ell = \frac{4\pi f_{\text{sky}} T_\sigma^2}{n_B \ell}.$$  

Importantly, this equation is dependent on the value of $f_{\text{sky}}$, and so, as expected, for a noise-dominated regime it is no longer necessarily advantageous to measure over the largest fraction of the sky possible. Therefore, a trade-off between the increase in cosmic variance due to observing a smaller fraction of the sky and the corresponding reduction in the noise is needed. For any experiment, the most important quantity is the signal-to-noise ratio (S/N), which is ratio of the actual value to the error, i.e. $C_\ell/\Delta C_\ell$. For a given value of $\ell$, $(S/N)^{-1}$ as a function of $n_B$ and $f_{\text{sky}}$ can be found to be

$$\frac{N}{S} = \frac{\Delta C_\ell}{C_\ell} = \sqrt{\frac{2}{2l+1}} \left( \frac{1}{\sqrt{f_{\text{sky}}}} + \frac{4\pi T_\sigma^2 \sqrt{f_{\text{sky}}}}{n_B t B_\ell^2 C_\ell} \right).$$  

From this equation, it can be seen that there are two competing factors, the effect of cosmic variance (which is dependent on the number of modes observed and, thus, reduces as $f_{\text{sky}}$ is increased) and the effect of the noise (which is dependent on the time spent observing each pixel and, thus, increases as $f_{\text{sky}}$ increases, due to the observing time being spread over fewer pixels). The result of these two competing effects is that $\frac{N}{S} \propto \frac{A}{\sqrt{f_{\text{sky}}}} + B \sqrt{f_{\text{sky}}}$ and, thus, it is possible to find a value of $f_{\text{sky}}$ for which the ratio of noise to signal will be minimised.

If the survey strategy needed to be computed for only a single value of $\ell$, then this equation could be differentiated to find the optimal value of $f_{\text{sky}}$, and therefore find the maximum possible S/N ratio for a given noise, beam size and predicted $C_\ell$. However, what I am attempting to determine is the S/N ratio as a function of $r$. This requires finding some method of determining the error in $r$, given the error in $C_\ell$ for a range of values of $\ell$. To do this, I have to make two assumptions. The first assumption is that there are no sources of B-modes other than the primordial gravitational waves. Whilst this is not strictly correct, since B-modes can also be generated by lensing, it
simplifies the analysis considerably and, thus, is useful as a first approximation, to
demonstrate the method. The second assumption is that $r$ simply acts as a scaling
parameter for the primordial B-mode power spectrum, and does not affect the shape.
These two assumptions imply that $C_\ell = rC_\ell(r = 1)$, which turns Equation (3.11) into
an equation for S/N as a function of $r$ and an $\ell$-dependent parameter $C_\ell(r = 1)$. Note
that the optimal scan strategy will, in fact, depend on the value of $r$ and, thus, there
will be a trade-off between the possibility of detecting lower values of $r$ and providing
a better measurement of the spectrum for larger values of $r$. However, in practice the
aim will always be to obtain some kind of detection and, thus, this method can be used
to determine the value of $r$ which the experiment could reasonably detect which will,
in turn, allow us to define the optimal survey strategy.

Whilst this enables me to calculate the error in $r$ for a given value of $\ell$, using only
one value of $\ell$ would be an extremely non-optimal method of measuring $r$. Therefore,
some method of combining the errors for different values of $r$ to give a single S/N ratio
for $r$, as a function of the various parameters and $r$ itself, is required, and this will be
given in the next section (Equation (3.12)). The value $C_\ell$ only appears in Equation
(3.11) once, and as a result of this, the S/N ratio, for a given value of $\ell$, is a function
of the quantity $n_B r$ (as opposed to either $n_B$ or $r$ on their own). This property is very
useful since it means that the minimum detectable $r$ value can easily be given as a
function of $n_B$, which allows the potential performance of an experiment to be easily
characterised. For any given experiment, the system noise, $T_\omega$ is generally difficult to
reduce and, similarly, the total observation time cannot easily be increased. Thus, the
only way to significantly improve the sensitivity of any given experiment is to increase
the number of beams.

Errors for the QUIJOTE experiment

In order to demonstrate the use of the method described in the previous section, I will
now use it to calculate the optimal survey size and the detectable level of $r$ for a po-
tential future experiment. QUIJOTE is a planned experiment to observe polarisation
in the sky at a frequency of 30 GHz, with the major aim being to detect the primordial B-mode polarisation of the CMB. The original plan for QUIJOTE was to use 19 receivers to observe approximately 10000 deg$^2$, some of which will cover the galactic plane and thus will need to be masked out. This gives an $f_{\text{sky}}$ of about 0.15. However, this may not be the optimal proportion of the sky to observe, and thus I will attempt to calculate this using Equation (3.11), as well as computing the minimum value of $r$ (as a function of the number of beams) that the experiment could hope to detect, both with the planned $f_{\text{sky}}$ and if it were designed to observe the theoretically optimal proportion of the sky that I will compute.

In order to do this, I need to know various parameters. The value of $T_\sigma$ is taken to be 250 $\mu$Ks$^{\frac{1}{2}}$, which is a reasonable estimate of what the NET for the QUIJOTE experiment was intended to be (this corresponds to a system temperature of 25 K and a 10 GHz bandwidth), and two values of the beam width ($\theta_{\text{FWHM}}$) are used, 1 deg and 5 deg, which correspond to two different potential designs for the experiment.

For this calculation, I generated a set of $C_\ell$ values using a BB spectrum generated by the CAMB software (Lewis et al. 2000) with $r = 0.1$. Using these, I then generated an estimate for the error on $r$. To do this, I used a Maximum Likelihood (ML) estimator to determine the total error in $r$ from the error in each $\ell$ value. The specific method was taken from Riley et al. (2002). The Maximum-Likelihood estimator for $r$ is defined to be

$$\hat{r} = \frac{\sum_{\ell=2}^n \frac{k_\ell}{\sigma(\hat{C}_\ell)}}{\sum_{\ell=2}^n \frac{1}{\sigma(\hat{C}_\ell)^2}},$$

(3.12)

where $n$ is an arbitrary limit chosen such that the S/N ratio for $\ell > n$ is negligible and I have defined

$$\hat{C}_\ell = \frac{\hat{C}_\ell}{\hat{C}_\ell(r = 1)}$$

(3.13)

and

$$\sigma(\hat{C}_\ell) = \frac{\sigma(\hat{C}_\ell)}{\hat{C}_\ell(r = 1)}.$$

(3.14)
Since the $C_\ell$ values are simulated, the mean of the estimator is irrelevant. However, the error on it is not. Using the properties that $(\sigma(x + y))^2 = (\sigma(x))^2 + (\sigma(y))^2$ and $\sigma(Cx) = C\sigma(x)$ (where $x$ and $y$ are random variables, and $C$ is a constant), it can be shown that the S/N ratio for $r$ ($r/\sigma(r)$) is given by

$$\frac{S}{N} = \sqrt{\sum_{\ell=2}^{n} \left( \frac{C_\ell}{\sigma(C_\ell)} \right)^2}. \tag{3.15}$$

Thus, the signal-to-noise ratio of $r$ is found by summing the signal-to-noise ratios for each $\ell$ in quadrature. Thus, it is possible to determine the S/N ratio for a given value of $n_Br$, $f_{\text{sky}}$ and $\theta_{\text{FWHM}}$. For my program, I looped over values of $f_{\text{sky}}$ from $5 \times 10^{-6}$ up to 0.5 (since for a ground-based experiment, it is not possible to measure more than half of the sky) and over values of $n_Br$ from 0.01 to 100. For each value of $n_Br$, I output two different results. The first was the optimal $f_{\text{sky}}$ value for the given $n_Br$, together with the resultant S/N ratio, and the second was the S/N ratio for the fiducial QUIJOTE $f_{\text{sky}}$ value, 0.15. I did this for two different beam widths, $\theta_{\text{FWHM}} = 1^\circ$ and $\theta_{\text{FWHM}} = 5^\circ$, using a NET of 250$\mu$K$s^{-\frac{1}{2}}$, and plotted the results in Figure (3.1). The top panel shows the S/N as a function of $n_Br$ for the four different cases listed, and the lower panel shows the optimal $f_{\text{sky}}$ value for the cases where this is relevant.

Using these results, it is possible to set a limit on the possible values of $r$ that QUIJOTE could potentially detect. For a 1$\sigma$ measurement ($S/N = 1$), QUIJOTE requires $n_Br = 0.437$ for a $\theta_{\text{FWHM}} = 1^\circ$, and 0.490 for a $\theta_{\text{FWHM}} = 5^\circ$ for $f_{\text{sky}} = 0.15$. Using an optimal $f_{\text{sky}}$ value could in theory reduce these to 0.269 and 0.479 respectively. However, in the first of these cases, the optimal $f_{\text{sky}}$ value is about 0.0009, and in practice this error estimation method would break down (even for T) for $f_{\text{sky}}$ values that small due to the effect of mixing between different $\ell$ modes. For the 5 degree case, the optimal $f_{\text{sky}}$ is about 0.2, and so the QUIJOTE $f_{\text{sky}}$ is near optimal. For the detection to be significant, the $S/N$ ratio needs to be at least 2, and probably 3. For a $S/N$ ratio of 3, a 5 degree beam would require $n_Br$ to be 11.48, although this could be reduced to 6.026 by observing the whole available sky ($f_{\text{sky}} = 0.5$). Thus, to 3$\sigma$, a QUIJOTE experiment with 19 beams and $\theta = 5^\circ$ could only detect $r = 0.6$, which is already
3.1: THEORETICAL ERRORS FOR MASKED CMB MEASUREMENTS

Figure 3.1: QUIJOTE optimal errors: the top panel is the S/N and the bottom panel is the optimal $f_{\text{sky}}$. The solid red line is for $f_{\text{sky}} = 0.15$ (QUIJOTE $f_{\text{sky}}$) and a 1 degree beam, the dashed black line is for the calculated optimal $f_{\text{sky}}$ with a 1 degree beam, the dot-dashed green line is for the QUIJOTE $f_{\text{sky}}$, 5 degree beam and the dotted blue line is for the calculated optimal $f_{\text{sky}}$ with a 5 degree beam.
ruled out by WMAP. To detect $r = 0.1$, QUIJOTE would need about 100 beams. If QUIJOTE were to use a 1 degree beam, the $3\sigma$ limit improves to $n_B r = 4.07$, which would mean that the 19 beam experiment could still only detect $r = 0.21$. Decreasing the beam width any further is unlikely to have any significant effect (since the higher $\ell$ B-modes are negligibly small compared to the noise), and thus for QUIJOTE to set a competitive limit on the value of $r$, it requires considerably more than the planned 19 beams.

However, this is still not the full story. These errors are optimal, in the sense that they are derived using a Maximum-Likelihood estimator under the assumption that there is no correlation between particular modes, nor contamination from other sources. Therefore, the actual errors are likely to be significantly larger than those used in this calculation, especially for small $f_{\text{sky}}$ fractions where mixing between different $\ell$ values is likely to be significant. For the purposes of this thesis, I will concentrate on the effects of the non-optimality of the pure-$C_\ell$ method on this calculation.

### 3.2 Effects of survey size on pure-$C_\ell$ errors

In order to determine if the optimal error formula used in the last section is valid, I have performed simulations for a range of different sized masks, varying from one covering half the sky to one covering only a circular cap region of 500 deg$^2$. For each simulation, I used a cosine apodisation function (see Section 2.3.3 for details) with an apodisation length of $2.5^\circ$ and a bin size of $\Delta \ell = 20$. I have plotted the results in Figure (3.2).

From these results, it can clearly be seen that, in the case of an azimuthal mask, the pure-$C_\ell$ method is near-optimal except at very low values of $\ell$. Even the smallest cut, 500 deg$^2$, is worse than optimal by a factor of less than two for values of $\ell$ greater than 30, and since this mask only covers an area of around $12^\circ$, it would only be expected to be able to detect $\ell > 15$ anyway. This is to be likely due to the fact that an azimuthal mask has a short boundary relative to the total area covered and, thus, the potential problems that boundary effects can cause for the pure-$C_\ell$ method do not arise. At low-
3.2: EFFECTS OF SURVEY SIZE ON PURE-$C_\ell$ ERRORS

Figure 3.2: Signal to noise ratio $\Delta C_\ell / C_\ell$ (left) and error relative to the optimal errors $\Delta C_\ell / \Delta C_\ell (\text{optimal})$ (right) given by Equation (3.2) for azimuthal masks with a half sky cut (black line) and 10000 deg$^2$ (green dotted), 5000 deg$^2$ (red dashed), 1000 deg$^2$ (blue long dashed) and 500 deg$^2$ (magenta dot-dashed) spherical cap cuts $\ell$, the errors are between five and seven times worse than optimal for the lowest bin. Even here, however, the increase relative to optimal is only weakly sensitive to the size of the observed region, and this issue may, in fact, be partly due to the fact that the area covered by the mask is simply not large enough to allow for the power spectrum to be accurately measured at large scales.

The proximity of these results to optimality implies that the calculation carried out in the previous section should be a good approximation to the truth, although the fact that the standard deviation is substantially worse at very low-$\ell$ values (where the noise is lowest) will slightly increase the threshold value of $r$ that can be detected. In addition, the optimal value of $f_{\text{sky}}$ predicted in Section 3.1.2 will be lower than is actually the case.
3: SURVEY STRATEGIES AND FUTURE INSTRUMENTS

3.3 Calculation of variance for pseudo-$C_\ell$ methods

In order to understand better the effects of boundary terms on the errors for the pure-$C_\ell$ method, it is useful to consider an analytical estimate for the errors. In order to do this, I will need to extend the calculation performed in [Challinor and Chon (2005)] to cover the pure-$C_\ell$ method. First, however, I will demonstrate the effectiveness of the formula given there in the case of ordinary pseudo-$C_\ell$.

3.3.1 E-B mixing in the standard pseudo-$C_\ell$ approach

For the case of the standard pseudo-$C_\ell$ approach, the optimal error estimates given in (3.1.1) work reasonably well for T and E measurements, but are inadequate for estimating the errors in the BB spectrum, which are dominated by the effects of E-B mixing. Therefore, an analytical approximation to the errors in the BB spectrum is required. [Challinor and Chon (2005)] presents a calculation which models the increased variance due to E-B mixing. By assuming that $C_\ell$ is constant over the range of $\ell$ values for which mixing occurs, they find an approximate expression for the error in $C_{\ell}^{BB}$, in the limit that $C_{\ell}^{BB} \ll C_{\ell}^{EE}$. However, their calculation is only valid for the case where the mask is smooth. This means that apodisation must be applied to the edges of the unmasked region (i.e. the mask is no longer either 0 or 1 everywhere). As a result of this, carrying out the calculation for a non-azimuthal mask is non-trivial, and would require an implementation of a finite differencing scheme on the HEALpix grid. Thus, I have only considered the effect for azimuthal masks. In particular, I have considered the mask with the weight function (setting $\theta = 0$ to be the equator)

$$w(\theta, \phi) = \begin{cases} 
1 & \theta > 80^\circ \\
\cos^2\left(\frac{\pi(80-\theta)}{10}\right) & 80^\circ \geq \theta \geq 75^\circ \\
0 & \theta < 75^\circ \end{cases} .$$

(3.16)
3.3: CALCULATION OF VARIANCE FOR PSEUDO-\(C_\ell\) METHODS

Challinor and Chon (2005) find that

\[
\text{var}(\hat{C}_{\ell}^{EE}) \approx \frac{2w^{(4)}(C_{\ell}^{EE})^2}{(2\ell + 1)\Delta f_{\text{sky}}(w^2)^2} \quad (C_{\ell}^{BB} = 0) \quad (3.17)
\]

\[
\text{var}(\hat{C}_{\ell}^{BB}) \approx \frac{2(C_{\ell}^{EE})^2}{(2\ell + 1)\Delta f_{\text{sky}} \ell^2(\ell + 1)^2} \times \left( \frac{(\nabla w)^{(4)}}{(w^{(2)})^4} + \frac{2w^{(4)}((\nabla w)^{(2)})^2}{3(w^{(2)})^4} - \frac{4((\nabla w)^{(2)}(w\nabla w)^{(2)})}{3(w^{(2)})^3} \right) \quad (C_{\ell}^{BB} = 0), \quad (3.18)
\]

where \(4\pi f_{\text{sky}}X^{(i)} \equiv \int |X'| d\Omega\), with \(X = w, \nabla w\) or \(w\nabla w\).

Note that the equation for EE is in fact the generalised case of the optimal error formula for an apodised mask, and thus can be extended to include TT and the other spectra (not involving B). The error on BB, however, has an entirely different form. In particular, if I use a generalised version of Equation (3.16), the gradient terms are proportional to \(1/A\), where \(A\) is the width of the apodised region, and thus the predicted error in BB diverges as the apodisation goes to zero. In fact, at a certain point, the approximations used to calculate the formula break down. The calculation of the mask-dependant terms is very long and involved, with complicated answers, so I will not derive the result for an azimuthal mask here, nor will I state the result. However, I have carried out the calculation for the weight function in Equation (3.16) and I have plotted a graph of \(\Delta C_\ell/C_\ell\) for this, together with simulations carried out using a mask with the weight function given, and also with the non-apodised version of it (with \(w = 1\) everywhere that it is non-zero in Equation (3.16)). I have done this for two different values of \(r\) (0.1 and 0.05), and for two different bin sizes \((\Delta \ell = 10\) and \(\Delta \ell = 50\)), and these are plotted in Figure 3.3.

As can be seen from Figure 3.3, the errors predicted by Equation (3.18) are a very good fit to the Monte-Carlo simulation errors (for the apodised mask) for \(\ell > 50\). Thus, these error estimates can provide some insight into the problems caused by E-B mixing. In particular, since the size of the mask-dependent term is mainly a function of the gradient of the weight function, summed over all pixels, a weight function with large numbers of boundaries must have a larger number of regions with non-zero gradients and thus will have a larger E-B mixing term. This accounts for the much higher
Figure 3.3: Comparison of errors from Monte-Carlo simulations for unapodised (black) and apodised (blue) mask with unapodised optimal errors (green) computed using Equation (3.2), apodised optimal errors (red) computed using Equation (3.17) and errors due to E-B mixing (magenta) computed using Equation (3.18) for $r = 0.1$ (left) and $r = 0.05$ (right), using $\Delta l = 10$ (top) and $\Delta l = 50$ (bottom).
3.3: CALCULATION OF VARIANCE FOR PSEUDO-$C_\ell$ METHODS

errors for the non-azimuthal masks seen in the previous chapter. In general, the size of the gradient term will increase approximately as the size of the circumference of the observed region, and so the error on $C^B_B$ should be similarly dependent on the circumference.

3.3.2 Calculation of error for pure-$C_\ell$

In order to understand better the pure-$C_\ell$ method, it is useful to attempt to extend the above analysis to cover it, using the same methods as are used in Challinor and Chon (2005). I start from the (exact) formula for the covariance matrix

$$\text{Cov}(\tilde{C}_\ell^X, \tilde{C}_{\ell'}^X) = \frac{2}{(2\ell + 1)(2\ell' + 1)} \sum_{m,m'} \left| \sum_{\ell_1, m_1, Y} M_{\ell_1, mm_1}^{XY} (M_{\ell_1, mm_1}^{Y'})^* C_{\ell_1}^Y \right|^2,$$

where

$$\tilde{a}_{\ell m}^Y = \sum_{\ell_1, m_1, X} M_{\ell_1, mm_1}^{XY} a_{\ell_1 m_1}^X.$$  (3.20)

Now, in Smith and Zaldarriaga (2007), the definition for the pure-$C_\ell$ $a_{\ell m}$s is given as

$$\tilde{a}_{\ell m}^Y = \frac{1}{\sqrt{(\ell - 1)\ell(\ell + 1)(\ell + 2)}} \int \chi_B(\Omega) W(\Omega) Y_{\ell m}(\Omega) d\Omega,$$

where $\chi_B$ is defined such that

$$\chi_B(\Omega) = \sum_{\ell, m} \sqrt{(\ell - 1)\ell(\ell + 1)(\ell + 2)} a_{\ell m}^B Y_{\ell m}(\Omega).$$  (3.22)

Since this is spin zero, there is no mixing with the equivalent for the E-mode (which can, in principle, be defined in the same way). In this case, the covariance matrix is given by

$$\text{Cov}(\tilde{C}_\ell^{XY}, \tilde{C}_{\ell'}^{XY}) = \frac{2}{(2\ell + 1)(2\ell' + 1)} \sum_{m,m'} \left| \sum_{\ell_1} M_{\ell_1, mm_1}^{XY} (M_{\ell_1, mm_1}^{Y'})^* C_{\ell_1}^B \right|^2,$$

where $M^{XY}$ is given by

$$M_{\ell_1, mm_1}^{XY} = \sqrt{\frac{(\ell' - 1)\ell'(\ell' + 1)(\ell' + 2)}{(\ell - 1)\ell(\ell + 1)(\ell + 2)}} \int Y_{\ell_1 m_1}(\Omega) Y_{\ell m_1}(\Omega) W(\Omega) d\Omega.$$  (3.24)
3: SURVEY STRATEGIES AND FUTURE INSTRUMENTS

However, exact computation of Equation (3.23) is $O(\ell^4)$, and thus it is not computationally feasible to exactly calculate the variance. However, if I can find some approximation to the sum

$$\sum_{\ell_1, m_1} M_{\ell_1 m_1}^B (M_{\ell_1 m_1}^B)^\star C_{\ell_1}^{BB},$$  \hspace{1cm} (3.25)

then the remaining summation could be done. The major problem with doing this sum is the $C_{\ell_1}$ term. There are two possible approximations I can apply to eliminate this. I can either assume (as in Challinor and Chon 2005) that $C_{\ell} \approx \text{const.}$ or that $\ell (\ell + 1) C_{\ell} = \text{const.}$ Which of these two approximations is better depends on the nature of the power spectrum. For a spectrum that is mainly noise-dominated, the first approximation is clearly the more valid. However, in the case of signal domination it is not quite so simple, due to the fact that the B-mode power spectrum rises rapidly from $\ell = 10$ to around $\ell = 100$, with $C_{\ell} \approx \text{const.}$ in that range, as shown in Figure 1.2. Due to this rapid rise, the usual assumption for signal-dominated power spectra (that $\ell (\ell + 1) C_{\ell} = \text{const.}$) does not really hold in this range, and thus the $C_{\ell} \approx \text{const.}$ approximation is probably more valid. However, there is also a peak in the spectrum at very low-$\ell$ (below $\ell = 10$), and in fact $\ell (\ell + 1) C_{\ell}$ is the same at $\ell = 4$ as it is at around $\ell \sim 30$. Further, since the difference between the two methods is most pronounced at lower $\ell$, it could be argued that this effect makes the second approximation more valid than the first. Further, as shown in Figure 1.2 neither approximation is particularly valid on large scales, and since the second assumption makes the calculation substantially easier, I have used it in preference to the alternative.

Under the assumption explained above, the $C_{\ell_1}$ term can be removed from the summation by using the approximation $\ell_1 (\ell_1 + 1) C_{\ell_1} \approx \sqrt{\ell (\ell + 1) \ell' (\ell' + 1) C_{\ell} C_{\ell'}}$. This gives that

$$\text{Cov}(\tilde{C}_{\ell}^{\chi\chi}, \tilde{C}_{\ell'}^{\chi\chi}) \approx \frac{2 C_{\ell}^{BB} C_{\ell'}^{BB}}{(2\ell + 1)(2\ell' + 1)} \sum_{m, m'} \sum_{\ell_1, m_1} \sqrt{\ell (\ell + 1) \ell' (\ell' + 1)} \ell_1 (\ell_1 + 1) M_{\ell_1 m_1}^B (M_{\ell_1 m_1}^B)^\star \left(\sum_{\ell_1, m_1} \ell_1 (\ell_1 + 1) \ell_1 (\ell_1 + 1) M_{\ell_1 m_1}^B (M_{\ell_1 m_1}^B)^\star \right)^2 \hspace{1cm} (3.26)$$

104 ANALYSIS OF CMB POLARISATION
After rearrangement and application of integration by parts, I get

\[ \sum_{\ell_1,m_1} \sqrt{\frac{\ell_1(\ell_1+1)}{\ell_1(\ell_1+1)}} M^{B}_{\ell_1,m_1} (M^{B}_{\ell',m'})^* = \frac{1}{\sqrt{(\ell - 1)(\ell + 2)(\ell' - 1)(\ell' + 2)}} \]  

(3.27)

\[ \times \int \int Y^*_{\ell,m}(\Omega)Y^*_{\ell',m'}(\Omega')W(\Omega)W'(\Omega) \left( \sum_{\ell_1,m_1} (\ell_1 - 1)(\ell_1 + 2)Y_{\ell_1,m_1}(\Omega)Y_{\ell_1,m_1}'(\Omega') \right) \mathrm{d}\Omega \mathrm{d}\Omega'. \]

Using the definitions of the spin-weighted spherical harmonics, I can rewrite the sum in Equation (3.27) as

\[ \sum_{\ell_1,m_1} \delta (1Y_{\ell,m}(\Omega)) \delta (1Y^*_{\ell,m}(\Omega')) + 2\delta (\Omega - \Omega'). \]  

(3.28)

After rearrangement and application of integration by parts, I get

\[ \sum_{\ell_1,m_1} M^{B}_{\ell_1,m_1} (M^{B}_{\ell',m'})^* C^{BB}_{\ell_1} \approx \sqrt{\frac{C^{BB}_{\ell} C^{BB}_{\ell'}}{(\ell - 1)(\ell + 2)(\ell' - 1)(\ell' + 2)}} \]  

(3.29)

\[ \times \left( 2 \int Y^*_{\ell,m}(\Omega)Y^*_{\ell',m'}(\Omega) |W(\Omega)|^2 \mathrm{d}\Omega 
+ \int \delta (Y^*_{\ell,m}(\Omega)W(\Omega)) \delta (Y^*_{\ell',m'}(\Omega)W(\Omega)) \mathrm{d}\Omega' \right), \]

which can be expanded using the product rule to give

\[ \sum_{\ell_1,m_1} M^{B}_{\ell_1,m_1} (M^{B}_{\ell',m'})^* C^{BB}_{\ell_1} \approx \sqrt{C^{BB}_{\ell} C^{BB}_{\ell'}} \left( \sqrt{\frac{\ell(\ell + 1)}{(\ell - 1)(\ell + 2)(\ell' - 1)(\ell' + 2)}} \right. 
\times \int Y^*_{\ell,m}(\Omega)Y^*_{\ell',m'}(\Omega)W_{1}(\Omega)W(\Omega) \mathrm{d}\Omega 
+ \sqrt{\frac{1}{(\ell - 1)(\ell + 2)(\ell' - 1)(\ell' + 2)}} \times \int Y^*_{\ell,m}(\Omega)Y^*_{\ell',m'}(\Omega) \left( |W_{1}(\Omega)|^2 - 2 |W(\Omega)|^2 \right) \mathrm{d}\Omega 
\times \sqrt{\frac{\ell(\ell' + 1)}{(\ell - 1)(\ell + 2)(\ell' - 1)(\ell' + 2)}} \times \int Y^*_{\ell,m}(\Omega)Y^*_{\ell',m'}(\Omega)W_{1}(\Omega)W(\Omega) \mathrm{d}\Omega 
\times \sqrt{\frac{\ell(\ell + 1)\ell'(\ell' + 1)}{(\ell - 1)(\ell + 2)(\ell' - 1)(\ell' + 2)}} \times \int Y^*_{\ell,m}(\Omega)Y^*_{\ell',m'}(\Omega) |W(\Omega)|^2 \mathrm{d}\Omega'. \]  

(3.30)
3: SURVEY STRATEGIES AND FUTURE INSTRUMENTS

Using the definition

$$ W_{ab} (\Omega) = W_a (\Omega) W_b^* (\Omega) $$ \hspace{1cm} (3.31)

and the expression for integrals over three spherical harmonics in terms of $3j$-functions (Brown et al. [2005]), the right-hand side of Equation (3.30) can be rewritten as

$$ \sqrt{C_{\ell}^{BB} C_{\ell'}^{BB}} \sum_{\ell, m} (-1)^m \sqrt{(2\ell + 1)(2\ell' + 1)(2\ell_1 + 1)} \frac{1}{4\pi} \begin{pmatrix} \ell & \ell' & \ell'' \end{pmatrix} (-m \quad m' \quad m_1) $$ \hspace{1cm} (3.33)

Now, defining

$$ f(\ell, \ell', \ell_1, m_1) = \begin{pmatrix} \frac{w_{\ell m_1}^{(11)} - 2w_{\ell m_1}^{(00)}}{\sqrt{(\ell - 1)(\ell + 2)(\ell' - 1)(\ell' + 2)}} \frac{1}{0 \quad 0 \quad 0} \end{pmatrix} $$ \hspace{1cm} (3.34)

the orthogonality of the $3j$ functions can be used to show that

$$ \text{Cov}(C_{\ell}^{BB}, C_{\ell'}^{BB}) = \frac{C_{\ell}^{BB} C_{\ell'}^{BB}}{2\pi} \sum_{\ell, m_1} |f(\ell, \ell', \ell_1, m_1)|^2. $$ \hspace{1cm} (3.35)
3.3: CALCULATION OF VARIANCE FOR PSEUDO-$C_\ell$ METHODS

Now, the $3j$ function recursion relations can be applied to show that

$$f(\ell, \ell', \ell_1, m_1) = \left( \frac{w_{\ell_1m_1}^{(11)} - 2w_{\ell_1m_1}^{(00)} - \sqrt{\ell_1(\ell_1 + 1)}w_{\ell_1m_1}^{(10)} - \ell'(\ell' + 1)w_{\ell_1m_1}^{(00)}}{\sqrt{(\ell - 1)(\ell + 2)(\ell' - 1)(\ell' + 2)}} \right) \begin{pmatrix} \ell & \ell' & \ell_1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$+ \left( w_{\ell_1m_1}^{(00)} - \sqrt{\ell_1(\ell_1 + 1)}w_{\ell_1m_1}^{(00)} \right) \frac{\sqrt{\ell'(\ell' + 1)}}{\sqrt{(\ell - 1)(\ell + 2)(\ell' - 1)(\ell' + 2)}} \begin{pmatrix} \ell & \ell' & \ell_1 \\ 0 & -1 & 1 \end{pmatrix}$$

$$- \left( w_{\ell_1m_1}^{(10)} \sqrt{\ell'(\ell' + 1)} \right) \frac{\sqrt{(\ell - 1)(\ell + 2)(\ell' - 1)(\ell' + 2)}}{\sqrt{(\ell - 1)(\ell + 2)(\ell' - 1)(\ell' + 2)}} \begin{pmatrix} \ell & \ell' & \ell_1 \\ 0 & 1 & -1 \end{pmatrix}.$$  \hspace{1cm} (3.36)

3.3.3 Analytic expression for errors

The expressions given above are not particularly quick to calculate, and cannot be used to give an idea of the effect of different masks on the error. I will therefore now follow similar methods to those described in [Challinor and Chon (2005)] to find an approximation to the formula that can be easily calculated.

I have that

$$\text{Cov}(\tilde{C}_\chi^{yy}_\ell, \tilde{C}_\chi^{yy}_\ell') = \frac{C^{BB}_\ell C^{BB}_\ell'}{2\pi} \sum_{\ell_1, m_1} |f(l, l', l_1, m_1)|^2,$$  \hspace{1cm} (3.37)

with

$$f(l, l', l_1, m_1) = \sum_a f_a(l, l', l_1, m_1) \begin{pmatrix} \ell & \ell' & \ell'' \\ 0 & a & -a \end{pmatrix}.$$  \hspace{1cm} (3.38)

Where the $f_a$ are functions of the mask $a_{lm}$, as shown in equation (3.36). However, this alone is not sufficient to calculate an analytic formula for the pure-$C_\ell$ errors. Further approximations must be made. In order to do this, consider the process of calculating an unbiased estimate of the $C_\ell$. To do this, I must deconvolve the pure-$C_\ell$ estimator. From the definition of pure-$C_\ell$, the error in the deconvolved pure-$C_\ell$s can be shown to be given by

$$\text{Cov}(\hat{C}_\chi^{yy}_\ell, \hat{C}_\chi^{yy}_\ell') = \sum_{\ell_1, \ell_2} M^{yyBB}_{\ell_1} M^{yyBB}_{\ell_2} \text{Cov}(\hat{C}^{BB}_{\ell_1}, \hat{C}^{BB}_{\ell_2}).$$  \hspace{1cm} (3.39)

MICHAEL PREECE

107
Using the 3j function recursion relations, it is possible to show that

\[ M_{\ell' \ell''}^{BB} = \frac{(2\ell' + 1)(\ell' - 1)\ell'\ell' + 1)(\ell' + 2)}{4\pi(\ell - 1)(\ell + 1)(\ell + 2)} \sum_{\ell''} (2\ell'' + 1) \begin{pmatrix} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{pmatrix}^2 W_{\ell''}, \tag{3.40} \]

and \( W_{\ell} \) is the power spectrum of the mask (Equation (2.10)).

Now, suppose that binning is applied to the power spectrum, defining

\[ \tilde{C}_{\ell}^{XX} = \frac{1}{\Delta \ell} \sum_{\ell = \ell - \Delta \ell/2}^{\ell + \Delta \ell/2} \ell(\ell + 1) C_{\ell}^{XX}. \tag{3.41} \]

Then, I can write

\[
\text{Cov} \left( C_{\ell_1}^{XX}, C_{\ell_2}^{XX} \right) = \frac{1}{\Delta \ell_1} \sum_{\ell_1 = \ell - \Delta \ell_1/2}^{\ell + \Delta \ell_1/2} \sum_{\ell_2 = \ell - \Delta \ell_2/2}^{\ell + \Delta \ell_2/2} M_{\ell_1 \ell_2}^{BB} M_{\ell_1 \ell_2}^{BB} \text{Cov} \left( \tilde{C}_{\ell_1}^{BB}, \tilde{C}_{\ell_2}^{BB} \right).
\]

(3.42)

Now, assuming that the bin size is larger than \( \ell_{\text{max}} \) (the value of \( \ell \) where the mask power spectrum becomes negligibly small), the properties of the 3j functions imply that \( M_{\ell_1 \ell_2} = 0 \) for \( \ell_1 \neq \ell_2 \), and also that the limits of the sum over \( \ell_1 \) and \( \ell_2 \) can be extended to 0 and \( \infty \).

Doing this, and applying Equation (3.40), I obtain

\[
\sum_{\ell_1, \ell_2} \ell(\ell + 1)\ell'\ell' + 1) M_{\ell_1 \ell_2}^{BB} M_{\ell_1 \ell_2}^{BB} = \sum_{\ell_1, \ell_2, \ell'} \frac{(2\ell_1 + 1)(\ell_1 - 1)\ell_1 + 2)}{(\ell - 1)(\ell + 2)} \frac{(2\ell_2 + 1)(2\ell_2 + 1)(2\ell_2 + 1)(\ell_2 - 1)(\ell_2 + 2)}{(\ell' - 1)(\ell' + 2)} \times \left( \begin{array}{ccc} \ell & \ell_1 & \ell'' \\ 0 & 0 & 0 \end{array} \right)^2 \left( \begin{array}{ccc} \ell' & \ell_2 & \ell'' \\ 0 & 0 & 0 \end{array} \right)^2 W_{\ell''}. \tag{3.43} \]

Using the 3j function recursion relations, it is possible to show that

\[
\sum_{\ell_1} \frac{(2\ell_1 + 1)(\ell_1 - 1)\ell_1 + 2)}{(\ell - 1)(\ell + 2)} \left( \begin{array}{ccc} \ell & \ell_1 & \ell'' \\ 0 & 0 & 0 \end{array} \right)^2 = \ell(\ell + 1)\ell_1 + 1 \sum_{\ell_1} \frac{(2\ell_1 + 1)}{(\ell - 1)(\ell + 2)} \begin{pmatrix} \ell & \ell_1 & \ell'' \\ 1 & -1 & 0 \end{pmatrix}^2
\]

\[ + 2 \sqrt{\frac{\ell(\ell + 1)\ell''(\ell'' + 1)}{(\ell - 1)(\ell + 2)}} \sum_{\ell_1} (2\ell_1 + 1) \begin{pmatrix} \ell & \ell_1 & \ell'' \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell_1 & \ell'' \\ 0 & 1 & 1 \end{pmatrix} \tag{3.44} \]

\[ + \frac{\ell''(\ell'' + 1)}{(\ell - 1)(\ell + 2)} \sum_{\ell_1} (2\ell_1 + 1) \begin{pmatrix} \ell & \ell_1 & \ell'' \\ 0 & 1 & 1 \end{pmatrix} - \frac{2}{(\ell - 1)(\ell + 2)} \sum_{\ell_1} (2\ell_1 + 1) \begin{pmatrix} \ell & \ell_1 & \ell'' \\ 0 & 0 & 0 \end{pmatrix}^2. \]
3.3: CALCULATION OF VARIANCE FOR PSEUDO-$$C_\ell$$ METHODS

From Challinor and Chon (2005), I have that

$$\begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ n_1 & n_2 & n_3 \end{pmatrix} = \frac{1}{2} \int_{-1}^{1} d_{n_1m_1}(\beta) d_{n_2m_2}(\beta) d_{n_3m_3}(\beta) d(\cos \beta), \quad (3.45)$$

for $$n_1 + n_2 + n_3 = 0$$ and $$m_1 + m_2 + m_3 = 0$$, and, using the completeness relation (Chon et al. 2004)

$$\sum_{\ell} \frac{(2\ell + 1)}{2} d_{\ell m}(\beta_1) d_{\ell m}(\beta_2) = \delta(\cos \beta_1 - \cos \beta_2), \quad (3.46)$$

and the fact that $$d_{s,s}^{(3.44)}(0) = \delta_{s,s'}$$ (Wigner 1931), I get that

$$\sum_{\ell_2} (2\ell_2 + 1) \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ s_1 & s_2 & s_3 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ s_1' & s_2' & s_3' \end{pmatrix} = d_{s_1s_1'}^{(3.45)}(0) d_{s_2s_2'}^{(3.49)}(0) = \delta_{s_1s_1'} \delta_{s_2s_2'}, \quad (3.47)$$

provided that $$s_1 + s_2 + s_3 = 0$$ and $$s_1' + s_2' + s_3' = 0$$. Substituting this into Equation (3.44), it can be shown that

$$\text{Cov}(\tilde{C}_\ell^{xx}, \tilde{C}_{\ell'}^{xx}) \approx \frac{\text{Cov}^{BB} (\tilde{C}_\ell^{BB}, \tilde{C}_{\ell'}^{BB})}{16\pi^2} \sum_{\ell-\Delta/2}^{\ell+\Delta/2} \sum_{\ell'-\Delta/2}^{\ell'+\Delta/2} \sum_{\ell''} (2\ell'' + 1) \left( 1 + \frac{\ell''(\ell'' + 1)}{(\ell' - 1)(\ell + 2)} \right) W_{\ell''} \times \sum_{\ell'''} (2\ell''' + 1) \left( 1 + \frac{\ell'''(\ell''' + 1)}{(\ell' - 1)(\ell' + 2)} \right) W_{\ell'''} \right). \quad (3.48)$$

Using the fact that

$$W_s(\Omega) = \delta^s(W(\Omega)) = \sum_{\ell,m} \sqrt{\frac{(\ell + s)!}{(\ell - s)!}} W_{\ell m} Y_{\ell m}, \quad (3.49)$$

defining $$4\pi X^{(s)} f_{\text{sky}} = \int |X(\Omega)|^2 d\Omega$$, as in Challinor and Chon (2005), and using the fact that

$$\sum_{\ell} (2\ell + 1) s_W_{\ell} = \sum_{\ell,m} |s_{W_{\ell m}}|^2 = \int W(\Omega) W^{s'}(\Omega') \sum_{\ell,m} s Y_{\ell m}^s(\Omega) s Y_{\ell m}(\Omega') d\Omega d\Omega' = \int |W(\Omega)|^2 d\Omega, \quad (3.50)$$

MICHAEL PREECE

109
ANALYSIS OF CMB POLARISATION

Therefore, the above sum must be computed. Making the assumption that the bin size is large relative to $k$, where Equation (3.51) allows us to determine the value of the constant $\text{Cov}$, which can be inverted to find an approximate expression for the error in the binned, deconvolved $C_\ell$ values, given one for the error in the pure pseudo-$C_\ell$ values. Further, if $\ell$ is large, the $W_1$ terms can also be neglected.

Now, I can use Equation (3.39), and the definition of $\tilde{C}_\ell^{\chi \chi}$, in order to find an approximation for this error

$$\text{Cov} \left( \tilde{C}_\ell^{\chi \chi}, \tilde{C}_\ell^{\chi \chi} \right) \approx \frac{\text{Cov} \left( \tilde{C}_\ell^{BB}, \tilde{C}_\ell^{BB} \right)}{\Delta \ell^2} \times \sum_{\ell=\ell_{\Delta / 2}}^{\ell+\Delta / 2} \sum_{\ell'=\ell-\Delta / 2}^{\ell'+\Delta / 2} \left( W^{(2)} \right) \left( W^{(2)} \right) \left( \ell + 1 \right) \left( \ell' + 1 \right) \frac{C_{\ell}^{BB} C_{\ell'}^{BB}}{4 \pi^2 \Delta \ell^2},$$

(3.51)

which can be inverted to find an approximate expression for the error in the binned, deconvolved $C_\ell$ values, given one for the error in the pure pseudo-$C_\ell$ values. Further, if $\ell$ is large, the $W_1$ terms can also be neglected.

Now, I can use Equation (3.39), and the definition of $\tilde{C}_\ell^{\chi \chi}$, in order to find an approximation for this error

$$\text{Cov} \left( \tilde{C}_\ell^{BB}, \tilde{C}_\ell^{BB} \right) \approx k(\tilde{\ell}, \tilde{\ell}') \sum_{\ell=\ell_{\Delta / 2}}^{\ell+\Delta / 2} \sum_{\ell'=\ell-\Delta / 2}^{\ell'+\Delta / 2} \frac{\ell(\ell + 1) \ell'(\ell' + 1) C_{\ell}^{BB} C_{\ell'}^{BB}}{4 \pi^2 \Delta \ell^2} \frac{\pi}{2},$$

(3.52)

$$\times \sum_{\ell_{1, m_1}} f_a(\ell, \ell', \ell_1 m_1) f_b(\ell, \ell', \ell_1 m_1) \left( \ell \quad \ell' \quad \ell'' \right) \left( 0 \quad a \quad -a \right) \left( 0 \quad b \quad -b \right),$$

where Equation (3.51) allows us to determine the value of the constant $k$. Further making the assumption that the bin size is large relative to $\ell$, it can be assumed that $\ell \approx \tilde{\ell}$. Then, a term $(2\ell + 1)/(2\tilde{\ell} + 1)$ can be added into the sum. Now, supposing that the bin size is larger than $\ell_{\text{max}}$, I can extend the summation limits to 0 and $\infty$. This allows the orthogonality relations of the $3j$ functions to be applied, to give that

$$\text{Cov} \left( \tilde{C}_\ell^{BB}, \tilde{C}_\ell^{BB} \right) \approx \frac{k(\tilde{\ell}, \tilde{\ell}')} {\Delta \ell^2} \sum_{\ell'=\ell-\Delta / 2}^{\ell'+\Delta / 2} \frac{\ell'(\ell' + 1) C_{\ell}^{BB} C_{\ell'}^{BB}}{4 \pi^2 (2\tilde{\ell} + 1)} \sum_{\ell_{1, m_1}} \left| f_a(\tilde{\ell}, \tilde{\ell}', \ell_1 m_1) \right|^2,$$

(3.53)

where I have defined

$$C_{\ell}^{BB} = \frac{1}{\Delta \ell} \sum_{\ell=\ell_{\Delta / 2}}^{\ell+\Delta / 2} \frac{\ell(\ell + 1) C_{\ell}^{BB}}{2 \pi}.$$

(3.54)

Therefore, the above sum must be computed.

The form of the $f_a$ functions implies that it will be necessary to do sums of the form

$$\sum_{\ell_{1, m_1}} W_{\ell_{1, m_1}}^{(ab)} \left( W_{\ell_{1, m_1}}^{(cd)} \right)^*,$$

(3.55)
3.3: CALCULATION OF VARIANCE FOR PSEUDO-$C_\ell$ METHODS

where $a - b = c - d$ (ensuring the spin-weight of each function is the same). The definition of $w_{\ell m}$ allows us to write

$$
\sum_{\ell, m} W_{\ell m}^{(ab)} (W_{\ell m}^{(cd)})^* = \sum_{\ell, m} \int W_{ab}(\Omega)_{a-b} Y_{\ell m}(\Omega) d\Omega \int W_{cd}(\Omega')_{c-d} Y_{\ell m}(\Omega') d\Omega' = \int W_{ab}(\Omega) W_{cd}(\Omega)^* d\Omega = \int W_a(\Omega) W_b^*(\Omega) W_c(\Omega) W_d(\Omega) d\Omega,
$$

(3.56)

where the second formula follows from the first using the completeness relation, and the fact that $a - b = c - d$, and the $f_a$ functions are given by the various terms of Equation (3.36) to be

$$
f_0(\ell, \ell', \ell_1, m_1) = \frac{w_{\ell_1 m_1}^{(11)} - (\ell' (\ell' + 1) + 2) w_{\ell_1 m_1}^{(00)} - \sqrt{\ell_1(\ell_1 + 1) w_{\ell_1 m_1}^{(10)}}}{\sqrt{(\ell - 1) (\ell + 2) (\ell' - 1) (\ell' + 2)}} \quad \text{etc.}
$$

(3.57)

Now, assuming that $\ell$ and $\ell'$ are large, all the terms except for the $w_{000}$ term can be eliminated, leaving

$$
f_0(\ell, \ell', \ell_1, m_1) \approx \frac{-(\ell' (\ell' + 1)) w_{\ell_1 m_1}^{(00)}}{\sqrt{(\ell - 1) (\ell + 2) (\ell' - 1) (\ell' + 2)}}
$$

(3.58)

which allows us to show that

$$
\sum_{\ell, m_1} |f_0(\ell, \ell', \ell_1, m_1)|^2 \approx \frac{(\ell' (\ell' + 1))^2}{(\ell - 1) (\ell + 2) (\ell' - 1) (\ell' + 2)} \int |W_0(\Omega)|^4 d\Omega. \quad (3.62)
$$

Then, using the definition $4\pi X^{(i)} f_{\text{sky}} = \int |X(\Omega)|^4 d\Omega$, using the fact that, for moderately large values of $\ell$, $\ell (\ell + 1) \approx (\ell - 1) (\ell + 2)$ and assuming that $\ell \approx \ell'$ (which, of course, is completely true provided I am trying to compute the variance), this can be re-written as

$$
\sum_{\ell, m_1} |f_0(\ell, \ell', \ell_1, m_1)|^2 \approx 4\pi W_0^{(\ell)} f_{\text{sky}}.
$$
Then, finally, I can plug this into Equation (3.51) and then take the high-$\ell$ limit to get an approximate expression for the variance of the binned, deconvolved $C_{\ell}$ values:

$$
\text{Var}(\tilde{C}^{BB\bar{\ell}}_{\ell}) \approx \frac{2}{(2\ell + 1)\Delta f_{\text{sky}}} \left( \frac{W_0^{(4)}}{W_0^{(2)}} \right)^2, \quad (3.63)
$$

which is identical to the EE error formula for the ordinary pseudo-$C_{\ell}$ method given in Equation (3.17). This calculation thus shows that, at high-$\ell$, the pure-$C_{\ell}$ method is as good at detecting B-modes as the ordinary pseudo-$C_{\ell}$ method would be in the absence of E-B mixing.

However, unlike in the ordinary pseudo-$C_{\ell}$ case, there are also additional terms in the equation, which come into effect at low-$\ell$. Looking at the terms discarded from Equation (3.57), it can be seen that the assumption that these terms are small is only true if $\ell > \ell_{\text{mask}}$, where $\ell_{\text{mask}}$ is the maximum $\ell$ for which the mask power spectrum contains significant power. Thus, the pure-$C_{\ell}$ method is approximately optimal only for values of $\ell$ larger than the scale of the mask. This scale will depend on several things. It depends on the size of the mask, and also on the apodisation applied to it, since an unapodised mask will produce ‘ringing’, resulting in power in the mask spectrum even at the highest scales. On larger scales than this, there will be significant non-optimality in the errors output by the pure-$C_{\ell}$ method. This result explains why the boundary effects only seem to affect the pure-$C_{\ell}$ method in the lowest bin or two for almost all masks tested.

## 3.4 Conclusions

In this chapter, I have investigated optimal survey strategies for various experiments, using both theoretical error estimates and simulations. First, I looked at the usage of optimal error formulas to determine an optimal scan strategy and found that, when noise is taken into account, it is no longer necessarily optimal to scan the largest possible proportion of the sky, and that there is an experiment-dependent optimal $f_{\text{sky}}$ value, which also depends on the value of $r$ the experiment is aiming to detect. I then applied
3.4: CONCLUSIONS

this analysis to the QUIJOTE experiment and showed that, with a $1^\circ$ beam width, it would be able to detect a value of $r \approx 4/n_B$, where $n_B$ is the number of beams.

Then, I looked at the effect of issues with the pure-$C_\ell$ method on this analysis. Here, I found that, whilst there is a substantial increase in the variance at very low-$\ell$ values relative to the optimal case, for an azimuthal mask this effect is largely limited to $\ell < 20$ and, thus, it is unlikely to have a significant effect on the optimal survey strategy, although it will likely slightly increase the optimal $f_{\text{sky}}$ value. However, it would be useful to carry out an analysis using non-azimuthal masks. In particular, a mask made up of various numbers of non-intersecting circles would give a good indication of the effect of increasing the boundary length on the method, and this would be a good approach to take in the future.

Finally, I carried out an analytic calculation to attempt to find an approximation to the pure-$C_\ell$ errors. Firstly, I outlined the equivalent calculation for the pseudo-$C_\ell$ case, as given in Challinor and Chon (2005), and then I applied the method from this paper to the pure-$C_\ell$ method. In doing so, I was able to prove that, at high-$\ell$, the pure-$C_\ell$ method is equally as good as the ordinary pseudo-$C_\ell$ method without E-B mixing. However, at low-$\ell$, there are other terms which come into play, resulting in a higher variance than would be obtained from the ordinary pseudo-$C_\ell$. These additional terms are expected to come into play for $\ell < \ell_{\text{mask}}$, where $\ell_{\text{mask}}$ is the maximum value of $\ell$ for which there is significant power in the mask.
3: SURVEY STRATEGIES AND FUTURE INSTRUMENTS
Point source removal

Unlike the CMB, point sources have approximately equal E and B mode polarisation. As a result, the removal of point-source contamination could potentially be a great issue for future B-mode experiments. Battye et al. (2011) discusses the necessity of point-source removal for the purposes of detecting B-mode polarisation. Several important results about the distribution of point sources and their effects on B-mode polarisation are shown, and the level to which point source subtraction must be performed is calculated. However, that paper does not consider potential difficulties caused by the removal itself. For standard pseudo-$C_\ell$ measurements of E-mode polarisation, there is unlikely to be any significant issue, since the area covered by a point-source hole is usually negligible although, of course, standard pseudo-$C_\ell$ is of no use for detecting B-mode polarisation, as discussed in Section 2. As seen in previous chapters, in the case of the pure-$C_\ell$ method this is no longer true. Due to the nature of the counter-terms, the mask must be apodised, and this results in a much larger proportion of the map being affected by the point-source holes. Further, the error at low-$\ell$ depends on the length of the boundary as well as on the area covered by the mask. Since point-sources have a large boundary length relative to the area covered, the pure-$C_\ell$ method could potentially be disproportionately affected by them. Therefore, it is important to characterise the effect that such holes will have on the pure-$C_\ell$ method. In this chapter I will show, using simulations and theoretical calculations, what the expected effect of
removing point sources from CMB experiments would be. I have also considered the question of what the optimal level of point-source contamination to mask out is.

### 4.1 Effect of point sources on B-mode polarisation measurements

In order to study this issue further, I carried out simulations using various point-source masks. I created masks containing an average of one source per 2.5, 10, 25, 100 and 1000 square degrees, with each point-source removed by masking and apodised using a cosine apodisation function with an apodisation length of 1.5 deg. I also combined these point-source masks with the spherical cap masks covering 1000 deg$^2$ and 10000 deg$^2$, as used in Section 2.4. In this section, I will explain the method behind the point-source map generation and then outline the results obtained from these simulations.

#### 4.1.1 Generation of point-source maps

In order to carry out these simulations, a method for generating point-source maps was required. There are two ways in which a map with $N_s$ point-sources can be generated. Firstly, it can be generated by picking $N_s$ random pixels and defining them to be ‘sources’. This will result in there always being exactly $N_s$ sources in every such map created, however it is not entirely consistent with the formation of sources in real life. The alternative method, which I have used, is to set the probability, $p$ of a pixel being a source to be $p = N_s/N_{\text{pix}}$. Then, looping over every pixel in turn, a random number, $r$ is generated, and if $r \leq p$ then the pixel in question is defined to be a source. This method is more time-consuming than the other method (except for very large source counts), and also does not necessarily produce maps with exactly $N_s$ sources ($N_s$ instead being the mean number of sources). However, it is more closely analogous to how sources are created in the real universe and, thus, this is the method that I have chosen.
Once the source maps have been generated, they must be apodised, in order for the pure-$C_\ell$ method to be viable. Based on numerous simulations, I have concluded that the best apodisation length to use is 1.5°, and thus this is the length chosen for use in the rest of this chapter. In order to ensure that the mask is both continuous and differentiable, as required by the pure-$C_\ell$ method, I then applied the smoothing algorithm outlined in Section \ref{sec:smoothing}.

### 4.1.2 Simple study of point-source effects

First, I will look at how the point-source masking affects the ability to detect B-modes. In order to do this, I have carried out simulations at $N_{\text{side}} = 512$ using a B-mode power spectrum with $r = 0.01$. In order to demonstrate the results, I have plotted the signal/noise ratio of the simulations against $\ell$ for the masks with one source per 10 deg$^2$ and per 100 deg$^2$ and compared them to the expected standard deviation as given by Equation (3.63) and to simulations for the case where there are no sources. The results are shown in Figure \ref{fig:bmode_pointsource_effect}.

Looking at these plots, it is clear that the major effect of point-source contamination will be at low-$\ell$. When compared to the simulations with no point-sources added, there is very little difference between the simulations with one source removed per 100 square degrees (which corresponds to around 400 sources) for $\ell$ values above about 30 or 40. For the mask with one source per 10 square degrees, there is a visible effect even at moderate to high $\ell$ values, but even in this case the error is only a factor of 1.5 or so larger. Thus, for the rest of this chapter, I will focus on the low-$\ell$ case (values of $\ell$ below around 100). However, first I will make a brief comment on the effect seen at high-$\ell$. Here the error in all cases (including even the full-sky case) is substantially higher than would be expected from the theoretical calculation, and this should ideally be explained.
Figure 4.1: Noise-to-signal ratio of simulations of the effect of masking out point-sources for a full-sky mask (black), 10000 square degree spherical cap (red) and 1000 square degree spherical cap (blue) compared with theoretical estimates (dashed line) and simulations without sources (dotted line) in the case of one source per 100 deg$^2$ (left) and one source per 10 deg$^2$ (right).

High-$\ell$ discrepancy

As shown in the last section, there is a substantial increase in the error for high-$\ell$ values (larger than around $3/4$ of $\ell_{\text{max}} = 2N_{\text{side}}$). This is not entirely unexpected, since the pure-$C_\ell$ method is only able to perfectly remove E-B mixing in the continuous case and, thus, as explained in [Smith (2006)], some E-B mixing is expected due to pixelisation effects. If this is indeed the cause of the errors, then they should scale with the pixel size. If, on the other hand, they are more fundamental, then they should be largely independent of the pixel size. Therefore, I carried out simulations at three different $N_{\text{side}}$ values (128, 256 and 512) and plotted these results in two ways. First, I have plotted the ratio of the simulated error to the theoretical error, as given by Equation (3.1) against $\ell$, and in the second I have plotted it against the ratio $\ell/\ell_{\text{max}}$. 
4.1: EFFECT OF POINT SOURCES ON B-MODE POLARISATION MEASUREMENTS

Figure 4.2: Point-source-only simulations at three different values of $N_{\text{side}}$ with 6000 sources masked out. Optimal errors are shown by solid black lines, $N_{\text{side}} = 128$ by dotted green lines, $N_{\text{side}} = 256$ by dashed red lines and $N_{\text{side}} = 512$ by solid blue lines.

where $\ell_{\text{max}} = 2N_{\text{side}}$. These results are shown in Figure 4.2.

From these results, it can be seen that, whilst the low-$\ell$ errors are independent of the pixel size, as would be expected, the errors at high $\ell$ seem to scale with $\ell/\ell_{\text{max}}$. Whilst the error seems to be a little higher in the case of $N_{\text{side}} = 128$, this is likely because the smoothing algorithm used does not work at that resolution, due to the smoothing beam width being too small. This similarity is strong evidence that the errors are indeed caused by pixelisation effects and, thus, can be safely ignored, since all that is required to eliminate them is the usage of a sufficiently small pixel size (setting $N_{\text{side}} > \ell_{\text{max}}$ would be a sensible precaution). Therefore, throughout the rest of this chapter I will focus only on the low-$\ell$ effects.
4: POINT SOURCE REMOVAL

4.2 Large-scale effects of point-source removal

As can be seen from the last section, the only significant effect on the pure-$C_\ell$ method of the removal of point-sources is at low-$\ell$. This may seem somewhat counter-intuitive, since point-sources usually affect high-$\ell$ measurements most strongly (due to their rapidly rising (in $\ell(\ell + 1)C_\ell$) power spectrum). However, in the case of the pure-$C_\ell$ method, the effect is caused by the unusually large ratio of boundary length to mask area in the case of point-sources, which results in the error in the pure-$C_\ell$ method (which, at low-$\ell$, depends heavily on the boundary length of the mask as well as on its total area) being substantially less than optimal. However, in order to understand this and determine if it is a potential issue in B-mode measurements, the effect must be considered more closely.

Therefore, I have plotted the signal-to-noise ratio for each point-source mask listed earlier in the chapter over a full sky, and for a 10000 square degree and 1000 square degree spherical cap, as shown in Figure 4.3. It is clear from these results that point-source masking has a substantial effect on the performance of the pure-$C_\ell$ method at low-$\ell$ values, particularly in the case of an otherwise unmasked map. Even a relatively moderate cut (for example, one source per 100 square degrees, or about 400 sources in total) makes B-modes totally undetectable at very low $\ell$ values (below about $\ell = 15$) using this method, and more aggressive cuts have an even larger effect, although it is still mostly confined to values of $\ell$ below 50.

Furthermore, because this effect is intrinsic to the method rather than dependent on any noise that may be added or removed or on E-B mixing, it does not depend on the value of $r$ being measured. Thus, the pure-$C_\ell$ method is totally incapable of measuring the power spectrum at very low $\ell$ values with a mask containing a significant number of removed point-sources. The implications of this for B-mode experiments will be discussed further in Section 4.3. However, it is useful to look more closely at this effect, in order to better characterise it. In particular, it is useful to look at the relative increase in the error caused by the addition of the point-sources to the existing mask.
4.2: LARGE-SCALE EFFECTS OF POINT-SOURCE REMOVAL

Figure 4.3: Noise-to-signal ratio of simulations of the effect of masking out point-sources as a function of $\ell$ for masks containing no sources (black solid line) and one source per 1000 (magenta dot-dashed), 100 (cyan dashed), 25 (red dashed), 10 (blue dotted) and 2.5 (green solid) square degrees for a full sky survey (top left) and spherical cap masks of 1000 square degrees (top right) and 10000 square degrees (bottom).

MICHAEL PREECE

121
4: POINT SOURCE REMOVAL

Therefore, I have plotted the ratio of the error with sources to the error without them, one for a full sky map, one for a 10000 square degree cap and one for a 1000 square degree cap. These are shown in Figure 4.4. I have also plotted (in the same Figures)

\[
\frac{4000\Delta C_\ell (\text{with sources})}{N_s \Delta C_\ell (\text{without sources})} + 1,
\]

where the 4000 is an arbitrary factor chosen to scale the results such that they fit onto a logarithmic scale.

The results show that point-sources do indeed have a significant effect on the error, even in the case of a relatively small spherical cap. The most aggressive source masking considered here, masking out one source per 10 square degrees, has a substantial effect on the error, even out to relatively large \( \ell \) values. However, this is not overly surprising, since this level of masking results in a reasonable proportion of the mask being at least partially downweighted due to the large number of point-sources. However, what is rather more surprising is the effect of removing a much smaller number of sources, such as the 1500 sources that would be removed by masking one source per 25 square degrees, which reduces the measured \( f_{\text{sky}} \) (including apodisation) by only around 12% and, thus, would be expected to have relatively little effect. The error in the lowest bin is increased by a factor of ten even in the case of the smallest mask considered (the 1000 square degree cap), and the error is still a factor of two larger even out to \( \ell = 30 \). For the full-sky mask, the effect at low-\( \ell \) is even worse, with the error being a hundred times worse in the lowest \( \ell \) bin. For the full sky case, even removing only 40 sources causes an increase in the error of a factor of four for the lowest bin. This shows quite clearly that the effects of point-source masking need to be carefully considered when attempting to detect B-mode polarisation at low \( \ell \) values, especially for full-sky surveys.

The results on the right of Figure 4.4 are interesting because they clearly show that there is a close relationship between the number of point-sources and the error. These results quite clearly show that the increase in error caused by the addition of point-sources is proportional to the number of sources masked and, thus, it can be seen
4.2: LARGE-SCALE EFFECTS OF POINT-SOURCE REMOVAL

Figure 4.4: Comparison of simulations involving a point-source mask applied to a full sky survey (top), 10000 deg$^2$ spherical cap (middle) and 1000 deg$^2$ cap (bottom) with those from simulations without point-sources for one source per 1000 (magenta long-dashed), 100 (black solid), 25 (red dashed) and 10 (blue dotted) square degrees. On the left the ratio of the standard deviations of the power spectrum with and without sources is plotted, and on the right the results are scaled by an $N_s$ factor as explained in the text. Note the close agreement between the results in the right-hand plots.
that

\[
\frac{\sigma(C_{\ell}^{BB}, N_s)}{\sigma(C_{\ell}^{BB}, N_s = 0)} = (1 + k(l)N_s),
\]  

(4.2)

where the constant \(k(l)\) depends on the particular mask being used and, unfortunately, cannot be determined analytically. However, it appears that there are two regimes. At low-\(\ell\), the absolute error caused by a given number of sources seems to be approximately the same, regardless of the mask. The value of \(\Delta C_{\ell}/C_{\ell}\) for the \(\ell = 10\) bin is about 100 for the one source per 10 deg\(^2\) case for all three masks. On the other hand, at higher \(\ell\), the value of \(k\) seems to be the same for all three masks. In this case, the extra error caused by the point-sources scales with the effect of the mask. This formula is potentially very useful, since it means that, if there exist simulated errors for a given point-source mask, I can then estimate the error that would be caused by masking any other number of point-sources to a high level of accuracy. Thus, if I wish to determine the effect of varying the number of point-sources masked on our ability to observe B-modes, it is only necessary to do simulations with one particular source count, and the others can then be calculated using the simulations to calibrate the result. The three graphs in Figure 4.3 show that this formula holds for a wide range of source numbers, although it fails in the case of a very high number of sources due to the overlap between the different source functions. In particular, if the analysis is extended to the one source per 2.5 square degrees case, the formula begins to break down. However, in this case the fraction of sky covered by the apodised source functions is 75\%, and thus no reasonable survey would use such an aggressive source masking approach, even if it were possible to detect the 15000 or so sources that such a mask would be excluding. In any case, were it to be necessary to remove this large number of sources, a smaller apodisation length would need to be used in order to allow a larger proportion of the map to be observed, which would negate the effect of overlapping. Thus, for all reasonable choices for the number of sources to be masked, the formula given in Equation (4.2) will hold to a high level of accuracy. The consequences of these results will be discussed further in Section 4.3.
4.2: LARGE-SCALE EFFECTS OF POINT-SOURCE REMOVAL

Now, consider the effect of noise. Then, using Equation (3.6), it can be seen that the error can now be written as

$$\sigma(C_{\ell}^{BB}) \propto (1 + k'(\ell)N_s) \left(C_{\ell}^{BB} + N_{\ell}^{BB}\right). \quad (4.3)$$

Now, in general this would not be overly useful, since the addition of noise should not overly change the effects of the point-sources on the mask (other than, perhaps, a minor effect on the shape of the function $k$). However, as stated earlier, the effects of non-masked point-sources on the CMB measurements will be similar to those of white noise (although the sources do not necessarily follow a Gaussian distribution, and thus to apply this method more accurately the true variance of the source power spectrum measurement would have to be taken into account), and thus they can be considered in the same way. Therefore, if an estimate for the power spectrum of the point-sources as a function of $N_s$ can be found, then this formula can be used to compute the optimal number of sources to mask out. White and Majumdar (2004) gives the point-source power spectrum to be

$$C_{\ell} = c \int_0^{S_{\max}} S_\nu^2 dN_s dS_\nu, \quad (4.4)$$

where $S_\nu$ is the point-source flux density at the given frequency, $\frac{dN_s}{dS_\nu}$ is the differential source count as a function of flux density, $S_{\max}$ is the maximum flux density of sources that are not masked and $c$ is a frequency dependant constant given by

$$c = \left(\frac{dB}{dT}\bigg|_{T=T_{CMB}}\right)^{-2}, \quad (4.5)$$

where $B$ is the Planck function.

Therefore, I can write

$$\frac{d\sigma}{dS}\bigg|_{S=S_{\max}} = \sigma(N_s = 0) \left(c S_{\max}^2 (1 + k(l)N_s(S_{\max})) \frac{dN_s}{dS}\bigg|_{S=S_{\max}} \right) \quad (4.6)$$

$$+ k(l) \frac{dN_s}{dS}\bigg|_{S=S_{\max}} \left(C_{\ell}^{BB} + c \int_0^{S_{\max}} S_\nu^2 \frac{dN_s}{dS_\nu} dS_\nu\right),$$

which, given an expression for $N_s$ as a function of the source intensity (either in the form of an analytical formula or a numerical one), can be used to compute the minimum error as a function of the number of sources for a given value of $\ell$. Note that,
because of how the pure-$C_\ell$ method works, it is entirely possible to use different masks in order to analyse different $\ell$ values, and this is likely to be the optimal method of using the pure-$C_\ell$ method in practice.

### 4.3 Effect of point-sources on B-mode detection

In Section 4.2, the effect of masking sources on the pure-$C_\ell$ method was considered, and an analysis of the effect of the varying number of sources on the additional error produced in the pure-$C_\ell$ method was given. However, this by itself is not enough to understand the impact of point-source removal on B-mode detection, since there is another alternative. Rather than directly subtracting the sources, they can instead be removed by simply measuring the source power spectrum and treating it as an additional source of noise. When looking at detecting E-modes using the ordinary pseudo-$C_\ell$ method, it would be expected that masking out the sources would always be the optimal choice, since removing a few pixels from the map is unlikely to affect the method in any significant manner. However, in the case of the pure-$C_\ell$ method, this is no longer true, since the requirement to differentiate the mask to compute the counter-terms means that the mask must be apodised. This, coupled with the fact that the error at low-$\ell$ is dependent on the boundary length as well as the area, means that the effect of masking out sources must be considered rather more carefully for the pure-$C_\ell$ method than for the ordinary pseudo-$C_\ell$ method. In the previous section, a formula (Equation (4.6)) was given which allows the optimal level of source masking for a given experiment to be computed. However, in order to use this, I require an expression for the number of sources as a function of the cut level.

Battye et al. (2011) considers the properties of point-sources and the level to which they would need to be subtracted to detect B-mode polarisation. There are two results from this paper which are of great importance to this section of my thesis. The first is how the number of sources of a given intensity varies as a function of frequency, which is necessary to determine how many sources must be masked out for a given cut
level. Importantly, it is shown that the source count as a function of intensity is largely independent of frequency (with the difference between the 30 GHz case and the 220 GHz case being only a factor of two), which means that this variation can essentially be ignored when determining the effect of source masking on the pure-$C_\ell$ method. It is found that, in order to cut out sources above 1 Jy, $\sim 1$ source per 200 deg$^2$ would need to be removed, in order to cut out all sources above 100 mJy, $\sim 1$ source per 8 deg$^2$ will need to be removed and in order to cut out all sources above 10 mJy, $\sim 2$ sources per deg$^2$ would need to be removed.

The main result of Battye et al. (2011), however, is the analysis of the level of point-source subtraction needed to detect B-modes as a function of $r$ and of the observation frequency. It is found that, in order to detect the peak in the B-mode spectrum (at $\ell = 80$), source subtraction is required below 100 GHz for $r = 0.1$ and for all frequencies for $r = 0.01$. In general, the higher frequency observations are much less heavily affected by point-source contamination. At 150 and 220 GHz, even $r = 0.001$ could likely be detected with little or no source subtraction, whereas at 30 GHz even subtracting sources of 10 mJy would likely not be enough to detect anything other than the very lowest $\ell$ modes. However, Battye et al. (2011) does not consider the effects of issues with the point-source masking itself on the ability to detect B-modes. There are several potential problems with source subtraction. In particular, I must consider the size of the holes that are produced in the map by cutting out sources, and the effect that these holes will have on the pure-$C_\ell$ method. Since an ideal point-source can be described by a $\delta$–function, the size of the hole is, in general, determined by the beam-width of the experiment in question. This immediately suggests that low-frequency experiments with large beam-widths will not be particularly viable for the detection of B-modes when point-sources are taken into account. In particular, the QUIJOTE experiment described in Section 3.1.2 would be heavily affected by this, if it were to attempt to detect B-modes at a level much below $r = 0.1$.

However, even if the beam size is substantially smaller, so that the size of the holes themselves is not important, the pure-$C_\ell$ method will still require apodisation to be
4: POINT SOURCE REMOVAL

applied. In this case, the results from Section 4.2 are important. This effect, when combined with the limits caused by the confusion noise from unsubtracted sources, means that there is, in theory, a minimal level to which B-modes could be detected at a given frequency. As explained earlier, Battye et al. (2011) shows that point-sources are of little importance at high frequency, except if you are attempting to measure very low values of $r$. At 220 GHz, $r = 0.01$ could be detected at the peak in the B-mode spectrum ($\ell = 80$) with virtually no source subtraction, and subtracting only 1 Jy sources would result in the source spectrum being lower than the primordial spectrum up to $\ell = 100$. Looking at Figure 4.1 it can be seen that for a 1 Jy cut (which results in around one source per 200 deg$^2$), there is no noticeable effect at above $\ell \approx 50$, even for the full-sky case (and even at $\ell = 30$ the effect is minimal). Therefore, by using a hybrid approach, applying a mask with no subtraction (apart, perhaps, from the brightest few sources) for $\ell < 50$ and a 1 Jy mask for $\ell > 50$, point-source contamination would have little effect on the detection of $r = 0.01$. If I consider $r = 0.001$, however, the situation is a little different. Here, a 1 Jy cut would be needed to detect all but the very largest scales (even $\ell = 10$ would not be detectable without it) and, thus, using the pure-$C_\ell$ method the power spectrum would likely not be measurable at all for values of $\ell$ below about 10 (although it may be possible to detect these by not applying a cut), and the combination of the issues caused by masking and the confusion noise from the remaining sources will likely mean that no detection will be possible. However, a 100 mJy cut will allow for $r = 0.001$ to be detected to $\ell = 100$ and, whilst this level of cut will require the masking of around one source per 8 deg$^2$, which will cause the pure-$C_\ell$ method to be significantly affected even up to $\ell = 50$, the larger $\ell$ values could still be detected. Likely, however, some intermediate source cut level would allow for the effect of masking at low-$\ell$ to be mitigated whilst still eliminating enough sources to allow detection. The situation at 150 GHz is similar (although slightly worse) and, thus, point-sources will pose little problem for the detection of B-modes at high frequencies, except for very low values of $r$, and even then they will have a limited effect provided the application of source subtraction is done with care.
However, when lower frequencies are considered, this is no longer true. In particular, an experiment at 30 GHz (the frequency at which QUIJOTE was supposed to operate) will have serious difficulties in dealing with point-sources, especially with a large beam width. To detect $r = 0.1$ on even the largest scales will require the application of source masking. A 1 Jy cut will allow $r = 0.1$ to be detected (in theory) up to $\ell = 10$, but the results of this chapter clearly demonstrate that even masking to this level will render the pure-$C_\ell$ method useless at this large a scale, even ignoring issues with the survey size and other potential problems with observing these results. A more aggressive source-cutting approach will allow $r = 0.1$ to be observed, with a 100 mJy cut allowing for the peak at $\ell = 80$ to be just about detected, and this level of cutting will only cause issues with the pure-$C_\ell$ method out to $\ell = 40$, but attempting to observe significantly lower values of $r$ will require sources down to 10 mJy or even lower to be removed, which requires the removal of two sources per deg$^2$. Even if the immense technical issues with this level of masking are ignored, analysing such an experiment would likely not be feasible using the pure-$C_\ell$ method, due to the requirement for apodisation to be applied to the masks. However, it may be possible to carry out such a level of masking by using a higher resolution map and reducing the apodisation length significantly, although this would likely result in a significant degradation of the performance of the method for low $\ell$ values, which would probably render the B-mode spectrum undetectable. Of particular interest is the situation at 70 GHz, which is around the minimum in the CMB foregrounds. Here, again, detecting $r = 0.1$ is unlikely to be a problem, with little or no source masking required. However, to detect $r = 0.01$, cutting sources to 100 mJy may be required, and for the case of $r = 0.001$, a cut of 10 mJy may be required which is, as previously stated, not really feasible.

Overall, since Battye et al. (2011) focusses mainly on the effect of point-sources at $\ell = 80$ (the maximum of the B-mode spectrum), the conclusions reached in that paper are not really affected by the issues with the pure-$C_\ell$ method, since these mainly apply at low-$\ell$. However, it is clear that the requirement to apodise and the issue with large numbers of sources will most likely make the application of a 10 mJy source cut...
infeasible if the pure-$C_\ell$ method is to be used for analysing the results. Note, however, that the validity of this analysis relies on the validity of the point-source models used in Battye et al. (2011). At low frequency, where the well-characterised radio sources dominate, there is no reason to doubt the validity of the model, and thus the analysis is likely to be valid. However, at high-frequency, the intensity source spectrum, at least, is dominated by infra-red sources rather than radio ones, and the polarisation of these sources is not well known and, thus, they cannot easily be included in the analysis. However, it is stated in Battye et al. (2011) that, whilst “there is very little information available about the polarization of such objects”, “one would expect it to be relatively low”. Indeed, a calculation of the expected infra-red polarised source intensity indicates that the Poisson (randomly distributed) component of the source intensity, at least, should be sub-dominant to the radio intensity even at the higher frequencies, implying that the analysis made here is valid after all. There is, however, a caveat to this. There is another component to the infra-red intensity, caused by source clustering. However, whilst this effect is known to exist and to contribute a significant amount to the noise, the exact level of the effect is, as yet, unknown, and thus cannot be accounted for in this analysis. As a result, the performance of the pure-$C_\ell$ method at high frequencies may, in reality, be worse than suggested in this section. In order to determine if this is the case, however, further observations (in particular, from Planck) will be required to constrain its amplitude.

4.4 Analytic Calculation of point-source errors

In theory, the covariance matrix of the pure-$C_\ell$ method can be calculated exactly, by using its definition in terms of the mask $a_{\ell m}$s, which is

$$\text{Cov}(\tilde{C}_{\ell m}, \tilde{C}_{\ell' m'}) = \frac{2}{(2\ell + 1)(2\ell' + 1)} \times \sum_{\ell_1,\ell_2, m_1, m_2, m, m'} M^{B}_{\ell_1, \ell_2, m_1, m_2} \left(M^{B}_{\ell_1, \ell_1, m_1, m_1}\right)^* \left(M^{B}_{\ell_2, \ell_2, m_2, m_2}\right)^* M^{B}_{\ell_1, \ell_2, m_1, m_2} C^{BB}_{\ell_1} C^{BB}_{\ell_2}.$$
4.4: ANALYTIC CALCULATION OF POINT-SOURCE ERRORS

With the mask matrix given by

\[
M^{\ell\ell'' \ell' \ell''}_{\ell \ell' \ell'' \ell''} = \sqrt{\frac{(\ell' - 1)\ell' (\ell' + 1)(\ell' + 2)}{\ell + 1}(\ell + 1)(\ell + 2)} \times \sum_{\ell''} w_{\ell''} \sqrt{(2\ell + 1)(2\ell' + 1)(2\ell'' + 1)} \frac{4\pi}{\pi} \begin{pmatrix} \ell & \ell' & \ell'' \\ \ell & \ell' & \ell'' \end{pmatrix} - m \begin{pmatrix} m' \ m'' \ m'' \end{pmatrix}.
\] (4.8)

In practice, however, this formula is usually of little use, since its dependence on the mask \(w_{\ell \ell} \) means that the calculation speed scales extremely poorly with increasing \(\ell\) and, thus, this formula cannot, in practice, be used to compute the error in most scenarios. However, if the \(m\)-dependence of the mask \(w_{\ell \ell} \) could be removed, the calculation would become much more tractable. For any individual point-source mask, the mask \(w_{\ell \ell} \) are dependent on \(m\), and in a manner which cannot be predicted. However, if an average is taken over the ensemble of all possible point-source distributions, then it becomes possible to remove this dependence and, thus, produce a working formula.

First, consider how a point-source mask with \(N_s\) point-sources should be defined. One method to create such a mask is by picking \(N_s\) random pixels from a map and setting them to be point-sources (assuming, for convenience, that all sources are located at the centre of the relevant pixel, and are true point-sources, although this assumption is of limited importance when apodisation is applied).

Now, suppose I define a mask with \(N_s\) point-sources, by placing these sources in \(N_s\) different pixels with position vectors \(x_i\), and then apply apodisation, represented by a beam-smoothing function \(B\), which can be assumed to depend solely on the distance between the source in question and the point being measured. Then, the mask is given by

\[
M(\mathbf{x}) = 1 - \sum_{i=1}^{N_s} B(\mathbf{x}, \mathbf{x}_i).
\] (4.9)

Note here that I am ignoring the possibility of two sources being placed on the same pixel. However, provided that \(N_s \ll \sqrt{N_{\text{pix}}}\), this is an unlikely occurrence, and thus can safely be ignored. Using the assumption that the beam function depends solely on the
distance between \( x \) and \( x_i \) (which should be approximately true for most apodisation schemes), \( B \) can be decomposed in terms of Legendre polynomials to get

\[
B(x, x_i) = \sum_{\ell} \frac{2\ell + 1}{4\pi} b_{\ell} P_{\ell}(x, x_i),
\]

which allows it to be shown that the mask multipoles are given by

\[
w_{\ell m} = \sqrt{4\pi} \delta_{\ell 0} \delta_{m 0} - \sum_{i=1}^{N_s} b_{\ell} Y_{\ell m}^*(x_i).
\]

This formula is useful since the \( m \)-dependence comes only from a term involving the value of the spherical harmonics at the location of the point-sources, which is random. Thus, by averaging over all possible point-source distributions, the dependence on \( m \) can be averaged out and, thus, I can produce a formula that should be at least reasonably quick to compute.

Therefore, let us attempt to calculate the average of the covariance matrix over the ensemble of possible point-source distributions. This is given by

\[
\langle \text{Cov}(\tilde{C}_{\ell \ell'}^{\chi \chi}, \tilde{C}_{\ell \ell'}^{\chi \chi}) \rangle = \frac{2}{(2\ell + 1)(2\ell' + 1)} \times \sum_{\ell_1, \ell_2, m_1, m_2, m_1', m_2'} \langle M_{\ell_1 m_1 m_1'}^B \left( M_{\ell_2 m_2 m_2'}^B \right)^* \left( M_{\ell_2 m_2 m_2'}^B \right)^* \rangle C_{\ell_1 \ell_1} \tilde{C}_{\ell_2 \ell_2}.
\]

Using the definition of the mask matrix given in Equation (4.8), it can be seen that the only mask-dependent quantities are the mask \( w_{\ell m} \)s and, thus, to compute the average covariance matrix, I need to compute \( \langle w_{\ell_1 m_1} w_{\ell_1 m_1}' w_{\ell_2 m_2} w_{\ell_2 m_2}' \rangle \). Now, using Equation (4.11), I can expand this in terms of the averages of spherical harmonics over
4.4: ANALYTIC CALCULATION OF POINT-SOURCE ERRORS

all possible point-source distributions, which gives

\[
\left\langle w_{\ell_1 m_1} w_{\ell_2 m_2} w_{\ell_3 m_3} w_{\ell_4 m_4} \rightangle = 16\pi^2 \left( \delta_{\ell_1 0} \delta_{\ell_2 0} \delta_{\ell_3 0} \delta_{\ell_4 0} \delta_{m_1 0} \delta_{m_2 0} \delta_{m_3 0} \delta_{m_4 0} \right) + (4\pi)^{\frac{3}{2}} \left\{ b_{\ell_1} \delta_{\ell_1 0} \delta_{\ell_2 0} \delta_{\ell_3 0} \delta_{m_1 0} \delta_{m_2 0} \delta_{m_3 0} \sum_{i=1}^{N_i} \left\langle Y_{\ell_1 m_1}^* (x_i) \right\rangle \right\}
+ b_{\ell_2} \delta_{\ell_2 0} \delta_{\ell_3 0} \delta_{\ell_4 0} \delta_{m_1 0} \delta_{m_2 0} \delta_{m_3 0} \sum_{i=1}^{N_i} \left\langle Y_{\ell_2 m_2} (x_i) \right\rangle + \ldots \right)
+ 4\pi \left\{ b_{\ell_1} b_{\ell_2} \delta_{\ell_1 0} \delta_{\ell_2 0} \delta_{\ell_4 0} \delta_{m_1 0} \delta_{m_2 0} \delta_{m_3 0} \sum_{i,j=1}^{N_i} \left\langle Y_{\ell_1 m_1}^* (x_i) Y_{\ell_2 m_2} (x_j) \right\rangle \right\}
+ \ldots \right)
- \sqrt{4\pi} \left\{ b_{\ell_1} b_{\ell_2} b_{\ell_3} \delta_{\ell_1 0} \delta_{\ell_2 0} \delta_{\ell_3 0} \delta_{m_1 0} \sum_{i,j,k=1}^{N_i} \left\langle Y_{\ell_1 m_1}^* (x_i) Y_{\ell_2 m_2}^* (x_j) Y_{\ell_3 m_3} (x_k) \right\rangle \right\}
+ \ldots \right) \right.
+ \left. \sum_{i,j,k,l=1}^{N_i} \left\langle Y_{\ell_1 m_1}^* (x_i) Y_{\ell_2 m_2} (x_j) Y_{\ell_3 m_3}^* (x_k) Y_{\ell_4 m_4} (x_l) \right\rangle \right). \tag{4.13}

Therefore, I must calculate terms of the form

\[
\left\langle \sum_{l_1=1}^{N_i} Y_{l_1 m_1} (\Omega_1) \ldots \right\rangle. \tag{4.14}
\]

In order to expand this, I have to use the definition of the average of a function

\[
\left\langle f \right\rangle = \frac{\int f d\Omega_i}{\int d\Omega_i}, \tag{4.15}
\]

which gives the following expressions for the first and second order expansions

\[
\left\langle \sum_{l=1}^{N_i} Y_{l m} (\Omega_i) \right\rangle = \sum_{i=1}^{N_i} \frac{\int Y_{l m} (\Omega_i) d\Omega_i}{\int d\Omega_i}
= \frac{1}{4\pi} \sum_{i=1}^{N_i} \int Y_{l m} (\Omega_i) d\Omega_i
= \frac{N_i}{\sqrt{4\pi}} \delta_{l 0} \delta_{m 0}, \tag{4.16}
\]

MICHAEL PREECE

133
4: POINT SOURCE REMOVAL

\[ \left\langle \sum_{l,j=1}^{N_s} Y_{\ell_1 m_1}(\Omega_j) Y^{*}_{\ell_2 m_2}(\Omega_j) \right\rangle = \frac{1}{16\pi^2} \sum_{j=1}^{N_s} \sum_{k \neq j} \int Y_{\ell_1 m_1}(\Omega_j) Y^{*}_{\ell_2 m_2}(\Omega_j) d\Omega d\Omega_j + \frac{1}{4\pi} \sum_{j=1}^{N_s} \int Y_{\ell_1 m_1}(\Omega_j) Y^{*}_{\ell_2 m_2}(\Omega_j) d\Omega_j \]

\[ = \frac{N_s(N_s - 1)}{4\pi} \delta_{\ell_10} \delta_{m_10} \delta_{\ell_20} \delta_{m_20} + \frac{N_s}{4\pi} \delta_{\ell_1 \ell_2} \delta_{m_1 m_2}, \quad (4.17) \]

with similar expressions holding for higher-order functions.

Using these gives that

\[ \left\langle w_{\ell_1 m_1}w_{\ell_2 m_2}w^{*}_{\ell_3 m_3}w^{*}_{\ell_4 m_4} \right\rangle = \left( 16\pi^2 - 16\pi N_s b_0 + 6N_s(N_s - 1)b_0^2 \right) \]

\[ - \frac{N_s(N_s - 1)(N_s - 2)b_0^3}{\pi} + \frac{N_s(N_s - 1)(N_s - 2)(N_s - 3)b_0^4}{16\pi^2} \]

\[ \times \delta_{\ell_10} \delta_{\ell_20} \delta_{\ell_30} \delta_{\ell_40} \delta_{m_10} \delta_{m_20} \delta_{m_30} \delta_{m_40} + \left( N_s - \frac{N_s(N_s - 1)b_0}{2\pi} + \frac{N_s(N_s - 1)(N_s - 2)b_0^2}{16\pi^2} \right) \]

\[ \times (b_{\ell_1} b_{\ell_2} \delta_{\ell_1 \ell_2} \delta_{m_1 m_2} \delta_{\ell_30} \delta_{\ell_40} \delta_{m_30} \delta_{m_40}) + (-1)^{m_3} b_{\ell_1} b_{\ell_2} \delta_{\ell_1 \ell_2} \delta_{m_1 m_2} \delta_{\ell_30} \delta_{\ell_40} \delta_{m_30} \delta_{m_40} \]

\[ + b_{\ell_1} b_{\ell_2} \delta_{\ell_1 \ell_2} \delta_{m_1 m_2} \delta_{\ell_30} \delta_{\ell_40} \delta_{m_30} \delta_{m_40} + (-1)^{m_4} b_{\ell_1} b_{\ell_2} \delta_{\ell_1 \ell_2} \delta_{m_1 m_2} \delta_{\ell_30} \delta_{\ell_40} \delta_{m_30} \delta_{m_40} + \delta_{\ell_1 \ell_2} \delta_{\ell_3 \ell_4} \delta_{m_1 m_2} \delta_{m_30} \delta_{m_40} + O(b_0^3). \]

Now, Equation (4.7) gives that

\[ \left\langle M_{\ell_1 m_1} M_{\ell_2 m_2} \left( M_{\ell_3 m_3} M_{\ell_4 m_4} \right)^* \right\rangle = \frac{\sqrt{(\ell'_1 - 1)(\ell'_1 + 1)(\ell'_2 - 1)(\ell'_2 + 1)}}{\ell_11(\ell_1 + 1)(\ell_2 + 1)} \]

\[ \times \frac{\sqrt{(\ell'_3 - 1)(\ell'_3 + 1)(\ell'_4 - 1)(\ell'_4 + 1)}}{\ell_31(\ell_3 + 1)(\ell_4 + 1)} \]

\[ \times \sum_{\ell'_1', \ell'_2', \ell'_3', \ell'_4', m_1', m_2', m_3', m_4'} \left\langle w_{\ell_1' m_1'} w_{\ell_2' m_2'} w_{\ell_3' m_3'} w_{\ell_4' m_4'} \right\rangle \]

\[ B_{\ell_1 \ell_2 \ell_3 \ell_4}^{m_1 m_2 m_3 m_4} \]
where the $B$ coefficients are as defined in Equation (2.45), and, so, I must compute

$$
\sum_{m_1,m_2,m, m' \ell_1, \ell_2, \ell_1', \ell_2'} \sum_{m''} \sum_{m'''} \left\langle W_{\ell_1',m'''} W_{\ell_2',m''} W_{\ell_1,m}, W_{\ell_2,m'} \right\rangle B_{mm',mm''}^{m_1,m_2} B_{m''m,m'}^{m_2,m_1} B_{m,m}^{m_1,m_2} B_{m,m'}^{m_2,m_1}
$$

$$
= (1 - f_{int})^4 (2 \ell + 1) \delta_{\ell_1 \ell_2} \delta_{\ell_1' \ell_2'} \delta_{\ell_1' \ell_2} \delta_{\ell_1 \ell_2} (4.20)
$$

and

$$
\int \sum_{m_1,m_2,m, m' \ell_1, \ell_2, \ell_1', \ell_2'} B_{m,m}^{m_1,m_2} B_{m,m'}^{m_2,m_1} \delta_{\ell_1 \ell_2} \delta_{\ell_1' \ell_2'} \delta_{\ell_1' \ell_2} \delta_{\ell_1 \ell_2} + O(b_3^3)
$$

$$
= (1 - f_{int})^4 (2 \ell + 1) \delta_{\ell_1 \ell_2} \delta_{\ell_1' \ell_2'} \delta_{\ell_1' \ell_2} \delta_{\ell_1 \ell_2} (4.21)
$$

where $f_{int}$ is the integral of the weight function of the mask. I have dropped the spin-weight indices from the $B$ coefficients for simplicity, since they will be zero throughout this section and, further, I have removed all terms of order greater than $b_3^3$, again for simplicity, and because they appear to be negligible (although $b_3$ itself is not expected to be small).

Now, from the definition of the $B$ coefficients, and using the facts that

$$
\begin{pmatrix} \ell & \ell' & 0 \\ -m & m' & 0 \end{pmatrix}^2 = \frac{(-1)^{\ell+m} \delta_{\ell \ell'} \delta_{mm'}}{\sqrt{2 \ell + 1}} (4.22)
$$

and

$$
\sum_{m,m',m''} \begin{pmatrix} \ell & \ell' & \ell'' \\ -m & m' & m'' \end{pmatrix}^2 = \sum_{m=-\ell}^{\ell} \frac{1}{2 \ell + 1} = 1 (4.23)
$$

MICHAEL PREECE

135
and
\[ \delta m'' (-m''') \sum_{m'' = -\ell}^\ell \frac{1}{2\ell'' + 1} = \sum_{m'' = -\ell}^\ell \delta m'' (-m''') \sum_{m'' = -\ell}^\ell \frac{1}{2\ell'' + 1}, \] (4.24)

I have that
\[ \sum_{m, m', m''} (B_{\ell\ell'}^{m m''} B_{\ell'\ell''}^{m' m''})^2 = \frac{(2\ell + 1)(2\ell' + 1)(2\ell'' + 1)}{4\pi} \frac{1}{\ell''} \sum_{m, m', m''} \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{array} \right)^2 \] (4.25)

and that
\[ \sum_{m, m', m''} B_{\ell\ell'}^{m m''} B_{\ell'\ell''}^{m' m''} = \frac{(2\ell + 1)(2\ell' + 1)(2\ell'' + 1)}{4\pi} \frac{1}{\ell''} \sum_{m, m', m''} \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{array} \right)^2 \] (4.26)

Thus,
\[ \sum_{m_1, m_2, m_3, m_4} \sum_{\ell_1, \ell_2} \sum_{m_{1'}, m_{2'}, m_{1''}, m_{2''}} \left( w_{\ell_1, m_1, m_{1'}} w_{\ell_2, m_2, m_{2'}} w_{\ell_1, m_{1'}, m_{1''}} w_{\ell_2, m_{2'}, m_{2''}} \right) B_{\ell_1 \ell_2}^{m_{1'}, m_{2'}} B_{\ell_1 \ell_2}^{m_{1''}, m_{2''}} B_{\ell_1 \ell_2}^{m_{1'}, m_{2'}} B_{\ell_1 \ell_2}^{m_{1''}, m_{2''}} \] (4.27)

\[ = (1 - f_{\text{int}})^4 (2l + 1) \delta_{\ell_1 \ell_2} \delta_{\ell_1 \ell_2} \delta_{\ell_1 \ell_2} + \delta_{\ell_1 \ell_2} \delta_{\ell_1 \ell_2} (2l' + 1) \delta_{\ell_1 \ell_2} \delta_{\ell_1 \ell_2} + \delta_{\ell_1 \ell_2} \delta_{\ell_1 \ell_2} \delta_{\ell_1 \ell_2} \delta_{\ell_1 \ell_2} \] (4.27)
4.4: ANALYTIC CALCULATION OF POINT-SOURCE ERRORS

Now, I must compute the RHS of Equation (4.7), which is found to be

\[
\frac{2}{(2\ell + 1)(2\ell' + 1)} \sum_{\ell_1,\ell_2,m_1,m_2,m'} \left[ M_{\ell_1 \ell_2}^B \left( \left( M_{\ell_1 \ell_2}^B \right)^* \left( M_{\ell_1 \ell_2}^B \right)^* \right) M_{\ell_1 \ell_2}^B \right] C_{\ell_1} C_{\ell_2}^{BB} \\
= \frac{2}{(2\ell' + 1)} \frac{(1 - f_{\text{int}})^4}{(2\ell' + 1)} \frac{\delta_{\ell \ell'} C_{\ell''}^{BB}}{C_{\ell}^{BB}} \\
+ \frac{f_{\text{int}}(1 - f_{\text{int}})^2}{2\pi} \sum_{\ell''} \frac{b_{\ell''}^2}{b_0} \left[ (C_{\ell}^{BB} C_{\ell''}^{BB} + (2\ell'' + 1) \left( (C_{\ell}^{BB})^2 + (C_{\ell''}^{BB})^2 \right)) \right] \left( \begin{array}{ccc} \ell & \ell' & \ell'' \\ 0 & 0 & 0 \end{array} \right)^2 \\
+ \frac{1}{(2\ell + 1)(2\ell' + 1)} \sum_{\ell_1,\ell_2} \frac{2(\ell + 1)(\ell - 1)(\ell_1 + 1)(\ell_1 + 2)}{(2\ell + 1)(2\ell') \delta_{\ell \ell_1} \delta_{\ell' \ell_2}} \left( \begin{array}{ccc} \ell & \ell_1 & \ell'' \\ 0 & 0 & 0 \end{array} \right)^2 C_{\ell_1} C_{\ell''}^{BB} \right) 
\]

In order to calculate the covariance matrix of the deconvolved $C_{\ell}$s, I require

\[
\sum_{m_1,m_2,m,m'} \sum_{l'_1,l'_2,l} \sum_{l''_1,l''_2,l''} \left[ W_{l'_1,m_1} W_{l'_2,m_2} W_{l''_1,m_1} W_{l''_2,m_2} \right] B_{\ell'_1 \ell'_2 \ell_1}^{m_1,m_1} B_{\ell'_1 \ell'_2 \ell_2}^{m_2,m_2} B_{\ell'_1 \ell'_2 \ell_1}^{m_1,m_1} B_{\ell'_1 \ell'_2 \ell_2}^{m_2,m_2} \\
= (1 - f_{\text{int}})^4 (2\ell + 1)(2\ell' + 1) \delta_{\ell \ell_1} \delta_{\ell' \ell_2} \\
+ f_{\text{int}}(1 - f_{\text{int}})^2 \sum_{l''} \frac{b_{\ell''}^2}{b_0} \left( 4(-1)^{m_1+m_2} B_{\ell \ell_1 \ell_2}^{m_1,m_1} B_{\ell \ell_1 \ell_2}^{m_2,m_2} \delta_{\ell_1 \delta_{\ell_2}} \delta_{\ell_2 \delta_{m_1}} \delta_{m_2} \\
+ B_{\ell \ell_1 \ell_2}^{m_1,m_2} B_{\ell \ell_1 \ell_2}^{m_1,m_2} \delta_{\ell_1 \delta_{m_1}} + B_{\ell \ell_1 \ell_2}^{m_1,m_2} B_{\ell \ell_1 \ell_2}^{m_1,m_2} \delta_{\ell_2 \delta_{m_2}} \right) \\
= (1 - f_{\text{int}})^4 (2\ell + 1)(2\ell' + 1) \delta_{\ell \ell_1} \delta_{\ell' \ell_2} \\
+ 4\pi f_{\text{int}}(1 - f_{\text{int}})^2 \sum_{l''} \frac{b_{\ell''}^2}{b_0} \left( 4 \sum_{m,m'} (-1)^{m+m'} B_{\ell \ell_1 \ell_2}^{m_1,m_1} B_{\ell \ell_1 \ell_2}^{m_2,m_2} \delta_{\ell_1 \delta_{\ell_2}} \\
+ \sum_{m,m'} B_{\ell \ell_1 \ell_2}^{m_1,m_1} B_{\ell \ell_1 \ell_2}^{m_2,m_2} \delta_{\ell_1 \delta_{m_1}} + \sum_{m,m'} B_{\ell \ell_1 \ell_2}^{m_1,m_1} B_{\ell \ell_1 \ell_2}^{m_2,m_2} \delta_{\ell_2 \delta_{m_2}} \right). 
\]
4: POINT SOURCE REMOVAL

Therefore, I get that

\[
\frac{1}{(2\ell + 1)(2\ell' + 1)} \sum_{\ell_1,\ell_2,m_1,m_2,m'_2} |M_{\ell\ell_1m_1}^B|^2 |M_{\ell'\ell_2m'_2}^B|^2 \text{Cov}(\hat{C}_{\ell_1}^{BB}, \hat{C}_{\ell_2}^{BB})
\]

(4.31)

\[
= (1 - f_{\text{int}})^4 \text{Cov}(C_{\ell}^{BB}, C_{\ell}^{BB})
\]

\[
+ f_{\text{int}} (1 - f_{\text{int}})^2 \sum_{\ell''} \frac{b^{2}_{\ell''} b_0}{(2\ell + 1)(2\ell' + 1)} 4 \text{Cov}(\hat{C}_{\ell}^{BB}, \hat{C}_{\ell}^{BB}) \sum_{m} (-1)^m B_{\ell\ell' m m''}^B \sum_{m'} (-1)^{m'} B_{\ell\ell' m m'}^B
\]

\[
+ \sum_{\ell_1} \frac{(2\ell_1 + 1)(\ell_1 - 1)\ell_1(\ell_1 + 1)(\ell_1 + 2)}{4\pi} \text{Cov}(\hat{C}_{\ell}^{BB}, \hat{C}_{\ell}^{BB}) \left( \frac{g^{2}_{\ell'} g_0}{(\ell' - 1)\ell'((\ell' + 1)(\ell' + 2)} \left( \begin{array}{ccc} \ell' & \ell_1 & \ell'' \\ 0 & 0 & 0 \end{array} \right)^2 + \frac{\text{Cov}(\hat{C}_{\ell'}^{BB}, \hat{C}_{\ell}^{BB})}{(\ell - 1)\ell(\ell + 1)(\ell + 2)} \left( \begin{array}{ccc} \ell & \ell_1 & \ell'' \\ 0 & 0 & 0 \end{array} \right)^2 \right).
\]

Now,

\[
\sum_{m=-\ell}^{\ell} (-1)^m \left( \begin{array}{ccc} \ell & \ell & \ell'' \\ -m & m & 0 \end{array} \right) = \sum_{m=-\ell}^{\ell} (-1)^m \left( \begin{array}{ccc} \ell & \ell & 0 \\ -m & m & 0 \end{array} \right) \delta_{\ell''0}
\]

\[
= \sum_{m=-\ell}^{\ell} (-1)^{\ell} \delta_{\ell''0} \sqrt{2\ell + 1}
\]

\[
= (-1)^{\ell} \delta_{\ell''0} \sqrt{2\ell + 1},
\]

(4.32)

and, thus, I find that

\[
\sum_{m=-\ell}^{\ell} (-1)^m B_{\ell\ell' m m''}^B = \frac{(2\ell + 1)^3}{\sqrt{4\pi}} (-1)^{\ell} \delta_{\ell''0} \left( \begin{array}{ccc} \ell & \ell & 0 \\ 0 & 0 & 0 \end{array} \right)
\]

(4.33)

\[
= \frac{(2\ell + 1) \delta_{\ell''0}}{\sqrt{4\pi}}.
\]

(4.34)

138

ANALYSIS OF CMB POLARISATION
which gives that

\[
\frac{1}{(2\ell + 1)(2\ell' + 1)} \sum_{\ell_1, \ell_2, m_1, m_2, m_{1'}, m_{1''}} |M^{AB}_{\ell_1 m_1}|^2 |M^{AB}_{\ell' m_{1'} m_{1''}}|^2 \text{Cov}(\hat{C}_\ell^{BB}, \hat{C}_{\ell'}^{BB})
\]

\[
= (1 - f_{\text{int}})^4 \text{Cov}(\hat{C}_\ell^{BB}, \hat{C}_{\ell'}^{BB}) + f_{\text{int}} (1 - f_{\text{int}})^2 \sum_{\ell''} \frac{b^2_{\ell''}}{b_0} \delta_{\ell''0} \text{Cov}(\hat{C}_\ell^{BB}, \hat{C}_{\ell''}^{BB})
\]

\[
+ \sum_{\ell_1} \frac{(2\ell' + 1)(\ell_1 - 1)\ell_1(\ell_1 + 1)(\ell_1 + 2)}{4\pi} \text{Cov}(\hat{C}_\ell^{BB}, \hat{C}_{\ell_1}^{BB}) \left( \frac{\text{Cov}(\hat{C}_\ell^{BB}, \hat{C}_{\ell_1}^{BB})}{(\ell - 1)(\ell + 1)(\ell + 2)} \left( \begin{array}{ccc} \ell' & \ell_1 & \ell'' \\ 0 & 0 & 0 \end{array} \right) \right)^2
\]

\[
\left( \frac{\ell'}{\ell' + 1} \right)^2 \text{Cov}(\hat{C}_\ell^{BB}, \hat{C}_{\ell'}^{BB}) \right)
\]

\[
+ f_{\text{int}} (1 - f_{\text{int}})^2 \left( \frac{b_0}{\text{Cov}(\hat{C}_\ell^{BB}, \hat{C}_{\ell'}^{BB})} \right)
\]

\[
+ \sum_{\ell_1} \frac{(2\ell' + 1)(\ell_1 - 1)\ell_1(\ell_1 + 1)(\ell_1 + 2)}{4\pi} \left( \frac{(2\ell + 1)\text{Cov}(\hat{C}_\ell^{BB}, \hat{C}_{\ell_1}^{BB})}{(\ell - 1)(\ell + 1)(\ell + 2)} \sum_{\ell''} \frac{b^2_{\ell''}}{b_0} \left( \begin{array}{ccc} \ell' & \ell_1 & \ell'' \\ 0 & 0 & 0 \end{array} \right) \right)^2
\]

Putting these two together, I get

\[
\sum_{\ell_1, \ell_2} M^{AB}_{\ell_1 \ell_2} \text{Cov}(\hat{C}_{\ell_1}^{XX}, \hat{C}_{\ell_2}^{XX}) = \frac{2(1 - f_{\text{int}})^4 \delta_{\ell''0} C^{BB}_\ell C_{\ell'}^{BB}}{(2\ell + 1)} \left( \begin{array}{ccc} \ell' & \ell_1 & \ell'' \\ 0 & 0 & 0 \end{array} \right) \right)^2
\]

\[
+ \frac{f_{\text{int}} (1 - f_{\text{int}})^2}{2\pi} \sum_{\ell''} \frac{b^2_{\ell''}}{b_0} \left( C^{BB}_\ell C_{\ell'}^{BB} + (2\ell'' + 1) \left( \varepsilon^{BB}_\ell \right)^2 \right) \left( \begin{array}{ccc} \ell' & \ell_1 & \ell'' \\ 0 & 0 & 0 \end{array} \right) \right)^2
\]

\[
+ \sum_{\ell_1} \frac{2(2\ell + 1)(\ell_1 - 1)\ell_1(\ell_1 + 1)(\ell_1 + 2)}{(2\ell + 1)(\ell - 1)(\ell + 1)(\ell + 2)} \delta_{\ell''} \left( \begin{array}{ccc} \ell' & \ell_1 & \ell'' \\ 0 & 0 & 0 \end{array} \right) \right)^2
\]

\[
C^{BB}_\ell C_{\ell'}^{BB} + O(b^3_\ell),
\]
where the mask matrix $M_{\ell_1\ell_2}^{\ell_1\ell_2}$ can be approximated by

$$M_{\ell_1\ell_2}^{\ell_1\ell_2} = (1 - f_{\text{int}})^4 \delta_{\ell_1\ell_2} + f_{\text{int}} (1 - f_{\text{int}})^2 \left( \frac{4(-1)\ell^\prime \ell b_0}{\sqrt{(2\ell + 1)(2\ell^\prime + 1)}} \right)$$

$$+ \sum_{\ell_1} \frac{(2\ell + 1)(2\ell_1 + 1)(\ell_1 - 1)\ell_1(\ell_1 + 1)(\ell_1 + 2)\delta_{\ell_1\ell_2}}{4\pi(\ell - 1)\ell(\ell + 1)(\ell + 2)} \sum_{\ell^\prime} \frac{b^2_{\ell^\prime}}{b_0} \begin{pmatrix} \ell & \ell_1 & \ell^\prime \\ 0 & 0 & 0 \end{pmatrix}^2$$

$$+ \sum_{\ell_2} \frac{(2\ell^\prime + 1)(\ell^\prime - 1)\ell^\prime(\ell^\prime + 1)(\ell^\prime + 2)\delta_{\ell\ell_2}}{4\pi(\ell^\prime - 1)\ell^\prime(\ell^\prime + 1)(\ell^\prime + 2)} \sum_{\ell^\prime\prime} \frac{b^2_{\ell^\prime\prime}}{b_0} \begin{pmatrix} \ell & \ell_1 & \ell^\prime \prime \\ 0 & 0 & 0 \end{pmatrix}^2.$$  \(4.38\)

This formula should give the average error caused by a point-source mask, subject to several assumptions. Firstly, it must be assumed that the apodisation function is azimuthally symmetric. This will almost always be true in the initial implementation. However, if two sources come to overlap, then the interaction between their apodisation functions will cause both to cease to be symmetric. Thus, in the case where there are a large number of sources, the formula may not work. Secondly, the formula assumes that point-sources are randomly distributed across the sky. If this is not true, then the formula will not be entirely accurate, although it is likely that it could be modified to some extent to account for clustering.

### 4.4.1 Comparison of analytic formula to simulations

In order to test the effectiveness of this formula, I must compare it to simulated results. Due to the form of Equation (4.38), the mask matrix cannot easily be inverted exactly, since it is a 4-dimensional matrix rather than a usual two-dimensional one. Whilst this can be gotten around due to the structure of the matrix, the result is equivalent to being required to invert an $\ell_{\text{max}}^2 \times \ell_{\text{max}}^2$ matrix, which is simply too slow (and memory-intensive) to be done directly (given that $\ell_{\text{max}}$ is either 512 or 1024), and due to the properties of the matrix finding a suitable approximation to the inversion which works in all cases is non-trivial. Whilst it can in principle be done, I did not have time to correctly implement such a method. Therefore, in order to test Equation (4.37), I have compared instead the predictions from this formula to the convolved pure-$C_\ell$ variance.
4.4: ANALYTIC CALCULATION OF POINT-SOURCE ERRORS

\[ \text{Var}(\tilde{C}_{\ell}^{XY}) \text{ computed using simulations, thereby bypassing the issue with approximating the deconvolution process.} \]

Therefore, I have produced two different simulations, one using a point-source mask with one source per 100 deg\(^2\) and the other with one source per 10 deg\(^2\), in both cases using a Gaussian apodisation function with an apodisation length of 1.5°. Since there is a possibility of the different source masks overlapping, I computed the mask as

\[
M(\Omega) = \prod_{i=1}^{N_s} \left( 1 - \exp \left( \frac{|r - r_i|^2}{2\sigma^2} \right) \right),
\]

(4.39)

(where \(\sigma = r_c / \sqrt{8 \ln 2}\), which reduces to the required form

\[
M(\Omega) = 1 - \sum_{i=1}^{N_s} \exp \left( \frac{|r - r_i|^2}{2\sigma^2} \right),
\]

(4.40)

in the case where the source apodisation functions do not overlap, but unlike the latter formula cannot result in the mask becoming negative in certain scenarios. The results of this simulations are shown in Figure 4.5.

As can be seen, the theoretical predictions do not appear to fit the simulated results at all well. In both cases, the predicted error at low \(\ell\) is far too low, and in the case of the mask with the larger number of sources, the prediction at intermediate to high \(\ell\) values is actually too high (most likely, this is the case for both masks, but it is simply not visible in the case of the one source per 100 deg\(^2\) cases due to the relatively low number of sources). A potential explanation for the effect at low \(\ell\) is that the higher-order terms that were excluded from the calculation for simplicity have an effect on large scales. However, this hypothesis has been tested by including them in the computation, and these terms were found to be negligible, as expected. Therefore, the most likely explanation is that there is a numerical error somewhere in the calculation that I have been unable to detect. However, it is possible that the method itself is somehow flawed. Further investigation will be necessary to understand the deviation between the simulated and theoretical results.
4: POINT SOURCE REMOVAL

Figure 4.5: Comparison of the simulated noise-to-signal ratio (black solid line), without deconvolution applied, to the theoretical value as given by Equation (4.37) (red dotted line) for one source per 100 deg$^2$ (left) and one source per 10 deg$^2$ (right).

4.5 Conclusions

In this section, I have considered the effect of point-sources and their removal on B-mode measurements. In particular, I have looked at how the issues which arise at low-$\ell$ when using the pure-$C_\ell$ method might affect the ability to remove point-source contamination by cutting out point-sources above a certain flux level. In order to achieve this, I began by investigating the effect of point-sources on the pure-$C_\ell$ method. I found that, at low-$\ell$, there was a significant increase in the standard deviation of the estimator relative to the case where there was no point-source contamination. Further, I discovered that the additional error caused by the effects of the point-sources increases linearly with the number of sources. Further, it seems to depend little on the mask for very low $\ell$. Using this information, I was then able to derive a formula that would, in principle, allow the optimal point-source masking level to be determined as a function of the frequency and the value of $\ell$ being observed.

Then, using these results, I looked at the results of Battye et al. (2011) in order...
to determine what levels of $r$ might be detectable with and without source masking. In general, it was found that, at high frequencies, there would likely be little problem even with attempting to detect $r = 0.001$, since the source contamination there is relatively negligible. However, at low frequencies, the combination of the very high source fluxes (which result in a very high level of masking being necessary) coupled with the fact that masking out a large number of sources using the pure-$C_\ell$ method will render low-$\ell$ modes undetectable means that values of $r$ much below 0.1 will likely not be detectable.

Finally, I produced an analytic formula to attempt to determine the effect of point-sources on the pure-$C_\ell$ method. However, unfortunately this formula does not seem to work, likely due to some numerical error.
4: POINT SOURCE REMOVAL
Null test of CMB polarisation data using position angles

Measurement of the B-mode power spectrum will be technically challenging since a scalar-to-tensor ratio of \( r \approx 0.01 \) will require measurements with noise levels below 100 nK to be achieved. In addition, contamination has to be constrained to be well below the noise level. Contamination will come from three primary sources: astrophysical foregrounds (emission from diffuse components of the galaxy and extragalactic sources), the atmosphere and the telescope/receiver system. In order to ensure that such contamination has been successfully taken account of, a variety of null tests have been applied to CMB measurements in order to refine and test the quality and consistency of the data. This chapter describes a novel test which uses the CMB polarisation position angle, \( \alpha \), as defined in Section 1.2.2.

In this chapter of my thesis, I will outline the principles behind this method, determine and demonstrate the effectiveness of the method through numerical simulations and then attempt to theoretically derive the method’s properties as far as is possible. In order to understand the method, and the circumstances in which it will be useful, there are three steps. First, I must characterise the probability distribution of \( \alpha \) in the case where there are no systematic errors, including the expected level of accuracy to which
5: NULL TEST OF CMB POLARISATION DATA USING POSITION ANGLES

this distribution could be measured for any given experiment. Here, there are several
questions to be answered. Most importantly it is not clear under what circumstances
the null hypothesis of a uniform distribution for \( \alpha \) is valid, and this will be discussed
using simulations and, later, theoretical calculations. Another important question is
under what circumstances the standard deviation about the mean will be given by the
usual \( \sqrt{n} \) law, and this will also be considered and answered, using simulations and
theoretical results. Then, I must consider how systematic errors will affect the shape
of this distribution, as a function of the size and nature of the errors. Finally, I must
combine the two together to determine under what circumstances systematic errors
could be detected. In order to do this, I have used the \( \chi^2 \) method, and understanding
the workings of this for the case of correlated histogram bins is another important part
of this chapter. Finally, I will discuss the shortcomings of the method, and consider the
circumstances under which this method will be useful in practice.

5.1 Basic principles of the method

As stated in Section 1.2.2 the polarisation position angle is defined by

\[
\alpha = \frac{1}{2} \arctan \left( \frac{U}{Q} \right). \tag{5.1}
\]

Since this is a coordinate-dependent quantity, it is not often discussed in the context of
the CMB, as it is expected to be entirely random for CMB maps. However, this random
nature makes it ideal for use as a null test to detect certain forms of residual systematic
errors in the data and also, potentially, certain forms of foreground contamination,
since any observed deviation from the predicted flat distribution must be caused either
by correlated noise (which is known and, thus, can be accounted for) or else by some
unknown systematic contaminant.

Although polarisation position angles are not usually considered in CMB experi-
ments, they are commonly used in the context of other astrophysical sources. Battye et al. (2008), for example, uses similar techniques to those outlined later in this chap-
In order to establish the existence of biases (at the level of $\langle \Delta \alpha^2 \rangle^{1/2} \approx 0.3^\circ$) in the polarization position angles measured in the NRAO-VLA Sky Survey (NVSS) point-source catalogue. Briefly, it was found that there were highly significant biases in the measured position angles toward angles which were integer multiples of $45^\circ$ (that is, $Q = 0$ or $U = 0$), in complete contradiction to the null hypothesis that they should be uniformly distributed. It was argued that CLEAN bias (Högboom 1974) combined with small multiplicative and additive offsets were responsible for these effects.

Throughout this chapter, I will make the simplifying assumption that the observed Stokes’ parameters $(Q_{\text{obs}}, U_{\text{obs}})$ are just functions of the true values $(Q, U)$, that is, I will only allow for a restricted Muller matrix (see, for example, O’Dea et al. 2007). In this case, defining $Q_{\text{obs}} = Q'$ and $U_{\text{obs}} = U'$, I can write

$$
\begin{pmatrix}
Q' \\
U'
\end{pmatrix} = \begin{pmatrix}
Q \\
U
\end{pmatrix} + \begin{pmatrix}
\epsilon_{QQ} & \epsilon_{QU} \\
\epsilon_{UQ} & \epsilon_{UU}
\end{pmatrix} \begin{pmatrix}
Q \\
U
\end{pmatrix},
$$

(5.2)

where it is expected that $\epsilon$ will be small. Note that this is by no means the only source of error due to systematics, since leakage from intensity into polarisation could potentially also have a significant effect, which will be considered in future work. I will also assume, for simplicity, that $\epsilon$ is constant, and does not vary between pixels, although this assumptions may have to be relaxed if the method were to be applied to real experiments.

### 5.2 Properties of the probability distribution

In order to understand in what situations this method will be useful, and to determine its properties, I first need to consider the probability distribution of polarisation position angles in the absence of systematic errors, as a baseline for the case where systematics are present. All of the information required to calculate the number of pixels in a bin (hereafter denoted by $n$) is contained within the joint probability distribution $\mathcal{P}(\alpha_1, \ldots, \alpha_N)$. However, since the position angles for different pixels are, in general,
correlated, this is an extremely complicated distribution, and cannot usually be computed. Thus, it is usually necessary to resort to simulations to compute the probability distribution of $\alpha$, although it will be shown in Sections 5.7 and 5.8 that there are useful expressions which allow the mean and variance of the distribution to be calculated without computing the full probability distribution.

Since only a finite number of pixels can be observed, I cannot determine the exact probability distribution. Instead, the probability distribution is estimated using histograms produced by binning the data into bins corresponding to different ranges for the position angle. In order to make plotting the histograms more convenient, I have plotted the quantity

$$\hat{n} = n - \bar{n},$$

(5.3)

where $n$ is the number of pixels in the bin and I have defined $\bar{n} = N_{\text{pix}} / N_{\text{bins}}$, the average number of pixels in a bin. This quantity has the useful property that $\langle \hat{n} \rangle = 0$.

### 5.2.1 Distribution of $Q$ and $U$ using a multivariate Gaussian

In this section, I will first consider the probability distribution of $\alpha$ in a single pixel, before generalizing to a map with $N_{\text{pix}}$ pixels (or collection of $N_{\text{pix}}$ individual polarisation measurements). If the joint probability distribution for the Stokes’ $Q$ and $U$ parameters, $P(Q, U)$, is known, then one can convert from $P(Q, U)$ to $P(\alpha)$.

There are two steps required for the conversion. Firstly, $p(P, \alpha)$ must be calculated. To do this, I use the standard formula for co-ordinate transformations which gives

$$\mathcal{P}(\alpha) = 2 \int_0^\infty PdP \mathcal{P}(Q(P, \alpha), U(P, \alpha)).$$

Then, to find $\mathcal{P}(\alpha)$, I need to integrate over all possible values of $P$, which gives

$$\mathcal{P}(\alpha) = 2 \int_0^\infty PdP \mathcal{P}(Q(P, \alpha), U(P, \alpha)).$$

(5.4)

If $Q$ and $U$ are correlated, normally distributed random variables with zero mean, with $\text{Var}(Q) = \sigma_{QQ}$, $\text{Var}(U) = \sigma_{UU}$ and $\text{Cov}(Q, U) = \sigma_{QU}$ (where $\sigma_{QU} < \sqrt{\sigma_{QQ}\sigma_{UU}}$),
then

\[
P(Q, U) = \frac{1}{2\pi\sqrt{\sigma_{QQ}\sigma_{UU} - \sigma_{QU}^2}} \exp \left( -\frac{Q^2\sigma_{UU} + U^2\sigma_{QQ} - 2QU\sigma_{QU}}{2(\sigma_{QQ}\sigma_{UU} - \sigma_{QU}^2)} \right).
\]

(5.5)

Expressions for \( Q \) and \( U \) (in terms of \( P \) and \( \alpha \)) can be substituted in to this to obtain

\[
P(Q(P, \alpha), U(P, \alpha)) = \frac{1}{2\pi\sqrt{\sigma_{QQ}\sigma_{UU} - \sigma_{QU}^2}} \times \exp \left( -P^2\sigma_{UU} \cos^2 2\alpha + \sigma_{QQ} \sin^2 2\alpha - 2\sigma_{QU} \cos 2\alpha \sin 2\alpha \right),
\]

(5.6)

and using (5.4),

\[
P(\alpha) = \frac{2\sqrt{\sigma_{QQ}\sigma_{UU} - \sigma_{QU}^2}}{\pi((\sigma_{QQ} + \sigma_{UU}) + (\sigma_{UU} - \sigma_{QQ}) \cos 4\alpha - 2\sigma_{QU} \sin 4\alpha)}.
\]

(5.7)

In the case where \( \sigma_{UU} = \sigma_{QQ} \) and \( \sigma_{QU} = 0 \) then \( P(\alpha) = 1/\pi \), that is, the polarisation position angle is uniformly distributed.

One can generalize this treatment to an ensemble of measurements. The measurements of \( Q \) and \( U \) at each of \( N = N_{\text{pix}} \) pixels will be treated as a set of random variables (in this case, of non-zero mean), \( x = (x_1, \ldots, x_{2N}) = (Q_1, U_1, \ldots, Q_N, U_N) \). This set of variables have a multivariate Gaussian distribution

\[
P(x_1, \ldots, x_{2N}) = \frac{e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}}{(2\pi)^N|\Sigma|^{1/2}},
\]

(5.8)

where \( \Sigma \) is the covariance matrix of \( x \) (which must not be singular) and \( \mu \) is the mean of \( x \) (which will usually (but not always) be taken to be zero). However, in this thesis, I am dealing with the polarisation position angle, and not \( Q \) and \( U \) directly. Thus, I must calculate the distribution \( P(\alpha_1, \ldots, \alpha_N) \). In general this is non-trivial to calculate, since the values of \( Q \) and \( U \) can be correlated between pixels.

### 5.2.2 Uncorrelated pixels

As noted above, the probability distribution is usually too complex to be computed exactly. However, in the case where the pixels are not correlated, it becomes considerably
simpler. Here, the probability distribution is separable, meaning that $\mathcal{P}(\alpha_1, \ldots, \alpha_N) = \mathcal{P}(\alpha_1) \cdots \mathcal{P}(\alpha_N)$ and, thus, the general solution can be found by considering each pixel individually.

Using standard results, I find that, if the probability distribution for $\alpha$ is constant over all pixels, then

$$\langle n \rangle = N_{\text{pix}} \mathcal{P}(\alpha_1 \leq \alpha < \alpha_2),$$

where $N_{\text{pix}}$ is the total number of pixels and $n$ the number of pixels in the bin with $\alpha_1 \leq \alpha < \alpha_2$. I can also prove (see Section 5.7) that if the probability distribution for each pixel is different,

$$\langle n \rangle = \sum_{i=1}^{N_{\text{pix}}} \mathcal{P}(\alpha_1 \leq \alpha_i < \alpha_2).$$

Furthermore, the standard deviation can be predicted, and is given approximately by the formula $\sigma_n = \sqrt{n}$ (see Section 5.8.2). This formula also fits well with the simulated results. It is also useful to define

$$\hat{n} = \frac{n - \bar{n}}{\bar{n}},$$

where I have defined $\bar{n} = N_{\text{pix}}/N_{\text{bins}}$, the average number of pixels in a bin. This number has the useful property that $\langle \hat{n} \rangle = 0$. The variance is approximately given by

$$\langle \hat{n}^2 \rangle = 1/\bar{n}.$$

For the case of a distribution which can be described fully by a power spectrum (such as a Gaussian CMB), I have found that the pixels are uncorrelated (and hence the formula given here works) iff $C_\ell = \text{const.}$ (i.e. the map consists of white noise). In this case, it is found that $\langle \hat{n} \rangle = 0$ for all bins.

### 5.2.3 Correlated pixels

Whilst most noise-dominated maps will have fairly limited inter-pixel correlations, this cannot be assumed to be true for other types of map. In particular, CMB-dominated
5.2: PROPERTIES OF THE PROBABILITY DISTRIBUTION

maps will be correlated. In Section 5.7, it will be shown that inter-pixel correlations do not, in fact, affect the mean number of pixels in the bin and, thus, Equation (5.10) still gives the mean even for the correlated case. However, the probability distribution for $\alpha$ is affected by correlations, and in particular, the simple $\sqrt{n}$ formula for the standard deviation does not in general apply to maps with correlated pixels. For CMB-dominated maps, it is found, in fact, the standard deviation of the number of points in each histogram bin depends on the particular CMB power spectrum used and, further, that it varies between bins. To investigate the error on the measurement of $\alpha$ for CMB maps, I have carried out several simulations.

I have carried out simulations using four different power spectra, using a Monte-Carlo method with 1000 realisations at $N_{\text{side}} = 512$ (which corresponds to $N_{\text{pix}} \approx 3 \times 10^6$). The four spectra are a constant-$C_\ell$ power spectrum (white noise) and three CMB spectra, one using a standard E-mode polarisation spectrum and B-modes with $r = 1$ (but no lensing), one using a standard E-mode spectrum with $r = 0$ (no B-modes at all) and the final one using $r = 1$ again, but with $C_{\ell}^{EE}$ set to 0 artificially to produce pure B-mode maps. The results are presented in Figure 5.1.

In all four cases used here, $\langle \hat{n} \rangle$ is consistent with zero for all bins. This makes sense from a qualitative viewpoint, since the CMB is a cosmological signal, and thus should not be dependent on the choice of coordinate system, whereas the Stokes’ parameters are coordinate dependent. In fact, it can be proven (see Section 5.9.1) that the probability distribution of $\alpha$ is flat for all CMB-like maps.

From the histograms, however, it is obvious that the formula for the standard deviation when pixels are uncorrelated does not apply when pixel correlations are included. The first histogram, which is equivalent to white noise, follows the expected $\sqrt{n}$ law. However, the other three, which are CMB realisations, do not. The standard deviations in these are up to 10 times larger for certain bins, and furthermore, it varies between bins, in a manner which is dependent on the power spectrum used. Although an analytic formula for these variations has not yet been computed, they can be described and explained, to some extent, qualitatively. It can be seen that a power spectrum that
Figure 5.1: Histogram of the $\hat{n}$ in each bin for white noise ($C_\ell = \text{const}$.) (top left) and three CMB-like power spectra, one with $r = 1$ (top right), one with $C^{BB}_\ell = 0$ (bottom left) and the final one with $C^{EE}_\ell = 0$ (and $r = 1$) (bottom right), with the solid line being the mean of 1000 simulations and the error bars being the expected variance for one realisation, calculated from the simulations. Note that the scales are not the same.

It can be seen that, whilst the white noise case follows the standard $\sqrt{n}$ error formula, the CMB-like spectra have errors which are considerably larger, vary between bins and also depend on the particular power spectrum provided.
is dominated by E-modes (such as the CMB itself) has errors that oscillate following the equation

\[ \sigma_N(\alpha) = A + B \cos 4\alpha , \]

(5.12)

where \( A \) and \( B \) are measurable parameters that depend on the actual input power spectrum and that cannot currently be calculated theoretically. Similarly, it appears that power spectra dominated by B-modes oscillate following

\[ \sigma_N(\alpha) = A - B \cos 4\alpha . \]

(5.13)

In the case where both are equal, \( B = 0 \), and the error becomes constant.

From these results, it can be seen that the standard deviation for a pure B-mode CMB spectrum is considerably larger (approximately a factor of 10) than that for either the pure-E or the combined E and B spectra. The combined E and B spectrum has a similar maximum standard deviation to the pure-E spectrum. However, the minimum is considerably larger due to the addition of the B-modes. A CMB spectrum with \( r \ll 1 \) has a similar distribution to the pure E-mode case (since B-modes are then negligible).

The increase in the standard deviation of the histogram bins as compared to the uncorrelated case is due to the effect of the inter-pixel correlations. These correlations act to reduce the effective number of pixels being measured and, thus, reduce the level of accuracy to which the probability distribution can be measured. The oscillations in the magnitude between bins are probably caused by differences in the correlation function for \( Q \) and \( U \) (in the general case), which results in a difference in the level of accuracy of measurements of \( \alpha \) depending on which of the two is dominant for that particular value. At first glance, this might seem odd. Since \( E \) and \( B \) are supposed to, in general, be independent of the orientation of the axes, whereas \( \alpha \) is specifically dependent on it, it seems unusual that a quantity which is axis dependent should have a variance which changes according to the relative size of the E and B modes, which do not vary in such a way. However, the simulations have been repeated, and the same result is found. Further, the calculations done in Section 5.9.1 show that there
exists a difference between the covariance matrix of $Q$ (Equation (5.179)) and that of $U$ (Equation (5.180)) that depends on the nature of the EE and BB power spectra which, whilst not proof of the formulae given here, is an indication that they are at least plausible. Finally, whilst it would indeed be exceptionally dubious if the mean number of pixels in each histogram bin were not the same, it can be seen from these results that, to within experimental error, they are the same, and, indeed, it can be mathematically proven (Section 5.9.1) that the probability distribution of $\alpha$ is indeed flat for any map with statistical properties that can be described solely by a standard power spectrum. The variance being different between each bin is, whilst still perhaps unexpected, less obviously incorrect, at least given the strong evidence to the contrary presented here.

It is also possible to consider how the errors change when the number of pixels change. For the uncorrelated noise case, this is trivial (since the errors just go approximately as $\sqrt{n}$). However, in the case of the CMB and other correlated sets of pixels, it is non-trivial. In Figure (5.2), I have plotted the maximum and minimum errors for different values of $N_{\text{pix}}$ for a pure-E mode CMB and for a pure-B mode CMB.

These results show a substantial difference between the two cases. Whereas the error on the pure-E modes changes approximately as $1/\sqrt{N_{\text{pix}}}$, the error for the B-mode case hardly falls at all after about $4 \times 10^4$ pixels. This is because, whereas the E-mode polarisation has a reasonable amount of small-scale power, the primordial B-mode polarisation (excluding lensing) has most of its power at large scales. As a result, increasing the resolution beyond a certain point just results in creating many highly-correlated clusters of pixels, meaning that the effective number of pixels does not increase much.

## 5.3 Effect of systematic errors

The previous section gives some idea of the properties of histograms generated from considering the polarisation position angle of both CMB signals and noise. However, the main purpose of this chapter is to consider how well systematic errors can
5.3: EFFECT OF SYSTEMATIC ERRORS

Figure 5.2: Fractional error ($\sigma_n/n$) plotted against $N_{\text{pix}}$ for pure-E mode (left) and pure-B mode (right) CMB runs. The dashed lines are $\frac{1}{\sqrt{n}}$, the error for uncorrelated pixels, the dotted lines show the fractional error in the bin with the lowest error, and the solid lines show the error in the bin where the error is largest.

be detected using position angle histograms, and so I will now consider the effects of systematic errors on $\alpha$.

5.3.1 Definition and properties of systematic errors

Equation (5.2) gives the effect of systematic errors in terms of the Muller matrix. However, the $\epsilon$ parameters given in this equation do not form a particularly natural set of parameters and, thus, it is useful to re-parameterise this equation as follows

$$
\begin{pmatrix}
Q' \\
U'
\end{pmatrix} = 
\begin{pmatrix}
Q \\
U
\end{pmatrix} + 
\begin{pmatrix}
\epsilon_0 \begin{pmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{pmatrix} \\
\epsilon_1 \begin{pmatrix}
1 & 0 \\
0 & -1
\end{pmatrix} \\
\epsilon_2 \begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\end{pmatrix} \begin{pmatrix}
Q \\
U
\end{pmatrix}.
$$

(5.14)
Using Equation (5.2), I can find expressions for these parameters. I find that

\[
\tan \theta = \frac{\epsilon_{UQ} - \epsilon_{QU}}{(\epsilon_{QQ} + \epsilon_{UU})} \quad (5.15)
\]

\[
\epsilon_0 = \frac{1}{2} \sqrt{(\epsilon_{QQ} + \epsilon_{UU})^2 + (\epsilon_{UQ} - \epsilon_{QU})^2} \quad (5.16)
\]

\[
\epsilon_1 = \epsilon_{QQ} - \epsilon_{UU} \quad (5.17)
\]

\[
\epsilon_2 = \epsilon_{QU} + \epsilon_{UQ} \quad (5.18)
\]

Note that these parameters all have a physical meaning. The parameter \( \theta \) corresponds to a global rotation of the position angles, and \( \epsilon_1 \) and \( \epsilon_2 \) correspond to shearing. This parametrisation is useful because varying each of these four parameters individually gives a simple result, which will be shown by simulations in Section 5.3.4, and proven analytically in Section 5.10. Setting \( \epsilon_0 \neq 0 \) does not change the probability distribution at all in the absence of other systematic errors, and \( \theta \neq 0 \) (in conjunction with a non-zero \( \epsilon_0 \)) can only result in a shift in the histogram bars, which will not be detectable unless said distribution is not flat (noise with different variances for \( Q \) and \( U \) or with \( Q-U \) correlations seems not to have a flat distribution, but the CMB does).

### 5.3.2 Detection of systematic errors in experiments

For a real experiment, there will be a limit on the level of systematic errors that can be detected, due to the intrinsic error in the measurement of \( \hat{n} \). To evaluate the usefulness of this method, this limit must be determined. This requires the use of some statistical test on the data, in order to determine the likelihood of obtaining any particular value by chance. Since the number of data points in a histogram bin is (for a large enough bin size) very well approximated by a normal distribution, the most obvious test to use is the \( \chi^2 \) test. However, there are some complications with this method, due to the fact that the histogram bins are not uncorrelated.

First, consider the case where there are no correlations. Then, if there are \( N \) bins, the variable

\[
\chi^2_N = \sum_{i=1}^{N} \frac{(n_i - \bar{n}_i)^2}{\sigma_i^2},
\]

(5.19)
will, if the null hypothesis is correct, follow a chi-squared distribution with \( N \) degrees of freedom. However, the situation where the null hypothesis is in fact not correct, and \( \bar{n}_i \neq \langle n_i \rangle \) (where the first is the assumed mean, and the second is the actual mean) must also be considered. This will, for example, be the case where there exist unknown systematic errors. Even though the null hypothesis in this case will be false, if the difference between the predicted and actual means is small enough, this may not be detectable, due to the random fluctuations in the \( \hat{n}_i \). I therefore need to consider the probability distribution of the \( \chi^2 \) value in the presence of systematic errors.

It can be shown that, in the general case, the value of \( \chi^2 \) is given by a non-central chi-squared distribution, which depends on the number of degrees of freedom (which is \( N_{\text{bins}} \)) and the parameter

\[
\lambda = \sum_{i=1}^{N_{\text{bins}}} \left( \frac{\langle n_i \rangle - \bar{n}_i}{\sigma_i} \right)^2,
\]

(5.20)

which depends only on the properties of the distribution which can be calculated from theoretical values, or from simulations. Using this distribution, it is possible to calculated the likelihood of obtaining a \( \chi^2 \) that is higher than a given critical value, and thus determine the probability that the null hypothesis could be rejected at a given confidence level. To do this, I have calculated and plotted the probability of obtaining a \( \chi^2 \) greater than the value required to disprove the null hypothesis at the 50%, 75%, 90%, 95%, 99%, 99.9% and 99.99% confidence levels as a function of \( \lambda/N_{\text{bins}} \) for \( N_{\text{bins}} = 20 \). These results are presented in Figure 5.3 and these will be important for the following two sections.

It can be seen that a detection is virtually certain for values of \( \lambda \) over about 4, even at the 99.99% confidence level. Similarly, there is not a significant deviation from the probability of detection expected by chance for \( \lambda \ll 1 \). This agrees well with what would be intuitively expected. Using the 95% confidence level as an indication of a detection, a 50% chance of detection requires \( \lambda = 0.61 \), and a 90% chance of detection requires \( \lambda = 1.31 \). Thus, a detection is very likely for values of \( \lambda \) of order 1. The manner in which this method works makes it good as a null test, since whilst
the absence of a detection cannot rule out the presence of systematics below values of $\lambda$ of around 2, there is a good probability that even values of $\lambda$ of around 0.5 will be detectable.

In reality, this analysis is not sufficient to deal with histograms, since the constraint on the total number of pixels means that the number of pixels in each bin cannot possibly be uncorrelated (since increasing the number in one bin means that there are fewer pixels remaining to place in the other bins). However, in the case where the histogram bins are uncorrelated, this approach will produce an approximately correct answer. Understanding the $\chi^2$ test in the case of a histogram, in particular one where there are inter-pixel correlations (which, in turn, results in correlations existing between the number of pixels in different bins) is the focus of Section 5.6, but I have as yet not determined exactly how this will affect the analysis in this section.
5.3.3 Simulations of uncorrelated noise

I have carried out simulations of noise without correlations between different pixels. I computed 1000 pure, constant noise realisations for several values of $\sigma_{QQ}$, $\sigma_{UU}$ and $\sigma_{QU}$ (see Section 5.2.1 for definition) on a full sky using the HEALPix (Górski et al. 2005) pixelisation with $N_{\text{side}} = 512$ and binned the resultant position angles into 20 equally-sized bins. I then plotted a histogram of $\hat{n}$, as defined in Equation (5.11), for each different set of realisations, together with predictions, and these results are shown in Figures 5.4 and 5.5. The predictions are normalised by subtracting the flat probability distribution ($P(\alpha) = \frac{1}{\pi}$) from the predicted probability distribution, and then dividing by this. This has the effect of normalising the predictions in the same way as the simulations. In Section 5.2.1, it was shown that the expected distribution of the position angles (suitably modified to take account of the normalisation) is given by

\[
P(\alpha) - \bar{P} = \frac{2\sqrt{\sigma_{QQ}\sigma_{UU} - \sigma_{QU}^2}}{((\sigma_{QQ} + \sigma_{UU}) + (\sigma_{UU} - \sigma_{QQ}) \cos 4\alpha - 2\sigma_{QU} \sin 4\alpha) - 1},
\]  

(5.21)

and, so, this is what I plotted. Further, the error bars on the histograms should be given by

\[
\text{Var}(\hat{n}) \approx \frac{N_{\text{bins}}}{N_{\text{pix}}}. 
\]  

(5.22)

Since $N_{\text{bins}} = 20$, and $N_{\text{pix}} = 12 \times 512^2 = 3,145,728$, the 1$\sigma$ error on $\frac{n - \bar{n}}{\hat{n}}$ is approximately given by $\sqrt{\frac{20}{3145728}} \approx 0.0025$. This agrees well with the error bars shown in Figures 5.4 and 5.5 (this is most obvious on the control). There is a slight discrepancy due to the use of the $\sqrt{n}$ approximation, but it is not significant. Thus, it appears that Equation (5.22) well for uncorrelated noise. However, in the more general case of correlated noise or CMB, this formula will not apply.

From Figures 5.4 and 5.5 it can also be seen that the predictions from Equation (5.21) agree extremely well with the simulated results, with the predicted line running exactly through the middle of the histogram peaks. This is as expected, since the predictions should hold in this situation.
Figure 5.4: Histogram of number of pixels in each bin, plus error bars showing the standard deviation of the number of pixels in the bin, for different values of $\sigma_{QQ}$ and $\sigma_{UU}$, with $\sigma_{QU} = 0$, with theoretical probability distributions also included.
5.3: EFFECT OF SYSTEMATIC ERRORS

Figure 5.5: Histogram of number of pixels in each bin, plus error bars showing the standard deviation of the number of pixels in the bin, for different values of $\sigma_{QU}$, with $\sigma_{QQ} = \sigma_{UU} = 1$, with theoretical probability distributions also included.
5.3.4 Signal-dominated regime

Now, I shall consider the effects of systematic errors on maps of the CMB, and in particular, how large these effects have to be to be observed. I have created 100 CMB maps and then applied different systematic errors to these, setting precisely one of $\epsilon_1$ or $\epsilon_2$ to be non-zero in each case. As a control, I also carried out simulations with only $\epsilon_0$ and $\theta$ being non-zero. I have then plotted histograms of the results (see Figures 5.6 and 5.7). I find, as expected, that changing $\epsilon_0$ has no effect whatsoever on the distribution of $\alpha$, and changing $\theta$ simply changes $P(\alpha)$ to $P(\alpha+\theta)$, which in the case of the CMB could not be observed to within experimental error using this method (although it could, of course, be detected by the effect on the power spectrum). Thus, these errors should not be detectable in the signal-dominated regime (they may be detectable if $\epsilon_1$ or $\epsilon_2$ are large, however these cases are unlikely to be of importance).

Systematic effects which result in non-zero $\epsilon_1$ or $\epsilon_2$, however, are potentially detectable, as can be seen from Figures 5.6 and 5.7. From these results, it can be seen empirically that, in the presence of systematic errors, a flat initial probability distribution for $\alpha$ is transformed approximately into the following

$$P(\alpha) = P_{\text{orig}}(\alpha)(1 + 2\epsilon_1 \cos 4\alpha + 2\epsilon_2 \sin 4\alpha),$$  

(5.23)

which will be proven analytically in Section 5.10.

However, this in itself is not useful, since the important quantity here is the likelihood of a detection. This, of course, will depend on both the CMB power spectrum (which affects the error bars) and also the number of pixels observed (since a larger number of observed pixels gives smaller errors). From looking at the plots, it appears that the fluctuations start becoming detectable when $\epsilon$ is somewhere between 0.005 and 0.01. However, it is useful to have a more rigorous indicator of this. To do so, I would carry out a $\chi^2$ test to determine if the deviation from the control is statistically significant. Using the results from Section 5.3.2, I can define

$$\frac{A}{N_{\text{bins}}} = \frac{1}{N_{\text{bins}}} \sum_{i=1}^{N_{\text{bins}}} \frac{(\langle N_i \rangle_{\text{test}} - \langle N_i \rangle_{\text{control}})^2}{\sigma_i^2},$$  

(5.24)
5.3: EFFECT OF SYSTEMATIC ERRORS

Figure 5.6: Histogram of $\hat{n}$ against $\alpha$ including systematic effects for different values of $\epsilon_1$: (top left) $\epsilon_1 = 0.1$; (top right) $\epsilon_1 = 0.05$; (bottom left) $\epsilon_1 = 0.01$; (bottom right) $\epsilon_1 = 0.005$. I estimate that it should be possible to detect values of $\epsilon_1$ of around 0.01.
Figure 5.7: Histogram of \( \hat{n} \) against \( \alpha \) including systematic effects for different values of \( \epsilon_2 \): (top left) \( \epsilon_2 = 0.1 \); (top right) \( \epsilon_2 = 0.05 \); (bottom left) \( \epsilon_2 = 0.01 \); (bottom right) \( \epsilon_2 = 0.005 \). I estimate that it should be possible to detect values of \( \epsilon_2 \) of around 0.01.

where \( \langle N_i \rangle_{\text{test}} \) is the expected number of pixels in bin \( i \) for the test inputs (with the systematics added), \( \langle N_i \rangle_{\text{control}} \) is the average number of pixels in bin \( i \) for the control (without systematics) and \( \sigma_i \) is the error in the number of pixels in each bin on the
5.4: APPLICATION TO WMAP 5

control. I find

\[
\frac{\lambda}{N_{\text{bins}}} = (143 \epsilon_1)^2, \quad (5.25)
\]

\[
\frac{\lambda}{N_{\text{bins}}} = (100 \epsilon_2)^2. \quad (5.26)
\]

Thus, using the results of Section 5.3.2, I can show that to have a 50% chance of a
detection at a 95% confidence level, \( \lambda = 0.61 \) is needed, which gives \( \epsilon_1 = 0.0055 \) and
\( \epsilon_2 = 0.0078 \), and for a 90% chance of detection, \( \lambda = 1.31 \) is required, and this requires
\( \epsilon_1 = 0.008 \) and \( \epsilon_2 = 0.011 \). Notice that systematic errors with \( \epsilon_1 \neq 0 \) are slightly easier
to detect than those with \( \epsilon_2 \neq 0 \). This is because the oscillations in \( \epsilon_1 \) are in phase with
those in the errors, which means that the minimum error coincides with the point at
which the oscillations in \( P(\alpha) \) are largest, whereas in the case of \( \epsilon_2 \neq 0 \), the minimum
error occurs where there is no difference between \( P(\alpha) \) in the control and test cases.
Note, however, that this analysis assumes that the pixels and, thus, the histogram bins
are uncorrelated, which is manifestly not true. In practice, this correlation will reduce
the effective number of bins and, thus, will reduce the level to which systematic errors
can be detected by this method.

5.4 Application to WMAP 5

In this section, I will consider the properties of the WMAP five-year polarisation maps
(see Bennett et al. (2003b), Verde et al. (2003), Page et al. (2007), Jarosik et al. (2007),
Hinshaw et al. (2009) and Limon et al. (2008) for further details). The full (N_{\text{side}} = 512)
WMAP polarisation maps are strongly dominated by noise and foregrounds, with the
CMB signal being barely detectable. It is possible to obtain a map which contains a
larger proportion of CMB signal by combining the measurements for each difference
assembly (DA) at a given frequency, but the resultant maps are still noise and fore-
ground dominated at the single pixel level. It is also possible to reduce the amount of
foreground contamination by producing foreground-reduced maps (see Hinshaw et al.
(2009) and Limon et al. (2008) for details) or by masking the galactic plane, where
most of the foreground contamination is found. The WMAP experiment has five fre-
quency bands (Hinshaw et al. 2009) – K band (23 GHz), Ka band (33 GHz), Q band
(41 GHz), V band (61 GHz) and W band (94 GHz), but the foreground-reduced maps
only exist for Q, V and W bands, since the K and Ka bands are used to construct
foreground templates.

5.4.1 Properties of WMAP polarisation measurements

In order to do a useful test, it is first necessary to consider what the spectrum should
look like. Since the maps are noise-dominated, Equation (5.21) should give the prob-
ability distribution for $\alpha$ for any given pixel, once the noise correlation matrix for that
pixel is known. However, unlike in Section 5.2.2, the noise is not constant across the
sky, since each pixel has been measured a different number of times (Jarosik et al.
2007).

An $N_{\text{obs}}$ map is provided with the WMAP data set, and this, together with the given
noise level for each DA, allows the noise correlation matrix for each pixel to be found
(see Limon et al. (2008) and Jarosik et al. (2007) for more details). These can then
be combined together (Jarosik et al. 2007) to produce noise correlation matrices for
the combined frequency map. Once this is known, Equation (5.10) could in theory be
used to determine the mean number of pixels in each histogram bin, by summing the
probabilities of each pixel being in the bin. In practice, it is simpler to carry out Monte
Carlo simulations in order to determine the mean and variance of the number of pixels
in each bin. This can then be compared with the histogram generated from the actual
WMAP maps and any discrepancies can then be detected.

5.4.2 Simulations of WMAP noise

The simplest way to determine the expected histogram produced by WMAP noise is
to carry out simulations. Therefore, I have used the $N_{\text{obs}}$ matrix for each pixel from
each combined frequency map to determine the noise correlation matrix for each pixel.
Then, using this, I generated 1000 noise realisations with the correct statistical properties (using an adaptation of the method I have used for generation of $TE$ correlations (as defined in Section 2.1.1) with $TT$, $TE$ and $EE$ replaced by $\sigma_{QQ}$, $\sigma_{QU}$ and $\sigma_{UU}$ respectively and $a_{lm}^T$ and $a_{lm}^E$ replaced by $Q(x)$ and $U(x)$), added a CMB realisation with $r = 0.1$ to each realisation and then calculated a histogram of $\hat{n}$ in each bin (see Equation (5.11) for the definition of this), for two cases – the case where no mask was applied and the case where the WMAP five-year polarisation mask was applied. These histograms are shown in Figure 5.8.

In both cases, the histograms show an approximately sinusoidal oscillation with a period of $4\alpha$. This is expected from the theoretical prediction from Equation (5.21). The exact shape is more complicated than in Figures (5.4) and (5.5), due to the correlations varying between pixels. However, qualitatively they are similar to those deduced in more idealised situations.

### 5.4.3 WMAP without foreground removal

Once I have simulations of the WMAP noise, these need to be compared to the results generated from the actual WMAP maps. WMAP has a total of 10 DAs at five different frequencies, and each DA has five years of data currently available. I have produced combined maps for each frequency band using the process outlined in Jarosik et al. (2007). I have then generated a histogram by plotting, for each bin,

$$\hat{n}_{\text{sim}} = \frac{n - n_{\text{sim}}}{\bar{n}},$$

(5.27)

where $n_{\text{sim}}$ is the average number of pixels in the bin from simulations, and $\bar{n} = N_{\text{pix}}/N_{\text{bins}}$. These are plotted in Figure 5.9. For the non-foreground-reduced, unmasked case, it is found that, after subtraction of the predicted signal from the noise, there is a large peak at $\alpha = 0^\circ$, which seems to be caused by foregrounds. In particular, a look at the WMAP polarisation maps (Page et al. 2007) shows that around the North Polar Spur, there is a large region where $\alpha \approx 0^\circ$ (indicated by the polarisation vectors on the map being approximately vertical). This is likely to be the source of the peak. Looking
Figure 5.8: \( \hat{n} \) vs. \( \alpha \) for simulated WMAP single-frequency maps.
at the histograms shows that the foreground effect is dominant for the K band map, and has a strong effect on the Ka band map. It is, however, also noticeable in the other bands (Q, V and W). If the reduced chi-squared of $\hat{n}_{\text{sim}}$ is calculated, I find that the lowest (W band) is around 25 (as opposed to around 1, which would be expected if the data was well described by the noise and CMB model), indicating that the data in this case clearly cannot be described by noise and CMB alone.

However, it is possible to attempt to remove this foreground contamination in two ways. Firstly, by attempting to apply foreground-removal algorithms to the maps. This will be considered in the next section. Another method is to apply a mask to cut out the most highly contaminated regions. In particular, the PO2 WMAP five-year polarisation mask (which is more aggressive than the standard PO6 mask used for WMAP polarisation analysis, and thus will reduce the likelihood of residual foreground contamination) will cut out the galactic plane and also the North Polar Spur. By applying this mask to the map I can generate another set of histograms in the same manner as before, which are also shown in Figure [5.9]. In this case, the pattern that was seen before, with a large peak at $\alpha = 0^\circ$ caused by the North Polar Spur for low frequencies (particularly K band), is no longer present. However, the histograms now show clear peaks at around $\alpha = -20^\circ$ and $\alpha = 70^\circ$, which appear to exist at all frequencies. If I calculate the $\chi^2$ as before, I find that the $\chi^2$ values have reduced substantially, but are still mostly far greater than would be expected by chance. In this instance, I find that for the Q and V bands, the reduced $\chi^2$ is around 6, while the K and Ka bands are even worse. For the W band it is now 1.5, and the probability of obtaining a $\chi^2$ greater than this by chance is around 0.07. However, whilst this is not low enough to reject the null hypothesis outright, the clear indication of structure in the histogram and the correlation with the other frequencies, coupled with the relatively low probability (only around 7%) of this result occurring by chance, is a strong indication that it is not, in fact, purely random.

Most likely, this is due to foreground contamination. Whilst foregrounds are not expected to have any preferred direction, they are mainly caused by galactic emission,
Figure 5.9: Comparison of $\hat{n}$ vs. $\alpha$ for the simulated and observed WMAP single-frequency maps. Note the peaks at $\alpha = -20^\circ$ and $70^\circ$ in the masked maps, which appear to exist at all frequencies.
which is mainly large-scale and, thus, is likely to have high levels of correlation in the position angles between pixels. As a result, the real sky is substantially more correlated in the presence of foregrounds than the simulations are, and so the error in the histogram bins is larger. Therefore, the presence of substantial foreground contamination will make the chi-squared considerably larger than would be expected by chance. Further, since the foregrounds are likely to be present at all frequencies (albeit with varying strengths), this explains the apparent correlation between the peaks at different frequencies. Further, the peaks seem to become clearer as the frequency increases (V and W bands), which suggests that they are, in fact, caused by large areas of foreground contamination with position angles around $-20^\circ$ and $70^\circ$ which emit more strongly at the higher frequencies, but which are also active at low frequencies.

### 5.4.4 WMAP with foreground removal

As can be seen from the previous section, the standard WMAP maps still have substantial foreground contamination, hence the need for foreground-reduction techniques. I have applied the same approach to the foreground-reduced WMAP maps in order to see if these maps are consistent with the predictions from the noise and CMB maps, and hence whether these maps can be considered to be clean, in the sense that they have no residual foreground contamination. I have produced histograms in a similar way to in the previous section, using the foreground-reduced maps. However, there are no WMAP K and Ka band foreground-reduced maps, because these maps are used to produce the foreground templates, and have sufficiently high foreground contamination that any reduced maps that were constructed would not be considered to be reliable. Therefore, here I have only used 3 bands, the Q, V and W bands. The resultant histograms are shown in Figure 5.10.

To determine if the histograms are consistent with chance, the $\chi^2$ values must be used. These are given in Table 5.1. There is still a small level of residual foreground contamination detectable in the unmasked, foreground-reduced map, with the reduced
\( \chi^2 \) being between 2 and 3.5 for the three bands, showing that the noise and CMB signal alone cannot explain the observed map. This is not unexpected, since the foreground-reduction algorithm is not expected to work very well in heavily contaminated areas, due to the foregrounds there being many times larger than the CMB signal. However, when I look at the masked, foreground-reduced values, it appears that for two of the three bands, the \( \chi^2 \) is consistent with chance. However, the \( \chi^2 \) for the V band, even with masking, does not seem to be consistent with the simulated data fully explaining the observations. Further, the histograms look visually as if there is some correlation between the values in each bin. This is particularly true in the Q band map, and could be due to correlated noise (which has not been accounted for in the simulations). However, it is difficult to explain the high V-band \( \chi^2 \). If it were due to foregrounds, then
5.5: EFFECT OF SYSTEMATIC ERRORS ON MEASUREMENTS OF B-MODES

<table>
<thead>
<tr>
<th>Band</th>
<th>Unmasked reduced $\chi^2$</th>
<th>$P(\chi^2 &gt; \chi^2_{\text{obs}})$</th>
<th>Masked reduced $\chi^2$</th>
<th>$P(\chi^2 &gt; \chi^2_{\text{obs}})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>2.24</td>
<td>0.0014</td>
<td>1.10</td>
<td>0.35</td>
</tr>
<tr>
<td>V</td>
<td>2.27</td>
<td>0.0012</td>
<td>2.85</td>
<td>$3 \times 10^{-5}$</td>
</tr>
<tr>
<td>W</td>
<td>3.43</td>
<td>$6 \times 10^{-7}$</td>
<td>0.59</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Table 5.1: Table of reduced $\chi^2$ values for the difference between the five-year WMAP foreground-reduced maps and simulations of the noise and CMB signal expected in these maps.

the effect would be expected to be strongest in the Q-band map, which has the largest foreground signal, as opposed to the V-band, which in fact has the lowest level of foreground contamination. However, foregrounds cannot entirely be ruled out, since the effect may depend on the foreground-removal algorithm. One possibility is that it is in fact due to correlated noise, since correlations between pixels reduce the effective number of pixels measured, and thus increase the likelihood of getting more extreme results. This could plausibly also explain the rather low W-band value.

5.5 Effect of systematic errors on measurements of B-modes

In order to consider the usefulness of our method, I need to know how these systematic errors affect B-mode measurements. One way to look at this is to consider the r.m.s. change in $\alpha$ in two different situations, as shown in Figure 5.11. Firstly, I have looked at the change between a pure-E mode CMB realisation, and the same realisation but with B-modes added, using simulations with $N_{\text{side}} = 512$. I found that there is a simple formula connecting $\sqrt{\langle (\Delta \alpha)^2 \rangle}$ with $r$. It is also possible to compare a CMB realisation both with and without systematic errors, and I have found an approximate formula for this situation. Although I have only considered $\epsilon_{\text{QQ}}$ here, these results will apply equally to the other terms in the Muller matrix.
Figure 5.11: RMS difference in $\alpha$ vs. $r$ (left) and $\epsilon_{QQ}$ (right). The red lines are the simulated values, and the green lines (not visible on the left hand plot due to absolute agreement with the simulations) are the formulae given in Equations (5.28) and (5.29), respectively.

Using Equation (5.11), together with these simulations, I can find empirical relations giving $\sqrt{\langle \Delta \alpha^2 \rangle}$ as a function of $r$ and $\epsilon_{QQ}$ (the formula used is marked on the second graph with a green line). These formulae are

$$\sqrt{\langle \Delta \alpha^2 \rangle} \approx 10 r^{1/4}, \quad (5.28)$$

$$\sqrt{\langle \Delta \alpha^2 \rangle} \approx 10 |\epsilon_{QQ}|. \quad (5.29)$$

From these, it can be seen that a value of $\epsilon_{QQ} = 0.1$ causes an equivalent change in the RMS of $\alpha$ to the addition of B-modes with $r = 10^{-4}$.

A further possible test is to see what effect adding systematics to a map has on the E and B-mode spectra of those maps. As an indication of how this would affect the B-mode spectrum, I have plotted the spectra obtained from setting $\epsilon_1 = 0.1$ and $\epsilon_2 = 0.1$ in Figure 5.12. As can be seen from this, adding contamination at this level will totally swamp the BB spectrum due to leakage from the E-mode spectrum caused by...
5.5: EFFECT OF SYSTEMATIC ERRORS ON MEASUREMENTS OF B-MODES

Figure 5.12: Effect on E and B mode spectra of addition of systematic errors of $\epsilon_1 = 0.1$ (left) and $\epsilon_2 = 0.1$ (right). The red line is the input spectrum and the black points are the output spectrum, with the error bars derived from the simulations. Note how the EE spectrum is unchanged and the output BB spectrum looks like a scaled-down version of the input EE spectrum.

the systematic errors.
5.6 Chi-squared for correlated random variables

As briefly mentioned in Section 5.3, in order to properly understand the level to which systematic errors could be detected using the polarisation position angle method, the properties of the $\chi^2$ test in the case where the data points are correlated need to be understood. There are two cases to consider. The first is the case where the only correlation comes from the fact that histogram bins are intrinsically correlated due to the sum of the number of pixels in the bins being fixed, and the second is the case where there are additional inter-pixel correlations.

Here, I will begin by outlining the basic method in the case of the ordinary $\chi^2$ distribution, and derive an integral which allows the probability distribution for the $\chi^2$ statistic to be computed in the case of correlated bins. Then, I will consider how to solve this integral in certain cases. Finally, I will extend this approach to the case of the non-central $\chi^2$ distribution.

5.6.1 Ordinary $\chi^2$ distribution

Now, suppose there are a set of variables $x_i$ which follow a multivariate Gaussian distribution, with a covariance matrix given by $\text{Cov}(x_i, x_j) = \sigma_{ij}$. The aim is to find the probability distribution of the $\chi^2$ function

$$\chi^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{x}_i)^2}{\sigma_i^2},$$

(5.30)

where in this case $\sigma_i = \sigma_{ii} = \text{variance of variable } i$. Calculating the probability distribution of this directly is difficult, because the variables are not independent, so the usual formula for combining probabilities does not apply. However, suppose I define

$$\frac{x_i - \bar{x}_i}{\sigma_i} = \sum_{j=1}^{N} A_{ij} z_j$$

(5.31)

such that $\text{Cov}(z_i, z_j) = \delta_{ij}$. Then, $\chi^2$ can be written as

$$\chi^2 = \sum_{i,j} A_{ij} A_{jk} z_j z_k,$$

(5.32)
5.6: CHI-SQUARED FOR CORRELATED RANDOM VARIABLES

where the \( z_i \) are independent random variables. Now, from the definition of \( z_i \), I can show that

\[
\text{Cov}\left( x_i \sigma_i, x_j \sigma_j \right) = \sum_{k,l} A_{ik} A_{jl} \text{Cov}(z_k, z_\ell),
\]

(5.33)

and, since \( \text{Cov}(x_i \sigma_i, x_j \sigma_j) = \text{Corr}(x_i, x_j) \), and I require \( \text{Cov}(z_i, z_j) = \delta_{ij} \), I get that

\[
\sum_k A_{ik} A_{jk} = \sigma'_{ij},
\]

(5.34)

where I have defined \( \sigma'_{ij} = \text{Corr}(x_i, x_j) \). Now, the matrix \( \sigma' \) can be diagonalised by finding an orthogonal matrix \( P \) such that

\[
P^T \sigma' P = \Lambda,
\]

(5.35)

where \( \Lambda_{ij} = \lambda^{(i)} \delta_{ij} \), where the \( \lambda^{(i)} \) are the eigenvalues of the correlation matrix. Thus, the matrix \( P \) must satisfy the equation

\[
\sum_j \sigma'_{ij} P_{jk} = \lambda^{(k)} P_{ik}.
\]

(5.36)

Now, if I set

\[
A_{ij} = \sqrt{\lambda^{(i)} P_{ij}},
\]

(5.37)

then I find that \( A \) is a solution to Equation (5.34). Now, writing Equation (5.32) in terms of this, I get

\[
\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sqrt{\lambda^{(i)} \lambda^{(k)}} P_{ij} P_{ik} z_j z_k.
\]

(5.38)

However, \( P \) is orthogonal, so

\[
\chi^2 = \sum_{i=1}^{N} \lambda^{(i)} z_i^2,
\]

(5.39)

where the \( z_i \) are independent Gaussian random variables, and the \( \lambda^{(i)} \) values are the eigenvalues of the correlation matrix. However, the \( z_i \) are ill-defined in the case where \( \lambda^{(i)} = 0 \). In this situation, I can get around the issue by defining

\[
\frac{x_i - \bar{x}_i}{\sigma_i} = \sum_{j=1}^{N} P_{ij} t_j.
\]

(5.40)
In this case, I have that Cov($t_i, t_j$) = $\Lambda_{ij}$, with $P_{ij}$ defined as before. From the covariance matrix definition, it can be seen that, if $\lambda_k = 0$, then Var($t_k$) = 0, and thus $t_k = \text{const}$. Since $\langle x_i \rangle = 0$ for all $i$, I can easily show that $t_k = 0$, and thus this technicality is not exceptionally important here. However, it will become relevant when discussing the non-central chi-squared distribution later.

Of note is the fact that, for a zero eigenvalue to occur, the random variables must obey a condition of the form

$$\sum_{i=1}^{N} a_i \frac{x_i - \bar{x}_i}{\sigma_i} = 0,$$

which implies that exactly zero eigenvalues can only occur when there exists some linear combination of the $x_i$ that is constrained exactly by the definition of the random variables. Note that, if this is true, then $P_{ij} = a_i$ (with $\lambda^{(j)} = 0$), and that if there exist $k$ linearly independent sets of $a_i$ for which this expression holds true, then there will be $k$ zero eigenvalues. In particular, in the case of the histograms used in this chapter,

$$\sum_{i=1}^{N} \bar{y}_i = 0,$$

and, thus, these histograms will have precisely one zero eigenvalue.

Now, it is necessary to consider how to determine the probability distribution of the $\chi^2$ for this general case. To do this, I first define $y_i = \lambda_i z_i^2$. Then, since $z_i$ follows a standard Gaussian distribution, I can show that

$$\mathcal{P}(y_i) = \frac{e^{-\frac{y_i}{2\lambda_i}}}{\sqrt{2\pi\lambda_i y_i}}.$$

I can then define

$$y = \sum_{i=1}^{N} y_i,$$

and the probability distribution of this can be found using the formula

$$\mathcal{P}(Y = y) = \int_{y_1=-\infty}^{\infty} \cdots \int_{y_N=-\infty}^{\infty} \mathcal{P}\left( X_N = y - \sum_{i=1}^{N-1} y_i \right) \prod_{i=1}^{N-1} \mathcal{P}(Y_i = y_i) \ dy_i.$$

178

ANALYSIS OF CMB POLARISATION
Using the above formula, it is found that

\[
P(\chi^2 = y) = \frac{e^{-\frac{y}{2V}}}{(2\pi)^N/2 \prod_{i=1}^{N} \sqrt{\lambda_i}} \int_{y_{y_i=0}}^{y} \int_{y_{y_j=0}}^{y} \cdots \frac{\prod_{i=1}^{N-1} e^{-\frac{y_{(i+1)}}{2}}}{\prod_{i=1}^{N} \sqrt{y - \sum_{i=1}^{N-1} y_i \sqrt{k \prod_{i=1}^{N-1} y_i}}} dy_i \quad (5.46)
\]

### 5.6.2 Case of equal eigenvalues

Now, suppose that \( k \) of the eigenvalues are zero, and the remaining \( N - k \) all take the same value, \( \lambda \). Then, the integrals involving \( y_i \) which correspond to zero eigenvalues become \( \delta \)-functions and drop out, and after relabelling I am left with

\[
P(\chi^2 = y) = \frac{e^{-\frac{y}{2V(N-k)}}}{(2\pi)^{(N-k)/2} \prod_{i=1}^{N-k} \sqrt{\lambda_i}} \int_{y_{y_i=0}}^{y} \int_{y_{y_j=0}}^{y} \cdots \frac{\prod_{i=1}^{N-k-1} e^{-\frac{y_{(i+1)}}{2}}}{\prod_{i=1}^{N-k} \sqrt{y - \sum_{i=1}^{N-k-1} y_i \sqrt{k \prod_{i=1}^{N-k-1} y_i}}} dy_i \quad (5.47)
\]

However, all the eigenvalues are now the same, and so the terms cancel to give

\[
P(\chi^2 = y) = \frac{e^{-\frac{y}{2V(N-k)}}}{(2\pi)^{(N-k)/2} \sqrt{\lambda(N-k)}} \int_{y_{y_i=0}}^{y} \int_{y_{y_j=0}}^{y} \cdots \frac{\prod_{i=1}^{N-k-1} dy_i}{\prod_{i=1}^{N-k} \sqrt{y - \sum_{i=1}^{N-k-1} y_i \sqrt{k \prod_{i=1}^{N-k-1} y_i}}} \quad (5.48)
\]

To evaluate this, I need integrals of the form

\[
\int_{0}^{a} \frac{(a-y)^n}{\sqrt{y}} \, dy \quad (5.49)
\]

Substituting \( y = a \cos^2 \theta \), I find that

\[
\int_{0}^{a} \frac{(a-y)^n}{\sqrt{y}} \, dy = 2a^{n+\frac{1}{2}} \int_{0}^{\frac{\pi}{2}} \sin^{2n+1} \theta \, d\theta \quad (5.50)
\]

which can then be integrated to get

\[
\int_{0}^{a} \frac{(a-y)^n}{\sqrt{y}} \, dy = a^{n+\frac{1}{2}} \frac{(2n)!!}{(2n+1)!!} \times \left\{ \begin{array}{cl} 1 & n \in \mathbb{N}_0 \\ \frac{\pi}{2} & n + \frac{1}{2} \in \mathbb{N}_0 \end{array} \right. \quad (5.51)
\]

Now, taking the integral over \( y_{N-k-1} \) from \((5.48)\), by setting \( a = y - \sum_{j=1}^{N-k-1} y_j \), I get

\[
\int_{y_{y_i=0}}^{y_{y_{y_{(i+1)}}=0}} \frac{dy_{N-k-1}}{\sqrt{(y - \sum_{i=1}^{N-k-2} y_i - y_{N-k-1}) \sqrt{k(N-k-1)}}} = \int_{y_{y_i=0}}^{a} \frac{dy_{N-k-1}}{\sqrt{(a - y_{N-k-1}) \sqrt{k(N-k-1)}}} = \frac{\pi}{2} \quad (5.52)
\]
5: NULL TEST OF CMB POLARISATION DATA USING POSITION ANGLES

and thus that

\[
\frac{e^{-\frac{x^2}{2}}}{(2\pi \lambda)^{(N-k)/2}} \int_{y_1=0}^{y} \cdots \int_{y_{N-k-1}=0}^{y} \frac{\prod_{i=1}^{N-k-1} dy_i}{\sqrt{(y - \sum_{i=1}^{N-k-1} y_i)}}^N 
\]

\[
= \frac{\pi e^{-\frac{x^2}{2}}}{2(2\pi \lambda)^{(N-k)/2}} \int_{y_1=0}^{y} \cdots \int_{y_{N-k-1}=0}^{y} \frac{\prod_{i=1}^{N-k-1} dy_i}{\sqrt{\prod_{i=1}^{N-k-1} y_i}}. 
\]

In general, I can show, using Equation (5.49), that

\[
\frac{e^{-\frac{x^2}{2}}}{(2\pi \lambda)^{(N-k)/2}} \int_{y_1=0}^{y} \cdots \int_{y_{N-k-1}=0}^{y} \left(y - \sum_{i=1}^{n-1} y_i\right)^m \frac{\prod_{i=1}^{n-1} dy_i}{\sqrt{\prod_{i=1}^{n-1} y_i}} 
\]

\[
= \frac{k_m(2m)!}{(2m + 1)!! (2\pi \lambda)^{(N-k)/2}} \int_{y_1=0}^{y} \cdots \int_{y_{N-k-1}=0}^{y} \left(y - \sum_{i=1}^{n-2} y_i\right)^{m+\frac{1}{2}} \frac{\prod_{i=1}^{n-2} dy_i}{\sqrt{\prod_{i=1}^{n-2} y_i}}, 
\]

where \(k_m = 1\) if \(m \in \mathbb{N}\), and \(k_m = \frac{n}{2}\) if \(m + \frac{1}{2} \in \mathbb{N}\). From this, I can therefore find that

\[
P(\chi^2 = y) = \frac{e^{-\frac{y}{2}}}{(2\pi \lambda)^{(N-k)/2}} \prod_{i=0}^{N-k-2} k_i/2!! (i + 1)!! = \frac{\frac{y^{N-k-2}}{2} e^{-\frac{y}{2}}}{(N-k-1)!! (2\pi \lambda)^{(N-k)/2}} \left(\frac{\pi}{2}\right)^{(N-k)/2}, 
\]

which, when simplified, gives the correct expression for the chi-squared

\[
P(\chi^2 = y) = \frac{y^{N-k-2}}{(2\pi \lambda)^{(N-k)/2}} \frac{e^{-\frac{y}{2}}}{\Gamma\left(\frac{N-k}{2}\right)}, 
\]

which means that the quantity

\[
\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - \bar{y}_i}{\lambda \sigma_i^2}\right)^2 
\]

follows a chi-squared distribution with \(N - k\) degrees of freedom. Furthermore, the sum of the eigenvalues is given by the trace of the correlation matrix, which must be \(N\) (since the diagonal elements are 1 by definition), and so

\[
\sum_{i=1}^{N} \lambda_i = (N - k) \lambda = N, 
\]

and thus \(\lambda = N/(N - k)\). From this, it is found that the variable \(\chi^2\), if defined as

\[
\chi^2 = \sum_{i=1}^{N} \frac{N - k}{N} \left(\frac{x_i - \bar{x}_i}{\sigma_i^2}\right)^2, 
\]
has a $\chi^2$ distribution with $N - k$ degrees of freedom. This allows the standard $\chi^2$-test for a histogram where the errors are not correlated between each measurement to be derived, which is that

$$\chi^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{x}_i)^2}{\bar{x}_i}.$$  

(5.60)

follows a chi-squared distribution with $N - 1$ degrees of freedom.

Note that this only applies when the errors are all the same (i.e. when $\bar{x}_i$ is the same for all values of $i$). If this is not the case, the difference between the expressions given in Equations (5.59) and (5.60) can be considered. The mean difference is zero, but the distributions are not quite the same. However, provided that the difference in errors is relatively small (which it usually will be) this is unlikely to be a problem.

5.6.3 Solution in general case

In general, however, the result is considerably more complex. In particular, in the case where the pixels are correlated, the assumption that all but one of the eigenvalues are equal will no longer hold. In this case, the integral (5.46) seemingly becomes intractable and, thus, the general solution can probably only be found numerically. However, in certain cases approximations that show the likely effect of inter-pixel correlations on the method can be found.

Two-variable case

In particular, in the case where there are only two non-zero eigenvalues, the integral (5.46) becomes

$$P(\chi^2 = y) = \frac{e^{-\frac{y}{\lambda_2}}}{2\pi \sqrt{\lambda_1 \lambda_2}} \int_0^\infty \frac{e^{-\frac{y_1}{\lambda_1}}}{\sqrt{(y - y_1)y_1}} \, dy_1.$$  

(5.61)

This can again be solved using the substitution $y_1 = y \cos^2 \theta$, which simplifies the above equation to

$$P(\chi^2 = y) = \frac{e^{-\frac{y}{\lambda_2}}}{\pi \sqrt{\lambda_1 \lambda_2}} \int_0^{2} e^{-\frac{\cos^2 \theta}{\lambda_1 \lambda_2}} \, d\theta.$$  

(5.62)
Using the identity $\cos 2\theta = 2\cos^2 \theta - 1$, this can be written as
\[
\mathcal{P}(\chi^2 = y) = e^{-\frac{y}{\lambda_1\lambda_2}} e^{\frac{i}{2\lambda_1\lambda_2}} \int_0^{2\pi} e^{-\frac{y}{2\lambda_1\lambda_2} \cos^2 \left(\frac{\theta}{2}\right)} d\theta. \tag{5.63}
\]

Now, it can be shown that
\[
\int_0^{2\pi} e^{y\cos \theta} d\theta = 2\pi I_0(y), \tag{5.64}
\]
where $I_0$ is a modified Bessel function of the first kind, and Equation (5.63) can be manipulated into this form by first using the fact that
\[
\int_0^{2\pi} e^{y\cos^2 \theta} d\theta = \frac{1}{2} \int_0^{\pi} e^{y\cos^2 \theta} d\theta, \tag{5.65}
\]
using the cosine identity and then making the substitution $\phi = 2\theta$, to get that
\[
\int_0^{\pi} e^{y\cos^2 \theta} d\theta = \frac{e^y}{4} \int_0^{2\pi} e^{y\cos 2\theta} d\theta = \frac{\pi e^y}{2} I_0(y), \tag{5.66}
\]
and thus that
\[
\mathcal{P}(\chi^2 = y) = e^{-\frac{y}{2\lambda_1\lambda_2}} I_0 \left( \frac{y}{4} \left( \frac{1}{\lambda_2} - \frac{1}{\lambda_1} \right) \right). \tag{5.67}
\]

However, whilst this formula is exact, it is not particularly informative. In particular, it is difficult to see how this relates to the ordinary $\chi^2$ distribution. In order to compare the two, it is necessary to use series expansions for the Bessel function. There are two relevant expansions, depending on the particular regime.
\[
I_0(z) = \frac{e^z}{\sqrt{2\pi z}} \left( 1 + \frac{1}{8z} + \frac{9}{128z^2} + O \left( \frac{1}{z^3} \right) \right) \quad z \to \infty \tag{5.68}
\]
\[
I_0(z) = 1 + \frac{z^2}{4} + \frac{z^4}{6} + O \left( z^6 \right) \quad z \to 0 \tag{5.69}
\]

Now, Equation (5.67) shows that these two cases correspond, respectively, to the situation where $\frac{1}{\lambda_2} - \frac{1}{\lambda_1} \ll 1$ (which means that either $\lambda_1 \approx \lambda_2$, or that $\lambda_1 \gg 1$ and $\lambda_2 \gg 1$), and to the situation where $\frac{1}{\lambda_2} - \frac{1}{\lambda_1} \gg 1$ (which means that $\lambda_2 \ll 1$). To make the calculation simpler, I will assume that the eigenvalues are ordered by size, such that
the largest value is $\lambda_1$. Then, there are two cases. Now, consider the situation where
\[ \frac{1}{\lambda_2} - \frac{1}{\lambda_1} \ll 1. \]
Then,
\[ \mathcal{P}(\chi^2 = y) \approx \frac{e^{-\left(\frac{y}{2}\right)}}{2 \sqrt{\lambda_1 \lambda_2}} \left(1 + \frac{y^2}{64} \left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right)^2 + O\left(y^4 \left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right)^4\right)\right). \] (5.70)

From the definition of the chi-squared distribution, this can be written as
\[ \mathcal{P}(\chi^2 = y) \approx \frac{1}{\sqrt{\lambda_1 \lambda_2}} \left(\chi^2_0 \left(\frac{y}{2} \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right)\right) + \frac{1}{2} \left(\lambda_1 - \lambda_2\right)^2 \chi^2_6 \left(\frac{y}{2} \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right)\right) + O\left(\left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^4\right)\right), \] (5.71)
where $\chi^2_n(x)$ is the chi-squared distribution with $n$ degrees of freedom. Then, I can re-define the chi-squared in a similar manner to in Section 5.6.2, to be
\[ \chi^2 = \frac{\lambda_1 + \lambda_2}{\lambda_1 \lambda_2} \sum_{i=1}^{2} \frac{(y_i - \bar{y})^2}{\sigma_i^2}. \] (5.72)

Then, using this definition of chi-squared,
\[ \mathcal{P}(\chi^2 = y) \approx \frac{2 \sqrt{\lambda_1 \lambda_2}}{\lambda_1 + \lambda_2} \left(\chi^2_0(y) + \frac{1}{2} \left(\lambda_1 - \lambda_2\right)^2 \chi^2_6(y) + O\left(\left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^4\right)\right). \] (5.73)

Since it is a probability distribution, if the calculation were exact, then the integral over the domain of the function would have to be 1. However, since it is not exact, this will not necessarily hold. Integrating the function will give some idea of the relative importance of the different terms, and thus the situations under which the approximation is valid. Using the fact that the chi-squared distribution is a probability distribution (and thus integrates to 1), the integrals are trivial, and I get that
\[ \int_0^{\infty} \mathcal{P}(\chi^2 = y) dy = \frac{2 \sqrt{\lambda_1 \lambda_2}}{\lambda_1 + \lambda_2} \left(1 + \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2 + O\left(\left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^4\right)\right), \] (5.74)
which is indeed equal to 1 to $O((\lambda_1 - \lambda_2)^2)$. Furthermore, this shows that the approximation used is valid provided that $\lambda_1 - \lambda_2 \ll \lambda_1 + \lambda_2$, which will be true iff $\lambda_1 \approx \lambda_2$. 

MICHAEL PREECE 183
**Three-variable case**

Now, it would be useful to extend the calculation done in the previous section to the case where there are more than two variables. Unfortunately, this leads to a non-trivial integral. For the case where there are three variables, the probability distribution is given by

\[
P(\chi^2 = y) = \frac{e^{-\frac{y}{12}}}{(2\pi)^{\frac{3}{2}}} \sqrt{\lambda_1 \lambda_2 \lambda_3} \int_{y_1=0}^{y} \int_{y_2=0}^{y-y_1} e^{-\frac{y}{12} \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_3} \right)} e^{-\frac{y}{12} \left( \frac{1}{\lambda_2} - \frac{1}{\lambda_3} \right)} \frac{dy_1 dy_2}{\sqrt{(y - y_1 - y_2)y_1 y_2}}.
\]  

The integral over \( y_2 \) can be carried out in a manner analogous to before, using the substitution \( y_2 = (y - y_1) \cos^2 \theta \), which gives

\[
P(\chi^2 = y) = \frac{e^{-\frac{y}{12}}}{(2\pi)^{\frac{3}{2}}} \sqrt{\lambda_1 \lambda_2 \lambda_3} \int_{y_1=0}^{y} \frac{e^{-\frac{y}{12} \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_3} \right)}}{\sqrt{y_1}} \int_{\theta=0}^{\frac{\pi}{2}} e^{-\frac{(y-y_1) \cos^2 \theta}{12} \left( \frac{1}{\lambda_2} - \frac{1}{\lambda_3} \right)} \frac{dy_1}{\sqrt{(y - y_1)^2 (1 - \cos^2 \theta) \cos^2 \theta}} \cos \theta \sin \theta d\theta d\theta,
\]  

which can then be simplified, using the relation (5.65), to get

\[
P(\chi^2 = y) = \frac{e^{-\frac{y}{12}}}{(2\pi)^{\frac{3}{2}}} \sqrt{\lambda_1 \lambda_2 \lambda_3} \int_{y_1=0}^{y} \frac{e^{-\frac{y}{12} \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_3} \right)}}{\sqrt{y_1}} \int_{\theta=0}^{\pi} e^{-\frac{(y-y_1) \cos^2 \theta}{12} \left( \frac{1}{\lambda_2} - \frac{1}{\lambda_3} \right)} \frac{dy_1 d\theta}{(y - y_1) 2 \cos \theta \sin \theta}.
\]  

The same method as used in the two-variable case can then be applied to the integral over \( \theta \), leaving

\[
P(\chi^2 = y) = \frac{e^{-\frac{y}{12} \left( \frac{1}{\lambda_1} + \frac{1}{\lambda_3} \right)}}{2 \sqrt{\lambda_1 \lambda_2 \lambda_3}} \int_{y_1=0}^{y} \frac{e^{-\frac{y}{12} \left( \frac{1}{\lambda_1} + \frac{1}{\lambda_3} \right)}}{\sqrt{y_1}} I_0 \left( \frac{(y-y_1)}{4 \left( \frac{1}{\lambda_2} - \frac{1}{\lambda_3} \right)} \right) dy_1.
\]  

At this point, there is a problem. The second integral is considerably more difficult than the first, since it involves integrating a combination of a Bessel function, an exponential and a square root. The substitution \( y_1 = y \cos^2 \theta \) can still be applied, but it leaves a \( \sin^2 \theta \) function inside the Bessel function, with no obvious way to integrate the result. Thus, it is not possible to proceed further in this case, except by the use of approximations.
5.6.4 Alternate method for deriving formulae

For the more general case, the calculation can be somewhat simplified by using the characteristic function (denoted by $\phi$) of the probability distribution, which is defined as the Fourier transform of the probability distribution, such that

$$
P(\chi^2 = x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \phi_{\chi^2}(t) \, dt.
$$

(5.79)

Since this is a Fourier transform, the convolution theorem applies to it, resulting in the very useful result that, if I define a random variable

$$
X = \sum_{k=1}^{N} a_i x_i,
$$

(5.80)

that the characteristic function is given by

$$
\phi_X(t) = \prod_{k=1}^{N} \phi_{X_i}(a_i t).
$$

(5.81)

The characteristic function for a chi-squared distribution is given by

$$
\phi_{\chi^2_k} = (1 - 2it)^{-\frac{k}{2}}
$$

(5.82)

and, thus, defining the chi-squared as in Equation (5.39), and using the summation and inversion formulae, it is found that

$$
P(\chi^2 = x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \prod_{k=1}^{N} (1 - 2i\lambda_k t)^{-\frac{1}{2}} \, dt.
$$

(5.83)

This formula can, in principle, be integrated (either analytically or numerically) to give a probability distribution for the chi-squared variable as it is defined here.

5.6.5 Extension of method to non-central chi-squared distribution

In order to determine the level to which systematic errors can be detected, an analog of the non-central $\chi^2$ distribution is required. Thus, the calculations from the previous
parts of this section must be extended to the case where the mean value of \( x_i \) is not equal to the assumed value \( \mu_i \). To do this, I must consider the expression

\[
\chi^2 = \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2},
\]

(5.84)

where, in general, \( \langle x_i \rangle \neq 0 \). As in the case of the ordinary \( \chi^2 \) distribution considered in Section [5.6.1](#) it is necessary to find the probability distribution of this in the case where the set of \( x_i \) are correlated Gaussian random variables with a correlation matrix \( \sigma'_{ij} = \text{Corr}(x_i, x_j) \). Defining

\[
\frac{x_i}{\sigma_i} = \sum_{i=1}^{N} A_{ij} z_j,
\]

(5.85)

then it can easily be shown that the calculation carried out previously does not depend on the mean of the summation terms and, thus, that the non-central \( \chi^2 \) distribution is, like the ordinary version considered in Section [5.6.1](#) given by

\[
\chi^2 = \sum_{i=1}^{N} \lambda^0_i z_i^2,
\]

(5.86)

where the \( z_i \) are again uncorrelated Gaussian random variables. However, since the means of the \( x_i \) are no longer zero, \( \langle z_i \rangle \neq 0 \). This, in order to compute the probability distribution, I must find an expression for \( \langle z_i \rangle \).

Inverting the definition of the \( z_i \), and using the expression for \( A \) I find that

\[
\sqrt{\lambda^0}(z_i) = \sum_j P_{ji} \left( \frac{x_j}{\sigma_j} \right).
\]

(5.87)

Now, if I suppose that all the eigenvalues are non-zero, then this can be re-arranged to get

\[
z_i = \frac{1}{\sqrt{\lambda^0_i}} \sum_j P_{ji} \frac{\tilde{x}_j}{\sigma_j}.
\]

(5.88)

However, suppose that there exist zero eigenvalues. Then the \( z_i \) will be ill-defined, and I must again define \( t_i = \sqrt{\lambda^0_i} z_i \). Similarly to Section [5.6.1](#) it is found that \( t_i = \text{const} \). In this case, though, the constant is no longer zero. From the definition of \( t_i \),

\[
\langle t_i \rangle = \sum_j P_{ji} \left( \frac{x_j}{\sigma_j} \right).
\]

(5.89)
and since \( t_i = \text{const, } t_i \equiv \langle t_i \rangle \). Further, it can also be seen that

\[
\lambda_i = 0 \Leftrightarrow \sum_{k=1}^{N} P_{ki} \frac{x_k}{\sigma_k} \equiv c , \quad (5.90)
\]

and thus that, once again, zero eigenvalues occur if there exist constraints on some linear combination of the random variables being measured.

Now, I wish to find an expression for the probability distribution for the quantity defined by Equation (5.84). The probability distribution for \( z_i \) is given by

\[
P(z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z_i - \bar{z}_i)^2}{2}} . \quad (5.91)
\]

If I define

\[
y_i = \lambda_i z_i^2 \quad (5.92)
\]

\[
y = \sum_{i=1}^{N} y_i , \quad (5.93)
\]

in a similar manner to Equation (5.44), then this formula can be rearranged. There are two different values of \( z_i \) which lead to the same value of \( y_i \) \((\pm \sqrt{y_i/\lambda_i})\), and thus

\[
P(y_i) = \frac{1}{\sqrt{2\pi}} \frac{\left( e^{\frac{-(z_i - \bar{z}_i)^2}{2}} + e^{\frac{-(z_i - \bar{z}_i)^2}{2}} \right)}{2\lambda_i z_i} , \quad (5.94)
\]

which can be simplified to get that

\[
P(y_i) = \frac{1}{\sqrt{2\pi \lambda_i y_i}} e^{\frac{-(y_i - \bar{z}_i)^2}{2\lambda_i}} \cosh \left( \frac{\bar{z}_i}{2} \sqrt{\frac{y_i}{\lambda_i}} \right) . \quad (5.95)
\]

Note that if an eigenvalue is zero, I simply get a delta-function here, with \( y_i = \lambda_i \frac{z_i^2}{\lambda_i} (= \bar{t}_i^2) \).

**Calculation through characteristic functions**

Now, consider the more general situation, where there are more than two variables. The characteristic function of the non-central chi-squared distribution is given by (Patnaik 1949)

\[
\varphi_{\chi^2_k}(\Lambda, t) = \frac{e^{i\langle \Lambda \rangle}}{(1 - 2it)^{k/2}} . \quad (5.96)
\]

MICHAEL PREECE
where the parameter $\Lambda$ (equivalent to the $\lambda$ defined in Equation (5.20)) is defined as

$$\Lambda = \sum_{i=1}^{k} \left( \frac{\bar{x}_i}{\sigma_i} \right)^2. \quad (5.97)$$

Now, suppose I define $y = \sum_i y_i$, with $y_i = \lambda_i z_i^2$ as before. Then,

$$\varphi_{Y_i}(t) = \frac{e^{-\frac{\bar{y}^2 t}{2}}}{(1 - 2i\lambda t)^{\frac{1}{2}}}, \quad (5.98)$$

and, using Equation (5.81), I thus get that

$$\varphi_{Y}(t) = \frac{\exp\left(\sum_{k=1}^{N} \frac{\bar{y}_k^2 t}{1 - 2i\lambda_k t} \right)}{\prod_{k=1}^{N} (1 - 2i\lambda_k t)^{\frac{1}{2}}}, \quad (5.99)$$

where

$$\bar{y}_i = \sum_{j=k+1}^{N} P_{ji} P_{kj} \frac{\bar{x}_j \bar{x}_k}{\sigma_j \sigma_k}. \quad (5.100)$$

In general, the inverse transformation can be applied to the characteristic function, and thus an expression for the probability distribution can be found. Unfortunately, this is very unlikely to be integrable, in the general case.

However, in case looked at in Section 5.6.2, where there are $N - k$ equal eigenvalues (with $k$ being zero), progress can be made. In this case,

$$\varphi_{Y}(t) = \frac{\exp\left(\frac{it}{1 - 2i\lambda t} \sum_{j=k+1}^{N} \bar{y}_j^2 \right)}{(1 - 2i\lambda t)^{\frac{N-k}{2}}} \times \exp\left(it \sum_{j=k+1}^{N} \bar{y}_j^2 \right). \quad (5.101)$$

This can be split into two parts. The second part will, as discussed earlier, lead to a delta function probability distribution, and results in a shift in the probability distribution of $\sum_{j=k+1}^{N} \bar{y}_j^2$. The first part is equal to the formula given in Equation (5.96), with $t \to \lambda t$ and the parameter $\Lambda = \sum_{j=1}^{N} \bar{z}_j^2$. Once again, $\lambda = N/(N - k)$. Thus, if I define a new chi-squared value (by analogy with Equation (5.59)) as

$$\chi^2 = \sum_{i=1}^{N} \frac{N - k}{N} \frac{x_i^2}{\sigma_i^2}, \quad (5.102)$$

then I get that

$$P(\chi^2 = y) = \chi_{N-k}^2 \left( y - \sum_{j=k+1}^{N} \bar{y}_j^2 \right), \quad (5.103)$$
where $\chi_{N-k}^2(y; \Lambda)$ is a non-central chi-squared distribution with $N - k$ degrees of freedom and a non-centrality parameter given by $\Lambda$, applied to the random variable $y$.

Using the definition of the $z_i$ in terms of the $x_i$, an expression for $\Lambda$ can be found. I get that

$$\Lambda = \sum_{i=1}^{N-k} \left( \sum_{j=1}^{N} \frac{1}{\sqrt{\lambda_i}} (x_j) P_{ji} \right)^2$$

(5.104)

$$= \frac{N}{N} \sum_{j=1}^{N} \sum_{\ell=1}^{N-k} \sum_{i=1}^{N} \frac{P_{ji} P_{li}}{\lambda_i} .$$

(5.105)

Now, $P_{ij}$ is an orthogonal matrix, so, using the fact that $\lambda_i = N/(N - k)$ (for $i \leq N - k$), an expression for $\Lambda$ can be found

$$\Lambda = \frac{N - k}{N} \sum_{j=1}^{N} \frac{x_j \bar{x}_\ell}{\sigma_j \sigma_\ell} \left( \delta_{ij} - \sum_{i=N-k+1}^{N} P_{ji} P_{li} \right)$$

(5.106)

$$= \frac{N - k}{N} \sum_{j=1}^{N} \left( \frac{x_j}{\sigma_j} \right)^2 - \frac{N - k}{N} \sum_{i=N-k+1}^{N} \left( \sum_{j=1}^{N} P_{ji} \frac{x_j}{\sigma_j} \right)^2 .$$

(5.107)

Thus, it is sufficient, in order to calculate $\Lambda$, to know the eigenvectors corresponding to the zero eigenvalues. Furthermore, it is shown earlier that, in order to have a zero eigenvalue, it is necessary (and sufficient) to have

$$\sum_{j=1}^{N} P_{ji} \frac{x_j}{\sigma_j} \equiv c_i,$$

(5.108)

for some constant $c_i$, which will be defined beforehand (and can usually be set to zero, by sensible definition of the $x_i$). Therefore,

$$\Lambda = \frac{N - k}{N} \left( \sum_{j=1}^{N} \left( \frac{x_j}{\sigma_j} \right)^2 - \sum_{i=1}^{k} c_i^2 \right),$$

(5.109)

where

$$\sum_{i=1}^{N} a_i^{(j)} x_i \equiv c_j,$$

(5.110)

for $k$ linearly independent sets of constants $a_i^{(j)}$ (given by $a_i^{(j)} = \frac{P_{ji}}{\sigma_i}$) and for all possible values of the $x_i$. 
A useful specific case of this is, once again, the case of a histogram with errors not correlated between each measurement. Here, there is only one constraint, that

$$\sum_{i=1}^{N} x_i = c,$$  \hspace{1cm} (5.111)

and it is also the case that $\sigma_i = \sqrt{N-1}$. I can then show that the quantity

$$\chi^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{\langle x_i \rangle},$$  \hspace{1cm} (5.112)

(where $\bar{x}_i$ is some nominal mean value ($\neq \langle x_i \rangle$, in general) defined such that $\sum_i \bar{x}_i = N_{\text{bins}}$), will follow a non-central $\chi^2$ distribution with $N - 1$ degrees of freedom and a non-centrality parameter given by

$$\lambda = \frac{1}{N} \sum_{j=1}^{N} \langle x_j \rangle^2.$$  \hspace{1cm} (5.113)

### 5.7 Mean of the sum of Bernoulli distributed variables

In this section, I will show how to derive the mean and variance of the number of pixels in a histogram bin when the probability distributions for each pixel are not identical. To do this, I will use proof by induction.

First, consider the mean number of pixels in a bin, for a given value of $N_{\text{pix}}$. This is given by

$$\langle N \rangle(N_{\text{pix}}) = \sum_{N_i=1}^{N_{\text{pix}}} N_i P_{N_{\text{pix}}}(N = N_i),$$  \hspace{1cm} (5.114)

where $P_{N_{\text{pix}}}(N = N_i)$ is the probability that there are $N_i$ pixels in the bin, given that there are $N_{\text{pix}}$ pixels in total. Now, I wish to show that

$$\langle N \rangle(N_{\text{pix}}) = \sum_{i=1}^{N_{\text{pix}}} p_i,$$  \hspace{1cm} (5.115)

where $p_i$ is the probability that pixel $i$ is in the bin. First, I can show trivially that for a single pixel ($N_{\text{pix}} = 1$), $\langle N \rangle = p_1$. Therefore, to prove Equation (5.115), it is sufficient
to show that

\[ \langle N \rangle_{\text{pix}} = \sum_{i=1}^{N_{\text{pix}}} p_i \Rightarrow \langle N \rangle_{\text{pix}} + 1 = \sum_{i=1}^{N_{\text{pix}}+1} p_i. \] (5.116)

Now, from Equation (5.114), it is found that

\[ \langle N \rangle_{\text{pix}} (N_{\text{pix}} + 1) = \sum_{N_i=1}^{N_{\text{pix}}+1} N_i \mathcal{P}_{N_{\text{pix}}+1}(N = N_i), \] (5.117)

and hence I need to find \( \mathcal{P}_{N_{\text{pix}}+1}(N = N_i) \). Now, consider the general case where the pixels are correlated. Then, if I define \( S \) to be the set of pixels index numbers for which the associated pixel is in the bin, then

\[
\mathcal{P}_{N_{\text{pix}}+1}(N = N_i) = \mathcal{P}_{N_{\text{pix}}}(N = N_i|N_{\text{pix}} + 1 \notin S) \mathcal{P}(N_{\text{pix}} + 1 \notin S) \\
+ \mathcal{P}_{N_{\text{pix}}}(N = N_i - 1|N_{\text{pix}} + 1 \in S) \mathcal{P}(N_{\text{pix}} + 1 \in S),
\] (5.118)

and it is found that

\[
\mathcal{P}_{N_{\text{pix}}}(N = N_i) = \mathcal{P}_{N_{\text{pix}}}(N = N_i|N_{\text{pix}} + 1 \notin S) \mathcal{P}(N_{\text{pix}} + 1 \notin S) \\
+ \mathcal{P}_{N_{\text{pix}}}(N = N_i|N_{\text{pix}} + 1 \in S) \mathcal{P}(N_{\text{pix}} + 1 \in S). \] (5.119)

Now using substitution and the fact that \( \mathcal{P}(N_{\text{pix}} + 1 \in S) = p_{N_{\text{pix}}+1} \), I get that

\[
\mathcal{P}_{N_{\text{pix}}+1}(N = N_i) = \mathcal{P}_{N_{\text{pix}}}(N = N_i) + p_{N_{\text{pix}}+1}(\mathcal{P}_{N_{\text{pix}}}(N = N_i - 1|N_{\text{pix}} + 1 \in S) \\
- \mathcal{P}_{N_{\text{pix}}}(N = N_i|N_{\text{pix}} + 1 \in S)). \] (5.120)

From Equations (5.117) and (5.120), I then get that

\[
\langle N \rangle_{\text{pix}} (N_{\text{pix}} + 1) = \sum_{N_i=1}^{N_{\text{pix}}+1} N_i \mathcal{P}_{N_{\text{pix}}}(N = N_i) \\
+ p_{N_{\text{pix}}+1} \left( \sum_{N_i=1}^{N_{\text{pix}}+1} N_i \mathcal{P}_{N_{\text{pix}}}(N = N_i - 1|N_{\text{pix}} + 1 \in S) \\
- \sum_{N_i=1}^{N_{\text{pix}}+1} N_i \mathcal{P}_{N_{\text{pix}}}(N = N_i|N_{\text{pix}} + 1 \in S) \right), \]

MICHAEL PREECE
which can be rewritten as

\[
\langle N \rangle (N_{\text{pix}} + 1) = \sum_{i=1}^{N_{\text{pix}}+1} p_i + p_{N_{\text{pix}}+1} \left( \sum_{i=1}^{N_{\text{pix}}+1} (N_i - 1) \mathcal{P}_{N_{\text{pix}}} (N = N_i - 1 | N_{\text{pix}} + 1 \in S) \right) + \sum_{i=1}^{N_{\text{pix}}+1} \mathcal{P}_{N_{\text{pix}}} (N = N_i - 1 | N_{\text{pix}} + 1 \in S) - \sum_{j=1}^{N_{\text{pix}}+1} N_j \mathcal{P}_{N_{\text{pix}}} (N = N_j | N_{\text{pix}} + 1 \in S), \tag{5.122}
\]

which gives

\[
\langle N \rangle (N_{\text{pix}} + 1) = \sum_{i=1}^{N_{\text{pix}}} p_i + p_{N_{\text{pix}}+1} \sum_{N_j=0}^{N_{\text{pix}}} (N_j) \mathcal{P}_{N_{\text{pix}}} (N = N_j | N_{\text{pix}} + 1 \in S) + \sum_{i=1}^{N_{\text{pix}}+1} \mathcal{P}_{N_{\text{pix}}} (N = N_i - 1 | N_{\text{pix}} + 1 \in S) - \sum_{j=1}^{N_{\text{pix}}+1} N_j \mathcal{P}_{N_{\text{pix}}} (N = N_j | N_{\text{pix}} + 1 \in S)) = \sum_{i=1}^{N_{\text{pix}}} p_i, \tag{5.123}
\]

which is what I set out to prove. Therefore, if there are a set of \(N_{\text{pix}}\) (potentially) correlated pixels where the probability of each being in a given bin is given by \(p_i\), then the mean number of pixels in each bin is given by

\[
\langle N \rangle = \sum_{i=1}^{N_{\text{pix}}} p_i. \tag{5.124}
\]

This result is extremely important since it shows that inter-pixel correlations are unimportant for the purposes of calculating the mean number of pixels in the bin. Provided that the probability distribution of \(Q\) and \(U\) for each measured pixel is known, it is possible to compute the distribution of \(\alpha\) for each pixel and then sum over all pixels. In particular this result, in conjunction with the result in Section 5.9.1 allows me to show that the probability distribution for CMB-like maps (ones which can be described solely by a power spectrum) must necessarily be flat.
5.8 Variance of number of pixels in histogram bins

In this section, I will derive an expression for the variance of the number of pixels in each histogram bin in both the uncorrelated and correlated cases. In particular, it will be shown that the variance depends only on the covariance matrix of the pixels and not on the full probability distribution. I will also prove the standard $\sqrt{N}$ formula for the uncorrelated case.

5.8.1 Formulae for variance in uncorrelated and correlated case

I will begin this section by stating the formulae for the variance. These will be proven in the next two subsections. By assuming that inter-pixel correlations are negligible, it is possible to come up with an expression for the variance of $N$, and in the case of identical pixels, a simple expression for the probability distribution can also be deduced. First consider the case of identically-distributed pixels. Then $N$ follows a binomial distribution, and

$$p(N) = \binom{N_{\text{pix}}}{N} p(\alpha_1 \leq \alpha < \alpha_2)^N (1 - p(\alpha_1 \leq \alpha < \alpha_2))^{(N_{\text{pix}} - N)}.$$  \hfill (5.125)

Thus, it can be trivially seen that $\text{Var}(N) = N_{\text{pix}} p(\alpha_1 \leq \alpha < \alpha_2) (1 - p(\alpha_1 \leq \alpha < \alpha_2))$, by standard properties of the binomial distribution. Writing this in terms of the mean number of pixels in the bin, it is found that

$$\text{Var}(N) = \langle N \rangle \frac{(N_{\text{pix}} - \langle N \rangle)}{N_{\text{pix}}},$$  \hfill (5.126)

and if there are a large number of bins (so $\langle N \rangle \ll N_{\text{pix}}$), then $\sigma_N \approx \sqrt{\langle N \rangle}$.

Now, consider the case where the pixels have different (but still uncorrelated) distributions. Then, I can show (see Section 5.8.2) that

$$\text{Var}(N) = \sum_{i=1}^{N_{\text{pix}}} p_i (1 - p_i),$$  \hfill (5.127)

where $p_i$ is the probability that the $i$th pixel is in the bin. Under the assumption that $p_i$ is small (i.e., there are a lot of bins), I can once again show that $\sigma_N \approx \sqrt{\langle N \rangle}$. 

MICHAEL PREECE 193
5.8.2 Uncorrelated case

Now, I wish to prove that the variance,

\[ \sigma^2_N(N_{\text{pix}}) = \sum_{N_i=1}^{N_{\text{pix}}} N_i^2 p_{N_{\text{pix}}}(N = N_i) - \left( \sum_{i=1}^{N_{\text{pix}}} N_i p_{N_{\text{pix}}}(N = N_i) \right)^2 \]  

(5.128)

is given by the expression

\[ \sigma^2_N(N_{\text{pix}}) = \sum_{i=1}^{N_{\text{pix}}} p_i(1 - p_i), \]  

(5.129)

in the case where the pixels are uncorrelated. To do this, I will once again use proof by induction. First, consider the case of a single pixel. Then the problem is a binomial distribution with a single trial, and the variance is given by

\[ \sigma_N = p_i(1 - p_i), \]  

(5.130)

as required. Therefore, it is sufficient to show that

\[ \sigma^2_N(N_{\text{pix}}) = \sum_{i=1}^{N_{\text{pix}}} p_i(1 - p_i) \Rightarrow \sigma^2_N(N_{\text{pix}} + 1) = \sum_{i=1}^{N_{\text{pix}}+1} p_i(1 - p_i). \]  

(5.131)

In the case of uncorrelated pixels, Equation (5.120) simplifies to

\[ p_{N_{\text{pix}}+1}(N = N_i) = p_{N_{\text{pix}}}(N = N_i) \]  

(5.132)

\[ + p_{N_{\text{pix}}+1}(p_{N_{\text{pix}}}(N = N_i - 1) - p_{N_{\text{pix}}}(N = N_i|N_{\text{pix}} + 1)), \]

and substituting this into Equation (5.128), I get

\[ \sigma^2_N(N_{\text{pix}} + 1) = \sum_{N_i=1}^{N_{\text{pix}}+1} N_i^2 (p_{N_{\text{pix}}}(N = N_i) \]  

(5.133)

\[ + p_{N_{\text{pix}}+1}(p_{N_{\text{pix}}}(N = N_i - 1) - p_{N_{\text{pix}}}(N = N_i))) \]

\[ - \left( \sum_{i=1}^{N_{\text{pix}}+1} N_i(p_{N_{\text{pix}}}(N = N_i) + p_{N_{\text{pix}}+1}(p_{N_{\text{pix}}}(N = N_i - 1) - p_{N_{\text{pix}}}(N = N_i))) \right)^2. \]
5.8: VARIANCE OF NUMBER OF PIXELS IN HISTOGRAM BINS

I can then follow the same approach as in Section 5.7 to get

\[
\sigma^2_N(N_{\text{pix}} + 1) = \sum_{N_i=1}^{N_{\text{pix}}} N_i^2 P_{N_{\text{pix}}}(N = N_i) + p_{N_{\text{pix}}+1} \sum_{N_j=1}^{N_{\text{pix}}} (((N_j + 1)^2 - N_j^2)P_{N_{\text{pix}}}(N = N_j)) \\
- \left( \sum_{N_i=1}^{N_{\text{pix}}} N_i P_{N_{\text{pix}}}(N = N_i) \right)^2 \\
- 2p_{N_{\text{pix}}+1} \left( \sum_{N_i=1}^{N_{\text{pix}}} N_i P_{N_{\text{pix}}}(N = N_i) \right) \sum_{N_j=1}^{N_{\text{pix}}} (((N_j + 1) - N_j)P_{N_{\text{pix}}}(N = N_j)) \\
+ p_{N_{\text{pix}}+1}^2 \left( \sum_{N_j=1}^{N_{\text{pix}}} (((N_j + 1) - N_j)P_{N_{\text{pix}}}(N = N_j)) \right)^2, \tag{5.134}
\]

which can be simplified to

\[
\sigma^2_N(N_{\text{pix}} + 1) = \sigma^2_N(N_{\text{pix}}) + p_{N_{\text{pix}}+1} \sum_{N_j=1}^{N_{\text{pix}}} ((2N_j + 1)P_{N_{\text{pix}}}(N = N_j)) \\
- 2p_{N_{\text{pix}}+1} \left( \sum_{N_j=1}^{N_{\text{pix}}} N_j P_{N_{\text{pix}}}(N = N_j) \right) + p_{N_{\text{pix}}+1}^2 \\
= \sigma^2_N(N_{\text{pix}}) + p_{N_{\text{pix}}+1}(1 - p_{N_{\text{pix}}+1}), \tag{5.135}
\]

which is what I needed to prove. Therefore, if there is a set of \(N_{\text{pix}}\) uncorrelated pixels where the probability of each being in a given bin is \(p_i\), then the variance of the number of pixels in each bin is given by

\[
\sigma^2_N = \sum_{i=1}^{N_{\text{pix}}} p_i(1 - p_i). \tag{5.137}
\]

5.8.3 Extension to correlated pixels

I will now consider errors in the measurement of the probability distribution in the general case where the polarisation is correlated between different pixels. First, I define a set of \(N_{\text{pix}}\) Bernoulli random variables \(B_i\), and then further define

\[
P(B_i = 0) = 1 - p_i, \tag{5.138}
\]
\[
P(B_i = 1) = p_i. \tag{5.139}
\]
Then, I define \( N = \sum_{i=1}^{N_{\text{pix}}} B_i \), and define \( S = \{ B_i \text{ such that } B_i = 1 \} \). Using this, I get that \( \mathcal{P}(B_i \in S) = p_i \). I can then further define

\[
p_{ij} = \mathcal{P}(B_i \in S \land B_j \in S).
\] (5.140)

This expression has the following properties

\[
p_{ij} = p_{ji},
\] (5.141)

\[
p_{ii} = p_i,
\] (5.142)

\[
p_{ij} = p_i p_j \iff B_i \text{ and } B_j \text{ are uncorrelated}.
\] (5.143)

Using these definitions, I find that the variance of \( N \) when the variables \( B_i \) are correlated is given by

\[
\sigma^2_{N_{\text{pix}}} = \sum_{i,j=1}^{N_{\text{pix}}} (p_{ij} - p_i p_j).
\] (5.144)

The easiest way to prove this is by induction. First, consider the case where \( N_{\text{pix}} = 1 \). Then, the only terms in the expression are \( p_{11} \) and \( p_1 \), and from Equation 5.142, \( p_{11} = p_1 \), and therefore

\[
\sigma^2_1 = p_1(1 - p_1),
\] (5.145)

which is the correct expression (since inter-pixel correlations require there to be more than one pixel). Therefore, it is sufficient to prove that

\[
\sigma^2_{N_{\text{pix}}} = \sum_{i,j=1}^{N_{\text{pix}}} (p_{ij} - p_i p_j) \Rightarrow \sigma^2_{N_{\text{pix}}+1} = \sum_{i,j=1}^{N_{\text{pix}}+1} (p_{ij} - p_i p_j).
\] (5.146)

Now,

\[
\sigma^2_{N_{\text{pix}}+1} = \sum_{i=1}^{N_{\text{pix}}+1} N_i^2 \mathcal{P}_{N_{\text{pix}}+1}(N = N_i) - \left( \sum_{i=1}^{N_{\text{pix}}+1} N_i \mathcal{P}_{N_{\text{pix}}+1}(N = N_i) \right)^2.
\] (5.147)
where $\mathcal{P}_{N_{\text{pix}}}(N = N_i)$ is the probability that $N$ takes the value $N_i$ given that there are $N_{\text{pix}}$ pixels. It can be shown (see Section 5.8.2) that

$$
\mathcal{P}_{N_{\text{pix}}+1}(N = N_i) = \mathcal{P}_{N_{\text{pix}}}(N = N_i) + p_{N_{\text{pix}}+1}(p_{N_{\text{pix}}}(N = N_i - 1|N_{\text{pix}} + 1 \in S) - \mathcal{P}_{N_{\text{pix}}}(N = N_i|N_{\text{pix}} + 1 \in S)).
$$

(5.148)

I can substitute this into Equation (5.147) and simplify to get

$$
\sigma^2_{N_{\text{pix}}+1} = \sigma^2_{N_{\text{pix}}} + p_{N_{\text{pix}}+1} \sum_{i=1}^{N_{\text{pix}}} \left( ((N_i + 1)^2 - N_i^2) \mathcal{P}_{N_{\text{pix}}}(N + N_i|B_{N_{\text{pix}}+1} \in S) \right) - 2p_{N_{\text{pix}}+1} \mathcal{P}(B_{N_{\text{pix}}+1} \in S) \sum_{i=1}^{N_{\text{pix}}} N_i \mathcal{P}_{N_{\text{pix}}}(N = N_i) - p^2_{N_{\text{pix}}+1} + 2p_{N_{\text{pix}}+1} \sum_{i=1}^{N_{\text{pix}}} N_i \left( \mathcal{P}(N = N_i|B_{N_{\text{pix}}+1} \in S) - \mathcal{P}(N = N_i) \right). 
$$

(5.149)

(5.150)

Now, from before, I have that

$$
\sum_{i=1}^{N_{\text{pix}}} N_i \mathcal{P}(N = N_i) = \sum_{i=1}^{N_{\text{pix}}} \mathcal{P}(B_i \in S),
$$

(5.151)

and this can trivially be extended to show that

$$
\sum_{i=1}^{N_{\text{pix}}} N_i \mathcal{P}(N = N_i|N_{\text{pix}}+1 \in S) = \sum_{i=1}^{N_{\text{pix}}} \mathcal{P}(B_i \in S|B_{N_{\text{pix}}+1} \in S),
$$

(5.152)

since the proof of Equation (5.151) is not affected by dependence on another pixel.

Hence, I can write that

$$
\sigma^2_{N_{\text{pix}}+1} = \sigma^2_{N_{\text{pix}}} + p_{N_{\text{pix}}+1}(1 - p_{N_{\text{pix}}+1}) + 2p_{N_{\text{pix}}+1} \sum_{i=1}^{N_{\text{pix}}} \left( \mathcal{P}(B_i \in S|N_{\text{pix}} + 1 \in S) - p_i \right),
$$

(5.153)

and, using the definition of conditional probability I can rewrite this as

$$
\sigma^2_{N_{\text{pix}}+1} = \sigma^2_{N_{\text{pix}}} + p_{N_{\text{pix}}+1}(1 - p_{N_{\text{pix}}+1}) + 2p_{N_{\text{pix}}+1} \sum_{i=1}^{N_{\text{pix}}} \left( \frac{\mathcal{P}(B_i \in S \cap N_{\text{pix}} + 1 \in S)}{p_{N_{\text{pix}}+1}} - p_i \right),
$$

(5.154)

which gives us

$$
\sigma^2_{N_{\text{pix}}+1} = \sigma^2_{N_{\text{pix}}} + p_{N_{\text{pix}}+1}(1 - p_{N_{\text{pix}}+1}) + 2 \sum_{i=1}^{N_{\text{pix}}} \left( p_i(N_{\text{pix}}+1) - p_iP_{N_{\text{pix}}+1} \right).
$$

(5.155)
Now, suppose that Equation (5.144) holds for \( N_{\text{pix}} \) pixels or less. Then,

\[
\sigma^2_{N_{\text{pix}}+1} = \sum_{i,j=1}^{N_{\text{pix}}} (p_{ij} - p_ip_j) + (p_{N_{\text{pix}}+1} - p_{N_{\text{pix}}+1}^2) + 2 \sum_{i=1}^{N_{\text{pix}}} (p_{i(N_{\text{pix}}+1)} - p_ip_{N_{\text{pix}}+1}) .
\] (5.156)

Using the fact that \( i \) is a dummy variable in the summation, and Equations (5.141) and (5.142), this can be rewritten as

\[
\sigma^2_{N_{\text{pix}}+1} = \sum_{i,j=1}^{N_{\text{pix}}} (p_{ij} - p_ip_j) + (p_{N_{\text{pix}}+1} - p_{N_{\text{pix}}+1}^2)
\]

\[
+ \sum_{i=1}^{N_{\text{pix}}} (p_{i(N_{\text{pix}}+1)} - p_ip_{N_{\text{pix}}+1}) \sum_{j=1}^{N_{\text{pix}}} (p_{(N_{\text{pix}}+1)j} - p_{N_{\text{pix}}+1}p_j) ,
\] (5.157)

and this can be simplified to give

\[
\sigma^2_{N_{\text{pix}}+1} = \sum_{i,j=1}^{N_{\text{pix}}+1} (p_{ij} - p_ip_j) .
\] (5.158)

Which is what I set out to prove. Thus, by induction, Equation (5.144) gives the variance in the number of pixels in a histogram bin for the case where the pixel position angles are correlated.

### 5.9 Determination of probability distribution of \( \alpha \)

As stated in Section 5.2.1, assuming that \( Q \) and \( U \) are normally distributed but not necessarily independent, the probability distribution can be written as a multivariate Gaussian. Further, in the specific case where \( Q \) and \( U \) are not correlated between pixels and have zero mean, then Equation (5.4) holds. This equation is particularly useful because, if \( Q \) and \( U \) do in fact follow a multivariate normal distribution (possibly with correlation between pixels), then only the mean and covariance matrix of \( Q \) and \( U \) are needed to completely determine the probability distribution. This is much simpler than trying to calculate the probability distribution directly, since, it only requires the well-known two-point correlation functions for \( Q \) and \( U \) to be calculated, as opposed to the full probability distribution needing to be computed by hand.
5.9: DETERMINATION OF PROBABILITY DISTRIBUTION OF $\alpha$

5.9.1 Calculation of covariance matrix for a Gaussian CMB

Since, as stated in the previous section, the probability distribution of $\alpha$ depends on the two-point correlation function for $Q$ and $U$, it is useful to attempt to compute this for the case of the CMB. From [Chon et al. (2004)], it can be seen that

$$
\epsilon_- (\beta) = \left\langle \tilde{P}(\theta, \phi) \tilde{P}(\theta', \phi') \right\rangle = \sum_\ell \frac{2\ell + 1}{4\pi} \left( C^{EE}_\ell - C^{BB}_\ell - 2iC^{EB}_\ell \right) d_{2(-2)}^\ell (\beta), \quad (5.159)
$$

$$
\epsilon_+ (\beta) = \left\langle \tilde{P}^*_r (\theta, \phi) \tilde{P}(\theta', \phi') \right\rangle = \sum_\ell \frac{2\ell + 1}{4\pi} \left( C^{EE}_\ell + C^{BB}_\ell \right) d_{22}^\ell (\beta), \quad (5.160)
$$

where $d_{m^r}$ are the Wigner-$d$ matrices [Wigner 1931], and I have defined

$$
\tilde{P}(\theta, \phi) = e^{2i\alpha} P(\theta, \phi) \quad (5.161)
$$

$$
\tilde{P}(\theta', \phi') = e^{2i\gamma} P(\theta', \phi') \quad (5.162)
$$

with $\alpha$, $\beta$ and $\gamma$ being the Euler angles, defined by

$$
cot \alpha = \frac{\cot \theta \sin \theta' - \cos \theta' \cot (\phi' - \phi)}{\sin (\phi' - \phi)}, \quad (5.163)
$$

$$
cos \beta = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos (\phi' - \phi), \quad (5.164)
$$

$$
cot \gamma = \cos \theta \cot (\phi' - \phi) - \frac{\cot \theta \sin \theta'}{\sin (\phi' - \phi)}. \quad (5.165)
$$

Now, from the definition of $\tilde{P}$, $Q$ and $U$,

$$
Q(\theta, \phi) = \tilde{Q}(\theta, \phi) \cos 2\alpha + \tilde{U}(\theta, \phi) \sin 2\alpha, \quad (5.166)
$$

$$
U(\theta, \phi) = \tilde{U}(\theta, \phi) \cos 2\alpha - \tilde{Q}(\theta, \phi) \sin 2\alpha, \quad (5.167)
$$
with the obvious extension applying to $Q(\theta', \phi')$. Therefore, I get that

$$
\langle Q(\theta, \phi)Q(\theta', \phi') \rangle = \langle \tilde{Q}(\theta, \phi)\tilde{Q}(\theta', \phi') \rangle \cos 2\alpha \cos 2\gamma + \langle \tilde{U}(\theta, \phi)\tilde{U}(\theta', \phi') \rangle \cos 2\alpha \sin 2\gamma 
$$

(5.168)

$$
+ \langle \tilde{U}(\theta, \phi)\tilde{Q}(\theta', \phi') \rangle \sin 2\alpha \cos 2\gamma + \langle \tilde{Q}(\theta, \phi)\tilde{U}(\theta', \phi') \rangle \sin 2\alpha \sin 2\gamma 
$$

$$
\langle Q(\theta, \phi)U(\theta', \phi') \rangle = \langle \tilde{Q}(\theta, \phi)\tilde{U}(\theta', \phi') \rangle \cos 2\alpha \cos 2\gamma - \langle \tilde{Q}(\theta, \phi)\tilde{Q}(\theta', \phi') \rangle \sin 2\alpha \sin 2\gamma 
$$

(5.169)

$$
+ \langle \tilde{U}(\theta, \phi)\tilde{Q}(\theta', \phi') \rangle \cos 2\alpha \sin 2\gamma + \langle \tilde{Q}(\theta, \phi)\tilde{U}(\theta', \phi') \rangle \sin 2\alpha \cos 2\gamma 
$$

$$
\langle U(\theta, \phi)Q(\theta', \phi') \rangle = \langle \tilde{U}(\theta, \phi)\tilde{Q}(\theta', \phi') \rangle \cos 2\alpha \cos 2\gamma - \langle \tilde{Q}(\theta, \phi)\tilde{U}(\theta', \phi') \rangle \sin 2\alpha \sin 2\gamma 
$$

(5.170)

$$
+ \langle \tilde{U}(\theta, \phi)\tilde{Q}(\theta', \phi') \rangle \cos 2\alpha \sin 2\gamma - \langle \tilde{Q}(\theta, \phi)\tilde{Q}(\theta', \phi') \rangle \sin 2\alpha \cos 2\gamma 
$$

$$
\langle U(\theta, \phi)U(\theta', \phi') \rangle = \langle \tilde{U}(\theta, \phi)\tilde{U}(\theta', \phi') \rangle \cos 2\alpha \cos 2\gamma + \langle \tilde{Q}(\theta, \phi)\tilde{Q}(\theta', \phi') \rangle \sin 2\alpha \sin 2\gamma 
$$

(5.171)

$$
- \langle \tilde{U}(\theta, \phi)\tilde{Q}(\theta', \phi') \rangle \cos 2\alpha \sin 2\gamma - \langle \tilde{Q}(\theta, \phi)\tilde{U}(\theta', \phi') \rangle \sin 2\alpha \cos 2\gamma .
$$

Chon et al. (2004) gives that

$$
\langle \tilde{Q}(\theta, \phi)\tilde{Q}(\theta', \phi') \rangle = \frac{1}{2} (\epsilon_+ (\beta) + \mathbb{R} (\epsilon_- (\beta))) ,
$$

(5.172)

$$
\langle \tilde{U}(\theta, \phi)\tilde{U}(\theta', \phi') \rangle = \frac{1}{2} (\epsilon_+ (\beta) - \mathbb{R} (\epsilon_- (\beta))) ,
$$

(5.173)

$$
\langle \tilde{Q}(\theta, \phi)\tilde{U}(\theta', \phi') \rangle = \langle \tilde{U}(\theta, \phi)\tilde{Q}(\theta', \phi') \rangle = \frac{1}{2} (\mathbb{I} (\epsilon_- (\beta))) ,
$$

(5.174)
and, thus,

\begin{align}
\langle Q(\theta, \phi)Q(\theta', \phi') \rangle &= \left( \frac{\cos 2\alpha \cos 2\gamma + \sin 2\alpha \sin 2\gamma}{2} \right) \epsilon_+(\beta) \\
&+ \left( \frac{\cos 2\alpha \cos 2\gamma - \sin 2\alpha \sin 2\gamma}{2} \right) \Re (\epsilon_-(\beta)) \\
&+ \left( \frac{\cos 2\alpha \sin 2\gamma + \sin 2\alpha \cos 2\gamma}{2} \right) \Im (\epsilon_-(\beta)) , \quad (5.175)
\end{align}

\begin{align}
\langle U(\theta, \phi)U(\theta', \phi') \rangle &= \left( \frac{\cos 2\alpha \cos 2\gamma + \sin 2\alpha \sin 2\gamma}{2} \right) \epsilon_+(\beta) \\
&+ \left( \frac{\sin 2\alpha \sin 2\gamma - \cos 2\alpha \cos 2\gamma}{2} \right) \Re (\epsilon_-(\beta)) \\
&- \left( \frac{\cos 2\alpha \sin 2\gamma + \sin 2\alpha \cos 2\gamma}{2} \right) \Im (\epsilon_-(\beta)) , \quad (5.176)
\end{align}

\begin{align}
\langle Q(\theta, \phi)U(\theta', \phi') \rangle &= \left( \frac{\sin 2\alpha \cos 2\gamma - \cos 2\alpha \sin 2\gamma}{2} \right) \epsilon_+(\beta) \\
&- \left( \frac{\sin 2\alpha \cos 2\gamma + \cos 2\alpha \sin 2\gamma}{2} \right) \Re (\epsilon_-(\beta)) \\
&+ \left( \frac{\cos 2\alpha \cos 2\gamma - \sin 2\alpha \sin 2\gamma}{2} \right) \Im (\epsilon_-(\beta)) , \quad (5.177)
\end{align}

\begin{align}
\langle U(\theta, \phi)Q(\theta', \phi') \rangle &= \left( \frac{\cos 2\alpha \sin 2\gamma - \sin 2\alpha \cos 2\gamma}{2} \right) \epsilon_+(\beta) \\
&- \left( \frac{\sin 2\alpha \cos 2\gamma + \cos 2\alpha \sin 2\gamma}{2} \right) \Re (\epsilon_-(\beta)) \\
&+ \left( \frac{\cos 2\alpha \cos 2\gamma - \sin 2\alpha \sin 2\gamma}{2} \right) \Im (\epsilon_-(\beta)) , \quad (5.178)
\end{align}
which can be written, using the formula for $\epsilon_+$ and $\epsilon_-$ and the trigonometric addition formulae as

$$
\langle Q(\theta, \phi) Q(\theta', \phi') \rangle = \sum_l \frac{2l + 1}{4\pi} \left( (C^E l + C^B l) d^2_{22}(\beta) \cos (2(\alpha - \gamma)) + (C^E l - C^B l) d^l_{2(2-2)}(\beta) \cos (2(\alpha + \gamma)) \right),
$$

$$
\langle U(\theta, \phi) U(\theta', \phi') \rangle = \sum_l \frac{2l + 1}{4\pi} \left( (C^E l + C^B l) d^2_{22}(\beta) \cos (2(\alpha - \gamma)) - (C^E l - C^B l) d^l_{2(2-2)}(\beta) \cos (2(\alpha + \gamma)) \right),
$$

$$
\langle Q(\theta, \phi) U(\theta', \phi') \rangle = \sum_l \frac{2l + 1}{4\pi} \left( (C^E l + C^B l) d^2_{22}(\beta) \sin (2(\gamma - \alpha)) - (C^E l - C^B l) d^l_{2(2-2)}(\beta) \cos (2(\alpha + \gamma)) \right).
$$

In principle, given these expressions, it is possible to find the covariance matrix of $Q$ and $U$, and hence (if it can be inverted) the probability distribution of $Q$ and $U$. However, since correlations exist between $Q$ and $U$ across the whole sky, this calculation will be non-trivial.

However, of interest is the case where $\theta = \theta'$ and $\phi = \phi'$. In this case, $\beta = 0$ and $\alpha - \gamma = 0$, with $\alpha + \gamma$ being undefined. From the properties of the Wigner $d$-functions \cite{Chon2004}, this gives that $d^l_{s's'}(0) = \delta_{s's'}$, and thus the above formulae become

$$
\langle Q(\theta, \phi) Q(\theta, \phi) \rangle = \langle U(\theta, \phi) U(\theta, \phi) \rangle = \sum_l \frac{2l + 1}{4\pi} \left( C^E l + C^B l \right),
$$

$$
\langle Q(\theta, \phi) U(\theta, \phi) \rangle = \langle U(\theta, \phi) Q(\theta, \phi) \rangle = 0.
$$

A useful thing to prove is that the probability distribution of $\alpha$ is flat for any probability distribution that can be characterised solely by a power spectrum (such as the CMB). Equation (5.124) shows that the number of pixels in each bin depends only on the probability distribution for each individual pixel (and not on the correlation between different pixels). This means that if the probability distribution of $\alpha$ is flat for all pixels, then the mean number of pixels in each bin will also be constant.
5.10: PROBABILITY DISTRIBUTION FOR SYSTEMATIC ERRORS

In Section 5.9, it was shown that

\[ P(\alpha) = \frac{2 \sqrt{\sigma_{QQ} \sigma_{UU} - \sigma_{QU}^2}}{\pi((\sigma_{QQ} + \sigma_{UU}) + (\sigma_{UU} - \sigma_{QQ}) \cos 4\alpha - 2\sigma_{QU} \sin 4\alpha)}, \]  

(5.185)

and therefore, the probability distribution of \( \alpha \) is flat iff \( \sigma_{QQ} = \sigma_{UU} \) and \( \sigma_{QU} = 0 \). Supposing that pixels are sufficiently small, I can approximate that \( \sigma_{AB}(\theta, \phi) = \langle A(\theta, \phi)B(\theta, \phi) \rangle \), and thus in the case of any probability distribution characterised by a power spectrum, the number of pixels in each bin must be flat, as is observed in my simulations.

5.10 Probability distribution for systematic errors

Suppose the probability distribution for the real value of \( \alpha \) (i.e. that which would be observed in the absence of systematic errors) can be calculated. Then, temporarily define

\[
\begin{pmatrix}
Q' \\
U'
\end{pmatrix} = \begin{pmatrix}
\epsilon_{QQ} & \epsilon_{QU} \\
\epsilon_{UQ} & \epsilon_{UU}
\end{pmatrix} \begin{pmatrix}
Q \\
U
\end{pmatrix},
\]

(5.186)

with \( Q \) and \( U \) being the real Stokes parameters and \( Q' \) and \( U' \) being the observed parameters (note the slight alteration in the definition relative to Equation (5.2)), and further define the observed position angle to be

\[ \alpha' = \frac{1}{2} \arctan \left( \frac{U'}{Q'} \right). \]  

(5.187)

Then, I can show that

\[ p(\alpha') = p_\alpha(\alpha) \frac{d\alpha}{d\alpha'}, \]

(5.188)

where \( p_\alpha(\alpha) \) is the real probability distribution, and \( p(\alpha') \) is the observed probability distribution.

Thus, to determine the probability distribution for the observed position angle, I
need to calculate \( \frac{d\alpha}{d\alpha'} \). From the definitions given above,

\[
\tan 2\alpha' = \frac{U'}{Q'} = \frac{\epsilon_{UU}U + \epsilon_{UQ}Q}{\epsilon_{QQ}Q + \epsilon_{QU}U} = \frac{\epsilon_{UU} \tan 2\alpha + \epsilon_{UQ}}{\epsilon_{QQ} + \epsilon_{QU} \tan 2\alpha},
\]

by using the chain rule, I can differentiate this to get

\[
\frac{d\alpha'}{d\alpha} = \frac{(\epsilon_{UU}\epsilon_{QQ} - \epsilon_{UQ}\epsilon_{QU}) \cos^2 2\alpha' \sec^2 2\alpha}{(\epsilon_{UU} \tan 2\alpha + \epsilon_{QQ})^2}.
\]

By inverting the expression for \( Q' \) and \( U' \), I can get an expression for \( \tan 2\alpha \),

\[
\tan 2\alpha = \frac{\epsilon_{UU} \tan 2\alpha' - \epsilon_{UQ}}{\epsilon_{UU} - \epsilon_{QU} \tan 2\alpha'}.
\]

This can then be substituted into the expression for the differential to give

\[
\frac{d\alpha}{d\alpha'} = \frac{(\epsilon_{UU}(\epsilon_{UU} \tan 2\alpha' - \epsilon_{UQ}) + \epsilon_{QQ}(\epsilon_{UU} - \epsilon_{QU} \tan 2\alpha'))^2}{(\epsilon_{UU} \tan 2\alpha + \epsilon_{QQ})^2((\epsilon_{UU} - \epsilon_{QU} \tan 2\alpha')^2 + (\epsilon_{UQ} - \epsilon_{QQ} \tan 2\alpha')^2)},
\]

which can then be simplified to give

\[
\frac{d\alpha}{d\alpha'} = \frac{2(\epsilon_{UU}\epsilon_{QQ} - \epsilon_{UQ}\epsilon_{QU})}{\epsilon_{UU}^2 + \epsilon_{QQ}^2 + \epsilon_{UQ}^2 + \epsilon_{QU}^2 + \cos 4\alpha'(\epsilon_{UU}^2 - \epsilon_{QQ}^2 + \epsilon_{UQ}^2 - \epsilon_{QU}^2) - 2 \sin 4\alpha'(\epsilon_{UU}\epsilon_{QU} + \epsilon_{QQ}\epsilon_{UQ})}.
\]

Now, returning to the original definition, as used in Equation 5.2, and setting

\[
\begin{pmatrix}
Q' \\
U'
\end{pmatrix} = \begin{pmatrix}
Q \\
U
\end{pmatrix} + \begin{pmatrix}
\epsilon_{QQ} & \epsilon_{UQ} \\
\epsilon_{UQ} & \epsilon_{UU}
\end{pmatrix} \begin{pmatrix}
Q \\
U
\end{pmatrix},
\]

I can write

\[
p(\alpha') = \frac{(1 + K_a)p_0(\alpha', \epsilon)}{1 + K_b}.
\]

where I have defined

\[
K_a = 2(\epsilon_{QQ} + \epsilon_{UU} + \epsilon_{UU}\epsilon_{QQ} - \epsilon_{QU}\epsilon_{UQ})
\]

\[
K_b = 2(\epsilon_{QQ} + \epsilon_{UU}) + \epsilon_{UU}^2 + \epsilon_{QQ}^2 + \epsilon_{QU}^2 + \epsilon_{UQ}^2 + \cos 4\alpha'(2(\epsilon_{UU} - \epsilon_{QQ}) + \epsilon_{UU}^2 - \epsilon_{QQ}^2 + \epsilon_{UQ}^2 - \epsilon_{QU}^2) - 2 \sin 4\alpha'((\epsilon_{QU} + \epsilon_{UQ}) + \epsilon_{UU}\epsilon_{QU} + \epsilon_{QQ}\epsilon_{UQ}),
\]
5.10: PROBABILITY DISTRIBUTION FOR SYSTEMATIC ERRORS

Now, in the case where $\epsilon$ is small, and so $K_a$ and $K_b$ are also small, it is possible to carry out a Taylor expansion. To first order in $\epsilon$ (but, retaining second-order terms in $K$ for the moment), I get (noting that $\epsilon = 0 \Rightarrow \alpha = \alpha'$)

$$p(\alpha') = (1 - K_b + K_b^2 + K_a(1 - K_b)) \left( p_\alpha(\alpha)|_{\alpha=\alpha'} + \sum_{A,B=Q,U} \frac{\partial p_\alpha}{\partial \varepsilon_{AB}}|_{\alpha=\alpha'} \varepsilon_{AB} \right). \quad (5.200)$$

Now, using the chain rule, I can write

$$\frac{\partial p_\alpha}{\partial \varepsilon_{AB}} = \frac{\partial p_\alpha}{\partial \alpha} \frac{\partial \alpha}{\partial \varepsilon_{AB}}, \quad (5.201)$$

and, thus, in the case where the probability distribution of $\alpha$ is flat, $\frac{\partial p_\alpha}{\partial \varepsilon_{AB}} = 0$. Further, to second order in $\epsilon$

$$K_b^2 = 4(\epsilon_{QQ} + \epsilon_{UU})^2 + 2(\epsilon_{U}^2 - \epsilon_{QQ}^2) \cos 4\alpha' - 2(\epsilon_{UU} + \epsilon_{QQ})(\epsilon_{QQ} + \epsilon_{UU}) \sin 4\alpha' \quad (5.202)$$

$$+ (\epsilon_{UU} - \epsilon_{QQ})^2 \cos 2^2 4\alpha' + (\epsilon_{UU} + \epsilon_{QQ})^2 \sin 2^2 4\alpha' + 2(\epsilon_{UU} - \epsilon_{QQ})(\epsilon_{QQ} + \epsilon_{UU}) \cos 4\alpha' \sin 4\alpha'$$

$$K_aK_b = 4(\epsilon_{QQ} + \epsilon_{UU})^2 + 4(\epsilon_{U}^2 - \epsilon_{QQ}^2) \cos 4\alpha' - 4(\epsilon_{UU} + \epsilon_{QQ})(\epsilon_{QQ} + \epsilon_{UU}) \sin 4\alpha', \quad (5.203)$$

and, hence, I get

$$1 - K_b + K_b^2 + K_a - K_aK_b = 1 + (\epsilon_{UU} - \epsilon_{QQ})^2 \cos 2^2 4\alpha' + (\epsilon_{UU} + \epsilon_{QQ})^2 \sin 2^2 4\alpha' \quad (5.204)$$

$$+ 2(\epsilon_{UU} - \epsilon_{QQ})(\epsilon_{UU} + \epsilon_{QQ}) \cos 4\alpha' \sin 4\alpha' + \epsilon_{QQ}\epsilon_{UU} - \epsilon_{UU}\epsilon_{QQ}$$

$$+ (\epsilon_{QQ}^2 + \epsilon_{UU}^2)(\cos 4\alpha' - 1) - (\epsilon_{U}^2 - \epsilon_{UU}^2)(\cos 4\alpha' + 1)$$

$$+ 2(\epsilon_{UU} + \epsilon_{QQ} + \epsilon_{UU}\epsilon_{QQ} + \epsilon_{UU}\epsilon_{QQ}) \sin 4\alpha' + 2(\epsilon_{UU} - \epsilon_{UU}) \cos 4\alpha',$$

which, to first order in $\epsilon$, can be simplified to

$$1 - K_b + K_b^2 + K_a - K_aK_b = 1 + 2(\epsilon_{QQ} - \epsilon_{UU}) \cos 4\alpha' + 2(\epsilon_{UU} + \epsilon_{QQ}) \sin 4\alpha'. \quad (5.205)$$

Then, plugging this into Equation (5.200), the probability distribution of $\alpha'$ can be determined to be, to first order in $\epsilon$

$$p(\alpha') = p_\alpha(\alpha)|_{\alpha=\alpha'} + p_\alpha(\alpha)|_{\alpha=\alpha'}((\epsilon_{QQ} - \epsilon_{UU}) \cos 4\alpha' + (\epsilon_{UU} + \epsilon_{QQ}) \sin 4\alpha'), \quad (5.206)$$

and, rewriting the above expression in terms of the parameters as defined in Section 5.3.1 I find that

$$p(\alpha') = p_\alpha(\alpha)|_{\alpha=\alpha'} + p_\alpha(\alpha)|_{\alpha=\alpha'}(2\epsilon_1 \cos 4\alpha' + 2\epsilon_2 \sin 4\alpha'). \quad (5.207)$$
From this, I can prove that, as shown in the simulations in Section 5.3.4, for an approximately flat probability distribution, systematic errors of the form $\epsilon_1 \neq 0$ produce oscillations of the form $\cos 4\alpha$, those of the form $\epsilon_2 \neq 0$ produce oscillations of the form $\sin 4\alpha$, those of the form $\theta \neq 0$ simply rotate the existing probability distribution (by mapping $\alpha$ to $\alpha'$) and those of the form $\epsilon_0 \neq 0$ have no effect. However, note that this only applies for an approximately flat probability distribution, and only to first order in $\epsilon$. If the distribution is not flat (as in the case of many noise maps), then the differential terms that I neglected result in the errors being more complex. Note also that this does not take account of the more realistic situation where the systematic errors themselves can vary over time. This results in a significantly more complicated result, which I have not had time to compute.

5.11 Conclusions

In this chapter, I have outlined how polarisation position angles could be used as a null test to detect systematics in CMB polarisation experiments, and characterised the properties of the probability distribution of the position angles. It has been proven that the mean number of pixels in each bin can be calculated simply by summing up the probabilities of each pixel having a value of $\alpha$ lying in that bin and this, coupled with the fact that the probability distribution of $\alpha$ for a CMB-like map is flat, proves that the histogram produced from a map that is generated solely from a power spectrum must necessarily be flat. Further, a formula has been found which allows for the probability distribution of $\alpha$ for correlated noise to be computed.

Also, the standard deviation of the number of pixels in a histogram bin has also been computed. For the case where the pixels are uncorrelated, it can be shown that the error follows a standard $\sqrt{n}$ law. However, in the case where the polarisation is correlated between different pixels (as in the case of the CMB), it can be shown that the error is considerably larger, due to the inter-pixel correlations reducing the effective number of pixels that are being measured. In the case of a CMB map, the error takes the
form $\sigma = A + B \cos 4\alpha$, where $A$ and $B$ are constants that depend on the power spectrum of the map in question, with $A$ being $O(10^{\sqrt{n}})$ (except for a pure B-mode map, where it is considerably larger) and with $B$ being positive for a E-mode dominated map and negative for a B-mode dominated map. The massively increased error in the case of pure CMB will restrict the detection of systematics using this method.

Then, I investigated the effect of systematics on the probability distribution, and considered which types of systematic effect would be detectable by this method. It was shown, both analytically and numerically, that only shear-type systematic errors (ones with either $\epsilon_1 = \epsilon_{QQ} - \epsilon_{UU} \neq 0$ or $\epsilon_2 = \epsilon_{QU} + \epsilon_{UQ} \neq 0$) would realistically be detectable using this method. Using the non-central $\chi^2$ distribution, I have shown that there would be a 50% chance of detecting $\epsilon_1 = 0.0055$ and $\epsilon_2 = 0.0078$ at a 95% confidence level. Thus, this method is able to detect systematic errors of certain types at below the 1% level. Further, it is well-suited to being used as a null test since, whilst it would be very difficult to conclusively rule out the presence of systematic errors using this method, there is nevertheless a high probability that they will be detected even at relatively low levels. However, these results may be altered by the effects of inter-pixel correlations, which I have not had time to explore properly.

Overall, this method appears to be rather promising as a null-test for the detection of systematic errors in CMB polarisation experiments, although it needs more work on understanding the probability distributions in the case where the pixels are correlated.
Conclusions and future work

In this thesis, I have looked at the pure-$C_\ell$ method and its ability to detect B-mode polarisation. I have written my own implementation of the algorithm, proven that it does indeed eliminate E-B mixing and shown that it is generally considerably better than the standard pseudo-$C_\ell$ method. However, I have also shown that there are potential issues with the method. In particular, I have demonstrated that, in the case of low-$\ell$, the effects of removing the ambiguous modes which cause E-B mixing manifest themselves as an increase in the standard deviation of the estimator, which causes the accuracy of the method to degrade significantly at very low-$\ell$. In the case of an azimuthal mask, this effect is not too large except for the lowest bin. However, in the case of a non-azimuthal mask, this may not be true. In particular, a point source mask can be shown to have a considerably larger effect. Further, I have proven analytically that the pure-$C_\ell$ method variance is the same as that for the pseudo-$C_\ell$ method without E-B mixing for high-$\ell$.

In studying point source masks, I found that there is a close relationship between the number of sources and the additional error on top of the error from the underlying mask. In particular, this error is proportional to the number of sources. This is a very useful relationship, since it allows the effect of any number of point sources to be determined from a single simulation. I have also looked into the effect of point-source subtraction, using the results of Battye et al. (2011). For high-frequency observations,
there is little noticeable effect, since the point-source spectrum is too low to have any significant effect on the detection of B-modes, even for \( r = 0.001 \). On the other hand, at 30 GHz, detection of much lower than \( r = 0.1 \) would require masking sources to the level of 10 mJy, and this is unlikely to be possible using the pure-\( C_\ell \) method.

I also looked at optimal survey strategies. Using an optimal error formula, I found an expression which can be used to find the optimal fraction of the sky to observe as a function of the noise level, and also the level of \( r \) that it could observe. Using this, I showed that the QUIJOTE experiment could only detect \( r \approx 0.2 \). I then used simulations of the pure-\( C_\ell \) method to determine how much the effect of boundary terms would have on the result. However, for the case of an azimuthal mask, the effects are minimal, except in the lowest \( \ell \) bin.

Overall, it appears that the pure-\( C_\ell \) method is perfectly acceptable in the majority of situations. However, if the measurement of very low values of \( \ell \) is required, then the estimator will be significantly sub-optimal and, thus, some alternate method will need to be used. Further, this method is likely not suited to a mask with many small patches, since the requirement for apodisation will reduce the available sky area substantially in this case.

I also attempted to calculate a formula to compute analytically the variance produced on the pure-\( C_\ell \) estimator by a point-source mask. However, I was unable to produce a formula which agreed with the simulated results and, thus, I will need to find a way to correct it. As further future work, I would also like to investigate the effects of non-azimuthal masks on the pure-\( C_\ell \) method in a systematic manner, and in order to do this I would like to attempt producing masks made up of large numbers of disconnected circular regions, and other masks where the boundary length is well-known, in order to test the effect of changing the boundary length and mask shape on the low-\( \ell \) estimator variance.

Finally, I outlined a method for detecting systematic errors in polarisation experiments through the use of the polarisation position angle. To do this, I produced histograms of the number of pixels with a given polarisation position angle for the case of
various CMB realisations and also the case of pure noise realisations, and showed that, for the CMB realisations, the histogram is flat, and for the noise realisations, the standard deviation of the number of pixels in each bin follows a standard $\sqrt{n}$ law. Then, I produced simulations of the effect of systematic errors on these histograms, showing that shear-type errors (those with $\epsilon_1 = \epsilon_{QQ} - \epsilon_{UU} \neq 0$ or $\epsilon_2 = \epsilon_{QU} + \epsilon_{UQ} \neq 0$) could potentially be detectable to the 1% level using this method. Based on this, I can conclude that this method has substantial potential for being used as a null-test for systematic errors.
6: CONCLUSIONS AND FUTURE WORK
References


REFERENCES


REFERENCES


A. R. Liddle and D. H. Lyth. *Cosmological Inflation and Large-Scale Structure*. Cos-
REFERENCES


216 ANALYSIS OF CMB POLARISATION
REFERENCES


