



**Università
degli Studi
di Ferrara**

**DOCTORAL COURSE IN
PHYSICS**

CYCLE **XXXIII**

DIRECTOR PROF. LUPPI ELEONORA

**Distilling information from present and
future CMB datasets: the cases of
large-scale polarization and lensing**

SCIENTIFIC/DISCIPLINARY SECTOR (SDS) FIS/05

Candidate

NATALE Umberto

(signature)

Supervisor

Prof. NATOLI Paolo

(signature)

2017/2020

Contents

	Page
Abstract	1
Sommario	2
Introduction	3
1 Standard cosmological model	5
1.1 Homogeneous background	6
1.1.1 Parametrisation of the homogeneous background	7
1.2 Thermal history of the Universe	9
1.3 Inflationary paradigm	12
1.3.1 Quantum fluctuation of the Inflaton	16
1.4 Cosmological distances	19
2 Cosmic Microwave Background Radiation	22
2.1 CMB temperature anisotropies	22
2.2 Temperature power spectrum	23
2.2.1 Secondary anisotropies	26
2.3 Polarization power spectrum	27
2.3.1 The reionization fingerprint	30
2.4 Choice of cosmological parameters	32
2.5 Foregrounds	33
3 Overview on data analysis techniques	34
3.1 Some concepts of probability	34
3.2 Frequentist interpretation of probability	37
3.2.1 Hypothesis testing	38
3.2.2 The principle of Maximum Likelihood	38
3.2.3 Fisher information and the minimum variance limit	39
3.2.4 The Fisher matrix analysis of CMB	40
3.3 Bayesian interpretation of probability	41
3.3.1 Parameter estimation	42
3.3.2 Choice of prior	43
3.3.3 Markov Chain Monte Carlo	44
3.3.4 Likelihood analysis of CMB	46
3.4 Surprise, uncertainties and entropy	48
3.4.1 Maximum Entropy Principle	50
3.4.2 Kullback-Leibler divergence	52
4 Combined WMAP and Planck LFI likelihood	54
4.1 Datasets	55
4.2 Polarization masks	56

4.3	Methods	58
4.4	Foreground cleaning	61
4.5	Power Spectra	64
4.6	Likelihood and validation	66
4.7	Reionization constraints	68
5	Lack of power and Galactic plane orientation	74
5.1	Data set and simulations	75
5.1.1	CMB maps and masks	75
5.1.2	Sets of simulations	76
5.1.3	Angular power spectrum estimator	76
5.2	Analysis in Λ CDM framework	77
5.2.1	Variance analyses including rotations	79
5.3	Analysis of Λ CDM simulations with low variance	82
5.3.1	Variance analyses including rotations	83
A	Appendix	85
A.1	Generating the rotations	85
A.2	Comparison between 2018 and 2015 <i>Planck</i> release	87
A.3	Dependence on threshold	88
6	A small-scale estimate of the reionization optical depth and its interplay with ΛCDM extensions	92
6.1	Dataset and forecasts	93
6.2	Methods	94
6.3	Information content of different CMB probes	96
6.3.1	The effect induced by the presence of a missing parameter	101
	Conslusions	104
	Acknowledgements	107

Abstract

The level of precision expected by future cosmic microwave background experiments makes it necessary to refine the techniques used to analyse the data. It becomes essential to understand how to maximise their information content. This can be done, either by developing techniques to reduce the noise level present in the data or by systematically studying the significance of different observations for a given model. In addition, it is possible to define new estimators to test their statistical properties. In this thesis, we initially show how it is possible to construct a pixel-based dataset that combines the WMAP and low-frequency *Planck* large-scale polarization maps. After demonstrating its robustness, we derive constraints on the optical depth obtaining $\tau = 0.069^{+0.012}_{-0.011}$ (68% CL). Adding small-scale data, BAO and lensing we find $\tau = 0.0714^{+0.0087}_{-0.0096}$ (68% CL). As a further topic, we show how it is possible to define new estimators to study the correlation between the orientation of the Galactic plane and the low-variance anomaly shown by large angle CMB temperature data. Through the use of random rotations, we show the stability of this anomaly at high Galactic latitudes, finding a significance of $\sim 3\sigma$. Finally, we compare two main observables in CMB experiments: lensing and large-scale polarization. We show how the information carried by these two probes affects our ability to constrain the base Λ CDM parameters. We extend the analyses considering also some of its most debated extensions, quantifying which future probe will play a crucial role in their characterisation.

Sommario

Il livello di precisione atteso dai futuri esperimenti sulla radiazione cosmica di fondo rende necessario il perfezionamento delle tecniche utilizzate per analizzare i dati. Diviene indispensabile capire come massimizzare il loro contenuto informativo. Questo può essere fatto sia sviluppando tecniche per ridurre del livello di rumore presente nei dati, sia tramite lo studio sistematico della significatività delle diverse osservazioni per un determinato modello. In aggiunta, è possibile definire nuovi estimatori per testarne le proprietà statistiche. In questo lavoro di tesi, inizialmente mostriamo come sia possibile costruire un dataset nello spazio dei pixel che combini le mappe di polarizzazione su larga scala ottenute dalle misure a bassa frequenza di WMAP e *Planck*. Dopo averne dimostrato la robustezza, deriviamo i vincoli sullo spessore ottico ottenendo un valore pari a $\tau = 0.069^{+0.012}_{-0.011}$ (68% CL). Aggiungendo misure derivanti dalle piccole scale, BAO e lensing troviamo un valore pari a $\tau = 0.0714^{+0.0087}_{-0.0096}$ (68% CL). Come ulteriore argomento, facciamo vedere come sia possibile definire nuovi estimatori per studiare la correlazione tra l'orientazione del piano Galattico e l'abbassamento anomalo della varianza visto nei dati di temperatura della CMB su larga scala angolare. Tramite l'uso di rotazioni random mostriamo la stabilità di questa anomalia ad alte latitudini Galattiche, trovando una significatività di $\sim 3\sigma$. Infine, compariamo due osservabili principali negli esperimenti di CMB: quella del lensing e quella della polarizzazione su larga scala. Mostriamo come l'informazione contenuta in queste due sonde influenzi la nostra capacità di vincolare i parametri base del modello Λ CDM. Estendiamo l'analisi considerando anche alcune delle sue estensioni più dibattute, quantificando quale sarà la sonda che giocherà un ruolo cruciale nella loro caratterizzazione.

Introduction

The Michelson and Morley experiment in 1887 marked the change in our perception of the Universe. They found that the speed of light is independent of reference frame used to measure it. This result radically changed the human beliefs about the laws of nature. However, a coherent formulation of a cosmological theory had to wait until the beginning of the 20th century. The mathematical formalism that describes the gravitational framework of the Universe came to life only in 1915, with the formulation of General Relativity by A. Einstein. Nevertheless, only with the implicit hypothesis that on a large scale, for any observer, the Universe appears isotropic and homogeneous, can the foundations be laid for any cosmological model. This statement was formally named and formalised by Edward Milne only in the 1930s.

Although the establishment of a new mathematical formalism, nobody could predict the number of scientific discoveries that would come shortly. The years that followed were a succession of predictions and experimental confirmations. In the twenties, A. Friedmann finds a solution to the Einstein equations, which suggests an expansion of space. A few years later, G. Lemaitre comes independently to the same conclusion, showing a relation between its solutions and the recession of nearby galaxies. In the same decade, E. Hubble shows that galaxies are moving away from the Earth, confirming the expansion of the Universe. In the late 1940s, G. Gamow formulates the theory of “Big Bang”, which explains, in a rapidly expanding and cooling universe, the synthesis of light elements. And there were also many others. But what pushed physicists into space was the accidental discovery of an omnidirectional, homogeneous and isotropic noise in the microwave bands, whose power was consistent with a black body with a temperature around 3 K, by Arno Penzias and Robert Wilson in 1964.

This observation paved the way for a sequence of space-borne missions such as COBE (1989-1992), WMAP (2001-2008), Planck (2009-2018), and more than a dozen other sub-orbital experiments (e.g., BICEP, Keck Array, ACT, Spider). This rapid series of experiments transformed cosmology into a precision science, allowing for the studying of the tiny anisotropies in the temperature and polarization of the Cosmic Microwave Background radiation (CMB). The new generation of ground-based experiments (e.g., Simons Observatory), the balloon experiments (LSPE, Piper), and the future space missions (CORe, LiteBIRD, Pixie) are going to increase the precision of the observations, in particular for the polarised anisotropies which constitute a pool of information to test the inflationary era and put constraints on several parameters of non-standard physics.

All the promised achievements promote two sides of the same coin. If on the one side we are going through an increasing precision era of the observations, on the other side, we are going to face the increase of the amount of data. As a consequence of this, more effort is required in the development of analysis techniques.

In this PhD thesis, we are going to present three applications of statistical analyses techniques on CMB data. To this purpose, one needs an introduction of the formalism that allows us to characterise the CMB fluctuations, to quantify information from data, and to forecast errors on cosmological parameters from a future experiment. The thesis is structured as follows. In Chapter 1, we review the basics of the currently-accepted Stan-

standard Model of Cosmology. We describe the homogeneous background parametrisation, reviewing the thermal history of the Universe and highlighting the connections between Inflation and the CMB radiation. In Chapter 2, we move our attention to the CMB, which is the main observable used in this thesis. We describe the methods commonly used in CMB cosmology for isotropic Gaussian random fields, characterising both temperature and polarization anisotropies. We briefly review the fingerprint of reionization on CMB and highlights the choice for cosmological parameters. In Chapter 3, we review some concepts of data analysis techniques, focussing the attention on three different frameworks in which interpreting probability. In Chapter 4, we start to present one of the original work of this thesis. We present a novel CMB polarization likelihood package for large angular scales built from combined WMAP and Planck LFI legacy maps. The content of this chapter was accepted on 29 July 2020 in the *Astronomy & Astrophysics* journal and produced by the work in collaboration with L. Pagano, M. Lattanzi, M. Migliaccio, L. P. Colombo, A. Gruppuso, P. Natoli, and G. Polenta (Natale et al. 2020). In Chapter 5, we analyse the low-variance anomaly, that is a feature of the CMB temperature anisotropy pattern present in both WMAP and *Planck* data, which results to be dependent from the high Galactic latitude data. The content of this chapter was published on 17 December 2019 in the *JCAP* journal and produced by the work in collaboration with A. Gruppuso, D. Molinari and P. Natoli (Natale et al. 2019). In Chapter 6, we apply some Information Theory-based concepts to study the effects of the experimental-setup improvement on different cosmological parameter. In particular, we focus on the comparison between the information carried by lensing and large scale polarization measurements. We perform this analyses in both the standard cosmological model and some of its extensions. The content of this chapter is a work in preparation produced in collaboration with A. Gruppuso, M. Lattanzi, P. Natoli, L. Pagano, and myself.

Chapter 1

Standard cosmological model

The standard cosmological model (SCM), as we know it today, is the result of years of theoretical insights and observation. Nevertheless, it is possible to date its starting point with the formulation of *General Relativity* (GR) by A. Einstein, in 1915. The GR represents the mathematical framework that describes the connection between the geometry of the Universe and its energy content. The space-time geometry is identified by a four-dimensional pseudo-Riemannian manifold \mathcal{M} with metric g . The connection with its energy content is determined through a set of differential field equations

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G T_{\mu\nu} + \Lambda g_{\mu\nu}. \quad (1.1)$$

Here $c = \hbar = 1$, G is the universal constant of gravitation, $T_{\mu\nu}$ is the energy momentum tensor, Λ is the cosmological constant¹, and $G_{\mu\nu}$ is the Einstein tensor, in which appear the Ricci tensor $R_{\mu\nu}$ and the Ricci scalar R .

Despite the beauty of this mathematical formulation, this set of field equations alone are not enough to determine the evolution of a gravitational system in many cases. However, with some peculiar assumptions, they admit an elegant solution that represents the building block of modern cosmology. This assumption states that on large scales, and for any observer, the Universe appears to be isotropic and homogeneous (Durrer 2008; Carroll 2019). Homogeneity is the property of being identical everywhere in space, while isotropy is the property of looking the same in every direction. This is often called the *cosmological principle*: the universe looks statistically the same in all directions and no direction should be distinguished. One can argue on the validity of this statement if we limit the attention at our galaxy and the nearby Universe. However, observations on scales larger than ~ 100 Mpc (where 1 Mpc = 3.086×10^{22} m, or equivalently 1 Mpc = 3.2615×10^6 light years) all show isotropy and homogeneity (Mukhanov 2005; Hogg et al. 2005; Pandey and Sarkar 2015). This implicitly push the cosmologists to define homogeneity in an average sense, where the Universe is taken to be identical in different places when one looks at sufficiently large pieces.

These two basic assumptions allow us to slice the space-time into homogeneous and isotropic 3-dimensional hyper-surfaces characterized by a time coordinate t , called the *cosmic time* or *proper time*. All these hyper-surfaces are such that the 3-spaces of constant time, $\Sigma_t = \{\mathbf{x} | (\mathbf{x}, t) \in \mathcal{M}\}$, are maximally symmetric spaces, hence spaces of constant curvature. Therefore, the metric g takes the form

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -dt^2 + a^2(t)\sigma_{ij}^K dx^i dx^j, \quad (1.2)$$

¹Einstein originally introduced the concept in 1917 to counterbalance the effects of gravity and achieve a static Universe. Today is used in the current standard model of cosmology to parametrize the dark-energy.

where $a(t)$ is the scale factor, which measure the expansion rate of the Universe. In other word, the scaling factor tell us how large the space-like slice Σ_t is at the moment t . The coordinate used here are known as *comoving coordinates*, and σ_{ij}^K is the spatial part of the metric of constant curvature K . Depending on the sign of K this space is locally isometric to a 3-sphere ($K > 0$), a three dimensional pseudo-sphere ($K < 0$) or flat Euclidean space ($K = 0$). We shall usually normalise the scale factor such that $a(t_0) = a_0 = 1$, today t_0 . Note that, sometime, when $K \neq 0$, the scale factor is normalised such that $K = \pm 1$. In this case, we have no normalisation constant left and a_0 has the dimension of a length. The form of the metric σ_{ij}^K that we often use is given in polar coordinate (r, θ, ϕ) , and takes the form

$$\sigma_{ij}^K dx^i dx^j = \frac{dr^2}{1 - Kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2), \quad (1.3)$$

Inserted in Eq. (1.2), it constitutes the *Friedman-Lemaître-Robertson-Walker metric* (FLRW). It will sometimes be useful to change the time variable we use from proper time to *conformal time*

$$\eta = \int \frac{dt}{a(t)}, \quad (1.4)$$

from which the FLRW metric becomes

$$ds^2 = a^2(\eta) \left[-d\eta^2 + \frac{dr^2}{1 - Kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right]. \quad (1.5)$$

This metric is often expressed as

$$ds^2 = a^2(\eta) \left[-d\eta^2 + dr^2 + \chi(r)^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right], \quad (1.6)$$

where

$$\chi(r) = \begin{cases} r & \text{in the Euclidean case, } K = 0 \\ \frac{1}{\sqrt{K}} \sin(\sqrt{K}r) & \text{in the spherical case, } K > 0 \\ \frac{1}{\sqrt{|K|}} \sinh(\sqrt{|K|r}) & \text{in the hyperbolic case, } K < 0 \end{cases}. \quad (1.7)$$

1.1 Homogeneous background

Imposing the validity of the cosmological principle, we are allowed to plug the FLRW metric into Einstein's equations. This choice allows us to find a simplified form of the Eq. (1.1) suitable for an homogeneous and unperturbed Universe. This solution characterises the evolution of the background. As a first step, we can observe that, due to the symmetry of spacetime, one can safely make the assumption that the content of the universe is a perfect fluid, fully described by its energy-momentum tensor

$$T_{\mu\nu} = (\rho(t) + P(t)) u_\mu u_\nu + P(t) g_{\mu\nu}, \quad (1.8)$$

where ρ and P represent, respectively, the fluid energy density and pressure. In the case in which more species are present in the Universe we have

$$P = \sum_i P_i \quad (1.9)$$

$$\rho = \sum_i \rho_i. \quad (1.10)$$

The four-vector u_μ , instead, is the four-velocity in a frame that is comoving with the fluid, defined as

$$u_\mu \equiv (1, 0, 0, 0), \quad (1.11)$$

with the normalisation condition $u_\mu u^\mu = -1$.

Collecting these results in Eq. (1.1), we find the three (dependent) solutions for the Einstein equations

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{K}{a^2}. \quad (1.12)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3}, \quad (1.13)$$

$$\dot{\rho} = -3H(\rho + P). \quad (1.14)$$

Here Eq. (1.12) comes from the time-time component, G_{00} , Eq. (1.13) comes from the space-space component, G_{ij} , and Eq. (1.14) follows from the time component of the Bianchi identity $\nabla^\mu T_{\mu\nu} = 0$, where ∇^μ is the covariant derivative (see, e.g. Carroll 2019). This set of three equations are known as *Friedmann equations*. Here H is the *Hubble function*, usually defined in terms of the *dimensionless Hubble parameter* h as

$$H = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}. \quad (1.15)$$

These equations are not independent. Precisely, we have two independent equations and one dependent from the other. Thus we are facing a problem in which we have two equations and three unknowns, (ρ, P, a) . To overcome this issue, we can assume a constant equation of state,

$$w = \frac{P}{\rho} = \text{const.}, \quad (1.16)$$

to solve Eq. (1.14). The resulting form for the energy density parameter is

$$\rho(t) = \rho_0 a(t)^{-3(1+w)}, \quad (1.17)$$

where ρ_0 denotes the value of the energy density at present time, t_0 (we are imposing $a_0 = 1$). Having this set of equations we can analyse one by one the contributions of the different constituents of the Universe. In particular, relying on current observations (see, e.g., Planck Collaboration VI 2018), we can classify the different energy densities into three main families: *non-relativistic matter* having $w_m = 0$ (baryon, dark matter), *relativistic matter* with $w_r = 1/3$ (photons, light neutrinos) and *fluids having negative pressure* with $w_\Lambda = -1$ (dark energy in the form of a Cosmological constant). This classification, due to the scaling dependency outlined in Eq. (1.17), allows us to identify three different eras, in which a single component dominates. Precisely, in a universe dominated by radiation with $\Lambda = 0$ we have

$$\rho_r(t) = \rho_{r,0} a_r(t)^{-4}. \quad (1.18)$$

In a universe dominated by pressureless matter with $\Lambda = 0$ we have

$$\rho_m(t) = \rho_{m,0} a_m(t)^{-3}. \quad (1.19)$$

In a universe dominated by cosmological constant

$$\rho_\Lambda(t) = \rho_{\Lambda,0} = \text{const.}. \quad (1.20)$$

1.1.1 Parametrisation of the homogeneous background

The previous classification, in which each component is treated individually, represents a huge simplification. Indeed, even if it is possible to distinguish between the different single-component-dominated eras, a rigorous analysis cannot exclude the subdominant contribution to the energy content of the Universe. We need an equation that describes the

expanding homogeneous Universe, which has an explicit dependence from the different components present today ($t = t_0$) in the Universe. A useful quantity that can be used to properly take into account each component is the *density parameter* for each species Ω , defined as the ratio

$$\Omega(t) = \frac{\rho(t)}{\rho_c}, \quad (1.21)$$

where ρ_c is the *critical density*

$$\rho_c = \frac{3H^2}{8\pi G}. \quad (1.22)$$

The critical density is defined as the density necessary to have a flat universe. Thus, manipulating Eq. (1.12), and defining

$$\Omega_k = -\frac{3K}{8\pi G a^2 \rho_c}, \quad \Omega_\Lambda = \frac{\Lambda}{8\pi G \rho_c}, \quad (1.23)$$

for the different components of the Universe, we have

$$H(t) = H_0 \sqrt{\sum_i \Omega_{i,0} a^{-3(1+w_i)}(t) + \Omega_{k,0} a^{-2}(t) + \Omega_{\Lambda,0}}, \quad (1.24)$$

where H_0 is the Hubble constant, $H(t = t_0)$. Here we implicitly separate the contributions of different components to the total density, see Eq. (1.10). The Hubble function $H(t)$ measures the expansion rate at any particular time t for any model obeying the Cosmological Principle. The Eq. (1.24) shows how it vary with time in a way that depends upon the contents of the Universe today. That is, the Eq. (1.24) represents the equation we are looking for. It is important to observe that, since $\Omega_k \propto a^{-2}$, the curvature is always negligible in the early universe because a is a monotonic function of time. Furthermore, evaluating Eq. (1.24) for $t = t_0$ we have

$$1 = \sum_i \Omega_{i,0} + \Omega_{k,0} + \Omega_{\Lambda,0}, \quad (1.25)$$

which imply that, for a spatially flat Universe ($K = 0$) $\rho_0 = \rho_c$, i.e. the definition of critical density.

As we mention above, current observations (see, e.g., [Planck Collaboration VI 2018](#)) show that the best-fitting picture for describing the statistics of the Universe on large scales is a model in which the matter is mostly cold and dark (i.e. effectively collisionless and with no electromagnetic interactions, CDM), with the bulk of the energy density of the Universe behaving like vacuum energy (i.e. like the cosmological constant of general relativity, Λ), and, for this reason, it is known as Λ CDM model. Within this context, we can rewrite Eq. (1.24) in a more explicit form by using the equations of state for each component. Removing the subscript to the density parameter, this leads to

$$H^2(z) = H_0^2 \left[\Omega_r(z+1)^4 + \Omega_m(1+z)^3 + \Omega_k(z+1)^2 + \Omega_\Lambda \right], \quad (1.26)$$

where we used the definition of cosmological redshift $1+z = 1/a$, which is directly linked to the measurable Doppler shift of spectral lines of objects in the sky via $z = \Delta\lambda/\lambda_0$. Note that, since the Hubble constant is used as the unit of measurement to describe the expansion of the Universe, the cosmological parameters that enter in Eq. (1.26) can be interpreted as the five parameters determining the background homogeneous spacetime.

1.2 Thermal history of the Universe

As we have seen, according to the Friedmann equations, the expansion rate of the universe is determined by the energy density and equation of state of its constituents. When we go back in time, from Eqs. (1.18)-(1.20), we can see that the dark energy density has a negligible impact on the dynamics of the universe, leaving space to the cold matter-dominated era. Even if dark-energy cannot begin to dominate too early, because a substantial period of matter domination is needed for structure formation, we know that there is this phase transition in the history of the Universe. Going back further, we have a period in which the Universe was radiation-dominated. This is characterised by an equation of state $P = w\rho$ ($w = 1/3$), that inserted in Eq. (1.13), assuming $\Lambda = 0$, imply that the second derivative of the scaling factor $a(t)$ is negative, that is, $a(t)$ is a concave curve. Thus, we expect the scale factor of the Universe to cross the $a = 0$ line in a finite amount of time. The moment when this happens is called the *Big Bang*, which represents a singularity in the coordinates (the spatial metric vanishes for $a = 0$), in the Ricci scalar, and in the density.

Even if we can identify this epoch in which radiation dominates, to properly describe the physical processes in an expanding universe we need, strictly speaking, a full kinetic theory. Fortunately, the situation greatly simplifies in the very early universe, when the particles are in a state of *local equilibrium* with each other, where by local equilibrium we simply mean that matter has maximal possible entropy. If within a typical cosmological time the particles scatter from each other many times, their entropy reaches the maximal possible value before the size of the universe changes significantly. Thus, even if the system is far from equilibrium, we can describe it through its entropy (always definable and not-decreasing). This state of local equilibrium is then characterised by comparison between the *collision time* and the *expansion time*. Precisely, defining the *reaction rate*, $\Gamma = \sigma nv$, where σ is the cross-section of the process, n is the number density of the particles and v is their relative velocity, the time that establish equilibrium (the collision time) is

$$t_c \sim \frac{1}{\sigma nv}. \quad (1.27)$$

Since the expansion time is given by $t_H \sim 1/H$, and local equilibrium is reached before expansion becomes relevant, the condition for local equilibrium translates in

$$t_c \ll t_H. \quad (1.28)$$

This condition implies that the *entropy per comoving volume* elements remain constant, ie the expansion is adiabatic. From the second law of thermodynamics:

$$dS = \frac{dE + PdV}{T} = \frac{d(\rho(T)V) + P(T)dV}{T} = \frac{1}{T} \left[\frac{\partial S}{\partial V} dV + \frac{\partial S}{\partial T} dT \right], \quad (1.29)$$

where

$$\frac{\partial S}{\partial V} = \frac{P(T) + \rho(T)}{T}, \quad \frac{\partial S}{\partial T} = \frac{V}{T} \frac{d\rho}{dT}. \quad (1.30)$$

Since the entropy is a differentiable function, the Schwarz condition

$$\frac{\partial^2 S}{\partial V \partial T} = \frac{\partial^2 S}{\partial T \partial V}, \quad (1.31)$$

holds. That is,

$$\frac{\partial}{\partial T} \left[\frac{P(T) + \rho(T)}{T} \right] = \frac{\partial}{\partial V} \left[\frac{V}{T} \frac{d\rho}{dT} \right] \Rightarrow \frac{dP}{dT} = \frac{P(T) + \rho(T)}{T}, \quad (1.32)$$

that relates the energy density and the pressure. Using this relation we get

$$dS = \frac{1}{T}d[(\rho + P)V] - \frac{V}{T}dP = \frac{1}{T}d[(\rho + P)V] - \frac{V}{T^2}[P + \rho]dT, \quad (1.33)$$

that is

$$dS = d\left[\frac{(P + \rho)V}{T} + \text{const.}\right]. \quad (1.34)$$

From the Friedmann Eq. (1.14)

$$a^3\dot{P} = \frac{d}{dt}\left[a^3(\rho + P)\right], \quad (1.35)$$

and the energy density and the pressure relation follows the definition of the conserved quantity

$$\frac{d}{dt}\left[\frac{a^3}{T}(\rho + P)\right] = 0. \quad (1.36)$$

But it is easy to see that, defining the *density entropy* $s = S/V$, the quantity in the square bracket is exactly the entropy per comoving volume, that is

$$\frac{d}{dt}\left[\frac{a^3}{T}(\rho + P)\right] = \frac{d}{dt}\left[a^3s\right] = 0. \quad (1.37)$$

The entropy density is dominated by the contribution of relativistic particles, so that, since the total radiation density can be written as

$$\rho_r = \frac{\pi^2}{30}T^4\left[\sum_{i=b}g_i\left(\frac{T_i}{T}\right)^4 + \frac{7}{8}\sum_{i=f}g_i\left(\frac{T_i}{T}\right)^4\right] = \frac{\pi^2}{30}T^4g_*, \quad (1.38)$$

where the b index stays for bosons while f for fermions, and g_i is the internal degrees of freedom, using again the equation of state $P = w\rho$, we get

$$s = \frac{2\pi^2}{45}g_{*S}(T)T^3. \quad (1.39)$$

Here

$$g_{*S}(T) = \sum_{i=b}g_i\left(\frac{T_i}{T}\right)^3 + \frac{7}{8}\sum_{i=f}g_i\left(\frac{T_i}{T}\right)^3. \quad (1.40)$$

If all the relativistic species are in thermal (local) equilibrium, $g_* = g_{*S}$. Note that, the entropy in a comoving volume is $S = sa^3 \propto a^3T^3$, which implies that the temperature is a decreasing function with respect to the expansion of the universe

$$T \propto a^{-1}. \quad (1.41)$$

What we have computed so far is valid as long as the speed of the interactions between the particles is greater than or equal to the Hubble parameter, and then the conditions of thermal equilibrium are maintained. If, instead, $\Gamma < H$ the species considered decouples from plasma. Thus, fixing the range of temperature, we can identify different transitions that characterise the thermal history of our Universe. Precisely (see, e.g. [Ellis et al. 2012](#)):

- **Planck Era** ($T \sim 10^{19}$ GeV - $t \sim 10^{-43}$ s) Near the Planckian scale, general relativity can no longer be trusted, even if at energies slightly below this scale classical spacetime still makes sense. At this energy, we expect that GR breaks down and gravity should become a quantum interaction. Non-perturbative quantum gravity dominates.

- **Grand Unification Theory** ($T \sim 10^{16}$ GeV - $t \sim 10^{-36}$ s) There is no reason to expect that non-perturbative quantum gravity plays any significant role below 10^{19} GeV. Therefore, we can still use General Relativity to describe the dynamics of the universe. The main uncertainty here is the matter composition of the universe. There are candidate Grand Unified Theories (GUT), mainly based on supersymmetry, which relates bosons to fermions, so that each fermion has a boson superpartner, and vice versa. There are good reasons to expect that a Grand Unification of the electroweak and strong interactions takes place at energies about 10^{16} GeV
- **Inflation & reheating** ($T \lesssim 10^{15}$ GeV - $t \gtrsim 10^{-34}$ s) Currently the most successful phenomenology we have for understanding the very early universe is inflation, which is discussed in the following section. This is typically expected to take place at an energy scale 10^{15} GeV. Inflation provides a framework for understanding how the apparently causally disconnected regions of the observable universe happen to have the same temperature, and it also predicts the generation of fluctuations that seed the growth of large-scale structure.
- **Baryogenesis** ($T > 1$ TeV - $t < 10^{-10}$ s) At the end of inflation, the observable universe is cold and essentially empty of matter: the universe is reheated and populated with particles via the decay of the inflaton field. Between reheating and the electroweak transition, a number of crucial processes are expected to occur, all of them beyond the reach of the Standard Model of particle physics, and all remaining uncertain at the time of writing. They include the problem of identifying the dark matter particle and the problem of baryogenesis. Baryogenesis is one of the major problems in cosmology, and consists in accounting for the matter/anti-matter asymmetry, i.e. the fact that we only observe matter in stars and galaxies (apart from high-energy collisions that can produce anti-particles, which rapidly annihilate). That is, the problem of baryogenesis consists in finding a mechanism that generated a baryon asymmetry which led to the baryonic structures that we observe.
- **Electroweak transition** ($T \sim 0.1 - 1$ TeV - $t \sim 10^{-10}$ s) The temperature of the universe falls below values which correspond to energies of the mass of bosons mediators of the weak interactions Z^0 , W^\pm , the electromagnetic force is separated from the weak.
- **Quark-gluon transition** ($T \sim 200$ MeV - $t \sim 10^{-5}$ s) The quark-gluon transition takes place: free quarks and gluons become confined within baryons and mesons.
- **Neutrino decoupling** ($T \sim 1$ MeV - $t \sim 1$ s) The primordial neutrinos decouple from the other particles and propagate without further scatterings.
- **Electron-positron annihilation** ($T \sim 0.5$ MeV - $t \sim 1$ s) The typical energy at this time is of order the electron mass. The numerous electron-positron pairs present in the very early universe begin to annihilate when the temperature drops below their rest mass and only a small excess of electrons over positrons, roughly one per billion photons, survives after annihilation. The photons produced are in thermal equilibrium and the radiation temperature increases compared to the temperature of neutrinos, which decoupled earlier.
- **Nucleosynthesis** ($T \sim 0.05$ MeV - $t \sim 200$ s) Nuclear reactions become efficient at this temperature. As a result, free protons and neutrons form helium and other light elements. The abundances of the light elements resulting from primordial nucleosynthesis are in very good agreement with available observation data and this strongly supports our understanding of the universe's evolution back to the first second after the big bang.

- **Matter-radiation equality** ($T \sim \text{eV} - t \sim 10^{11} \text{ s}$) This time corresponds to matter-radiation equality which separates the radiation-dominated epoch from the matter-dominated epoch. The exact value of the cosmological time at equality depends on the constituents of the dark component and, therefore, is known at present only up to a numerical factor of order unity.
- **Recombination** ($T \sim 0.25 \text{ eV} - t \sim 10^{12} \text{ s}$) After nucleosynthesis, the main ingredients of the cosmic plasma are γ , e , $p \equiv \text{H}^+$ and fully ionized helium, He^{2+} (other ionized light nuclei play a negligible role). Photons are strongly coupled to baryons via Thomson ($e - \gamma$) and Coulomb ($p - e$) interactions. As the temperature drops, the ionised nuclei begin to capture free electrons. For $T \lesssim 5000 \text{ K}$, the reaction $p + e \leftrightarrow \text{H} + \gamma$ keeps the plasma in equilibrium. As the temperature drops further, this interaction becomes less effective, and the probability grows of electrons being captured by protons to form hydrogen. The universe becomes transparent to the background radiation.
- **The Dark Ages and the epoch of reionization** ($T < \mathcal{O}(1) \text{ eV} - t \sim 10^{12} - 10^{15} \text{ s}$) After recombination, the baryonic matter is effectively all in the form of neutral hydrogen and helium. From the decoupling redshift of $z = 1100$ down to a redshift $z \sim 200$, the gas temperature follows the *Cosmic Microwave Background* (CMB) temperature field² since the residual ionisation, although very small, is enough to maintain sufficient coupling via Compton scattering (Ellis et al. 2012)

$$T_{\text{gas}} = T_{\gamma} = T_{\gamma 0}(1 + z), \quad z \sim 200. \quad (1.42)$$

Expansion and cooling eventually break this coupling and the gas temperature drops below the CMB temperature, evolving adiabatically as

$$T_{\text{gas}} \propto (1 + z)^2, \quad 200 \gtrsim z \gtrsim 20. \quad (1.43)$$

For $z \lesssim 20$, the gas begins to be heated by emissions from the first stars, and eventually exceeds the CMB temperature. After recombination, the baryonic pressure drops towards zero and gravity overcomes the counterbalancing effect of pressure. The baryonic gas falls into dark matter haloes, and overdensities grow as $\delta \sim a$. Because of the weakness of gravitational instability in an expanding background, it takes of the order of a few 100 Myr before the first stars form. Thus there is a period after recombination, the so-called *Dark Ages*, when baryonic matter is dark. The “backlight” of the CMB radiation leads to emission and absorption features of the neutral hydrogen 21 cm hyperfine spin flip transition. Ionizing radiation from luminous sources convert the cold and neutral gas into a warm and highly ionised medium, starting the *epoch of reionization*.

1.3 Inflationary paradigm

The standard cosmological model that we have described so far is in great agreement with current observations. With the proper choice of the initial conditions, both the anisotropy seen in the CMB temperature field, that the *Large Scale Structure* (LSS) are predicted by the Λ CDM model. These initial conditions are not arbitrary. Indeed, not only these are chosen in such a way to reproduce the structures that we observe today in the Universe, but also to face certain puzzling features of a decelerating early universe: e.g. the *horizon problem* and the *flatness problem*.

²See Chapter 2 for further details.

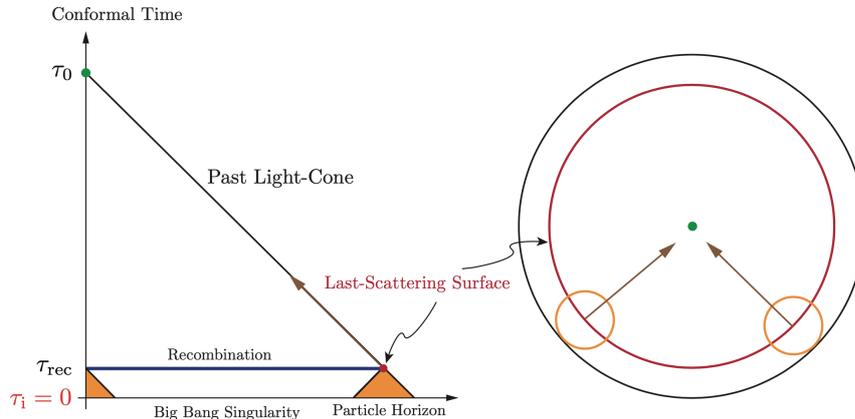


Figure 1.1: Conformal diagram of Big Bang cosmology. Here τ is the conformal time. Each spacetime point in the conformal diagram has an associated past light cone which defines its causal past. Two points on a given constant τ -surface are in causal contact if their past light cones intersect at the Big Bang, $\tau_i = 0$. This means that the surface of last-scattering (τ_{rec}) consisted of many causally disconnected regions that won't be in thermal equilibrium. This diagram makes the horizon problem evident since the uniformity of the CMB on large scales becomes a serious puzzle (Baumann 2011).

If we look in opposite directions on the sky and measure the CMB temperature, we find it is the same to 1 part in $\sim 10^{-5}$. This suggests that a thermalization process operated before decoupling. However, in a decelerating radiation universe, thermalization could not have taken place across the CMB sky. Precisely, the *particle horizon* at recombination is

$$R_{\text{rec}} \simeq \frac{1}{H_{\text{rec}}}. \quad (1.44)$$

This is the distance that light travels from the “beginning” of the Universe at $t = 0$ and represents the limit of causal interaction at the time of last scattering, i.e. particles that are separated by more than R_{rec} can never have been in causal communication. Points on the last scattering surface at opposite ends of the sky are separated today by a distance equal to the distance to the last scattering surface, $D_{\text{rec}} \sim 1/H_0^{-1}$, which is much greater than the maximal causal separation, $R_{\text{rec}} \ll D_{\text{rec}}$. And yet, the particles at these locations at the time when the CMB distribution was frozen had never been in causal communication, see Fig. 1.1. This is what we call the horizon problem.

The second mentioned problem is, instead, a fine-tuning problem related to the temporal behaviour of the curvature parameter Ω_k in the context of Λ CDM model. Indeed, today $|\Omega_k| \lesssim 10^{-2}$ (95% CL) (Planck Collaboration VI 2018), and since $\Omega_k \sim a^{-2}$ we had immediately after the Big Bang a value of $|\Omega_k| \sim 10^{-60}$. Thus, to explain the low value of Ω_k today we have to fix the initial value with an inexplicable precision.

Before going further, we would like to stress that these are not real problems of the standard cosmological model since, by imposing peculiar initial conditions, it is possible to reproduce the current observations. However, it was this peculiarity that leads A. Guth to formulate the theory of inflation (Guth 1987).

The founding idea beyond the theory of inflation is that the Universe, in its early stages, undergo into an accelerated expansion. From Eq. (1.13), it is clear that this happens if $P < -\rho/3$, implying the presence of a fluid with negative pressure different from Λ , that starts to dominates in a more recent epoch. Thus we have to introduce a scalar field (or more) that drive the expansion of the Universe. Note that here, since we are in the primordial phase of the Universe, we are implicitly assuming $\Lambda=0$. The simplest models

of inflation involve a single scalar field ϕ , the *inflaton*, minimally coupled with gravity. That is, its dynamics is governed by the action

$$S = \int d^4x \sqrt{-g} \left[\frac{1}{2}R - \frac{1}{2}g_{\mu\nu}\partial^\mu\phi\partial^\nu\phi - V(\phi) \right]. \quad (1.45)$$

The first term is the Einstein-Hilbert action. The second term describes a canonical scalar field coupled to gravity through the metric $g_{\mu\nu}$. From the variational principle we get the Klein-Gordon (KG) equation

$$\square\phi = \frac{\partial V}{\partial\phi}, \quad (1.46)$$

where \square is the D'Alembert covariant operator

$$\square\phi = \frac{1}{\sqrt{-g}}\partial_\nu(\sqrt{-g}g^{\mu\nu}\partial_\mu\phi). \quad (1.47)$$

In a flat FLRW Universe the KG equation becomes

$$\ddot{\phi} + 3H\dot{\phi} - \frac{\nabla^2\phi}{a^2} + V_\phi(\phi) = 0, \quad (1.48)$$

where $V_\phi = \partial V/\partial\phi$. The energy-momentum tensor is given by

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S_\phi}{\delta g^{\mu\nu}} = \partial_\mu\phi\partial_\nu\phi + g_{\mu\nu}\left(-\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi - V(\phi)\right), \quad (1.49)$$

from which, restricting to the case of a homogeneous field $\phi(t, \mathbf{x}) \equiv \phi(t)$, the scalar energy-momentum tensor takes the form of a perfect fluid (1.8) with

$$\rho_\phi = -T_0^0 = \frac{\dot{\phi}^2}{2} + V(\phi) \quad (1.50)$$

$$P_\phi = \frac{1}{3}T_i^i = \frac{\dot{\phi}^2}{2} - V(\phi). \quad (1.51)$$

The equation of state

$$w_\phi = \frac{P_\phi}{\rho_\phi} = \frac{\frac{\dot{\phi}^2}{2} - V(\phi)}{\frac{\dot{\phi}^2}{2} + V(\phi)}, \quad (1.52)$$

shows that a scalar field can lead to negative pressure ($w_\phi < 0$) and accelerated expansion ($w_\phi < -1/3$) if the potential energy V dominates over the kinetic energy $\dot{\phi}^2/2$. Thus, if $V(\phi) \gg \dot{\phi}^2/2$, the inflaton ϕ “slowly” go through the minimum of the potential. This phase is called *Slow-Roll* (SR), and take place when the inflaton is in a region in which the potential is sufficiently flat ($V(\phi) \sim \text{const.}$), see Fig. 1.2. Under these conditions, for the homogeneous background, the equations of motion became

$$H^2 = \frac{8\pi G}{3}\left(\frac{\dot{\phi}^2}{2} + V(\phi)\right), \quad (1.53)$$

$$\dot{H} = -4\pi G\dot{\phi}^2, \quad (1.54)$$

$$\ddot{\phi} + 3H\dot{\phi} + V_\phi(\phi) = 0. \quad (1.55)$$

The first equation determines the Hubble parameter H , which is the expansion rate of the universe. The second equation is the continuity condition. The third equation describes the evolution of the inflaton. Only two of them, again, are independent.

From Eq. (1.53), if the condition of SR is respected, we have also $H \sim \text{const}$, thus the *comoving Hubble radius*, $r_H \equiv \frac{R_H}{a(t)} = \frac{1}{a(t)H}$, results a decreasing function during inflation.

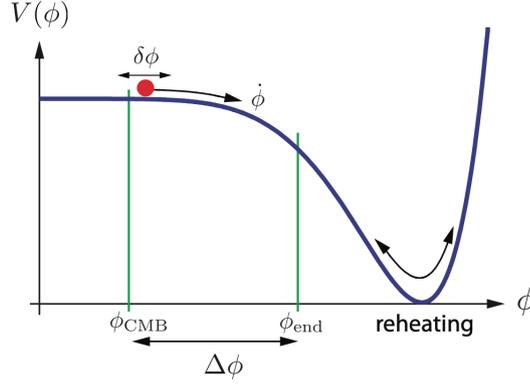


Figure 1.2: Example of an inflaton potential. Acceleration occurs when the potential energy of the field, $V(\phi)$, dominates over its kinetic energy, $\frac{1}{2}\dot{\phi}^2$. Inflation ends at ϕ_{end} when the kinetic energy has grown to become comparable to the potential energy, $\frac{1}{2}\dot{\phi}^2 \simeq V$. CMB fluctuations are created by quantum fluctuations $\delta\phi$ about 60 e -folds before the end of inflation. At reheating, the energy density of the inflaton is converted into radiation (Baumann 2011).

That is, region causally connected after the Big-Bang can move away and become not-causally connected. Furthermore, since the scale factor can be written as (in conformal coordinates)

$$a(\eta) \simeq -\frac{1}{H\eta}, \quad (1.56)$$

the Big-Bang singularity ($a = 0$) is pushed to the infinite past, see Fig. 1.3. Note that this relation breaks down near the end of inflation. That is, $\eta = 0$ does not correspond to the Big-Bang but to the end of inflation.

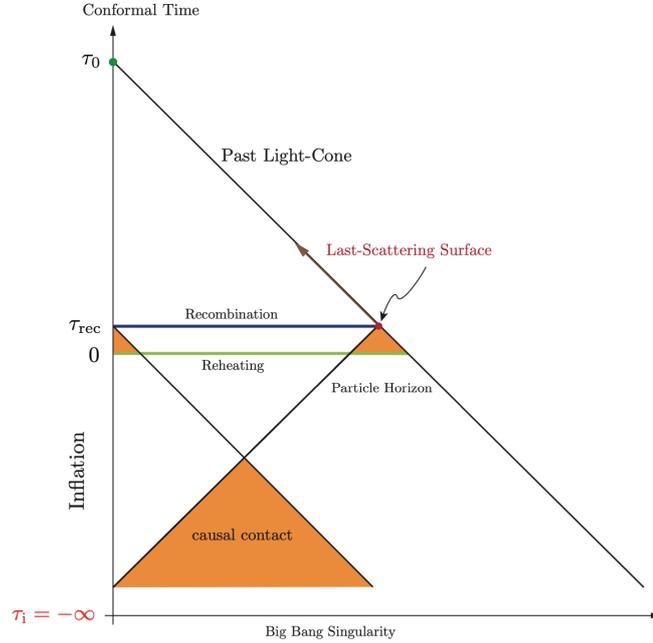


Figure 1.3: Conformal diagram of inflationary cosmology. Here τ represent the conformal time. Inflation push the Big-Bang at $\tau = -\infty$, and there is no singularity at $\tau = 0$. The light cones intersect at an earlier time if inflation lasts for at least 60 e -folds (Baumann 2011).

However, even if the horizon problem seems to be resolved, to take into account also the flatness problem we have to impose some constraints on the duration of inflation. To

do that it is useful to introduce the number of *e-foldings*

$$N = \int_{t_i}^{t_e} dt H(t) = \ln \left(\frac{a(t_e)}{a(t_i)} \right), \quad (1.57)$$

where t_i and t_e are the time in which the inflation begins (initial) and the time in which it ends. It can be shown that both the discussed problems are resolved if $N \gtrsim 60 - 70$ (see, e.g., [Lyth and Liddle 2009](#)) (the precise value depends on the details of reheating and the post-inflationary thermal history of the universe).

It is customary to introduce, in the inflationary context, a parametrization of the SR condition. First, to have this amount of inflation, $N \sim \mathcal{O}(60)$, the Hubble function cannot change much within a Hubble time H^{-1} . This defines the condition

$$\epsilon_{\text{SR}} \equiv -\frac{\dot{H}}{H} \ll \mathcal{O}(1). \quad (1.58)$$

We also need that the parameter ϵ_{SR} does not change much within a Hubble time, that is

$$\eta_{\text{SR}} \equiv \frac{\dot{\epsilon}_{\text{SR}}}{H\epsilon_{\text{SR}}} \ll \mathcal{O}(1). \quad (1.59)$$

In principle, η_{SR} can be close to $\mathcal{O}(1)$ but ϵ_{SR} kept small. In such a case, ϵ_{SR} grows exponentially with e-folds and the inflation period tends to be shorter. More importantly, such a case will not generate a scale-invariant spectrum, thus cannot be responsible for the CMB.

1.3.1 Quantum fluctuation of the Inflaton

Until now, we have discussed the evolution of the homogeneous background $\phi(t)$. However, to relate the inflation with the initial condition needed to produce the CMB anisotropy pattern, we need to introduce perturbations to this background dynamics. This is possible introducing quantum fluctuations

$$\phi(t, \mathbf{x}) = \phi_0(t) + \delta\phi(t, \mathbf{x}), \quad (1.60)$$

where $\phi_0(t)$ is the homogeneous background part used so far. Here, initially, we are going to ignore the perturbations in the gravity sector, only perturbing the inflaton as in Eq. (1.60). Instead, terms suppressed by the slow-roll parameters will be always ignored, e.g. the mass of the inflaton $V_{\phi\phi} \sim \mathcal{O}(\epsilon)H^2$ (see, e.g. [Lyth and Liddle 2009](#), for a complete treatment). The Lagrangian for the perturbed inflaton is

$$L = \int d^3x \left[\frac{a^3}{2} \dot{\delta\phi}^2 - \frac{a}{2} \partial_i \phi \partial^i \phi \right]. \quad (1.61)$$

The resulting equation of motion, written in the comoving momentum space

$$\delta\phi(\mathbf{k}, t) = \int d^3x \delta\phi(\mathbf{x}, t) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (1.62)$$

is the following

$$\delta\ddot{\phi}(\mathbf{k}, t) + 3H\dot{\delta\phi}(\mathbf{k}, t) + \frac{k^2}{a^2}\delta\phi(\mathbf{k}, t) = 0. \quad (1.63)$$

Its solution can be written in terms of its *mode function* $u_{\mathbf{k}} \equiv u(\mathbf{k}, t)$

$$\delta\phi(\mathbf{x}, t) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \left[u_{\mathbf{k}}(t) a_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} + u_{-\mathbf{k}}^*(t) a_{-\mathbf{k}}^\dagger e^{-i\mathbf{k}\cdot\mathbf{x}} \right]. \quad (1.64)$$

It is not difficult to check that

$$a^3 u_{\mathbf{k}}(t) u_{\mathbf{k}}^*(t) - \text{c.c.} = t\text{-independent const.} \quad (1.65)$$

According to the canonical commutation relations between $\delta\phi$ and its conjugate $\delta\pi \equiv \partial L / \partial \dot{\delta\phi}$, we have

$$[\delta\phi(\mathbf{x}, t), \delta\pi(\mathbf{y}, t)] = i\delta(\mathbf{x} - \mathbf{y}) \quad (1.66)$$

$$[\delta\phi(\mathbf{x}, t), \delta\phi(\mathbf{y}, t)] = 0 = [\delta\pi(\mathbf{x}, t), \delta\pi(\mathbf{y}, t)] \quad (1.67)$$

If we use the decomposition

$$\delta\phi = u_{\mathbf{k}} a_{\mathbf{k}} + u_{-\mathbf{k}}^* a_{-\mathbf{k}}^\dagger \quad (1.68)$$

$$\delta\pi = a^3 \dot{u}_{\mathbf{k}} a_{\mathbf{k}} + a^3 \dot{u}_{-\mathbf{k}}^* a_{-\mathbf{k}}^\dagger, \quad (1.69)$$

we can rewrite the commutation relation in the following form

$$[a_{\mathbf{k}}, a_{\mathbf{p}}^\dagger] = \delta^{(3)}(\mathbf{k} - \mathbf{p}) \quad (1.70)$$

$$[a_{\mathbf{k}}^\dagger, a_{\mathbf{p}}^\dagger] = 0, \quad (1.71)$$

The explicit solution of the mode function from Eq. (1.63) is

$$u_{\mathbf{k}}(\eta) = C_+ \frac{H}{\sqrt{2k^3}} (1 + ik\eta) e^{-ik\eta} + C_- \frac{H}{\sqrt{2k^3}} (1 - ik\eta) e^{ik\eta}, \quad (1.72)$$

where we use the conformal time η and the relation in Eq. (1.56). This mode function is a superposition of two linearly independent solutions with the normalisation condition

$$|C_+|^2 - |C_-|^2 = 1. \quad (1.73)$$

Consider the limit in which the mode is well within the horizon, that is, its wavelength a/k much shorter than the Hubble length $1/H$, and consider a time period much shorter than a Hubble time. In these limits only the first term in Eq. (1.72) survives. Furthermore, the mode effectively feels the Minkowski spacetime. We choose this component as our vacuum choice, and it is usually called the *Bunch-Davies state* (see, e.g. Lyth and Liddle 2009). The annihilation operator $a_{\mathbf{p}}$ annihilates the corresponding Bunch-Davies vacuum, $a_{\mathbf{p}} |0\rangle = 0$.

The mode function

$$u_{\mathbf{k}}(\eta) = C_+ \frac{H}{\sqrt{2k^3}} (1 + ik\eta) e^{-ik\eta}, \quad (1.74)$$

has the following important properties (Chen 2010):

- It is oscillatory within the horizon $k|\eta| \gg 1$.
- As it gets stretched out of the horizon $k|\eta| \ll 1$, the amplitude becomes a constant and frozen.

Physically the last item means that, if we look at different comoving patches of the universe that have the superhorizon size, and ignore the shorter wavelength fluctuations, they all evolve classically but with different $\delta\phi$. This difference makes them arrive at ϕ_f , the location of the end of inflation, at different times. This space-dependent time difference $\delta t \sim \delta\phi / \dot{\phi}_0$ leads to the space-dependent inflationary e-fold difference

$$\zeta \sim H\delta t \sim H \frac{\delta\phi}{\dot{\phi}_0}. \quad (1.75)$$

This e-fold difference is the conserved quantity after the mode exits the horizon, and remains so until the mode reenters the horizon sometime after the Big Bang. It is the physical quantity that we can measure, for example, by measuring the temperature anisotropy in the CMB, $\zeta \sim -5\Delta T/T$ (Lyth and Liddle 2009). The information about the primordial inflation is then encoded in the statistical properties of this variable. So we would like to calculate the correlation functions of this quantity.

It is important to mention that the presence of the inflaton field in the early Universe is not only responsible for seeding the density fluctuation, but also for generating a background of weak gravitational waves. This is possible because quantum fluctuations around the homogeneous solution for the inflaton field couple to metric fluctuations (gravitational perturbations) via Einstein's equations. Precisely, if we assume the *Newtonian gauge* and we ignore possible vector perturbations of the metric, the perturbed line element can be written as (see, e.g. Lyth and Liddle 2009):

$$ds^2 = -(1 + 2\Phi)dt^2 + a^2(t) [(1 - 2\Psi)\delta_{ij} + h_{ij}] dx^i dx^j, \quad (1.76)$$

where Φ and Ψ are known as *Bardeen potentials* and the term h_{ij} describes tensor fluctuations, which can propagate as gravitational radiation. For scalar perturbation we can define the gauge invariant quantity

$$\zeta = -\Psi + H \frac{\delta\phi}{\dot{\phi}}, \quad (1.77)$$

that is the gauge invariant extension of Eq. (1.75), known as *curvature perturbation*. Note that this quantity relates the Bardeen potential, the dynamics, and the initial quantum fluctuations $\delta\phi$. Under the assumption of a homogeneous and isotropic Universe, we have that the variance of ζ takes the form

$$\langle \zeta_{\mathbf{k}} \zeta_{\mathbf{p}} \rangle = (2\pi)^3 \delta(\mathbf{k} + \mathbf{p}) \frac{1}{2k^3} \left(\frac{H^2}{\dot{\phi}} \right)^2 \Big|_{k=aH} \equiv (2\pi)^3 \delta(\mathbf{k} + \mathbf{p}) \mathcal{P}_\zeta(k), \quad (1.78)$$

where the variance of each mode is defined at the horizon exit (i.e. $k = aH$). Here $\mathcal{P}_\zeta(k)$ is the *power spectra* of the scalar perturbations.

Similar calculations can be carried out for tensor perturbations. The tensor h_{ij} can be decomposed into two independent components h_+ and h_\times , and isotropy ensures that the amplitude of the tensor fluctuations is equally partitioned between these two components. This leads to

$$\langle h_{+, \mathbf{k}} h_{+, \mathbf{p}} \rangle + \langle h_{\times, \mathbf{k}} h_{\times, \mathbf{p}} \rangle = \langle h_{\mathbf{k}} h_{\mathbf{p}} \rangle = (2\pi)^3 \delta(\mathbf{k} + \mathbf{p}) \frac{1}{k^3} \left(\frac{4H^2}{M_p^2} \right) \Big|_{k=aH} \equiv (2\pi)^3 \delta(\mathbf{k} + \mathbf{p}) \mathcal{P}_h(k), \quad (1.79)$$

where, also this time, the variance of each mode is defined at the horizon exit (i.e. $k = aH$). Here $\mathcal{P}_h(k)$ is the *power spectra* of the tensor perturbations.

It is customary to define the adimensionless power spectra

$$P_s(k) = \frac{k^3}{2\pi^2} \mathcal{P}_\zeta(k), \quad (1.80)$$

$$P_t(k) = \frac{k^3}{2\pi^2} \mathcal{P}_h(k). \quad (1.81)$$

A parametric description is often used for the scalar and tensor fluctuations power spectra

$$P_s(k) = A_s \left(\frac{k}{k_*} \right)^{n_s - 1}, \quad (1.82)$$

$$P_t(k) = A_t \left(\frac{k}{k_*} \right)^{n_t} = r A_s \left(\frac{k}{k_*} \right)^{n_t}. \quad (1.83)$$

Here r , A_s , n_s , and n_t , evaluated at the pivot scale k_* represent the inflationary parameter used to constrain the physics of inflation. In particular, r is the *tensor-to-scalar ratio* defined as

$$r \equiv \frac{A_t}{A_s}. \quad (1.84)$$

The n_s and n_t are the scalar and tensor *spectral index*, respectively,

$$n_s - 1 = \frac{d \ln P_s(k)}{d \ln k}, \quad (1.85)$$

$$n_t = \frac{d \ln P_t(k)}{d \ln k}. \quad (1.86)$$

The current CMB constraints are (Planck Collaboration VI 2018) $n_s = 0.9649 \pm 0.0042$, and $\ln(10^{10} A_s) = 3.044 \pm 0.014$ at 68% CL. These results indicate that the primordial power spectrum of the density perturbations is nearly scale-invariant, meaning that even on very large scales (i.e. small k) points in the sky are expected to be somewhat correlated. From the CMB B-mode polarization measurements by the BICEP2/Keck CMB polarization experiment, the current constraint on the tensor-to-scalar ratio results $r < 0.07$ at 95% CL (Ade et al. 2018a).

1.4 Cosmological distances

Before ending this chapter we would like to explore one of the most difficult task in cosmology: the measurement of distances in the Universe. It represents an issue to be faced, since if we know the distance of a source, we can extract, in principle, information about cosmological parameters. Indeed, distant objects can be observed through the light they emit, which takes a finite time to travel to us. We have seen that the dynamical evolution of the Universe changes under the effect of different matter and energy species contributions. Thus, if the mathematical framework is correct, we expect an effect on the propagation of light, induced by the expansion of the Universe. Therefore, once we know the physical properties of the source, and we have a theoretical prediction about the change of the geometry of the Universe, we can fit the data to extract the wanted information. However, for the same reason, we cannot make measurements along a surface of constant proper time, but only along with the set of light-paths traveling to us from the past. This leads to define distances that are, at least in principle, directly measurable.

A possible definition of distance is the *luminosity distance* d_L measured using the flux emitted by a known source, e.g. type Ia supernovae. This is defined in such a way it preserves the Euclidean inverse-square law for the diminution of light with distance from a point source. If L denote the power emitted by a source at a given point at time t (i.e. the luminosity), and l be the power received per unit area at time t_0 (i.e. the flux) by an observer, we then define

$$d_L = \sqrt{\frac{L}{4\pi l}}. \quad (1.87)$$

The observed flux from a source is by definition

$$l = \frac{Nh\nu_o}{A\Delta t_o}, \quad (1.88)$$

where N is the number of photons, h is the Planck constant, and ν_o is the observed frequency of the emitted photons in the time interval Δt_o . Here $A = 4\pi d_p^2$ is the area of a spherical surface centred on the emission point and passing through the observer at time t_o . The parameter at denominator, d_p , is the *proper distance*. It is defined as the distance measured by a chain of rulers held by observers which connect two points at fixed time t .

In particular, in a FLRW metric, if we take the origin of a set of polar coordinates r , θ and ϕ on one of the two point, since the time is fixed (i.e. $dt = 0$),

$$d_p = a(t) \int_0^r \frac{dr'}{\sqrt{1 - Kr'^2}}. \quad (1.89)$$

Since the luminosity of the source emitted with frequency ν_e at time Δt_e is given by

$$L = \frac{Nh\nu_e}{\Delta t_e}, \quad (1.90)$$

it follows that, since frequency and time intervals are redshifted, (i.e. $\nu_o = \nu_e(1+z)$ and $\Delta t_o = \Delta t_e(1+z)^{-1}$), the luminosity distance is

$$d_L = d_p(1+z). \quad (1.91)$$

Another useful distance is the *angular diameter distance* D_A of an object of known size (standard ruler). It represents the variation of the angular size of an object with its distance from an observer. To derive its expression we use the coordinates introduced in Eq. (1.6) setting, without loss of generality, $r = 0$ at the observer position. Let $D_p(\eta)$ be the (proper) diameter of a source placed at coordinate r at time η . It follows that, for a small source, $D_p = D_A\theta$, where D_A is the angular diameter distance. Suppose now that a flash is simultaneously emitting from the source at both of its ends. Since the light moves along null geodesics, the proper diameter of the source (the arc) seen by the observer is

$$ds^2 = a^2(\eta)\chi^2(r)d\theta^2 \Rightarrow D_p = a(\eta)\chi(r)\theta, \quad (1.92)$$

since the distance of the source is fixed respect to the observer (η is fixed). Now, observing that at the flashes we have $r = \eta_0 - \eta$, and thus $\chi(r) = \chi(\eta_0 - \eta)$, we can rewrite D_A as

$$D_A(z) = a(\eta)\chi(\eta_0 - \eta). \quad (1.93)$$

To go further, we can observe that from $1+z = 1/a$ we have

$$\frac{dz}{dt} = -\frac{\dot{a}}{a^2} = -\frac{H(z)}{a(t)} \Rightarrow \int d\eta = -\int \frac{dz}{H(z)}, \quad (1.94)$$

and then

$$\eta_0 - \eta = \int_0^z \frac{dz'}{H(z')}. \quad (1.95)$$

Finally, from (1.26) follows that

$$\eta_0 - \eta = \frac{1}{H_0} \int_0^z \frac{dz'}{[\Omega_r(z+1)^4 + \Omega_m(1+z)^3 + \Omega_k(z+1)^2 + \Omega_\Lambda]^{1/2}}, \quad (1.96)$$

from which

$$D_A(z) = \begin{cases} \frac{1}{1+z}\chi\left(\frac{1}{H_0} \int_0^z \frac{dz'}{[\Omega_r(z+1)^4 + \Omega_m(1+z)^3 + \Omega_k(z+1)^2 + \Omega_\Lambda]^{1/2}}\right) & \text{if } K \neq 0 \\ \frac{1}{(1+z)H_0} \int_0^z \frac{dz'}{[\Omega_r(z+1)^4 + \Omega_m(1+z)^3 + \Omega_\Lambda]^{1/2}} & \text{if } K = 0 \end{cases}. \quad (1.97)$$

This result shows that, if we know how to measure the angular diameter distance, we are in principle able to extract from Eq. (1.97) the values of the five parameters, namely the density parameters Ω_r , Ω_m , Ω_Λ and Ω_k with the Hubble constant H_0 (or equivalently the Hubble parameter h). This statement is true also for the other distances. The current

CMB plus lensing³ constraints at 68% CL on these cosmological parameters are (Planck Collaboration VI 2018) $\Omega_\Lambda = 0.6847 \pm 0.0073$, $\Omega_m = 0.3153 \pm 0.0073$, $H_0 = 67.36 \pm 0.54$ and $\Omega_r \sim 10^{-5}$. Note that, the radiation component is usually neglected, but it can be estimated from the black-body temperature of the CMB spectrum, $\Omega_r = \Omega_{\text{CMB}}$. These values are obtained with the constraint of a flat universe (i.e. $\Omega_m + \Omega_\Lambda = 1$). However, a 1-parameter extension to the Λ CDM model can be used to constrain $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$, resulting in $\Omega_k = 0.0007 \pm 0.0019$. It is important to observe that this constraint is obtained by combining the *Planck* data with Baryonic Acoustic Oscillation (BAO) data, step necessary to breaks the internal geometric degeneracy. This joint result suggests our Universe looks remarkably flat.

³See Sec. 2.2.1.

Chapter 2

Cosmic Microwave Background Radiation

The Cosmic Microwave Background was accidentally discovered in 1965 by Arno Penzias and Robert Woodrow Wilson. They measured an excess in temperature of few K that was, within the limits of their observations, isotropic, unpolarised, and free from seasonal variations (Penzias and Wilson 1965). Later, this radiation was interpreted by R. Dicke, P. Peebles, P. Roll, and D. Wilkinson as a signal of a “hot” primordial phase of our Universe (Dicke et al. 1965). The spectral energy distribution associated with this thermal emission follows the Planck law

$$B_\nu(T) = 2\nu^3 \left(e^{h\nu/T} - 1 \right)^{-1}, \quad (2.1)$$

This discovery pushed cosmologist into space, leading to the first space mission devoted to the measurement of the CMB spectrum. The spacecraft COBE (COsmic Background Explorer), in 1989 (Fixsen et al. 1994, 1996), measured the associated black-body temperature to be $T = 2.72548 \pm 0.00057$ K, finding the presence of small fluctuations of the order of

$$\frac{\Delta T}{T} \sim 10^{-5}. \quad (2.2)$$

The COBE collaboration constrained possible departures from the blackbody spectrum to be less than 1% (Fixsen 2009). After COBE, the CMB anisotropies have been observed by many ground-based and balloon-borne experiments, as well as by two other satellites, WMAP (Hinshaw et al. 2013) and *Planck* (Planck Collaboration VI 2018), launched in 2001 and 2009 respectively.

2.1 CMB temperature anisotropies

As we have mentioned in Sec. 1.3, the inflationary scenario allows us to relate the quantum fluctuation of the inflaton to the small fluctuation of the temperature field, known as anisotropies of the CMB temperature. In the standard scenario, the inflation ends when the inflaton starts to oscillate around the minimum of its potential. These oscillations comes after a phase in which the SR parameters became of the order of unity, allowing the inflaton to acquire mass (we have discarded this term, since during inflation $M_\phi \sim \mathcal{O}(\epsilon) \ll 1$). The massive inflaton decays into radiation and relativistic particles, starting the *reheating* phase, that marks the transition from an inflationary to a FLRW Universe.

To properly describe the formation and the evolution of CMB radiation it is necessary to make use of Boltzmann equation. It represents the mathematical framework which relates the evolution of the distribution function of a species and their interaction (gravitational and not) within the expanding Universe. Starting from the assumption that the distribution function of radiation $f(\mathbf{x}, \mathbf{p}, t)$ on large scale is a Planckian, with small

temperature perturbation $\Delta T/T$, it is possible to rewrite $f(\mathbf{x}, \mathbf{p}, t)$ as (Dodelson 2003)

$$f(\mathbf{x}, p, \hat{\mathbf{p}}, t) \simeq \left\{ \exp \left[\frac{p}{T(t) \left[1 + \frac{\Delta T}{T}(\mathbf{x}, \hat{\mathbf{p}}) \right]} \right] - 1 \right\}^{-1}. \quad (2.3)$$

Since the fluctuations are supposed to be small, we can separate the unperturbed evolution of $f(\mathbf{x}, p, \hat{\mathbf{p}}, t)$ from the perturbed one. In particular, focussing the attention only on the inhomogeneous component, in Fourier space, we can rewrite the ℓ -th order term as

$$\begin{aligned} \Theta_\ell(k, \eta_0) &\simeq [\Theta_0 + \Phi](k, \eta_*) j_\ell[k(\eta_0 - \eta_*)] \\ &+ 3\Theta_1(k, \eta_*) \left(j_{\ell-1}[k(\eta_0 - \eta_*)] - \frac{\ell+1}{k(\eta_0 - \eta_*)} j_\ell[k(\eta_0 - \eta_*)] \right) \\ &+ \int_0^{\eta_0} d\eta e^{-\tau} [\dot{\Phi} - \dot{\Psi}] j_\ell[k(\eta_0 - \eta)]. \end{aligned} \quad (2.4)$$

Here we have used the conformal time η , and defined the multipole expansion

$$\Theta_\ell(k, \eta) \equiv \frac{1}{(-1)^\ell} \int_{-1}^1 \frac{d\mu}{2} \mathcal{P}_\ell(\mu) \frac{\Delta T}{T}(k, \hat{\mathbf{k}}, \eta), \quad (2.5)$$

where $\mu = \hat{\mathbf{k}} \cdot \hat{\mathbf{p}}$ and \mathcal{P}_ℓ is the ℓ -th Legendre polynomial. Furthermore, η_* is the conformal time at the epoch of recombination, j_ℓ is the ℓ -th spherical Bessel function, τ is the optical depth

$$\tau(\eta) \equiv \int_\eta^{\eta_0} d\eta' n_e \sigma_T a(\eta'), \quad (2.6)$$

The equation is written in the conformal Newtonian gauge (cfr. Eq. (1.76)), where the scalar perturbations are characterized by two scalar potentials, Φ and Ψ :

$$ds^2 = a(\eta)^2 \left[-(1 + 2\Phi)d\eta^2 + (1 - 2\Psi)\delta_{ij}dx^i dx^j \right]. \quad (2.7)$$

The three terms in Eq. (2.4) are the *Sachs-Wolfe*, *Doppler*, and the *integrated Sachs-Wolfe* (ISW) term. The first represent an effective temperature given by the monopole Θ_0 and the gravitational contribution due to the overdensities on the last scattering surface. The dipole term is generally subdominant and produce Doppler anisotropies, and is due to peculiar velocities of the photo-baryonic fluid. The last term basically encodes the information about the gravitational potential that lays between us and the last scattering surface. In other words, it takes into account the whole story of CMB, since it is an integral from $\eta = 0$ to $\eta = \eta_0$.

2.2 Temperature power spectrum

The Eq. (2.4) represents the mathematical foundation of the phenomenological parametrization of the CMB temperature field

$$T_{obs}(\hat{\mathbf{n}}) = T_0 \left(1 + (\boldsymbol{\beta} \cdot \hat{\mathbf{n}}) + \frac{\Delta T}{T_0}(\hat{\mathbf{n}}) \right) + \mathcal{O}(\beta^2), \quad (2.8)$$

where T_0 is the blackbody temperature of the smooth component, $\boldsymbol{\beta} = \mathbf{v}/c$ is our proper velocity vector with respect to the CMB rest frame, and $\Delta T(\hat{\mathbf{n}})/T_0$ is the CMB temperature anisotropy field. However, even if this equation represents a complete picture of the CMB temperature sky, the stochastic nature of the quantum fluctuations during inflation does not allow us to develop a theory to exactly predict $\Delta T(\hat{\mathbf{n}})/T_0$. Nevertheless, this problem can be suitably approached from a statistical point of view. A CMB anisotropy

temperature map, $\Delta T(\hat{\mathbf{n}})/T_0$, can be uniquely decomposed in spherical harmonics $Y_{\ell m}(\hat{\mathbf{n}})$, which define an orthonormal basis on a complete sphere, such that

$$\frac{\Delta T}{T_0}(\hat{\mathbf{n}}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{\mathbf{n}}), \quad (2.9)$$

where

$$a_{\ell m} = \int d\Omega_{\hat{\mathbf{n}}} Y_{\ell m}^*(\hat{\mathbf{n}}) \frac{\Delta T}{T_0}(\hat{\mathbf{n}}). \quad (2.10)$$

Note that, the reality condition $\Delta T(\hat{\mathbf{n}})/T_0 \in \mathbb{R}$ imply that

$$a_{\ell-m} = (-1)^m a_{\ell m}^*. \quad (2.11)$$

We have seen that these anisotropies have an inflationary origin. Precisely, they are due to the quantum fluctuations of the fields that drive inflation (in the simplest model the inflaton). These quantum fluctuations are (near) Gaussian distributed around the vacuum state, and then we expect that also the $a_{\ell m}$ are Gaussian distributed with null average, i.e.

$$\langle a_{\ell m} \rangle = 0. \quad (2.12)$$

Note that, the vacuum expectation value can generally be different from zero. However, any constant value can be reabsorbed through a redefinition of the field in interest (a translation), identifying $\langle \delta\phi \rangle = 0$. Thus, the first non-vanishing contribution is the variance

$$\langle a_{\ell m} a_{\ell' m'}^* \rangle = \delta_{\ell\ell'} \delta_{m m'} C_{\ell}, \quad (2.13)$$

where C_{ℓ} represents the *angular power spectrum* (APS). Here the constraints imposed by the two Dirac delta functions follow from the $a_{\ell m}$ being independent random variables (diagonal covariance). Moreover, statistical isotropy ensures that the variance does not depend on m (rotational invariance of C_{ℓ}). The power spectrum is related to the two-point correlation function of the field $C(\theta) = \langle T(\hat{\mathbf{n}}_1) T(\hat{\mathbf{n}}_2) \rangle$ observed at two directions $\hat{\mathbf{n}}_1$ and $\hat{\mathbf{n}}_2$ in the sky such that $\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2 = \cos\theta$:

$$C(\theta) = \sum_{\ell} \frac{(2\ell+1)}{4\pi} C_{\ell} \mathcal{P}_{\ell}(\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2), \quad (2.14)$$

where \mathcal{P}_{ℓ} is the Legendre polynomial of order ℓ . The importance of the variance C_{ℓ} , or equivalently the two-point correlation function $C(\theta)$, emerges from the Wick's theorem. It ensures that, for a Gaussian distribution, odd momenta vanish and even momenta beyond the second can be recast as a function of the variance. That is, if a random variable is Gaussian distributed, all the statistical properties are encoded in its mean and variance, which are the only momenta of the distribution we need to know.

It is important to note that, since the $a_{\ell m}$'s follow a Gaussian distribution with zero mean and variance C_{ℓ} , the probability density function¹ $p(a_{\ell m}|C_{\ell})$ of the $a_{\ell m}$'s conditioned by the C_{ℓ} 's is

$$p(a_{\ell m}|C_{\ell}) = \frac{1}{\sqrt{2\pi C_{\ell}}} \exp\left\{-\frac{|a_{\ell m}|^2}{2C_{\ell}}\right\}. \quad (2.15)$$

Statistical isotropy of the C_{ℓ} 's allows us to rewrite Eq. (2.13) as:

$$C_{\ell} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{m=\ell} \langle |a_{\ell m}|^2 \rangle, \quad (2.16)$$

¹See Chapter 3 for a detailed definition of probability.

where the symbol $\langle \dots \rangle$ represents an ensemble average. This follows from the random nature of the CMB field, which implies that the statistical information about the outcome of such a process should be obtained by averaging over all possible realisations. In practice, however, we can only observe a single realisation of the CMB field. A way out is provided by the statistical homogeneity and isotropy of the CMB fluctuations, that in principle allows to substitute the ensemble average in Eq. (2.13) with an average over different positions and directions. According to this *ergodic hypothesis*, different regions that are widely separated in the sky are statistically independent from each other and can be considered as different statistical realisations of the same stochastic process. Since we only have access to the CMB field observed at x_0 and η_0 , i.e., the CMB field here and now, what we are really left is the average over different directions, or equivalently over different values of m . In other words, for a given ℓ , all the $a_{\ell m}$ are drawn from the same distribution, which can be therefore sampled by measuring all the $2\ell+1$ coefficients. These observations lead us to define an unbiased estimator² of the observed power spectrum

$$\hat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{m=\ell} |a_{\ell m}|^2. \quad (2.17)$$

This replacement induces an intrinsic source of inaccuracy known as *cosmic variance*

$$\begin{aligned} \left\langle \frac{\Delta C_\ell^2}{C_\ell^2} \right\rangle &= \frac{\langle C_\ell^2 \rangle - \langle C_\ell \rangle^2}{C_\ell^2} \\ &= \frac{1}{(2\ell+1)^2} (2\ell+1)(3+2\ell) - 1 \\ &= \frac{2}{2\ell+1}. \end{aligned} \quad (2.18)$$

Cosmic variance is an irreducible source of uncertainty in cosmological measurements of the CMB power spectrum, and one of the major sources of uncertainties especially at the largest scales (low- ℓ), where we have only a limited number of coefficients $a_{\ell m}$ to average over with respect to the small-scale (high- ℓ) regime.

The APS represents the meeting point between theory and experiments. In particular, Λ CDM model gives accurate prediction about the shape of the power spectrum, see Fig. 2.1. In the large-scale regime, $\ell < 30$, we find the Sachs-Wolfe plateau, since the large-scale anisotropies do not evolve significantly, according to the near-scale invariance imposed by inflation. At the intermediate scale, $30 \lesssim \ell \lesssim 1500$, we have the acoustic peaks due to the photo-baryonic interactions, that in this regime are relevant. Precisely, before recombination, the perturbations in gravitational potential (dominated by dark matter) compress the photo-baryonic fluid, which resists to these compressions due to the internal pressure (produced by the presence of radiation). These compression-decompression phases of the fluid give rise to the acoustic oscillations. After recombination, photons can freely travel to us, and the phase of the oscillations is frozen and it is projected on the sky as an harmonic series of peaks (the acoustic peaks). In the small-scale regime, $\ell \gtrsim 1500$, we can see the damping tail. Here the scales involved in the process are comparable with the mean distance of the photons during recombination. Thus, due to the duration of recombination, the last scattering surface acquires depths, and then, if the distance between a Thomson scattering and another is greater than this depth, there is a damping of the acoustic oscillations due to absorption processes.

²See Sec. 3.2 for the definition of estimator.

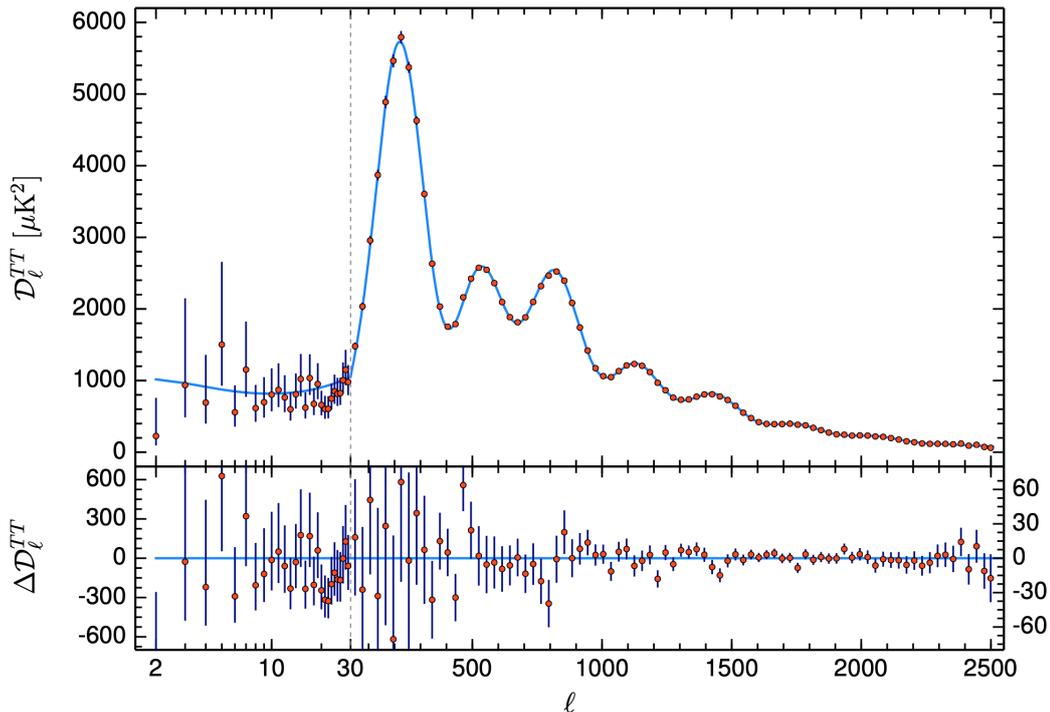


Figure 2.1: The base- Λ CDM theoretical spectrum best fit to the *Planck* TT,TE,EE+lowE+lensing likelihoods is plotted in light blue in the upper panel. Residuals with respect to this model are shown in the lower panel. The error bars show $\pm 1\sigma$ diagonal uncertainties, including cosmic variance (approximated as Gaussian) and not including uncertainties in the foreground model at $\ell \geq 30$ (Planck Collaboration VI 2018).

2.2.1 Secondary anisotropies

Several processes in the foreground of recombination could alter the anisotropy spectrum. Their existence is due to effects that happens well after the last scattering, and that generate what we generally call as *secondary anisotropies*. We can summarize the most important:

- **Gravitational effects:** Secondary anisotropies can be generated by gravitational redshift effects between recombination and today. As pointed out in Eq. (2.4), the scalar potentials difference $\dot{\Phi} - \dot{\Psi}$ has to be integrated along the trajectory of the photons. This represents an important effect due to the time-varying potential over cosmological time-scale, caused by the expansion of the Universe that impacts on its density composition.
- **Gravitational lensing:** The ISW effect is caused by the gravitational push given to the photons in a direction parallel to their motion. This changes the energy but not the direction. If the stress is in direction perpendicular to the motion, there is a change of direction, while the first order of the photon energy remains unchanged. If a pair of photons, moving towards the observer, are initially separated by an angle θ , due to this effect, they arrive to the observer with an angle $\theta + \delta\theta$. This is an effect of *weak gravitational lensing*. The effect on the spectrum of the anisotropies of the CMB is to smooth both peaks and valleys.
- **Reionization:** After recombination the universe is essentially neutral. However, measurements made on the absorption spectra of high-redshift quasars show no evidence of a uniform background of neutral hydrogen until we go back as least as far as $z \sim 6$. This shows that during a period with $z > 6$ the Universe had a new global reionization, which brings the CMB back in contact with electrons. If

these photons hit a region with optical depth τ , only a fraction $e^{-\tau}$ will escape and continue on their way to us. In addition to these, we will also get a fraction $1 - e^{-\tau}$ from the ionized region. All of these have the equilibrated temperature, T . So the temperature we see today is

$$T \left[1 + \frac{\Delta T}{T} \right] e^{-\tau} + T(1 + e^{-\tau}) = T \left[1 + \frac{\Delta T}{T} e^{-\tau} \right], \quad (2.19)$$

that is, inside the standard cosmological model, this effect is parametrised as a reionization layer that reduces the fluctuation amplitude on all scales by a factor $e^{-\tau}$.

- **Sunyaev-Zeldovich (SZ) effect:** The phenomenon of reionization can also happen on a local scale, involving limited regions of the universe (for example a cluster of galaxies). The CMB photons are scattered by high energy photons in intracluster gas. Electrons transfer energy to CMB photons through inverse Compton processes and simultaneously change their direction of propagation. However, on average, the statistical information remains unchanged. This effect impacts on the blackbody distribution, which results distorted in the high frequency part. Basically it move photons from the Rayleigh-Jeans part of the CMB spectrum in the Wien region. It is possible to calculate the change of intensity, which, of course, depends on the physical properties of the cluster,

$$\frac{\Delta I_{\nu}^{\text{RJ}}}{I_{\nu}^{\text{RJ}}} = -2 \int dl \sigma_T n_e \frac{kT}{m_e}, \quad (2.20)$$

where the integral is performed along the length of the cluster. This equation shows that the change in intensity is independent on the redshift of the observed cluster, thus the SZ effect assumes great importance also for the identification of high redshift clusters.

2.3 Polarization power spectrum

Till now we have described only the temperature fluctuation of the CMB, moreover Thomson scattering provides also a mild ($\sim 5\%$) polarization of the radiation. Its cosmological information content, which is complementary to the one extracted from CMB temperature statistics, can be obtained from the angular distribution of the linear polarization of the CMB photons (see, e.g., Kamionkowski et al. 1997; Seljak and Zaldarriaga 1996). The polarization of light is commonly described by *Stokes parameters* I , Q , U , and V . If we consider a monochromatic wave that propagates in the direction $\hat{\mathbf{z}}$ with pulse ω_0 , the corresponding electric field can be written as

$$E_x(t) = a_x(t) \cos(\omega_0 t + \phi_x(t)), \quad (2.21)$$

$$E_y(t) = a_y(t) \cos(\omega_0 t + \phi_y(t)), \quad (2.22)$$

where $a_{x,y}$ are the electric field amplitudes in the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ directions, with $\phi_{x,y}(t)$ phases. The four Stokes parameters are functions of the electric field amplitudes, such that:

$$I = \langle a_x^2 \rangle + \langle a_y^2 \rangle, \quad (2.23)$$

$$Q = \langle a_x^2 \rangle - \langle a_y^2 \rangle, \quad (2.24)$$

$$U = \langle 2a_x a_y \cos(\phi_x - \phi_y) \rangle, \quad (2.25)$$

$$V = \langle 2a_x a_y \sin(\phi_x - \phi_y) \rangle, \quad (2.26)$$

$$(2.27)$$

where $\langle \dots \rangle$ indicates the time average. The parameter I represents the intensity of the wave, whereas the polarization is described by a non-zero value of the remaining 3 parameters. In particular, Q and U describe the linear polarization, while V is a measure of the circular one that is not expected for the case of the CMB. Thomson scattering does not produce circular polarization, while the Q component correspond to the polarization in the $\hat{\mathbf{x}} - \hat{\mathbf{y}}$ direction and U rotated of 45 degrees.

The usually approach followed in CMB analysis is to consider two combinations for the polarization components $Q \pm iU$

$$(Q + iU)(\hat{\mathbf{n}}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{2,\ell m 2} Y_{\ell m}(\hat{\mathbf{n}}), \quad (2.28)$$

$$(Q - iU)(\hat{\mathbf{n}}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{-2,\ell m -2} Y_{\ell m}(\hat{\mathbf{n}}), \quad (2.29)$$

$$(2.30)$$

where the $a_{\pm 2,\ell m}$, are the expansion coefficients of the spin-2 spherical harmonics $_{\pm 2}Y_{\ell m}$. At this point it is useful to introduce two scalar quantities that describe the polarization

$$E(\hat{\mathbf{n}}) = \sum_{\ell m} a_{\ell m}^E Y_{\ell m}(\hat{\mathbf{n}}), \quad (2.31)$$

$$B(\hat{\mathbf{n}}) = \sum_{\ell m} a_{\ell m}^B Y_{\ell m}(\hat{\mathbf{n}}), \quad (2.32)$$

$$(2.33)$$

where

$$a_{\ell m}^E = -\frac{a_{2,\ell m} + a_{-2,\ell m}}{2}, \quad (2.34)$$

$$a_{\ell m}^B = -i \frac{a_{2,\ell m} - a_{-2,\ell m}}{2}. \quad (2.35)$$

$$(2.36)$$

The corresponding power spectra are defined as

$$\langle a_{\ell m}^E a_{\ell' m'}^{E*} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_{\ell}^{EE}, \quad (2.37)$$

$$\langle a_{\ell m}^B a_{\ell' m'}^{B*} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_{\ell}^{BB}, \quad (2.38)$$

$$\langle a_{\ell m}^T a_{\ell' m'}^{E*} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_{\ell}^{TE}, \quad (2.39)$$

and their theoretical best-fit to the current available data is plotted in Fig. 2.2. The TB and EB cross-correlation are expected to be null.

We now need to connect the measurable Stokes parameters to the physical mechanism that generates linear polarization of the CMB. Photons and electrons interact in the photo-baryonic plasma via Compton scattering, which does not induce polarization unless the intensity of the light scattering off of the electron is anisotropically distributed. The cross-section of the process can be written as

$$\frac{d\sigma}{d\Omega} = \frac{3\sigma_T}{8\pi} |\hat{\epsilon} \cdot \hat{\epsilon}'|^2, \quad (2.40)$$

where $\hat{\epsilon}' = (\hat{\epsilon}'_x, \hat{\epsilon}'_y)$ and $\hat{\epsilon} = (\hat{\epsilon}_x, \hat{\epsilon}_y)$ are the polarization vectors of the incident wave and the scattered one, respectively, defined in the plane perpendicular to the direction of propagation of the wave, $\hat{\mathbf{z}}$. The $\hat{\mathbf{z}}$ -direction changes after the scattering by an angle θ defined in the plane that contains the propagation directions of the incoming and scattered

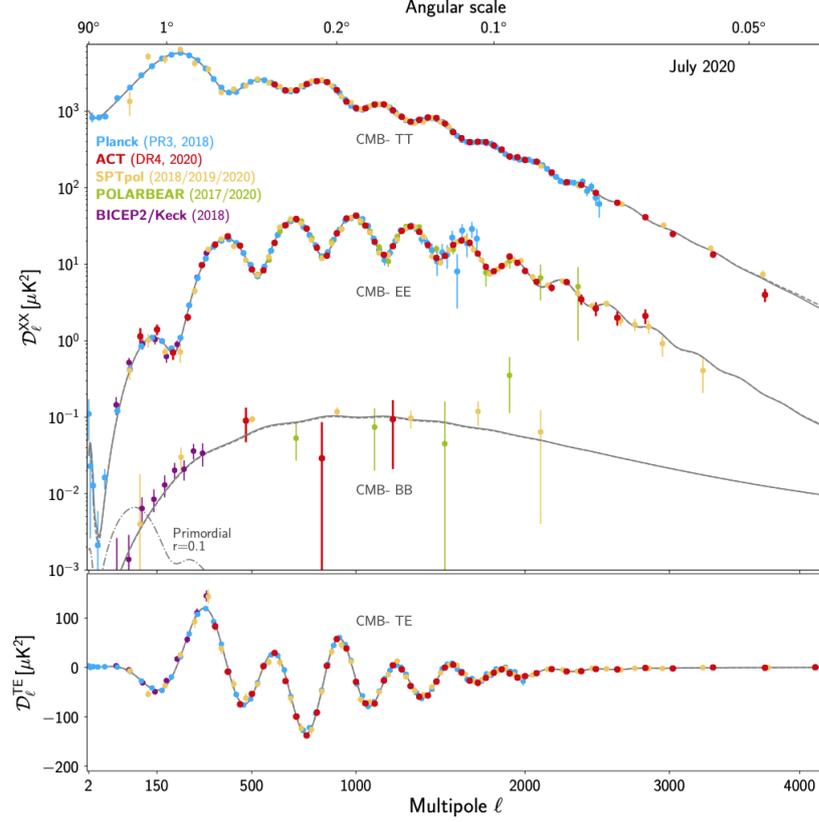


Figure 2.2: Recent measurements of the CMB temperature anisotropy and polarization (Choi et al. 2020). The two models, the thin nearly overlapping grey lines, are from *Planck* (dashed line) and from ACT plus WMAP (A20, solid line). The primordial BB signal with $r = 0.1$ is also shown with the dot-dashed line. For *Planck* here it is shown the 2018 results (Planck Collaboration VI 2018). For SPT it is shown Henning et al. (2018b) for 150 GHz TT $\ell < 2000$, TE and EE, and Sayre et al. (2020) for BB. For $\ell > 2000$ it is shown the SPT spectrum from George et al. (2015) which has been corrected for point source emission. It is visually indistinguishable from the more precise but uncorrected spectrum in Reichardt et al. (2020). For Polarbear/Simons Array it is shown EE from Adachi et al. (2020) and BB from pipeline Ade et al. (2017). For BICEP2/Keck it is used Ade et al. (2018b). All error bars are one sigma and points with no lower bound in TT and EE have been dropped at high ℓ . For ACT they also show preliminary EE results that were not used in the analysis.

waves. In this geometrical configuration, let us consider an initially unpolarised incident light, and let I' and I be the intensity of the incident and scattered light, respectively. For the scattered the intensity along the \hat{x} and \hat{y} directions can be written as $I_x = (I + Q)/2$ and $I_y = (I - Q)/2$, leading to:

$$I_x = \frac{3\sigma_T}{16\pi} \left[I'_x (\hat{e}'_x \cdot \hat{e}_x)^2 + I'_y (\hat{e}'_y \cdot \hat{e}_x)^2 \right] = \frac{3\sigma_T}{16\pi} I', \quad (2.41)$$

$$I_y = \frac{3\sigma_T}{16\pi} \left[I'_x (\hat{e}'_y \cdot \hat{e}_x)^2 + I'_y (\hat{e}'_y \cdot \hat{e}_y)^2 \right] = \frac{3\sigma_T}{16\pi} I' \cos^2 \theta, \quad (2.42)$$

which can be inverted to obtain the I and Q Stokes parameters of the scattered wave

$$I = I_x + I_y = \frac{3\sigma_T}{16\pi} I' (1 + \cos^2 \theta), \quad (2.43)$$

$$Q = I_x - I_y = \frac{3\sigma_T}{16\pi} I' \sin^2 \theta. \quad (2.44)$$

The U parameter can be calculated by rotating the reference frame by 45° , therefore substituting U with Q (that is $Q' = U$ and $U' = -Q$). The final expression for the three Stokes parameters of interest can be obtained by integrating over all possible incoming

directions, thus obtaining

$$I = \frac{3\sigma_T}{16\pi} \int d\Omega I'(\theta, \phi)(1 + \cos^2 \theta), \quad (2.45)$$

$$Q = \frac{3\sigma_T}{16\pi} \int d\Omega I'(\theta, \phi) \sin^2 \theta \cos(2\phi), \quad (2.46)$$

$$U = \frac{3\sigma_T}{16\pi} \int d\Omega I'(\theta, \phi) \sin^2 \theta \sin(2\phi), \quad (2.47)$$

where the dependence on ϕ in the integral for Q and U derives by the necessity to have a common frame for the various directions of incidence, than the initially chosen, convenient for calculations. Finally expanding $I'(\hat{\mathbf{n}})$ in spherical harmonics, $I'(\hat{\mathbf{n}}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{\mathbf{n}})$, is obtained:

$$I = \frac{3\sigma_T}{16\pi} \left[\frac{8}{3} \sqrt{\pi} a_{00} + \frac{4}{3} \sqrt{\frac{\pi}{5}} a_{20} \right], \quad (2.48)$$

$$Q - iU = \frac{3\sigma_T}{4\pi} \sqrt{\frac{2\pi}{15}} a_{22}. \quad (2.49)$$

These expressions show that the production of linear polarization is determined by the presence of a quadrupole term in the distribution of the intensity of the radiation around the electron. Note that, in a more rigorous treatment we have to consider the Compton cross section and the evolution of the Boltzmann equation. Even in this case, the result shows that the polarization part is homogeneous except for the source term given by the quadrupole radiation (see, e.g. [Dodelson 2003](#)). For the V -mode polarization, instead, we obtain an equation that is homogeneous and completely decoupled from other parameters, and, it has therefore no source term. Thus, if we initially have $V = 0$, we obtain a solution where V will not be produced. Note, however, that there are processes, other than the Thomson scattering, capable of producing circular polarization.

2.3.1 The reionization fingerprint

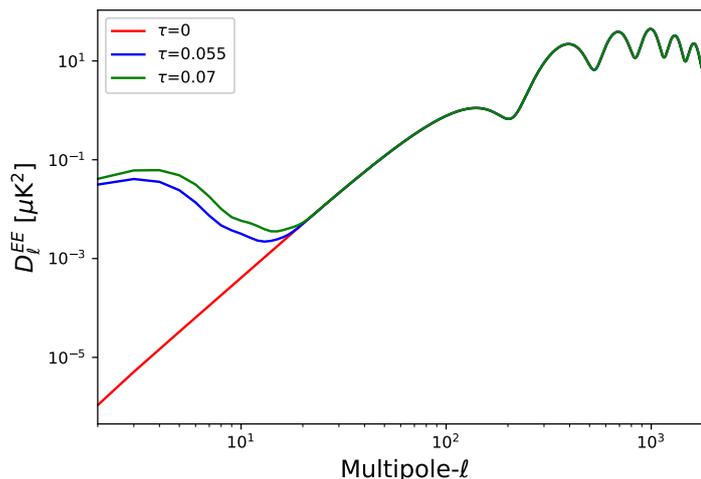


Figure 2.3: Shape of the normalized EE power spectrum, $\mathcal{D}_\ell^{EE} = \frac{\ell(\ell+1)}{2\pi} C_\ell^{EE}$, for three different values of the reionization optical depth, τ , in a flat Λ CDM model.

The physics behind the polarization production ensures that no polarization can be generated after decoupling if there is no new phase of ionisation. Precisely, before recombination, we have seen that photon and baryon form a tightly coupled system, in which the

damping scale is a few Mpc. This produces, in the electron rest frame, a nearly isotropic photon distribution function. As photons and electrons decouple, the mean-free path of the photons starts to grow and the temperature quadrupole moment is produced by free streaming. On the other hand, electrons are in different regions with slightly different velocities. Photons scattering off these electrons will have an intensity that, for a fixed wavelength, depends on the direction. For wavelengths longer than the width of the last scattering surface $\Delta\eta_d$ the polarization perturbation can be shown to be (Zaldarriaga and Harari 1995)

$$\Theta^P = 0.51(1 - \mu^2)e^{ik\mu(\eta_d - \eta_0)}k\Delta\eta_d\Theta_1^T(\eta_d), \quad (2.50)$$

where Θ_1^T is the temperature dipole term, $\mu = \hat{\mathbf{k}} \cdot \hat{\mathbf{n}}$ with $\hat{\mathbf{n}}$ equal to the direction of photon propagation, and η_d is the conformal time of decoupling. Thus, since for the standard adiabatic initial conditions Θ_1^T and the baryon velocity vanish (in the tight coupling regime $\Theta_1^T \propto v_b$) as $k\eta \rightarrow 0$, in absence of a reionization phase ($\tau = 0$) we have a dramatic fall of polarization for large angular scales, see Fig. 2.3. For smaller angular scales, $\ell \geq 100$, the same acoustic oscillations that generate the Doppler peaks in the temperature anisotropy cause the peaks in the polarization spectrum. The polarization peaks are located at different ℓ values with respect to temperature peaks because they occur for different wave vectors. Finally, also the polarization power spectra is damped due to the finite width of the last scattering surface.

However, on large angular scales, there is the presence of a peak in the polarization power spectrum that is not taken into account from the previous explanation. This is explained by the presence of an epoch reionization³. In this scenario, the visibility function, $g(\eta) = \dot{\tau}e^{-\tau}$, i.e. the probability that a photon reaching the observer last scattered between η and $\eta + d\eta$, has two peaks: one at recombination and the other due to reionization. This allows us to separate the formal line of sight solution for the polarization perturbation

$$\Theta^P = -\frac{1}{2} \int_0^{\eta_0} d\eta e^{ik\mu(\eta - \eta_0)} g(\eta) [1 - \mathcal{P}_2(\mu)] \Pi, \quad (2.51)$$

where $\Pi = \Theta_2^T + \Theta_2^P + \Theta_0^P$, in two pieces:

$$\Theta^P = -\frac{1}{2} [1 - \mathcal{P}_2(\mu)] \left\{ \int_0^{\eta_r} d\eta e^{ik\mu(\eta - \eta_0)} g(\eta) \Pi + \int_{\eta_r}^{\eta_d} d\eta e^{ik\mu(\eta - \eta_0)} g(\eta) \Pi \right\}, \quad (2.52)$$

where η_r is the conformal time of the start of reionization. The first integral just represents the polarization generated at recombination, and it is equal to $e^{-\tau_r} \Theta_{\text{NR}}^P$, with Θ_{NR}^P equal to the polarization that would be measured if there was no reionization. This contribution is damped because only a fraction $e^{-\tau_r}$ of the photons that arrive to the observer came directly from recombination without scattering again after reionization. The second integral is the new contribution coming from reionization. Since the temperature has a quadrupole term coming from the free streaming of the monopole at recombination, while the polarization terms do not grow after decoupling, as a first approximation we have $\Pi = \Theta_2^T$. In particular, the quadrupole at η_r arising from the free streaming of monopole at recombination is (Zaldarriaga 1997)

$$\Theta_2^T(\eta_r) = (\Theta_0^T + \Phi)(\eta_d) j_2 [k(\eta_r - \eta_d)], \quad (2.53)$$

where j_2 is the $\ell = 2$ spherical Bessel function. The first peak corresponds approximately at the first peak of the Bessel function, and its position is given by $\ell \sim 2\sqrt{z_r}$, where z_r is the redshift at which reionization starts. Note that, only the first peaks appear because the reionization scattering surface is very wide and thus the integrand in Eq. (2.52) for smaller wavelengths oscillates during its width and cancels out after integration. This

³See Sec. 1.2 and Sec. 2.2.1.

cancellation makes the new polarization small and thus hidden under the polarization generated at recombination, see Fig. 2.3.

Polarization provide the best probe for the characterisation of reionization in the context of CMB measurement. In particular, for the large angular scales, the EE power spectrum scales as $C_\ell^{EE} \propto \tau^2$, against the linear dependency of the TT and TE spectra (i.e. $C_\ell^{TT}, C_\ell^{TE} \propto \tau$).

2.4 Choice of cosmological parameters

Until now⁴, we have outlined five parameters determining the background homogeneous spacetime (matter density Ω_m , radiation density Ω_r , vacuum energy density Ω_Λ , curvature density Ω_k , and Hubble parameter h), four parameters determining the spectrum of primordial perturbations (scalar and tensor amplitudes A_s and A_t and power-law indices n_s and n_t), and a single parameter τ describing the total optical depth since reionization. However, the current data analyses done to constraint the cosmological model are driven by CMB anisotropy experiments, and in particular they rely on the analyses of CMB power spectra. There is one near-exact degeneracy (the geometric degeneracy) and several approximate degeneracies in the parameters describing the CMB power spectrum. These degeneracies increase the time needed to find the parameter values that fit the model on data. Indeed, a poor choice of parameter imply to spend time exploring degeneracy directions. This statement will be clear once the Markov Chain Monte Carlo techniques have introduced, see Sec. 3.3.3. The effects of these degeneracies are reduced by finding a combination of cosmological parameters that have essentially orthogonal effects on the angular power spectrum. As pointed out in Kosowsky et al. (2002), within the context of fitting a flat ($\Omega_k = 0$) Λ CDM model to a CMB power spectrum, six key parameters are primarily chosen to avoid degeneracies and thus speed convergence of the model fit to the data. There are two parameters for the *physical energy density today* of baryons, $\Omega_b h^2$, and cold dark matter, $\Omega_c h^2$. There is a parameter for the *characteristic angular scale of the acoustic peaks*,

$$\theta^* = r_s^*/D_A^*, \quad (2.54)$$

where the * indicate that the involved quantities are evaluated at the decoupling, r_s^* is he *sound horizon* at decoupling

$$r_s^* = \frac{c}{H_0 \sqrt{3}} \int_0^{a^*} \left\{ \left(1 + \frac{3\Omega_b}{4\Omega_\gamma} \right) \left[(1 - \Omega_m - \Omega_\Lambda) z^2 + \Omega_\Lambda z^{1-3w} + \Omega_m z + \Omega_r \right] \right\}^{-1/2} dz, \quad (2.55)$$

and D_A^* is the angular diameter distance at decoupling (see Eq. (1.97)). The reionization is parametrised through the *Thomson scattering optical depth due to reionization*, τ . The remaining two core parameters are the *scalar spectrum power-law index*, n_s , and the *log power of the primordial curvature perturbations*, $\ln(10^{10} A_s)$. Both are normalised at the pivot scale $k_0 = 0.05 \text{ Mpc}^{-1}$.

For more complex models we usually add other parameters. For example, we include the tensor-to-scalar ratio, r , to constraint the tensor perturbation. Any new physics that affects the damping tail of the CMB spectrum, such as additional relativistic particles, can alter the constraints on the scale-invariance of the spectrum. To study the scale dependence of primordial fluctuations we can include the running of the scalar spectral index, $dn_s/d \ln k$. The geometry of the Universe is parametrized adding as extra parameter the curvature density, Ω_k . We can also study the neutrino physics allowing the variation of its mass (the sum of neutrino masses) m_ν , or the effective number of neutrino species

⁴See Chapter 1.

N_{eff} . Another useful extension is the Δ -parametrisation of the lack-of-power anomaly, see Chapter 5.

2.5 Foregrounds

There are other astrophysical sources that are between the last-scattering surface and us. These are called *foreground*, and their presence, despite the great interest in other branches of physics, represents an obstacle for a CMB study. However, they have a different frequency behaviour with respect the CMB anisotropies, allowing us to distinguish them. To remove these emissions and clean the measured data maps, it is necessary to know what type of processes we are dealing with and how to parametrise them.

- **Synchrotron:** It is a diffuse emission due to spiraling electrons in the galactic magnetic fields. This radiation may be highly polarised, up to a 75%. The spectrum of synchrotron emission is basically flat at frequencies $\nu \leq 20$ GHz and then, for higher frequencies, follows an exponential law with a negative index $\beta_s \sim -3$.
- **Free-free:** It is bremsstrahlung emission coming from electron-ion collision. Its spectrum is close to a power law for frequencies greater than 1 GHz and presents a visible break at lower frequencies. This happens because the medium becomes optically thick, in addition the brightness temperature becomes equal to the electron temperature. Free-free spectrum is similar to the synchrotron one at low frequencies, but can be distinguished because its power-law index is flatter than the synchrotron one.
- **Spinning dust:** It is dust grain having non-zero dipole moment that rotates and emits in the microwave region of the electromagnetic spectrum. The frequency spectrum shows a peak between 25 and 30 GHz and then it follows a power law in analogy with the synchrotron and free-free case.
- **CO lines:** They are emission lines of carbon monoxide (CO). It is possible to separate these lines from the other diffuse components and to describe parametrically in terms of an amplitude $a(p)$ inside the corresponding detector map.
- **Thermal dust:** It is the dominant component at frequencies $\nu > 100$ GHz. Its characteristic spectrum is a modified black body with a free emissivity index β_d and a characteristic temperature T_d . Thermal dust gives its contribution also in polarization, because aspherical dust grains tend to distribute along the local magnetic field lines. This behaviour translates into a polarised emission in the microwave band with the same thermal-dust spectrum.
- **Thermal SZ:** The deviation from the black body spectrum of photons which undergo inverse compton scattering leaves an imprint increasing the brightness on high frequencies.

Chapter 3

Overview on data analysis techniques

The comparison between theories and observations is an essential part of physics. Extract information about physical quantities from data, often subject to various sources of uncertainty, is the crucial step needed to confirm or exclude a theory. The amount of data that needs to be analysed to achieve this goal is considerably high. Moreover, the complexity of our theories and experiments designed to test them increases with time. Nevertheless, limited resources require an excellent ability to make forecasts to obtain the highest scientific return.

The discipline that provides the tools to reach this purpose is Statistics. Its mathematical foundations were laid in the 17th century with the development of the probability theory, even if the first use of some basic concepts can be date back to the 8th century. There are many distinct interpretations of the word probability, but for the sake of brevity, we group them in two extreme interpretations: frequentist (or objective) and Bayesian (or subjective). Needless to say, both of them give the framework in which we can use the data to perform statistical inference about the underlying physical model.

Cosmologists often face difficult and computationally intensive inference problems (see, e.g., [Trotta 2008](#)). This led to push the usage of the Bayesian framework. However, there are varieties of analyses that are mostly done using the frequentist approach. Furthermore, there is also an increasing trend in the usage of information theory to make forecasts or to check the consistencies between datasets. In this chapter, we are going to explore these three different approaches, analysing the mathematical aspects needed to make inference. The specific application to cosmological problems is postponed to the subsequent chapters.

3.1 Some concepts of probability

The Bayesian and frequentist approaches are both based on their axiomatic definition, which makes them conceptually different. However, both are based on some basic notions and definitions that result independent from the particular framework in which we are working (see, e.g., [Ross \(1998\)](#)). It is clear that to deeply understand them, the notion of probability is necessary, which will be given in the following sections. Thus, the aim here is only to assimilate familiarity with the notation.

When we are performing an experiment, we define the *sample space* S as the collection of all its possible outcomes, and each of its subsets is indicated as an *event* (see, e.g., [Ross \(1998\)](#)). Thus, if we perform a coin tossing experiment, the sample space is the set of tail (T) and head (H), $S = \{T, H\}$. The two possible events are $A_1 = \{T\}$ and $A_2 = \{H\}$. Once we have an event A , we can define its complement with respect to S as $A^c \equiv S \setminus A$. With this nomenclature, we would like to quantify the probability that an event occurs. Usually, this is reached by assigning a measure that goes from the sample space S to the

real interval $[0, 1]$. We will denote the probability that A occurs as $P(A)$.

Given two events, A and B , the probability that both A and B occur is the *joint probability* (or *intersection probability*), $P(A, B)$ (or $P(A \cap B)$). The probability that A or B occur is, instead, the *union probability*, $P(A \cup B)$. In many situations, however, we are only concerned with those outcomes that are elements of B . This means that the original sample space S collapse in its subset B , that become our new sample space. Thus we define the *conditional probability*, $P(A|B)$, as the probability of A taking place given the occurrence of B . In other words, conditional probabilities are just the probabilities that we apply after revisions of our belief when someone gives us additional information (B occurs).

Two events are said to be *statistically independent* if the occurrence of one is independent on the occurrence of the other (Caticha 2008; Trotta 2008). This happens if and only if

$$P(A, B) = P(A)P(B). \quad (3.1)$$

It can be easily shown that this imply

$$P(A|B) = P(A). \quad (3.2)$$

If the two events are not independent, their joint probability is

$$P(A, B) = P(A|B)P(B). \quad (3.3)$$

Given that the two events enter symmetrically in the left-hand side of the above equation, we might as well write

$$P(A, B) = P(B|A)P(A). \quad (3.4)$$

Equating the two right-hand side of these last two equation, we get *Bayes' theorem*:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (3.5)$$

This theorem follows from the relation between joint and conditional probabilities, that is, it can be derived independently of the Bayesian interpretation of probability.

In many experiments, the elements of sample space are not necessarily numbers. For example, in our coin tossing experiment the sample space consists of tail and head. This is achieved through the notion of *random variable* X , that is a function from the sample space S into the set of real numbers \mathbb{R} (see, e.g., Ross 1998). If the random variable X is continuous, for every set of real numbers A we define the probability that $X \in A$ as

$$P(X \in A) = \int_A dx p(x), \quad (3.6)$$

where $p(x)$ is called the *probability density function* (pdf) of the continuous random variable X .

In physical application, in practice, possible events can always be mapped to numerical quantities. This means that the random variable corresponds to the physical quantity that we want to measure, while the measurement itself correspond to the possible outcome of an experiment (the event). This event can be one-dimensional or a vector, and in both cases, we are interested in using them (the data) to perform statistical inference about the underlying physical model. Such kind of problem forces us to introduce an *hypothesis* H , that is something testable on the basis of observations. Given the hypothesis, the natural question arises as to whether or not it (a theory) is supported by data. The Bayesian method always compares the probability of competing models, while frequentist hypothesis testing seeks to disprove a hypothesis by showing that the observed data would

not be likely if the hypothesis were true. From the frequentist point of view, there is no probability associated with parameters or models, it is only the data that are probabilistic.

At this stage, before entering in the details of the two main probabilistic point of view, it will be useful to define some quantities and distributions that we will use in the subsequent chapters. We can start defining the *cumulative distribution function* (CDF) as the function of x describing the probability of the measured value being lower than X , $P(X < x) \equiv F(x)$:

$$F(x) = \int_{-\infty}^x dx' p(x'), \quad (3.7)$$

where $p(x)$ is the pdf of the random variable X . The *expected value* of any function of a random variable X , $f_X(x)$ is defined as:

$$\mathbb{E}[f_X] = \langle f_X \rangle = \int_{-\infty}^{+\infty} dx f_X(x)p(x). \quad (3.8)$$

In particular, the expectation value of a random variable itself is the *mean*. Furthermore, if X is a random variable with mean μ , then its *variance* is defined as

$$\text{Var}[X] = \mathbb{E}[(X - \mu)^2], \quad (3.9)$$

and, in general, its *n-th moment* is given by $\mathbb{E}[X^n]$. If X is an m -dimensional random variable, then we can define the *covariance* between X_i and X_j as

$$\text{Cov}[X_i, X_j] = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]. \quad (3.10)$$

In the next subsections, we are going to introduce some of the probability distributions used in Cosmology and of essential importance for this thesis. We are also going to enunciate the Central Limit Theorem. These represent only a few of the most widely applicable distributions that come up very often in statistics. For a more complete overview on probability distributions see, e.g., [Gregory \(2005\)](#).

Multivariate normal distribution

As we will see, its importance also relies on the fact that it represents the maximum entropy distribution for a fixed variance (see Sec. 3.4.1). The functional form of a multivariate Gaussian distribution is

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \quad (3.11)$$

where \mathbf{x} is an n -dimensional vector, $\boldsymbol{\mu}$ is the mean and $|\cdot|$ indicate the determinant of the covariance matrix Σ . If \mathbf{x} follows a normal distribution with mean $\boldsymbol{\mu}$ and covariance Σ it is often used the notation $\mathbf{x} \in \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$.

The normal (Gaussian) distribution is the most important and widely used. One of the reasons why the normal distribution is so useful is because of the *Central Limit Theorem* (see below).

Central Limit Theorem

This theorem states that, if X_1, X_2, \dots, X_N are a collection of independent random variable, each with finite expectation value μ_i and finite variance σ_i^2 , then the variable

$$Y = \frac{\sum_{i=1}^N (X_i - \mu_i)}{\sum_{i=1}^N \sigma_i^2}, \quad (3.12)$$

is distributed as a Gaussian with expectation value 0 and unit variance, for $N \gg 1$.

χ^2 distribution

The χ^2 distribution with n degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables. That is, if X_1, \dots, X_n are independent normal random variables, each with finite expectation value μ_i and finite variance σ_i^2 , then the χ_n^2 is defined as

$$\chi_n^2 = \sum_{i=1}^n \frac{(X_i - \mu_i)^2}{\sigma_i^2}. \quad (3.13)$$

Its functional form is defined as

$$p(x = \chi_n^2) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} \\ 0 \end{cases} \quad \text{otherwise} \quad (3.14)$$

Here $\Gamma(\cdot)$ is the Gamma function is defined as

$$\Gamma(x) = \int_0^\infty dt e^{-t} t^{x-1}. \quad (3.15)$$

The mean of this distribution is $\langle x \rangle = n$ and the variance $Var[x] = 2n$.

Wishart distribution

The Wishart distribution arises in a natural way as a matrix generalization of the χ^2 distribution. When the X_1, \dots, X_n are p -dimensional random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$ rather than real-valued random variables, such that $\mathbf{X}_i \in \mathcal{N}(\mathbf{X}_i; \mathbf{0}, \mathbf{I}_p)$, one possible way to generalize the sum of squares is to form the $p \times p$ positive semidefinite matrix

$$\mathbf{S} = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top. \quad (3.16)$$

This $p \times p$ matrix \mathbf{S} follows a Wishart distribution with n degree of freedom. For $n \geq p$, its functional form is given by

$$p(\mathbf{S}) = \frac{1}{2^{np/2}\Gamma_p(n/2)} |\mathbf{S}|^{(n-p-1)/2} \exp\left\{-\frac{1}{2}\text{Tr}[\mathbf{V}^{-1}\mathbf{S}]\right\}, \quad (3.17)$$

where \mathbf{V} is the covariance matrix of size $p \times p$. Here $\Gamma_p(\cdot)$ is the multivariate gamma function

$$\Gamma_p(x/2) = \pi^{p(p-1)/4} \prod_{j=1}^p \Gamma\left(\frac{x-j+1}{2}\right). \quad (3.18)$$

3.2 Frequentist interpretation of probability

The main difference between the two schools of thought relies on the definition of probability. The frequentist definition basically depends on the exact reproducibility of an experiment. From this point of view, we can define the probability as the number of times an event occurs divided by the total number of events, in the limit of an infinite series of equiprobable trials (see, e.g., [Ross 1998](#)). It is clear that this definition is in a way unsatisfactory: nobody can reproduce an experiment in the same way twice, or an infinite number of times. However, it has the advantage that we can construct probabilities using rules and then easily interpret the results. Furthermore, it is objective in the sense that, once the exact experimental setup is defined, any individual can then decide if the null hypothesis is rejected or not.

This objective school uses basically the theory developed by the Russian mathematician Kolmogorov, which gave it a solid foundation by using measure theory. It is based on three axioms (Ross 1998). Let S be the sample space of an experiment, then

$$\begin{aligned} (A1) \quad & P(A) \geq 0, \quad \forall A \in S \\ (A2) \quad & P(S) = 1, \\ (A3) \quad & P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i), \\ & \text{if } A_1, \dots, A_k, \dots \text{ are mutually disjoint events of } S. \end{aligned}$$

These axioms allows us to derive some basic relation to construct probabilities. Specifically, if A is an event of the sample space S , then

$$P(A \cup A^c) = P(A) + P(A^c) = 1. \quad (3.19)$$

If A and B are any two events, then

$$P(A \cup B) = P(A) + P(B) - P(A, B). \quad (3.20)$$

3.2.1 Hypothesis testing

In any hypothesis test, the first step is to state the hypothesis as a well-posed true or false question. Once the hypothesis statement is done, we choose or invent a function of data points (a *statistic*) that should be affected by the truth of the hypothesis. The simplest examples are the arithmetic mean and the variance. Sometimes this function of data is meant to be an estimate of a parameter in the model, and in this case it is called an *estimator*. If the average of this estimator, over all possible data sets of the same size, is not equal to the “true value”, the estimator is said to be *biased*. If we increase the amount of data, this bias will become smaller if our estimator is a good one. Furthermore, if the bias goes to zero for an “infinitely large” data set, then we say it is *asymptotically unbiased*. At this point, we can determine, by analytic or numerical methods, the probability distribution of the statistic. The last step is to calculate the statistic with the data and determine if the measured value is improbable if the hypothesis is true. This is done by “counting” the p fraction of repeated trials under the condition that the null hypothesis is correct. The smaller it is the more evidence you have against the null hypothesis. This number p is known as *p-value* (or significance of the test). Unfortunately, p -values are often incorrectly viewed as the probability that the hypothesis is true. However, there is no objective means for deciding the latter without specifying an alternative hypothesis, H_1 , to the null hypothesis H_0 .

3.2.2 The principle of Maximum Likelihood

Many estimators can in principle be defined. One particular choice, and often the only reasonable one, is the *Maximum Likelihood Estimator* (MLE), defined as the value that maximizes the *likelihood function*. The likelihood is the probability of obtaining the observed data \mathbf{d} given the hypothesis H , regarded as a function of H :

$$L(H) \equiv P(\mathbf{d}|H). \quad (3.21)$$

It follows that, if the theory is characterized by the values $\boldsymbol{\theta}$ of the model parameters, the likelihood is expressed as a function of the parameters, $L(\boldsymbol{\theta})$. A likelihood $L(\boldsymbol{\theta})$ is not itself a probability for $\boldsymbol{\theta}$; it is a dimensionless numerical function which, when multiplied by a

prior probability and a normalization factor, may become a probability (Jaynes 2003). With this notation, we define the MLE value $\hat{\boldsymbol{\theta}}_{\text{ML}}$ as

$$\hat{\boldsymbol{\theta}}_{\text{ML}} \equiv \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}). \quad (3.22)$$

It is often more convenient to maximise the logarithm of the likelihood (the “log-likelihood”, $\mathcal{L} \equiv \log L$) instead. Since log function is monotonic, maximising the likelihood is the same as maximising the log-likelihood.

After an experiment, for each estimation we have to express uncertainty in our knowledge. In the frequentist approach, this is done through the *confidence interval*, that is a range of values designed to include the true value of the parameter with some minimum probability, (see, e.g., Trotta 2017). In particular, if the likelihood function can be approximated as a Gaussian (at least around the peak), we can use the results for a Gaussian distribution to approximate the probability content of an interval around the MLE estimate for the mean μ . In general, the interval $[\mu_{\min}, \mu_{\max}]$ is called a $100\alpha\%$ confidence interval for the mean μ if $P(\mu_{\min} < \mu < \mu_{\max}) = \alpha$. For instance, if $\alpha = 0.95$, one speaks of a 95% confidence interval. Note that this does not mean that the interval constructed from the observed data has 95% probability of containing the true value of the parameter. In order to speak of probability of the parameter we need the Bayesian framework.

3.2.3 Fisher information and the minimum variance limit

It is important to know that any estimator suffers from an important limitation: there exists an absolute lower bound on the variance of any estimator of a parameter. To derive this bound, let us consider the normalized likelihood function

$$\int d\mathbf{x} L(\mathbf{x}|\boldsymbol{\theta}) = 1. \quad (3.23)$$

Differentiating this equation and using the equality $\partial L/\partial\theta_i = L\partial \ln L/\partial\theta_i$ (see, e.g. Frieden and Gatenby 2006), it follows that

$$\int d\mathbf{x} \frac{\partial}{\partial\theta_i} L(\mathbf{x}|\boldsymbol{\theta}) = \int d\mathbf{x} L(\mathbf{x}|\boldsymbol{\theta}) \frac{\partial \ln L(\mathbf{x}|\boldsymbol{\theta})}{\partial\theta_i} \quad (3.24)$$

$$= \left\langle \frac{\partial \ln L(\mathbf{x}|\boldsymbol{\theta})}{\partial\theta_i} \right\rangle = 0, \quad (3.25)$$

where the expected value $\langle \dots \rangle$ follows from the definition in (3.8). Differentiating this again gives the relation

$$\mathcal{F}_{ij} \equiv \left\langle \frac{\partial \ln L}{\partial\theta_i} \frac{\partial \ln L}{\partial\theta_j} \right\rangle = - \left\langle \frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j} \right\rangle, \quad (3.26)$$

where \mathcal{F}_{ij} is known as the *Fisher information matrix*. It is the average of the curvature (or Hessian matrix) of the log-likelihood. If it is evaluated at the maximum of the likelihood, \mathcal{F}_{ij} measures the rate at which the posterior drops off from its maximum in parameter space on average, i.e. how “pointy” the peak is. Note that the Fisher matrix is not a function of any data set, but depends on the properties of the statistical model.

Let us now suppose that we have an estimator for the parameter θ_i , which we denote $\hat{\theta}_i$. Let us define the bias $b(\boldsymbol{\theta})$, which could be zero or not. It follows that

$$\int d\mathbf{x} \hat{\theta}_i(\mathbf{x}) L(\mathbf{x}|\boldsymbol{\theta}) = \theta_i + b(\boldsymbol{\theta}). \quad (3.27)$$

Taking the derivative with respect to θ_i gives

$$\int d\mathbf{x} \hat{\theta}_i(\mathbf{x}) \frac{\partial \ln L}{\partial \theta_i} L = 1 + \frac{\partial b(\boldsymbol{\theta})}{\partial \theta_i}. \quad (3.28)$$

Using the relation (3.25) follows

$$\left\langle \left(\theta_i - \hat{\theta}_i(\mathbf{x}) \right) \frac{\partial \ln L}{\partial \theta_i} \right\rangle = 1 + \frac{\partial b(\boldsymbol{\theta})}{\partial \theta_i}, \quad (3.29)$$

that jointly with the Cauchy-Schwarz inequality give us the relation (see, e.g. [Frieden and Gatenby 2006](#))

$$\left[\left\langle \left(\theta_i - \hat{\theta}_i(\mathbf{x}) \right) \frac{\partial \ln L}{\partial \theta_i} \right\rangle \right]^2 \leq \text{Var} [\hat{\theta}_i] \text{Var} \left[\frac{\partial \ln L}{\partial \theta_i} \right] = \text{Var} [\hat{\theta}_i] \mathcal{F}_{ij}. \quad (3.30)$$

In other word, we have

$$\text{Var} [\hat{\theta}_i] \geq \frac{\left(1 + \frac{\partial b(\boldsymbol{\theta})}{\partial \theta_i} \right)^2}{\mathcal{F}_{ij}}. \quad (3.31)$$

This relation is known as *Cramér-Rao limit*, and it is the absolute lower bound mentioned at the beginning of this subsection. An estimator that reaches this bound is called an *efficient estimator*.

The Fisher matrix and the Cramér-Rao limit on the variance are vastly used in cosmology as a way of forecasting errors. Nevertheless, this method suffers of several criticisms. Indeed, the Fisher matrix can be quite different for different fiducial parameter values. Furthermore, it does not account for degeneracies between parameters, although there are approximations that try to take this into account. In particular, one can use the expansion of the likelihood around the MLE $\hat{\boldsymbol{\theta}}$ to approximate the posterior of a future experiment as

$$p(\boldsymbol{\theta}) \simeq \frac{|\mathcal{F}|}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top \mathcal{F}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right\}, \quad (3.32)$$

where \mathcal{F}^{-1} represents the parameter covariance matrix. The variance of a single parameter after marginalizing over all the other parameters is

$$\sigma_i^2 \simeq \mathcal{F}_{ii}^{-1}. \quad (3.33)$$

In this approximation one can easily add priors on the parameters from other experiments. Lets C_{prior}^{-1} be the inverse covariance of the parameters (or the Fisher) from some previous experiment. Then, since the log of the posterior is the sum of the log of the likelihood and the prior it follows

$$\mathcal{F}^{\text{tot}} = \mathcal{F} + C_{\text{prior}}^{-1}. \quad (3.34)$$

3.2.4 The Fisher matrix analysis of CMB

When performing a parameter error forecast for future CMB experiments, it is customary to use the Fisher matrix formalism in which the formal error bar on a given parameter can be estimated from the derivatives of the observables with respect to the model parameters around the best-fit point. We have seen that, the Fisher matrix technique allows for a quick, analytic estimate of the confidence limits by approximating the likelihood function $L(\boldsymbol{\theta})$ as a multivariate Gaussian function of the theoretical parameters $\boldsymbol{\theta}$.

If we use the Gaussian approximation, Eq. (2.15) can be generalized as

$$L(\boldsymbol{\theta}) \propto \exp \left\{ -\frac{1}{2} \mathbf{a}^\dagger [\mathbf{C}(\boldsymbol{\theta})]^{-1} \mathbf{a} \right\}, \quad (3.35)$$

where $\mathbf{a} = (a_{\ell m}^T, a_{\ell m}^E, a_{\ell m}^d)^\top$, and $C(\boldsymbol{\theta})$ is the theoretical data covariance matrix. Here we consider lensing through the presence of the harmonic decomposition of the deflection field, $a_{\ell m}^d$. Furthermore, we ignore the presence of tensor perturbations. With this likelihood, Eq. (3.26) becomes

$$\mathcal{F}_{ij} = \sum_{\ell=2}^{\ell_{\max}} \sum_{PP', QQ'} \frac{\partial C_\ell^{PP'}}{\partial \theta_i} (\text{Cov}_\ell^{-1})_{PP' QQ'} \frac{\partial C_\ell^{QQ'}}{\partial \theta_j}, \quad (3.36)$$

where $PP', QQ' \in \{TT, EE, TE, dd, Td\}$, and the matrix Cov_ℓ is the power spectrum covariance matrix at the ℓ -multipole (for its explicit expression see [Perotto et al. 2006](#)).

The advantage of the Fisher matrix technique is that it is computationally tractable, and involves much less numerical machinery than a Markov Chain Monte Carlo exploration of the parameter space, see Sec. 3.3.3. However, it is important to note that, since $L(\boldsymbol{\theta})$ is generally a rather complicated function of $\boldsymbol{\theta}$, this approximation will likely lead to incorrect results, (see, e.g., [Perotto et al. 2006](#)). The Taylor expansion is valid only in regions close to the best fit point. Furthermore, the Fisher matrix is sensitive to small numerical errors in the computation of the derivatives $\partial C_\ell^{PP'}/\partial \theta_i$, and elements that are close to zero can be amplified significantly when inverting the matrix.

3.3 Bayesian interpretation of probability

The Bayesian interpretation defines probabilities as subjective assignments based on rational thought. In other words, in this framework the probability is a measure of the confidence of belief about a proposition. That is, the Bayesian approach allows us to directly compute the probability of any particular theory or particular value of a model parameter. This is in contrast with the frequentist statistical approach, that can address only indirectly this issue. Nevertheless, degrees of belief can be mapped onto probabilities if they satisfy simple consistency rules known as the Cox axioms ([Cox 1946](#)). Precisely, let the degree of belief in proposition A be denoted by $B(A)$. Since its negation corresponds to its complementary, NOT- A is written as A^c . Furthermore, the degree of belief in a conditional proposition, A , assuming proposition Y to be true, is represented by $B(A|Y)$. The Cox axioms can be stated as:

(A1) Degrees of belief can be ordered; if $B(A)$ is “greater” than $B(Y)$, and $B(Y)$ is “greater” than $B(Z)$, then $B(A)$ is “greater” than $B(Z)$. As a consequence beliefs can be mapped onto real numbers.

(A2) The degree of belief in a proposition A and its negation A^c are related, i.e. there is a function f such that

$$B(A) = f[B(A^c)]. \quad (3.37)$$

(A3) The degree of belief in a conjunction of propositions A, Y (A AND Y) is related to the degree of belief in the conditional proposition $A|Y$ and the degree of belief in the proposition Y . That is, there is a function g such that

$$B(A, Y) = g[B(A|Y), B(Y)]. \quad (3.38)$$

If a set of beliefs satisfy these axioms then they can be mapped onto probabilities satisfying $P(\text{FALSE}) = 0$, $P(\text{TRUE}) = 1$, and $0 \leq P(A) \leq 1$. The operations for manipulating probabilities that follow from these axioms are the sum and product rules ([Gregory 2005](#)):

$$P(A|B) + P(A^c|B) = 1 \quad (3.39)$$

$$P(A, B|C) = P(A|C)P(B|A, C) \quad (3.40)$$

$$= P(B|C)P(A|B, C). \quad (3.41)$$

In the Bayesian approach, probability calculations often fall into one of two categories: *forward probability* and *inverse probability*. Forward probability problems involve a generative model that describes a process that is assumed to give rise to some data; the task is to compute the probability distribution or expectation of some quantity that depends on the data. Like forward probability problems, inverse probability problems involve a generative model of a process, but instead of computing the probability distribution of some quantity produced by the process, we compute the conditional probability of one or more of the unobserved variables in the process, given the observed variables. The underlying idea behind this inversion follows from the Bayes' theorem in Eq. (3.5), when we replace the event A with the hypothesis H made on data, and the event B with the data \mathbf{d} itself:

$$P(H|\mathbf{d}) = \frac{P(\mathbf{d}|H)P(H)}{P(\mathbf{d})}. \quad (3.42)$$

This results in a peculiar interpretation of the various objects involved. Precisely, on the left, we have the *posterior probability* of the hypothesis, representing our degree of belief about the hypothesis after we have seen the data \mathbf{d} . As we have already seen in Eq. (3.21), $P(\mathbf{d}|H)$ is the likelihood function. The importance of knowing the functional form of the likelihood is highlighted in inverse probability problems (inference problems), where the parameters are usually estimated from the experimental data. The quantity $P(H)$ is the *prior probability distribution*, which represent our prior knowledge of H before we see the data. The denominator, $P(\mathbf{d})$, is a normalization called the *evidence* that, by the law of total probability, is given by

$$P(\mathbf{d}) = \int dH P(\mathbf{d}|H)P(H). \quad (3.43)$$

From Eq. (3.42), it is now clear what inversion of probability means. Furthermore, the relation between prior probability and posterior probability provides a natural way to update probabilities.

3.3.1 Parameter estimation

The most common use for Bayesian inference is parameter estimation. In this case, the model is assumed to be true and the hypothesis space of interest concerns the values of the model parameters $\boldsymbol{\theta}$. Parameter estimation concerns the use of Bayes' theorem to determine what one can learn about the values of parameters from data. This corresponds to the computation of the posterior $P(\boldsymbol{\theta}|\mathbf{d})$, which depends on the likelihood function and our prior knowledge on the parameters. Its information content for the model parameters is usually compressed in terms of a point estimate and/or an interval for each parameter $\theta_i \in \boldsymbol{\theta}$. Possible summaries of the best-fit values are the posterior mode (most probable value)

$$\hat{\theta}_i = \max_{\theta_i} P(\theta_i|\mathbf{d}), \quad (3.44)$$

or the posterior mean

$$\hat{\theta}_i = \langle \theta_i \rangle = \int d\theta_i \theta_i P(\theta_i|\mathbf{d}). \quad (3.45)$$

Here $P(\theta_i|\mathbf{d})$ corresponds to the *marginalization* with respect to the other parameters, i.e.

$$P(\theta_i|\mathbf{d}) = \int d\theta_1 \dots d\theta_{i-1} d\theta_{i+1} \dots d\theta_n P(\boldsymbol{\theta}|\mathbf{d}). \quad (3.46)$$

There are cases in which the mode and mean are very different, corresponding to a posterior pdf asymmetric to be adequately summarized by a single estimate. An allowed range for

a parameter with probability content C (e.g., $C = 0.95$ or 95%) is provided by a credible region, or highest posterior density region, R , defined by (Gregory 2005)

$$\int_R d\theta_i P(\theta_i | \mathbf{d}), \quad (3.47)$$

Frequently, a parameterized model have more than one parameter, but we want to focus the attention on a subset of them. In such problems, the uninteresting parameters are known as *nuisance parameters*. As always, the full Bayesian inference is the full joint posterior PDF for all of the parameters; but its implications for the parameters of interest can be simply summarized by integrating out the nuisance parameters. Note that this is one of the most important technical advantages of Bayesian inference over frequentist statistics. Indeed, there is no general frequentist method for dealing with such parameters; they are indeed a “nuisance” in frequentist statistics. It follows because the likelihood function is not a pdf, and therefore it does not makes sense to talk about marginalization (Jaynes 2003).

3.3.2 Choice of prior

The prior expresses the information one has about the parameters before using the current data to constrain them. This information might come from a previous experiment or observation in which case the prior would be the posterior of that experiment. The prior can also express the theoretically allowed range of a parameter. Sometimes, it needs to be given by the user.

In the problem of the Bayesian parameter estimation, the actual prior bounds on a parameter are often unimportant. This happens when the likelihood is “small” at the boundaries of parameter space. In this case, they do not affect the integral in the evidence, and the posterior cancels at these points. In other cases, the choice of the prior can largely affect the resulting posterior distribution. There is a vast literature about how to select a prior in an appropriate way (Jaynes 2003; Gregory 2005). The mostly used are the *uniform* and *Jeffreys* prior.

The uniform prior (or flat prior) is constant over a region of parameter space and zero outside of it. Precisely, if the prior is on the parameter θ follows

$$P(\theta) = \begin{cases} \frac{1}{\theta_{\max} - \theta_{\min}} & \text{for } \theta_{\min} \leq \theta \leq \theta_{\max} \\ 0 & \text{otherwise} \end{cases}. \quad (3.48)$$

With this choice of prior the posterior becomes functionally identical to the likelihood, up to a proportionality constant. It has the appearance of being unprejudiced in the sense that it do not favour one parameter value over another without the data supporting it. However, what is a uniform prior for one set of parameters do not result a uniform prior for another set. This is true even though they might describe the same model. This is because, under a change of variable $\gamma \equiv \gamma(\theta)$, the prior transforms as

$$P(\gamma) = P(\theta) \left| \frac{\partial \theta}{\partial \gamma} \right|. \quad (3.49)$$

It follows that, if the transformation is non-linear a flat prior in θ is no longer flat in γ .

The other widely used prior is the Jeffreys prior defined as

$$P(\theta) = \begin{cases} \frac{1}{\theta \ln(\theta_{\max}/\theta_{\min})} & \text{for } \theta_{\min} \leq \theta \leq \theta_{\max} \\ 0 & \text{otherwise} \end{cases}. \quad (3.50)$$

Note that, this prior gives equal weight to equal logarithmic ranges of θ , see Fig. 3.1.

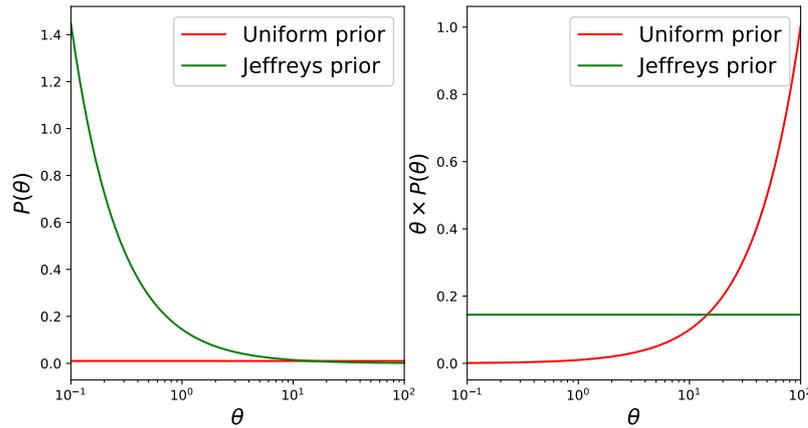


Figure 3.1: The left panel shows the prior probability density function, $P(\theta)$, for the uniform and Jeffreys priors. The right panel shows the same prior probability as the left panel, but per logarithmic interval, $\theta \times P(\theta)$.

Furthermore, it is invariant under a rescaling transformation, i.e. $\gamma = \theta^x$ for any x . Indeed, the probability for a parameter being in an infinitesimal region is

$$P(\gamma)d\gamma = P(\gamma)x\theta^{x-1}d\theta = xP(\theta)d\theta. \quad (3.51)$$

Note that the value of Jeffreys prior is infinite if the range is extended to $0 < \theta < \infty$. Similarly, the uniform prior is formally zero for the range $-\infty < \theta < \infty$. These ranges are routinely used when the posterior (likelihood times prior) has a well defined integral. In other word, sometime the integral of the prior values may not even need to be finite to get sensible answers for the posterior probabilities. These are examples of *improper prior distributions* that are not valid distributions by themselves, but make sense in a posterior.

3.3.3 Markov Chain Monte Carlo

Markov chain Monte Carlo (hereafter MCMC) methods are now becoming the standard tool to determine parameters from CMB data. It represents a very efficient mathematical tool to estimate the desired posterior distributions for high-dimensional models. Precisely, MCMC methods allow us to construct a sequence of points (or *samples*) in parameter space (called a *chain*). In statistics a chain is an ordered series of random variables, X_1, \dots, X_n, \dots , where the conditional probability of each element given the other elements is specified. In particular, a Markov chain is a chain where the conditional probability of any element X_n depends only on the previous element X_{n-1} , $P(X_n|X_0, \dots, X_{n-1}) = P(X_n|X_{n-1})$ (see, e.g. [Gregory 2005](#); [Ross 1998](#)). The probability $P(X_{n+1}|X_n)$ is known as *transitional kernel*, since it gives the probability of moving from point X_n to point X_{n+1} in parameter space. If the transition kernel is independent of n it is said to be *time-homogenous*.

Among all possible chains, we are interested in the so called *ergodic chains*. To be ergodic the chain must be *irreducible*, that is a chain starting at any state X_0 can reach any other state after a finite number of steps; *aperiodic*, that is the chain will not return to the same state after some fixed number of steps and all multiples of this number of steps; *positive recurrent*, that is the expectation value for the number of steps between any two states is finite. The most important consequence of ergodicity is that the chain has a unique stationary distribution $f(x)$ such that

$$\int_{-\infty}^{\infty} dx_n f(x_n)P(x_{n+1}|x_n) = f(x_{n+1}). \quad (3.52)$$

In other word, if we can find a transitional kernel that satisfies this requirement, we can produce chains whose samples are distributed according to $f(x)$. This stationary distribution is called *target distribution*, which in our case is the posterior $P(\boldsymbol{\theta}|\mathbf{d})$. Note that, once samples from the posterior have been gathered, obtaining Monte Carlo estimates of expectations for any function of the parameters becomes a trivial task thanks the low of large numbers:

$$\langle g(\boldsymbol{\theta}) \rangle = \int d\boldsymbol{\theta} g(\boldsymbol{\theta}) P(\boldsymbol{\theta}|\mathbf{d}) \simeq \frac{1}{N} \sum_{i=0}^{N-1} g(\boldsymbol{\theta}^i), \quad (3.53)$$

where $\boldsymbol{\theta}^i$ are generated from the posterior distribution $P(\boldsymbol{\theta}|\mathbf{d})$.

Metropolis-Hastings algorithm

The simplest sample generator algorithm is the Metropolis-Hastings (Metropolis et al. 1953; Hastings 1970). It produces a Markov chain whose equilibrium distribution is the target probability density, with a transition kernel that satisfies the *detailed balance* condition

$$P(\boldsymbol{\theta}_{n+1}|\boldsymbol{\theta}_n)P(\boldsymbol{\theta}_n|\mathbf{d}) = P(\boldsymbol{\theta}_n|\boldsymbol{\theta}_{n+1})P(\boldsymbol{\theta}_n|\mathbf{d}). \quad (3.54)$$

The core of this algorithm can be summarized in four points (see, e.g. Gregory 2005):

- i) Start the chain from a random point in parameter space, $\boldsymbol{\theta}_n$.
- ii) Propose a candidate point $\boldsymbol{\theta}_c$ extracted from an arbitrarily proposal density distribution $q(\boldsymbol{\theta}_n, \boldsymbol{\theta}_c)$.
- iii) Given the posterior evaluated at the candidate point, $P(\boldsymbol{\theta}_c|\mathbf{d})$, compute the ratio r given by

$$r = \frac{P(\boldsymbol{\theta}_c|\mathbf{d})q(\boldsymbol{\theta}_n, \boldsymbol{\theta}_c)}{P(\boldsymbol{\theta}_n|\mathbf{d})q(\boldsymbol{\theta}_c, \boldsymbol{\theta}_n)}. \quad (3.55)$$

- If $r \geq 1$, then we update the trial state by setting $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_c$
- If $r < 1$, then we draw a uniform deviate u between 0 and 1. If $r \geq u$, then we set $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_c$, otherwise $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n$. It corresponds to accept the proposal point with probability r .

- iv) Repeat all the previous point to fill the chain.

Although the MCMC is guaranteed to converge under the ergodic conditions, it might take a very long time. The chain moves around parameter space in a random walk and if it does not reach every region of significant probability many times it will not be a good approximation of an independent sampling from the target. This can happen if there is a strong degeneracy between parameters, which explain the needed of a suitable set of parameters to describe a model, see Sec. 2.4. Furthermore, if the rejection rate is too high the chain will have many duplicated points that will not fill parameter space in an even way. On the other hand, if it is too low the chain will move, but not fast enough to get around the space. The rate with a candidate point is rejected or not can be changed by adjusting the proposal function $q(\boldsymbol{\theta}_c|\boldsymbol{\theta}_n)$.

In the original algorithm (Metropolis et al. 1953) the proposal distribution is symmetric, $q(\boldsymbol{\theta}_n, \boldsymbol{\theta}_c) = q(\boldsymbol{\theta}_c, \boldsymbol{\theta}_n)$. In this case, the new step is always accepted if it improves on the posterior, otherwise it is accepted with probability $P(\boldsymbol{\theta}_c|\mathbf{d})/P(\boldsymbol{\theta}_n|\mathbf{d})$. Since we extract the propose candidate from $q(\boldsymbol{\theta}_n, \boldsymbol{\theta}_c)$, it makes sense to use a standard distribution with a well implemented random deviate generator. A popular choice is the multivariate Gaussian centred on the current point with an unspecified covariance C . Thus the new point is $\boldsymbol{\theta}_c = \boldsymbol{\theta}_n + \mathbf{y}$, where $\mathbf{y} \in \mathcal{N}(\boldsymbol{\theta}_n, C)$. The covariance of this Gaussian needs to be

adjusted until an acceptable rejection rate is found. Reducing it tends to decrease the rejection rate. When the covariance is large, the proposed new point likely belongs to regions that are far away from the peak in the probability, and thus are rejected. In this way, the chain will be attracted by the high probability regions, but might take a while to get there. This period where the chain is not near its stationary distribution is called *burn in*. Note that, for this reason one usually discards the first part of the chain.

Convergence

An important issue relates to the convergence of the chain. Assessing the convergence of the chain essentially means to know when we can stop. A sophisticated method that takes into account multiple chains is the Gelman-Rubin diagnostic (Gelman and Rubin 1992), \hat{R} . Precisely, if we have m independent chains each of length n and θ_i^α is the i -th parameter value of the α -th chain, we can define the following quantities:

$$\bar{\theta}^\alpha = \frac{1}{n} \sum_{i=1}^n \theta_i^\alpha \quad \bar{\bar{\theta}} = \frac{1}{m} \sum_{\alpha=1}^m \bar{\theta}^\alpha \quad (3.56)$$

$$s_\alpha^2 = \frac{1}{n-1} \sum_{i=1}^n (\theta_i^\alpha - \bar{\theta}^\alpha)^2 \quad B = \frac{n}{m-1} \sum_{\alpha=1}^m (\bar{\theta}^\alpha - \bar{\bar{\theta}})^2 \quad (3.57)$$

$$W = \frac{1}{m} \sum_{\alpha=1}^m s_\alpha^2 \quad V = \frac{n-1}{n} W + \frac{1}{n} B \quad (3.58)$$

The Gelman-Rubin diagnostic is defined as

$$\hat{R} = \sqrt{\frac{V}{W}}, \quad (3.59)$$

and it represent an estimate of the factor by which the variance in θ can be reduced by continuing the chains. A value $\hat{R} \sim 1$ is a good sign, and this should be compute for all the parameters of interest.

3.3.4 Likelihood analysis of CMB

The likelihood function represents a key ingredient to compare observed data with theoretical predictions in order to constrain the model parameters. Thus the choice of its functional form is an important task for cosmologists. Precisely, in the context of the standard cosmological model of the early universe we are able to determine an analytical form for the probability of the data given the theoretical model. Indeed, we know that primordial perturbations are Gaussian distributed, and so are CMB fluctuations (see Sec. 2.2). Therefore, as we have seen, all relevant physical information in the CMB field are contained in the variance of the distribution. This property makes the full-sky power spectra of CMB fluctuations a sufficient statistics. In this simple case of full-sky observations, the likelihood function is given by a Wishart distribution with $\nu = 2\ell + 1$ degree of freedom and dimension $p = 3$, see Sec. 3.1. Precisely, if we define the variable

$$\mathbf{X}_a = (a_{\ell m}^T, a_{\ell m}^E, a_{\ell m}^B)^\top, \quad (3.60)$$

distributed according to a multivariate-Gaussian distribution with covariance matrix

$$\text{cov}(\mathbf{X}_a, \mathbf{X}_a) = \begin{pmatrix} C_\ell^{TT} & C_\ell^{TE} & 0 \\ C_\ell^{TE} & C_\ell^{EE} & 0 \\ 0 & 0 & C_\ell^{BB} \end{pmatrix} \equiv \mathbf{V}_\ell, \quad (3.61)$$

in analogy with Eq. (2.17), we can define the estimator

$$\mathbf{S}_\ell \equiv \frac{1}{2\ell+1} \sum_m \mathbf{X}_a \mathbf{X}_a^\dagger = \begin{pmatrix} \hat{C}_\ell^{TT} & \hat{C}_\ell^{TE} & \hat{C}_\ell^{TB} \\ \hat{C}_\ell^{TE} & \hat{C}_\ell^{EE} & \hat{C}_\ell^{EB} \\ \hat{C}_\ell^{TB} & \hat{C}_\ell^{EB} & \hat{C}_\ell^{BB} \end{pmatrix}. \quad (3.62)$$

It follows that, with this notation, the pdf of the full set of power spectra is given by

$$p(\mathbf{S}_\ell | \mathbf{W}_\ell) = \mathcal{L}(\mathbf{W}_\ell) = \frac{|\mathbf{S}_\ell|^{\nu-p-1/2} \exp\left\{-\frac{\text{Tr}(\mathbf{S}_\ell \mathbf{W}_\ell^{-1})}{2}\right\}}{2^{p\nu/2} |\mathbf{W}_\ell|^{\nu/2} \Gamma_p(\nu/2)}, \quad (3.63)$$

were $\mathbf{W}_\ell = \mathbf{V}_\ell/\nu$, and

$$\Gamma_p(\nu/2) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma((\nu+1-i)/2). \quad (3.64)$$

Given the observed power spectra, \hat{C}_ℓ^{XY} , Eq. (3.63) represents the exact expression of the likelihood function of C_ℓ^{XY} . However, in general, realistic experimental conditions involves complications in the analysis. These require a likelihood analysis different from the simple case of full-sky. First of all, specific estimators of the power spectra should be defined in the partial-sky regime, which take into account spurious correlations between fields induced by the incomplete sky coverage. Secondly, the use of a Wishart distribution as a likelihood function is no longer possible. Either the new estimators are no longer distributed according to a Wishart, and therefore this choice is not exact anymore. Or, the use of the exact likelihood is unfeasible as one moves to the analysis of smaller scales (larger multipoles) and higher-resolution maps, due to the huge computational cost of inverting large covariance matrices. However, as we will see, at large scales and for low-enough angular resolutions, the exact likelihood in pixel space can still be adopted.

Here we briefly explore two likelihood approximation, referred to the small- and large-scale regimes, respectively. These represent the basic tools needed to understand the building procedure of the likelihood package presented in Natale et al. (2020), and summarized in Chap. 4. For a more complete review on CMB likelihood analysis we remand, for example, to Gerbino et al. (2020).

Small-scale regime

At small scales, the central limit theorem allows to approximate the Wishart distribution as a Gaussian in the power spectra. In general, quadratic forms in some functions of the CMB spectra have been adopted as approximate likelihood functions, with various choices of the covariance matrix. However, here we consider only the approximation used in the official analyses of the *Planck* (Planck Collaboration V 2019), the ACT (Louis et al. 2017) and SPT (Henning et al. 2018a) collaborations, known as *fiducial Gaussian approximation*. Precisely, the likelihood is quadratic in C_ℓ^{XY} :

$$-2 \ln \mathcal{L}(\mathbf{X}_C) = -\frac{1}{2} (\mathbf{X}_C - \hat{\mathbf{X}}_C)^\top \mathbf{Y}_{C,\text{fid}}^{-1} (\mathbf{X}_C - \hat{\mathbf{X}}_C) + \text{const.}, \quad (3.65)$$

where $\mathbf{X}_C = (C_\ell^{TT}, C_\ell^{TE}, C_\ell^{EE})^\top$. The covariance matrix, given by the curvature of the Wishart,

$$\mathbf{Y}_C^{-1} = \frac{\nu}{2 \left[\hat{C}_\ell^{TT} \hat{C}_\ell^{EE} - (\hat{C}_\ell^{TE})^2 \right]^2} \times \begin{pmatrix} (\hat{C}_\ell^{EE})^2 & -2\hat{C}_\ell^{TE} \hat{C}_\ell^{EE} & (\hat{C}_\ell^{TE})^2 \\ -2\hat{C}_\ell^{TE} \hat{C}_\ell^{EE} & 2 \left[\hat{C}_\ell^{TT} \hat{C}_\ell^{EE} + (\hat{C}_\ell^{TE})^2 \right] & -2\hat{C}_\ell^{TE} \hat{C}_\ell^{TT} \\ (\hat{C}_\ell^{TE})^2 & -2\hat{C}_\ell^{TE} \hat{C}_\ell^{TT} & (\hat{C}_\ell^{TT})^2 \end{pmatrix}, \quad (3.66)$$

is computed for a fixed fiducial model, assumed to be smooth and a close approximation to the underlying model under scrutiny. The constant term include also the determinant of the covariance matrix.

Large-scale regime

At large angular scales, the resolution of the map to be analysed is low enough to make a pixel-based approach computationally feasible. Furthermore, the low-resolution makes also possible to exploit the information encoded in harmonic space, and build the likelihood function from a simulation-based method.

Here only the pixel-based approach is considered. Its main advantage relies on the exactness of the likelihood function, even in the cut-sky regime. If we define the data vector as $\mathbf{m} = \mathbf{s} + \mathbf{n}$, where \mathbf{s} is the signal per pixel in temperature and polarization ($\mathbf{s} = (T, Q, U)$) and \mathbf{n} is the instrumental noise, the likelihood function can be written as

$$\mathcal{L}(C_\ell) = P(\mathbf{m}|C_\ell) = \frac{1}{\sqrt{2\pi|C|}} \exp\left(-\frac{1}{2}\mathbf{m}^\top C^{-1}\mathbf{m}\right). \quad (3.67)$$

Here the covariance matrix C is given by the sum of the signal and noise covariance matrices, $C = S + N$. The effect of beam smearing, also relevant for the large-scale data, is now taken into account when constructing the full covariance matrix in terms of the beam-weighted sum of Legendre polynomial. In evaluating the likelihood function, the data vector and the noise covariance matrix are fixed, while the signal covariance matrix, S , is recomputed for any given cosmological model to be compared against data. A detailed description of the full procedure to obtain the covariance matrix can be found in Appendix A of [Tegmark and de Oliveira-Costa \(2001\)](#).

The current resolution used in the large-scale pixel-based likelihood analysis is set to $N_{\text{side}} = 16^1$, that means a resolution up to a multipole $\ell \leq 4 \times N_{\text{side}} = 64$. This results in a $N_{\text{pix}} = 3 \times 12 \times N_{\text{side}}^2 = 3 \times 3072 = 9216$ pixels involved in the analyses of temperature and polarization signals, further reduced by the application of the analysis mask. In practice, only a subsection of S is recomputed, in particular that subsection corresponding to $\ell < 30$. The portion of S corresponding to multipoles $30 \leq \ell \leq 64$ is precomputed from a fixed fiducial model, which does not affect the performance of the likelihood.

The varying part of S can be further decomposed as $S = V^\top AV$, via a transformation V that effectively reduces the dimension of the actual evaluation cost from a $N_{\text{pix}} \times N_{\text{pix}}$ inversion to a $n_\lambda \times n_\lambda$ inversion, where $n_\lambda = 2\ell + 1$ is the dimension of the transformed matrix A . The latter is the only matrix that depends on the theoretical C_ℓ and, therefore, it is the only matrix to be recomputed and inverted. The fixed portions of the covariance matrix as well as the transformation matrix V can be pre-computed and stored. The inverse of S can be computed applying the Sherman-Morrison-Woodbury formula (see, e.g., Appendix B of [Planck Collaboration XI 2016](#)).

3.4 Surprise, uncertainties and entropy

The mathematical foundation of information theory relies on the publication of a landmark paper by Shannon ([Shannon 1948](#)). His attempt was to reproduce at one point either exactly or approximately a message selected at another point. More precisely, he showed how information could be quantified with absolute precision, and demonstrated the essential unity of all information media. Telephone signals, text, radio waves, and pictures, essentially every kind of communication, could be encoded in bits. The use of information

¹This corresponds to the HEALPix resolution (see [Gorski et al. 2005](#)).

theory was introduced in statistics only few years later by Kullback and Leibler (Kullback and Leibler 1951).

The tool by which we can quantify the uncertainty of a probability distribution is the *entropy*. The concept of entropy is intimately related with the concept of *surprise*. Consider an event A that can occur when an experiment is performed. A legitimate question is how surprised would we be to hear that A occurs. As it is reasonable to suppose, the amount of surprise generated by the information that A has occurred depend on the probability of A . For instance, if the experiment consists of tossing a coin, then we would not be too surprised to hear that the event “ $A = \text{get an head}$ ” ($P(A) = 1/2$), occurred. On the other hand, if the experiment consists of rolling a pair of dice, we would certainly be surprised to hear that the event “ $A = \text{the sum of the dice is 12}$ ” ($P(A) = 1/36$) occurred. Guided by this example, let us suppose that one feels upon learning that an event A has occurred depends only on the probability of A . Furthermore, let us denote with $s[p]$ the surprise evoked by the occurrence of an event with probability p . We can determine the functional form of the surprise $s[p]$ imposing the validity of four natural axioms (Ross 1998):

- (A1) $s[1] = 0$, that is there is no surprise in hearing that an event A sure to occur has indeed occurred.
- (A2) $s[p]$ is a strictly decreasing function of p ; that is, if $p < q$, then $s[p] > s[q]$. This states that more unlikely an event is to occur, greater is the surprise evoked by its occurrence.
- (A3) $s[p]$ is a continuous function of p .
- (A4) If A and B are two independent events, with probability $P(A) = p$ and $P(B) = q$, then $s[pq] = s[p] + s[q]$.

It can be shown that, if $s[\cdot]$ satisfies (A1) through (A4), then

$$s[p] = -K \log_2 p, \quad (3.68)$$

where K is an arbitrary positive integer. It is usual to let K equal 1. In this case the surprise is said to be expressed in unit of *bits*. If $K = -\log_2 e$ the surprise is said to be expressed in unit of *nats*.

Note that $-\log p$ represents the surprise evoked if a random variable X takes the value x , $P(X = x) = p(x)$. It follows that, if X can take one of the values x_1, \dots, x_n with probabilities p_1, \dots, p_n , the expected amount of surprise associated to the value of X is given by

$$S[p] = - \sum_{i=1}^n p_i \log p_i = \langle s[p] \rangle. \quad (3.69)$$

The quantity defined above represents the *entropy* of the random variable X . If X is a continuous random variable, follows

$$S[p] = - \int dx p(x) \log p(x). \quad (3.70)$$

In information theory, $S[p]$ is interpreted as the average amount of information received when the value of X is observed. Note that, the average surprise evoked by X , the uncertainties of X , or the average amount information yielded by X all represent the same concept viewed from three slightly different point of view.

3.4.1 Maximum Entropy Principle

The interpretation of the entropy $S[p]$ as a quantitative measure of the amount of missing information has one remarkable consequence: we have a method to build probabilities (see, e.g., [Caticha 2008](#)). This is a powerful tool since, in general, the knowledge of the probability distribution of the data as a function of the parameters (the likelihood $p(\mathbf{d}|\boldsymbol{\theta})$) is very limited. Usually, we face only with some statistical properties that are sensitive to the model parameters in order to predict the shape of $p(\mathbf{d}|\boldsymbol{\theta})$. However, we can select, among all possible probability distributions that agree with whatever we know (a set of constraints), the distribution that reflects maximum ignorance about everything else. Thus, since the lack of knowledge is measured by entropy, the method is mathematically implemented by selecting the distribution that maximizes entropy subject to the constraints imposed by the available information. This method of reasoning is called the method of Maximum Entropy, and is often abbreviated as MaxEnt.

The constraints can take any form whatsoever. However, the special case of constraints that are linear in the probabilities is of particular interest. That is, let us express the constraints F^k in the form of averages of some functions f_k

$$F^k = \langle f_k \rangle = \int dx p(x) f_k(x), \quad k = 1, 2, \dots \quad (3.71)$$

Introducing the Lagrange multipliers λ_k associated with the constraints F^k , and requiring that the probability $p(x)$ is normalized, the entropy maximization is achieved setting

$$\begin{aligned} 0 &= \frac{\delta}{\delta p(x)} \left[S[p] - \alpha \left(\int dy p(y) - 1 \right) - \sum_k \lambda_k F^k \right] \\ &= -\log p(x) - 1 - \alpha - \sum_k \lambda_k f^k(x). \end{aligned} \quad (3.72)$$

The solution is the so-called *canonical distribution*

$$p(x) = \exp \left\{ -\lambda_0 - \sum_k \lambda_k f^k(x) \right\}, \quad (3.73)$$

where we have set $\lambda_0 = 1 + \alpha$, determined by the normalization constraint

$$e^{-\lambda_0} = \int dx \exp \left\{ -\sum_k \lambda_k f^k(x) \right\} \equiv Z(\lambda_1, \lambda_2, \dots), \quad (3.74)$$

where we have introduced the *partition function* Z . Note that, the partition function Z is closely related to the entropy $S[p(x)]$ through

$$S[p] = Z + \sum_k \lambda_k F^k, \quad (3.75)$$

and the values of the multipliers can be explicitly written as a function of the entropy

$$\lambda_i = \frac{\partial S}{\partial F^i}. \quad (3.76)$$

Fisher information and Shannon entropy

The MaxEnt give us the functional form of the likelihood $p(\mathbf{d}|\boldsymbol{\theta})$, i.e. Eq. (3.73). It implies that the dependence on the model goes through the constraints, or, equivalently, through

their associated multipliers. Therefore, it follows that,

$$\begin{aligned} \frac{\partial \ln p(\mathbf{d}|\boldsymbol{\theta})}{\partial \theta_i} &= -\frac{\partial \ln Z}{\partial \theta_i} - \sum_j \frac{\partial \lambda_j}{\partial \theta_i} f^j(\mathbf{d}) \\ &= \sum_j \frac{\partial \lambda_j}{\partial \theta_i} (F^j - f^j), \end{aligned} \quad (3.77)$$

where the last line follows from the chain rule and the fact that

$$F_k = -\frac{\partial \ln Z}{\partial \lambda_k}. \quad (3.78)$$

Thus, since the covariance matrix of the constraints is given by

$$\langle (f_i - F_i)(f_j - F_j) \rangle = \frac{\partial^2 \ln Z}{\partial \lambda_i \partial \lambda_j}, \quad (3.79)$$

from the definition of the Fisher matrix, \mathcal{F}_{ij} , in Eq. (3.26) follows that (cfr. Carron et al. (2011)),

$$\mathcal{F}_{ij} = \sum_{\alpha\beta} \frac{\partial \lambda_\alpha}{\partial \theta_i} \frac{\partial^2 \ln Z}{\partial \lambda_\alpha \partial \lambda_\beta} \frac{\partial \lambda_\beta}{\partial \theta_j}. \quad (3.80)$$

Furthermore, it can be shown that, using Eq. (3.76) and (3.78),

$$\mathcal{F}_{ij} = -\sum_{\alpha\beta} \frac{\partial F_\alpha}{\partial \theta_i} \frac{\partial^2 \ln S}{\partial F_\alpha \partial F_\beta} \frac{\partial F_\beta}{\partial \theta_j}, \quad (3.81)$$

that represents the total amount of information on the model parameters $\boldsymbol{\theta}$ in the data \mathbf{d} , when the model predicts the set of constraints F_k . Note that, the amount of information given by the entropy S results identical to the Fisher information in a Gaussian distribution of the observables, despite in this approach Gaussian properties are not assumed (Carron et al. 2011).

MaxEnt fixing mean and variance

An important application of the MaxEnt principle consist in the maximization of the entropy when we know the mean μ and the variance σ^2 of the parameter in interest. This problem can be formalized introducing the Lagrange multipliers

$$\mathcal{L} = S[p] - \alpha \left(\int dy p(y) - 1 \right) - \lambda_1 \left(\int dy (\mu - y)^2 p(y) - \sigma^2 \right). \quad (3.82)$$

It follows that, imposing $\delta\mathcal{L}/\delta p(x) = 0$, the functional form of $p(x)$ is

$$p(x) = \exp \left\{ -1 - \alpha - \lambda_1 (\mu - x)^2 \right\}. \quad (3.83)$$

Maximizing \mathcal{L} with respect to α and λ_1

$$\frac{\partial \mathcal{L}}{\partial \alpha} = 0 \Rightarrow \int dy p(y) = 1 = e^{-1-\alpha} \int dy e^{-\lambda_1 (\mu - y)^2} \quad (3.84)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = 0 \Rightarrow \int dy (\mu - y)^2 \exp \left\{ -1 - \alpha - \lambda_1 (\mu - x)^2 \right\} = \sigma^2, \quad (3.85)$$

we get from Eq. (3.84) that $\exp(-1 - \alpha) = \sqrt{\lambda_1/\pi}$, and, by substituting it in Eq. (3.85), $\lambda_1 = 1/(2\sigma^2)$. Thus, the pdf $p(x)$ that reflects our knowledge of the variance (and thus the mean) of the parameter is a Gaussian

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}. \quad (3.86)$$

This shows that, among all distributions with a fixed variance, the Gaussian distribution maximises the entropy S . In particular, if the pdf is a one-dimensional Gaussian, the form of the Shannon entropy is given by

$$S = \frac{1}{2} \ln [2\pi\sigma^2 e^1] . \quad (3.87)$$

If the pdf is a multivariate normal distribution with dimension n , the Shannon entropy takes the form

$$S = \frac{1}{2} \ln [(2\pi)^n |\Sigma| e^n] , \quad (3.88)$$

where $|\Sigma|$ is the determinant of the covariance matrix in parameter space.

3.4.2 Kullback-Leibler divergence

A common problem in data analyses is to measure how similar, or different, two probability distributions are. This issue can be assessed, in the contest of information theory, by the *Kullback-Leibler divergence* (KL divergence).

Let us consider two probability density function, $p(x)$ and $q(x)$, of a continuous random variable X . We can construct the measure of the expected number of the extra nats (i.e. logarithmic difference) required to describe samples from $p(x)$ using the knowledge of $q(x)$, instead of $p(x)$ itself. This statement can be expressed in formula as

$$\Delta S[p(x)|q(x)] = \int dx p(x) \log \frac{p(x)}{q(x)} . \quad (3.89)$$

This quantity is sometimes read as the entropy of $p(x)$ relative to $q(x)$, and thus called “relative entropy”. It has several properties:

- It is always non-negative, $\Delta S[p(x)|q(x)] \geq 0$. The equality is reached only when $p(x) = q(x)$.
- It is invariant under parameter transformations. That is, if the transformation is made from variable x to variable $y(x)$ then

$$\begin{aligned} \Delta S[p(x)|q(x)] &= \int dx p(x) \log \frac{p(x)}{q(x)} \\ &= \int dy \frac{dx}{dy} p(y) \frac{dy}{dx} \log \left(\frac{p(y) \frac{dy}{dx}}{q(y) \frac{dy}{dx}} \right) \\ &= \int dy p(y) \log \frac{p(y)}{q(y)} . \end{aligned} \quad (3.90)$$

- It is additive for independent distributions.

Furthermore, if the involved pdf are normally-distributed, $p(\mathbf{x}) \sim \mathcal{N}(\mathbf{x}; \mu_1, \Sigma_1)$ and $q(\mathbf{x}) \sim \mathcal{N}(\mathbf{x}; \mu_2, \Sigma_2)$, then we can use the analytical expression given by

$$\begin{aligned} \Delta S [p(\mathbf{x})|q(\mathbf{x})] &= \frac{1}{2} (\mu_1 - \mu_2)^\top \Sigma_2^{-1} (\mu_1 - \mu_2) + \\ &\quad + \frac{1}{2} \left[\text{Tr} \left(\Sigma_1 \Sigma_2^{-1} \right) - d - \log \left(\frac{\det \Sigma_1}{\det \Sigma_2} \right) \right] . \end{aligned} \quad (3.91)$$

Updating probabilities

We have seen that Bayes' rule is the natural way to update probabilities. However, it is possible only when the information is in the form of data, i.e. constraints expressed in the form of data that can be plugged into a likelihood function. How we can update them when information is not in the form of data remains an open question. MaxEnt, allowed one to deal with information in the form of constraints on the allowed probability distributions. Anyhow, MaxEnt can handle arbitrary constraints but not arbitrary priors.

Our goal is to update from the prior distribution $q(x)$ to a posterior distribution $p(x)$ when new information (a set of constraints) becomes available. The relative entropy can be used to achieve this goal, since it describes a ranking of the distributions $p(x)$ relative to the given prior $q(x)$. Furthermore, it can be shown ([Caticha 2008](#)) that the only functional form for the relative entropy that can be used in inductive inference is the KL divergence defined in Eq. (3.89).

A novel CMB polarization likelihood package for large angular scales built from combined WMAP and Planck LFI legacy maps

In Sec. 2.3.1 we have studied the effect of reionization on the polarization power spectrum. The presence of a characteristic fingerprint of this epoch makes the measurement of CMB polarization at large angular scales a crucial step for the determination of the Thomson scattering optical depth to reionization, τ , which is currently the less constrained of the Λ CDM parameters. The optical depth, τ , is connected to the integrated amount of free electrons along the line of sight (cfr. Eq. (2.6)) and provides information on how and when the first stars and galaxies formed.

Remarkable advancements have been made in this field over the last 15 years. The WMAP (Hinshaw et al. 2013) and *Planck* (Planck Collaboration VI 2018) collaborations have continuously improved the quality of large-scale polarization measurements, which are known to be notoriously extremely tough to clean from contaminations coming from the foreground and instrumental systematic effects. The most constraining dataset currently available is provided by the *Planck* collaboration (Planck Collaboration I 2018), which uses the High Frequency Instrument (HFI) measurements at 100 and 143 GHz. Such results for the Legacy *Planck* release are presented in Planck Collaboration III (2018) and Planck Collaboration V (2019), while an improved post-*Planck* analysis is presented in Delouis et al. (2019) and in Pagano et al. (2019).

These HFI-based measurements are all specifically designed to determine the reionization optical depth and, thus, they are mainly dedicated to the characterization of the E-modes power spectrum. This approach, which is consistent with the corresponding likelihood codes delivered, is mainly driven by the difficulty of building reliable noise covariance matrices and by the relatively high level of residual systematic effects related to dipole and foreground temperature-to-polarization leakage. Such likelihoods, despite being the most sensitive to date, do not include the TE spectrum Planck Collaboration V (2019). Furthermore, they cannot be adapted to handle non-rotationally invariant cosmologies in a straightforward way and they might need tuned-up simulations for exotic models (see Planck Collaboration V 2019, Section 2.2.6).

For the Legacy data release, together with the HFI-based likelihood, the *Planck* collaboration has also delivered a map-based likelihood employing observations of the Low Frequency Instrument (LFI) in the 70 GHz channel. The sensitivity to the reionization optical depth of the LFI-based likelihood is inferior by more than a factor of two with respect to the HFI-based likelihood.

The possibility of combining the WMAP and *Planck* observations to build a “joint”

dataset that is more constraining was first explored in [Lattanzi et al. \(2017\)](#), using the data available at the time. However, a combined dataset using the WMAP and *Planck* legacy observations of the large-scale polarization is made publicly available only with the collaborative work of L. Pagano, M. Lattanzi, M. Migliaccio, L. P. Colombo, A. Gruppuso, P. Natoli, G. Polenta and myself ([Natale et al. 2020](#)). The aim of this chapter is to describe the combined real-space polarization dataset presented in [Natale et al. \(2020\)](#), which jointly considers the *Planck* 70 GHz channel and the WMAP Ka, Q, and V bands. Note that, because of the aforementioned difficulty in dealing with residual systematic effects in the pixel space ([Planck Collaboration V 2019](#)), the HFI CMB channels, such as 100 and 143 GHz, are not considered in [Natale et al. \(2020\)](#). However, the resulting dataset, despite still having an overall higher noise than the HFI-based one, allows for an independent estimation of the reionization optical depth. Moreover, being a real-space dataset, it is suitable for a number of studies that are not accessible for a spectrum-based likelihood (see, e.g., [Planck Collaboration XXIII \(2014\)](#); [Planck Collaboration XVI \(2016\)](#)) and it is capable of exploring non-rotationally invariant cosmologies.

4.1 Datasets

In this section, we describe the large-scale WMAP and *Planck* polarization maps that were used to build the combined dataset. As already mentioned, as CMB channels, we considered the 70 GHz channel from *Planck* LFI ([Planck Collaboration II 2018](#)) and the Ka, Q, and V bands from WMAP ([Bennett et al. 2013](#)). In the case of LFI 70 GHz, we used the full mission map after removing the bandpass and gain-mismatch-leakage correction maps. These maps, described in ([Planck Collaboration II 2018](#)), are part of the *Planck* 2018 legacy data release, and are publicly available through the Planck Legacy Archive¹. For WMAP, we use the raw nine-year frequency maps, available on the Lambda archive². In principle, we could have also considered the 44 GHz channel from *Planck* LFI and the W-band from WMAP as CMB channels. However, we found that both these channels show excess power, likely to be spurious in origin, after implementing the foreground cleaning procedure described in Sec. 4.3. For this reason, we decided not to include the 44 GHz and W-band channels in our analysis. We note that the *Planck* and WMAP collaborations made the same choice on similar grounds ([Planck Collaboration V 2019](#); [Page et al. 2007](#)).

We employed the K-band from WMAP, LFI 30 GHz, and HFI 353 GHz maps from *Planck* as tracers of Galactic foreground emission. These are used both to generate masks excluding regions dominated by Galactic emissions, and to mitigate the astrophysical foreground contamination in the remaining parts of the sky, as explained in detail in Secs. 4.2 and 4.3. At 30 GHz, we used the full-mission, bandpass leakage-corrected map. For the 353 GHz channel, we selected a map built only from data provided by polarization-sensitive bolometers (PSB) ([Planck Collaboration III 2018](#)), as done in the low- ℓ analysis presented in [Planck Collaboration V \(2019\)](#). The WMAP K band and *Planck* 30 GHz are used as a polarized synchrotron tracer, respectively, for the WMAP and *Planck* CMB channels. This follows the prescription of [Lattanzi et al. \(2017\)](#) and [Weiland et al. \(2018\)](#). The *Planck* 353 GHz is used as polarized thermal dust tracer for both WMAP and *Planck*.

Since we are mainly focused on the large angular scales, it appears convenient to work with low-resolution datasets. Thus, all the maps of the Stokes parameters, $\mathbf{m} = [Q, U]$, describing the measured linear polarization, were downgraded to a HEALPix resolution of $N_{\text{side}} = 16$ ([Górski et al. 2005](#)), which corresponds to a pixel size of ~ 3.7 degrees. A smoothing kernel was applied to the high-resolution maps prior to the downgrading, which is meant to avoid aliasing into the large angular scales of the high-frequency power present

¹<http://pla.esac.esa.int/pla/>

²https://lambda.gsfc.nasa.gov/product/map/dr5/m_products.cfm

in the maps. The smoothing was performed in harmonic space, using a cosine window function (Benabed et al. 2009; Planck Collaboration V 2019). This guarantees that the signal is left unaltered on the scales of interest, that is, up to multipoles of $\ell = N_{\text{side}} = 16$, while it is smoothly set to zero on smaller scales, $\ell > 3 \times N_{\text{side}} = 48$.

The instrumental noise properties of each low-resolution map are described by an associated pixel-pixel noise covariance matrix (NCVM). For the LFI channels, the covariance matrices are presented in Planck Collaboration II (2018). The 70 GHz covariance matrix has been rescaled in harmonic space in order to match the noise level of the half-difference of half-ring maps, following the procedure described in Planck Collaboration V (2019). For the HFI 353 GHz NCVM, we use a downgraded version of the map-making covariance matrix, which is instead generated at the native high-resolution of $N_{\text{side}} = 2048$. This NCVM only accounts for Q and U correlations within the same pixel, while correlations between different pixels are ignored. Finally, for WMAP we build the NCVMs starting from the polarization pixel-pixel inverse covariance matrices at $N_{\text{side}} = 16$ (Res 4) delivered by the WMAP team and described in (Page et al. 2007; Bennett et al. 2013). The cosine window function apodization is performed in harmonic space on the eigenvectors of these low resolution matrices. It is worth noting that although exchanging the order of the smoothing and downgrading operations is clearly not an option at the map level, due to the possible presence of sub-pixel structure, it can still be acceptable for the NCVMs.

Since all the (Q,U) NCVMs were convolved with a smoothing function, we added to them a white noise covariance matrix, with $\sigma^2 = (20 \text{ nK})^2$, in order to guarantee that they are numerically well-conditioned, as in Planck Collaboration V (2019). For consistency, noise with the same statistical properties has to be added to the corresponding maps. However, instead of adding a single noise realization to each smoothed data map, as in Planck Collaboration V (2019), we followed a different procedure, which is described in Sec. 4.3. This ensures that our results are not biased by a particular realization of the regularization noise.

Concerning the temperature (T) map, we always employ the *Planck* 2018 **Commander** solution (Planck Collaboration IV 2018) outside its confidence mask, which leaves 86% of the sky available. This map was filtered with a Gaussian beam of FWHM 440 arcmin and downgraded to $N_{\text{side}} = 16$. Since it is reasonable to assume that the temperature noise at large angular scales is negligible, we only need to include the regularization noise. Thus, we modeled the temperature NCVM as a white noise covariance matrix with $\sigma^2 = (2 \mu\text{K})^2$, as in Planck Collaboration V (2019). We consistently handle such regularization noise following the same procedure adopted for polarization. Finally, when building the NCVM of the full TQU maps, we neglect the correlation between temperature and polarization and set the corresponding off-diagonal blocks in the covariance matrix to zero (Planck Collaboration XI 2016).

4.2 Polarization masks

In order to efficiently perform the foreground cleaning and the cosmological parameter estimation, we must remove the pixels of the data map that are most affected by foreground contamination from the analysis. With regard to temperature, we always use the **Commander** 2018 confidence mask (Planck Collaboration V 2019) provided by the *Planck* collaboration. In this section, we describe how the polarization masks are produced.

In polarization, we built two different sets of masks for WMAP and LFI. For LFI 70 GHz, we used the 30 and the 353 GHz maps as, respectively, synchrotron (s) and dust (d) tracers, analogously to what is done in Planck Collaboration V (2019). We first applied a Gaussian smoothing with a full width half maximum (FWHM) of 7.5° to these input maps, taken at their native resolution of $N_{\text{side}} = 1024$ (30 GHz) and $N_{\text{side}} = 2048$

(353 GHz). Then we built maps of the polarization amplitude $P_s = \alpha \sqrt{Q_s^2 + U_s^2}$ and $P_d = \beta \sqrt{Q_d^2 + U_d^2}$, where the scaling coefficients are set to $\alpha = 0.063$ and $\beta = 0.0077$, as estimated in [Planck Collaboration XI \(2016\)](#). These two maps were subsequently downgraded to the HEALPix resolution $N_{\text{side}} = 16$. From these maps, two separate sets of masks for synchrotron and dust emission were built as follows. We excluded pixels where the relevant polarization intensity, P_s or P_d , is greater than a given threshold. This threshold is expressed in terms of excess intensity with respect to the corresponding mean value, $\langle P_s \rangle$ or $\langle P_d \rangle$, over the whole sky. Any pair of synchrotron and dust masks can then be combined to yield a single foreground mask. Varying the threshold, we were able to build foreground masks keeping a chosen fraction of the sky. We chose to build, for LFI 70 GHz, nine different masks with equally spaced sky fractions $f_{\text{sky}} = 30\%, 35\%, \dots, 65\%, 70\%$. We did not consider larger sky fractions because, as we show in [Sec. 4.4](#), we find an indication of excess residual power in the LFI maps after foreground removal for masks with $f_{\text{sky}} > 60\%$. A subset of the LFI masks is shown in the left panel of [Fig. 4.1](#).

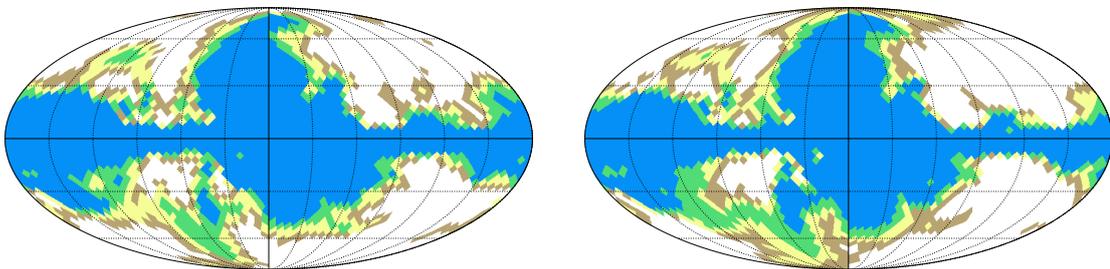


Figure 4.1: Left panel: Subset of the masks used in the analysis of the LFI 70 GHz data. Values of the available sky fraction f_{sky} in each mask are 30%, 40%, 50%, and 60%. Right panel: Subset of the masks used in the analysis of the WMAP data. Values of the available sky fraction f_{sky} in each mask are 30%, 40%, 50%, and 60%.

Table 4.1: Foreground scalings coefficients from WMAP K-band (α) and *Planck* 353 GHz (β) to the indicated WMAP channels.

Channel	α	β
Ka band	0.315	0.0031
Q band	0.163	0.0039
V band	0.047	0.0076

A corresponding set of masks for WMAP channels is built through a similar procedure. Here, we use the WMAP K-band as a tracer for synchrotron emission and *Planck* 353 GHz for dust. These are rescaled using the coefficients in [Lattanzi et al. \(2017\)](#); for completeness, these values are also reported in [Tab. 4.1](#) here. In this case, when the mask structure at intermediate and high latitudes is dominated by the synchrotron emission (i.e., by the K-band). Thus we decided to adopt the same mask for the three WMAP bands. This leads to a single set of ten masks, with a sky fraction ranging from 30% to 75% in steps of 5%. A subset of the masks is shown in the right panel of [Fig. 4.1](#).

Finally, with the aim of building a WMAP-*Planck* LFI combined dataset, we also produced another set of masks to be used in the analysis of the joint dataset. These were built by combining pairs of WMAP and LFI masks, taking the pixels that are left available in at least one of the two masks. In other words, if we think of a mask as the set of all pixels that can be used in the analysis, the “joint” masks are the union (in the set-theory meaning of the word) of the individual WMAP and LFI masks. For this reason, the sky fraction of each combined mask is always equal or larger than the sky fractions of the individual masks it is built from. For example, the union of the WMAP and *Planck* LFI 30% masks has $f_{\text{sky}} \simeq 35\%$. We then chose to produce a set of ten masks built as follows.

The first seven masks are the union of each pair of WMAP and *Planck* masks with the same sky fraction $f_{\text{sky}} = 30\%, 35\%, \dots, 55\%, 60\%$. The remaining three masks are the union of the LFI 60% mask with the WMAP 65%, 70% and 75% masks. The reason behind this choice is, as mentioned above and discussed in more detail in Sec. 4.5, that we do not consider the LFI masks with $f_{\text{sky}} > 60\%$ to be suitable for cosmological analyses. The sky fractions for the set of union masks, together with the individual masks used to produce them, are summarized in Table 4.2.

Table 4.2: Masks used in the analysis of the joint WMAP-*Planck* dataset. Each mask is built as the union of the individual masks, reported in the left column, and leaves the sky fraction reported on the right available for analysis.

Individual f_{sky} (WMAP \times <i>Planck</i> LFI)	total f_{sky}
30% \times 30% ...	35%
35% \times 35% ...	40%
40% \times 40% ...	45%
45% \times 45% ...	50%
50% \times 50% ...	54%
55% \times 55% ...	59%
60% \times 60% ...	63%
65% \times 60% ...	66%
70% \times 60% ...	70%
75% \times 60% ...	75%

4.3 Methods

In this section, we describe the cleaning procedure and the likelihood approximation used in cosmological parameter estimation. We pay particular attention to the impact of regularization noise on both scalings and cosmological parameters estimation and, at the end of the section, we discuss how it can be mitigated.

The cleaning procedure adopted here is based on fitting foreground templates at the map level (see, e.g., Page et al. 2007; Planck Collaboration XI 2016; Planck Collaboration V 2019). Denoting the linear polarization map at a given frequency, ν , with $\mathbf{m}_\nu^{\text{P}} = [\mathbf{Q}_\nu, \mathbf{U}_\nu]$, the corresponding foreground-cleaned map $\widetilde{\mathbf{m}}_\nu^{\text{P,fc}}$ is³

$$\widetilde{\mathbf{m}}_\nu^{\text{P,fc}} = \frac{\mathbf{m}_\nu^{\text{P}} - \alpha_\nu \mathbf{t}^{\text{s}} - \beta_\nu \mathbf{t}^{\text{d}}}{1 - \alpha_\nu - \beta_\nu}, \quad (4.1)$$

where \mathbf{t}^{s} (\mathbf{t}^{d}) and α_ν (β_ν) are the tracers and the scaling coefficient for synchrotron (dust) emission, respectively, described in Sec. 4.1.

If \mathbf{S}^{P} and $\mathbf{N}_\nu^{\text{P}}$ are, respectively, the signal and noise covariance matrices at frequency, ν , the fitted coefficients in Eq. (4.1) are estimated by minimization of the quantity:

$$\chi_\nu^2 = \left(\widetilde{\mathbf{m}}_\nu^{\text{P,fc}} \right)^\top \widetilde{\mathbf{C}}_\nu^{-1} \widetilde{\mathbf{m}}_\nu^{\text{P,fc}}, \quad (4.2)$$

where $\widetilde{\mathbf{C}}_\nu \equiv \left\langle \widetilde{\mathbf{m}}_\nu^{\text{P,fc}} \left(\widetilde{\mathbf{m}}_\nu^{\text{P,fc}} \right)^\top \right\rangle$ is the covariance matrix,

$$\widetilde{\mathbf{C}}_\nu = \mathbf{S}^{\text{P}}(C_\ell^{\text{fid}}) + \frac{\mathbf{N}_\nu^{\text{P}} + \alpha_\nu^2 \mathbf{N}^{\text{s}} + \beta_\nu^2 \mathbf{N}^{\text{d}}}{(1 - \alpha_\nu - \beta_\nu)^2}. \quad (4.3)$$

We note that χ_ν^2 is a χ^2 -distributed quantity when considered as a function of the map but not as a function of the scalings.

³Here “fc” stands for “foreground-cleaned”.

Here, N^s and N^d are the polarization parts of the NCVMs for the foregrounds tracers. The signal covariance matrix is built as described in Tegmark and de Oliveira-Costa (2001) and assumes a fiducial power spectrum, C_ℓ^{fid} , taken as the *Planck* legacy best-fit (Planck Collaboration VI 2018). The inversion of \tilde{C}_ν , needed to compute the χ^2 in Eq. (4.2), requires the addition of some regularization noise. In particular, we follow the approach used in the *Planck* legacy analysis (Planck Collaboration XI 2016; Planck Collaboration VI 2018) and consider white noise in polarization with rms $\sigma_r^P = 20$ nK. We thus sum a random white noise realization, \mathbf{n}_r^P , with this amplitude to $\tilde{\mathbf{m}}_\nu^{\text{P,fc}}$ and add a diagonal term, $N_r^P \equiv (\sigma_r^P)^2 \mathbf{I}$, to the covariance matrix (4.3) and then use these regularized objects to build the χ^2 in Eq. (4.2). In the following, we denote the cleaned map with regularization noise added as $\mathbf{m}_\nu^{\text{P,fc}} \equiv \tilde{\mathbf{m}}_\nu^{\text{P,fc}} + \mathbf{n}_r^P$ and the associated covariance matrix as $C_\nu \equiv \left\langle \mathbf{m}_\nu^{\text{P,fc}} (\mathbf{m}_\nu^{\text{P,fc}})^\top \right\rangle = \tilde{C}_\nu + N_r^P$.

Once α_ν and β_ν have been estimated through this minimization procedure, we can define the cleaned data vector $\mathbf{m}_\nu^{\text{fc}} \equiv [\mathbf{T}, \mathbf{m}_\nu^{\text{P,fc}}]$, with \mathbf{T} being the **Commander** map described in Sec. 4.1. We write down its likelihood function (cfr. Eq. (3.67)), $\mathcal{L}(C_\ell) \equiv P(\mathbf{m}_\nu^{\text{fc}}|C_\ell)$, as

$$-2 \log \mathcal{L}(C_\ell) = \log |\mathbf{S}(C_\ell) + N_\nu^{\text{fc}}| + (\mathbf{m}_\nu^{\text{fc}})^\top (\mathbf{S}(C_\ell) + N_\nu^{\text{fc}})^{-1} \mathbf{m}_\nu^{\text{fc}} + \text{const.} \quad (4.4)$$

The NCRM N_ν^{fc} used in the likelihood analysis is built as follows. The TT block is consistent with the **Commander** map having only white regularization noise with rms $\sigma_r^T = 2 \mu\text{K}$, while the TQ and TU blocks are vanishing. The polarization part $N_\nu^{\text{P,fc}}$ of the NCRM is instead given by

$$N_\nu^{\text{P,fc}} = \frac{N_\nu^P + \alpha_\nu^2 N_s + \beta_\nu^2 N_d + \sigma_{\alpha_\nu}^2 \mathbf{t}^s (\mathbf{t}^s)^\top + \sigma_{\beta_\nu}^2 \mathbf{t}^d (\mathbf{t}^d)^\top}{(1 - \alpha_\nu - \beta_\nu)^2} + N_r^P, \quad (4.5)$$

where σ_{α_ν} and σ_{β_ν} are the uncertainties in the estimates of foreground scaling coefficients and $\mathbf{t}^{s,d} (\mathbf{t}^{s,d})^\top$ is the outer product of the tracer maps.

The addition of regularization noise has a small, but not completely negligible, impact on the determination of the foreground scaling coefficients, and, consequently, on cosmological parameter estimates. In fact, the extra noise added to the map increases the scatter of point estimates (e.g., the posterior mean) of parameter values around the true value. Moreover, the extra term added to the NCRM increases parameter uncertainties. In what follows, we first assess the magnitude of the former effect at the level of both scaling coefficients and cosmological parameters. We then illustrate how we manage to avoid extracting a particular noise realization, which leads to non-negligible scatter (as compared to the one caused by instrumental noise).

In order to show and quantify the extra scatter in the estimates of α , β and cosmological parameters induced by regularization noise, we proceed as follows. We draw 1000 white noise realizations, $\mathbf{n}_{r,i}$ ($i = 1, \dots, 1000$), with an rms of $2 \mu\text{K}$ in temperature and 20 nK in polarization. We then estimate α and β on the *Planck* 70 GHz channel, following the procedure illustrated above, using each of the realization just described as the regularization noise map. For the sake of this test, we adopt a mask that retains 50% of the sky. This procedure results in 1000 Monte Carlo estimates, α_i and β_i . Once the scaling coefficients have been obtained, we further proceed with an estimation of the cosmological parameters $(\log(10^{10} A_s))_i$ and τ_i from the likelihood in Eq. (4.4). We note that in this last step, we consistently use the same regularization noise used when fitting the scaling coefficients.

Since the CMB signal and the instrumental noise are the same in each map belonging to this ensemble, the scatter in the recovered values of the parameters provides an estimate

of the dependence on the regularization noise realization, at the level of both scalings and cosmological parameters. The results of this procedure are shown in Fig. 4.2, where we show the distribution of the α_i 's, β_i 's, and τ_i 's with respect to the mean value, in units of the average uncertainty. We also show the χ^2 computed from Eq. (4.2) in units of $\sigma_{\chi^2} = \sqrt{2N_{\text{dof}}}$.

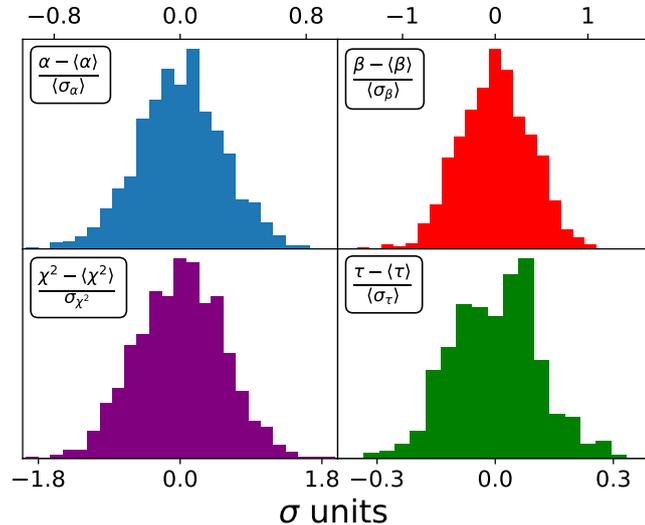


Figure 4.2: Histograms of the expected scatter in the recovered foreground scalings, in the χ^2 of the component separation and in the measured τ due to the regularization noise. For each quantity, we show the distance from the center of the empirical distribution in units of σ .

We then compute their standard deviation, that is, $\sqrt{\langle(\theta - \langle\theta\rangle)^2\rangle}/\langle\sigma_\theta\rangle$, where $\theta = \{\alpha, \beta, \chi^2, \tau\}$. For the synchrotron scaling coefficient, α , the scatter induced by the extra noise is, on average, 0.27 times the average parameter uncertainty. In other words, 68% of the α_i deviates from $\langle\alpha\rangle$ by less than $0.27 \times \langle\sigma_\alpha\rangle$. The corresponding value for β is 0.38 times the average parameter uncertainty. The χ^2 of the cleaned map is the most affected quantity by the particular realization of regularization noise. In fact, the impact is, at the 1σ level, at most 0.55 times the expected width of a χ^2 distribution with N_{dof} degrees of freedom. This extra scatter in the scaling estimates induces a smaller, but still non-negligible, effect on the final τ determination. The effect on τ , at one standard deviation of the distribution, equals 11% of its average uncertainty.

Thus, when we add regularization noise, we pay the price of an increased parameter uncertainty, and we might also be prone to unwanted parameter shifts caused by an unlucky choice of the actual noise realisation used. For example, a $3\text{-}\sigma$ noise realization can easily shift the scalings by $\sim 1\sigma$ and τ by 0.3σ . In fact roughly 1% of the noise realizations in our Monte Carlo resulted in shifts larger than 1 and 0.3σ 's for the scalings and τ , respectively.

A possible way to avoid large parameter shifts is to somehow average over different realizations of the regularization noise. In order to do so, we draw $N_{it} = 1000$ white noise realizations $\mathbf{n}_{r,i}$ ($i = 1, \dots, 1000$) with 20 nK rms. For given values of α and β , these are used to build as many cleaned polarization maps $\mathbf{m}_i^{\text{P,fc}} = \widetilde{\mathbf{m}}^{\text{P,fc}} + \mathbf{n}_{r,i}$ and the following quantity:

$$\overline{\chi^2} = \frac{1}{N_{it}} \sum_{i=1}^{N_{it}} \left(\mathbf{m}_i^{\text{P,fc}} \right)^\top \mathbf{C}^{-1} \mathbf{m}_i^{\text{P,fc}}. \quad (4.6)$$

We note that $\overline{\chi^2}$ does not follow a chi-square distribution. It is straightforward to show

that its expectation value over the regularization noise is

$$\langle \overline{\chi^2} \rangle_{\mathbf{n}_r} = \left(\widetilde{\mathbf{m}}^{\text{P,fc}} \right)^\top \mathbf{C}^{-1} \widetilde{\mathbf{m}}^{\text{P,fc}} + \text{Tr} \left(\mathbf{C}^{-1} \mathbf{N}_r^{\text{P}} \right), \quad (4.7)$$

which is the same as the expectation value of the χ^2 built from a single regularized map, $\chi^2 = \left(\mathbf{m}^{\text{P,fc}} \right)^\top \mathbf{C}^{-1} \mathbf{m}^{\text{P,fc}}$. Also, this expectation value is different from the value of the χ^2 on the regularization noise-free map, $\left(\widetilde{\mathbf{m}}^{\text{P,fc}} \right)^\top \widetilde{\mathbf{C}}^{-1} \widetilde{\mathbf{m}}^{\text{P,fc}}$. The variance associated to $\overline{\chi^2}$ is:

$$\text{Var} \left[\overline{\chi^2} \right]_{\mathbf{n}_r} = \frac{1}{N_{it}} \left\{ 4 \left(\widetilde{\mathbf{m}}^{\text{P,fc}} \right)^\top \mathbf{C}^{-1} \mathbf{N}_r^{\text{P}} \mathbf{C}^{-1} \widetilde{\mathbf{m}}^{\text{P,fc}} + 2 \text{Tr} \left[\left(\mathbf{C}^{-1} \mathbf{N}_r^{\text{P}} \right)^2 \right] \right\} \quad (4.8)$$

that, as should be expected, goes to 0 as the number of noise realizations, over which the average is performed, increases.

For these reasons, we chose to minimize the quantity in Eq. (4.6) to obtain estimates of the scaling coefficients that are less dependent on the particular realization of regularization noise. Similarly, when estimating cosmological parameters, we performed an analogous procedure by drawing $N_{it} = 1000$ noise realizations in temperature and polarization, and using the average of the quantity defined in Eq. 4.4 over these realizations. The results of these procedures are presented in the next sections.

4.4 Foreground cleaning

In this section, we discuss the results of the estimation of the synchrotron and dust scaling coefficients for the different channels in various masks. We also discuss how this leads to the choice of the “confidence” masks that are used to produce the foreground-cleaned maps for each channel and how inverse-noise-weighted combinations of these maps are built.

We clean independently the four cosmological channels (i.e., WMAP Ka, Q and V bands and Planck 70 GHz), following the template-fitting procedure described in Sec. 4.3. We thus minimize Eq. (4.6) to estimate the synchrotron, α_ν , and dust, β_ν , scaling coefficients for each map. The final polarization map, $\widetilde{\mathbf{m}}^{\text{P,fc}}$, and polarization noise covariance matrix, $\mathbf{N}^{\text{P,fc}}$, are given by Eqs. (4.1) and (4.5).

Figure 4.3 show the scaling coefficients computed for each cosmological channel in the masks described in Sec. 4.2. In the bottom panel of each figure, we also show the excess χ^2 in units of the expected dispersion, $\sqrt{2N_{\text{dof}}}$, that is: $\Delta\chi^2 = (\chi_\nu^2 - N_{\text{dof}}) / \sqrt{2N_{\text{dof}}}$, where χ_ν^2 is computed from Eq. (4.2).

We use the $\Delta\chi^2$ values to select the processing mask to be used in the template fitting. The rule of thumb is to use the mask with the largest f_{sky} among those with $\Delta\chi^2 \leq 2$. The only exception is represented by WMAP Ka band which shows a sudden change in both scalings, as well an noticeable increase in the excess χ^2 between the 55% and 60% masks (see Fig. 4.3). In this case, we cautiously choose to use the 55% mask, even though the excess χ^2 itself remains slightly below 2 also in the 60% mask. We note how similar jumps between the 55% and 60% masks are evident also in the scalings of the WMAP Q band, shown in the top-right panel of Fig. 4.3. An interesting case is represented by WMAP V band (bottom-left panel of Fig. 4.3), for which the $\Delta\chi^2$ reaches a maximum in the 55% mask before decreasing for larger masks, without ever reaching the threshold $\Delta\chi^2 = 2$. In this case we chose the 75% mask. The masks used in the foreground cleaning, together with the scaling coefficients obtained, are reported in Tab. 4.3. Note that, using

the scaling relations

$$\alpha_\nu = \left(\frac{\nu}{\nu_s} \right)^{\beta_s}, \quad (4.9)$$

$$\beta_\nu = \left(\frac{\nu}{\nu_d} \right)^{\beta_d-1} \frac{B_\nu(T_d)}{B_{\nu_d}(T_d)}, \quad (4.10)$$

it is possible to estimate the corresponding spectral index for synchrotron and dust. Here $T_d = 119.6$ K, and $B_\nu(T_d)$ is the spectrum of a black body, see (2.1). These are shown in Tab. 4.3 and are in agreement with the current estimations. It is worth to notice that, a simple template fitting does not take into account the impact of spatial variability of the foreground emissions (see e.g., Fuskeland et al. 2014; Krachmalnicoff et al. 2016). Even if the χ^2 shows the stability of our assumption of constant scalings, we perform a simple test using the V band as tester. We select the 75% mask and we downgrade it to $N_{side} = 2$. We then set to zero all pixel outside the mask with the exception of one of them. We repeat this procedure covering all pixel out of 48 available and estimate the scaling coefficients. The result is shown in Fig. 4.4. This test show that there is only one outlier with respect the assumption of constant scaling for the synchrotron. A further investigation for this behaviour is left for a future work.

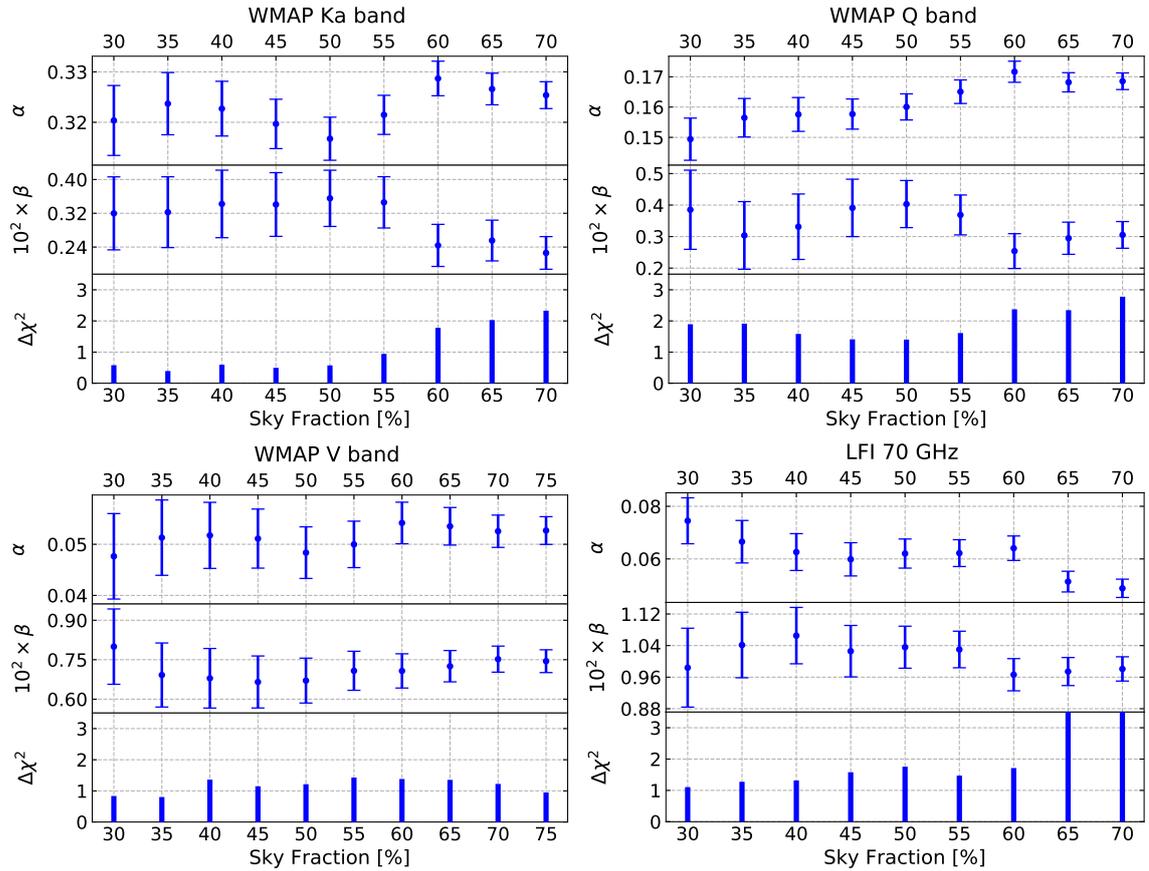
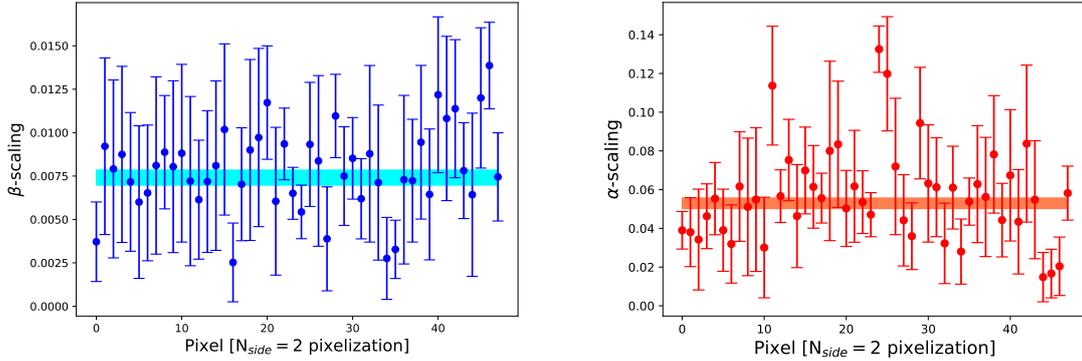


Figure 4.3: Scaling coefficients for synchrotron and dust (top and middle panels) estimated on different masks for WMAP Ka (top-left), Q (top-right), V (bottom-left) bands and for the *Planck* LFI 70 GHz channel (bottom-right). The bottom panel of each plot shows the excess χ^2 in units of $\sqrt{2N_{\text{dof}}}$, i.e., $\Delta\chi^2 = (\chi^2 - N_{\text{dof}}) / \sqrt{2N_{\text{dof}}}$.

The resulting cleaned maps can then be combined together to build inverse-noise-weighted maps. In particular, we build two combinations: the first is a “WMAP-only” map built from the Ka, Q, and V bands, while the second is a joint “WMAP+LFI” map, that uses the Ka, Q, and V WMAP bands and *Planck* LFI 70 GHz (hereafter WMAP+LFI).

Table 4.3: Masks used to produce foreground-cleaned maps for each channel, the corresponding estimates for the scaling coefficients, and their spectral index.

Channel	Mask	α	β_s	β	β_d
Ka band	55%	0.3215 ± 0.0039	-1.50 ± 0.02	0.00346 ± 0.00061	2.44 ± 0.16
Q band	55%	0.1651 ± 0.0039	-2.38 ± 0.03	0.00369 ± 0.00063	1.86 ± 0.15
V band	75%	0.0527 ± 0.0027	-3.91 ± 0.05	0.00744 ± 0.00043	1.37 ± 0.06
70 GHz	60%	0.0641 ± 0.0046	-3.59 ± 0.07	0.00966 ± 0.00041	1.21 ± 0.04

Figure 4.4: Scaling coefficients for synchrotron and dust (left and right panels) estimated on different masks for V at $N_{side} = 2$. Each pixel on the abscissa means that the corresponding pixel is the only one different from zero. The central shaded region correspond to the one sigma region for the scaling coefficient estimated at the same resolution using the original mask.

To provide more detail, if $\widetilde{\mathbf{m}}_\nu^{\text{P,fc}}$ is the final cleaned map of the band, ν , with corresponding noise-covariance matrix, $\mathbf{N}_\nu^{\text{P,fc}}$, the final noise weighted map, \mathbf{m}^{nw} , is built as:

$$\begin{aligned} \mathbf{m}^{\text{nw}} &= \left[\sum_\nu \left(\mathbf{N}_\nu^{\text{P,fc}} \right)^{-1} \right]^{-1} \sum_\nu \left(\mathbf{N}_\nu^{\text{P,fc}} \right)^{-1} \widetilde{\mathbf{m}}_\nu^{\text{P,fc}} \\ &= \overline{\mathcal{N}} \sum_\nu \left(\mathbf{N}_\nu^{\text{P,fc}} \right)^{-1} \widetilde{\mathbf{m}}_\nu^{\text{P,fc}}, \end{aligned} \quad (4.11)$$

where we define the total noise covariance matrix:

$$\overline{\mathcal{N}} = \left[\sum_\nu \left(\mathbf{N}_\nu^{\text{P,fc}} \right)^{-1} \right]^{-1}. \quad (4.12)$$

We note that in Eq. (4.11), we use the $\widetilde{\mathbf{m}}_\nu^{\text{P,fc}}$ with no regularization noise. This is because we do not want to “bring” the regularization noise into the noise-weighted map as we want to avoid possible biases in parameter estimates induced by particular realizations of the regularization noise, as explained in Sec. 4.3. However, we are forced to use the covariance matrices $\mathbf{N}_\nu^{\text{P,fc}}$ that do include regularization noise since, otherwise, we would not be able to invert them. For this reason, it is evident that $\overline{\mathcal{N}}$ would be the NCV of a noise-weighted combination built from the (un-tilded) $\mathbf{m}_\nu^{\text{P,fc}}$, but is not the NCV of \mathbf{m}^{nw} . The actual NCV can be computed by rewriting Eq. 4.11 as

$$\mathbf{m}^{\text{nw}} = \overline{\mathcal{N}} \sum_\nu \left(\mathbf{N}_\nu^{\text{P,fc}} \right)^{-1} \left(\mathbf{s} + \mathbf{n}_\nu^{\text{t}} - \mathbf{n}_\nu^{\text{r}} \right), \quad (4.13)$$

where $\mathbf{n}_\nu^{\text{t}}$ denotes the total noise, i.e., the sum of instrumental and regularization noise,

at each frequency. Taking the expectation value of $\mathbf{m}^{\text{nw}}(\mathbf{m}^{\text{nw}})^{\text{T}}$ yields

$$\begin{aligned} \langle \mathbf{m}^{\text{nw}}(\mathbf{m}^{\text{nw}})^{\text{T}} \rangle &= \mathbf{S} + \overline{\mathcal{N}} \sum_{\nu\nu'} \left(\mathbf{N}_{\nu}^{\text{P,fc}} \right)^{-1} \langle (\mathbf{n}_{\nu}^{\text{t}} - \mathbf{n}_{\nu}^{\text{r}})(\mathbf{n}_{\nu'}^{\text{t}} - \mathbf{n}_{\nu'}^{\text{r}})^{\text{T}} \rangle \left(\mathbf{N}_{\nu'}^{\text{P,fc}} \right)^{-1} \overline{\mathcal{N}} \\ &= \mathbf{S} + \overline{\mathcal{N}} - \left(\sigma_{\text{r}}^{\text{P}} \right)^2 \overline{\mathcal{N}} \sum_{\nu} \left[\left(\mathbf{N}_{\nu}^{\text{P,fc}} \right)^{-1} \right]^2 \overline{\mathcal{N}}, \end{aligned} \quad (4.14)$$

where we are assuming that the regularization noise rms $\sigma_{\text{r}}^{\text{P}}$ for all the involved maps is the same. In the last equation, we use the fact that $\langle \mathbf{n}_{\nu}^{\text{t}} \mathbf{n}_{\nu'}^{\text{r}} \rangle = \langle \mathbf{n}_{\nu}^{\text{r}} \mathbf{n}_{\nu'}^{\text{r}} \rangle = \left(\sigma_{\text{r}}^{\text{P}} \right)^2 \mathbf{I} \delta_{\nu\nu'}$. Thus, the final noise covariance matrix of the combined instrumental noise is

$$\mathcal{N} \equiv \overline{\mathcal{N}} - \left(\sigma_{\text{r}}^{\text{P}} \right)^2 \overline{\mathcal{N}} \sum_{\nu} \left[\left(\mathbf{N}_{\nu}^{\text{P,fc}} \right)^{-1} \right]^2 \overline{\mathcal{N}}. \quad (4.15)$$

4.5 Power Spectra

In this section, we present our results for the angular power spectra of the maps described in the previous sections. In particular, we use a QML code (Tegmark 1996; Tegmark and de Oliveira-Costa 2001) to extract the auto power spectra of the cleaned maps described in Sec. 4.4. In our analysis, power spectra are not directly used for the cosmological parameter extraction. We mainly use them as a probe of possible residual systematics in the maps and, consequently, for selecting the masks suitable for the likelihood analysis. The main tool for performing these consistency tests is the χ^2 in harmonic space, defined as:

$$\chi_{\text{h}}^2 = \sum_{\ell, \ell'=2}^{\ell_{\text{max}}} (C_{\ell} - C_{\ell}^{\text{th}}) \mathbf{M}_{\ell\ell'}^{-1} (C_{\ell'} - C_{\ell'}^{\text{th}}), \quad (4.16)$$

where C_{ℓ} is the power spectrum estimated from a given map-and-mask combination, $\mathbf{M}_{\ell\ell'}^{-1}$ is the Fisher matrix and C_{ℓ}^{th} is the power spectrum of a fiducial Λ CDM model with optical depth of $\tau = 0.065$ and logarithmic amplitude of primordial scalar fluctuations of $\ln(10^{10} A_{\text{s}}) = 3.0343$. We perform separate tests for the TE , TB , EE , EB , and BB power spectra. The quantity in Eq. (4.16) can be compared to the χ^2 distribution with $\ell_{\text{max}} - 1$ degrees of freedom, computing the corresponding probability-to-exceed (hereafter PTE). In Tables 4.4, 4.5, and 4.6, we report the PTEs for LFI, WMAP and WMAP+LFI for different sky fractions, corresponding to the masks presented in Sec. 4.2. As explained in that section, for the WMAP+LFI dataset, the masks are obtained by combining the individual LFI and WMAP masks. We refer the reader to Table 4.2 for further details.

Here, we consider $2 \leq \ell \leq 10$, which corresponds roughly to the multipole range affected by the reionization feature. For WMAP, the PTEs are nicely compatible with the theoretical model for all the sky fractions considered. For the LFI dataset we see $\sim 2\sigma$ deviations for the BB spectrum for intermediate sky fractions ($f_{\text{sky}} = 40\%$ and $f_{\text{sky}} = 45\%$), fluctuations reabsorbed in larger sky fractions. In the WMAP+LFI dataset, we do not see any particular failure in the PTEs. We define a “failure” as a PTE $< 1\%$.

We further perform additional consistency tests for the combined dataset. We compute the PTEs for different choices of ℓ_{max} , exploring the χ^2 consistency up to $\ell = 15$ and $\ell = 29$. For all the sky fraction we have considered, we do not observe any failure in the total PTEs as a function of ℓ_{max} . We also compute the ℓ -by- ℓ PTEs for all the polarization power spectra. The mask keeping a 54% fraction of the whole sky has the lowest number of outliers above 2.5σ : only 3 out of a total 140 analysed multipoles. As we explain in Sec. 4.6, the 54% mask also represents a robust choice for the likelihood analysis. The

Table 4.4: Probability to exceed χ_h^2 for LFI 70 GHz as a function of the sky fraction. The maximum multipole used to compute χ_h^2 is $\ell_{\max} = 10$.

Sky Fraction	PTE [%]				
	TE	EE	BB	TB	EB
30%	62.2	87.7	16.0	56.2	74.2
35%	49.9	75.3	38.5	32.8	51.6
40%	24.6	72.7	4.1	38.7	51.3
45%	15.6	56.2	5.5	46.7	60.6
50%	25.7	43.9	23.5	45.3	83.4
55%	23.1	34.8	34.8	23.4	97.4
60%	29.9	35.4	20.4	35.2	97.9

Table 4.5: Probability to exceed χ_h^2 for WMAP as a function of the sky fraction. The maximum multipole used to compute χ_h^2 is $\ell_{\max} = 10$.

Sky Fraction	PTE [%]				
	TE	EE	BB	TB	EB
30%	64.3	71.7	50.3	59.9	92.6
35%	91.3	32.5	81.3	29.0	97.2
40%	84.1	70.8	79.4	23.2	90.4
45%	82.2	91.0	92.6	19.5	61.0
50%	74.8	73.6	84.5	22.8	36.2
55%	81.7	95.3	65.3	21.3	50.0
60%	70.7	94.2	58.3	21.1	83.1
65%	57.5	92.2	66.2	20.4	75.2
70%	53.2	85.8	69.7	34.9	78.3
75%	55.7	85.4	67.7	37.5	63.5

Table 4.6: Probability to exceed χ_h^2 for the WMAP+LFI dataset as a function of the sky fraction. The value of maximum multipole used is fixed to $\ell_{\max} = 10$.

Sky Fraction	PTE [%]				
	TE	EE	BB	TB	EB
35%	41.8	80.6	11.2	15.9	11.4
40%	55.6	92.5	25.8	14.7	35.1
45%	37.6	93.5	17.4	18.0	12.7
50%	26.7	86.4	37.9	11.1	27.8
54%	31.9	91.6	36.6	5.8	33.5
59%	40.4	84.3	25.7	4.9	22.2
63%	35.3	85.3	37.1	12.4	50.2
66%	69.0	98.4	13.2	38.9	69.1
70%	72.3	97.8	26.9	50.3	74.1
75%	77.1	96.4	30.2	50.3	80.5

results of the PTEs computation for the combined dataset analysed in the 54% mask are reported in Table 4.7 and Table 4.8.

The spectra for WMAP, LFI and WMAP+LFI are shown in Fig. 4.5, in their own 50%, 50%, and 54% masks, respectively.

4.6 Likelihood and validation

In this section, we show the results of additional consistency tests performed at the level of parameter estimation. This allows us to test and validate both the datasets produced and the likelihood algorithm.

Table 4.7: Probability to exceed χ_h^2 for the combined dataset WMAP+LFI for different choices of ℓ_{\max} . Here the mask used to extract the power spectra is the combined mask with $f_{\text{sky}} = 54\%$.

Spectrum	PTE [%]		
	$\ell_{\max} = 10$	$\ell_{\max} = 15$	$\ell_{\max} = 29$
TE	31.9	54.2	66.5
EE	91.6	98.4	98.5
BB	36.6	32.8	14.8
TB	5.8	12.9	32.7
EB	33.5	58.7	56.2

Table 4.8: Probability to exceed χ_h^2 for the WMAP+LFI dataset ℓ -by- ℓ . Here the mask used to extract the power spectra is the combined mask with $f_{\text{sky}} = 54\%$.

Multipole	PTE [%]				
	TE	EE	BB	TB	EB
2	21.1	71.9	98.9	51.9	34.4
3	20.6	26.8	11.8	5.7	7.2
4	87.0	77.8	65.9	33.7	68.4
5	6.0	56.5	27.2	17.3	48.3
6	44.9	44.9	60.9	49.0	4.5
7	99.3	36.4	94.4	1.7	90.6
8	33.3	73.1	6.5	11.0	20.0
9	16.4	36.6	18.9	40.4	55.0
10	63.8	79.0	30.7	66.6	97.1
11	20.8	95.4	91.1	24.7	98.3
12	47.1	53.1	33.0	27.9	67.3
13	85.4	53.7	13.8	38.2	24.7
14	70.3	48.5	61.7	58.9	50.6
15	73.1	72.9	19.1	94.5	94.6
16	94.3	71.1	84.4	42.2	43.1
17	8.0	21.4	50.1	28.8	64.0
18	93.7	73.1	0.9	59.0	71.5
19	60.2	60.5	55.5	38.4	26.1
20	80.6	48.1	96.1	14.5	18.5
21	24.0	38.6	84.3	43.1	49.1
22	81.0	72.0	70.6	81.3	54.6
23	19.6	62.0	0.7	60.7	1.0
24	41.0	13.3	74.1	85.1	67.9
25	84.0	33.1	60.8	71.8	38.4
26	9.9	27.1	13.2	36.0	31.4
27	66.5	30.1	93.9	68.8	47.7
28	29.8	86.1	94.4	26.6	33.4
29	60.3	38.8	22.5	12.3	95.1

Parameter estimates are obtained from the likelihood function in Eq. (4.4). Since we are using low-resolution maps with $N_{\text{side}} = 16$, only the C_ℓ 's from $\ell = 2$ to $\ell_{\text{cut}} = 29$ are varied in accordance to the theoretical model that is being tested, when computing the signal covariance matrix; multipoles between $\ell_{\text{cut}} + 1 = 30$ and $\ell_{\max} = 64$ are instead fixed to a fiducial Λ CDM spectrum (Page et al. 2007; Planck Collaboration XI 2016; Planck Collaboration V 2019). We follow the procedure described in Sec. 4.3 in order to marginalize over the regularization noise.

As consistency test for the likelihood, we explore the stability of the reionization optical depth τ constraints with respect the mask used for cosmological parameter estimation. Thus, keeping fixed the underlying datasets (i.e., map and associated covariance matrix) we only change the cosmological parameter mask used in Eq. 4.4. The results of these test are reported in Fig. 4.6, respectively, for LFI, WMAP and WMAP+LFI. Visually

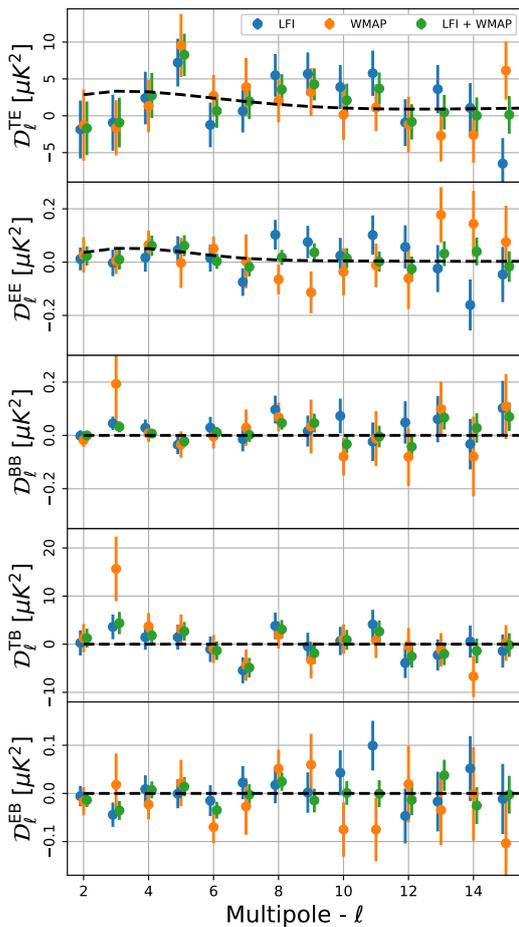


Figure 4.5: Polarization power spectra of the LFI 70 GHz, WMAP bands and WMAP+LFI. The sky fractions used are respectively 50%, 50%, and 54%. The dashed lines represent a Λ CDM power spectra corresponding to an optical depth value of $\tau = 0.065$.

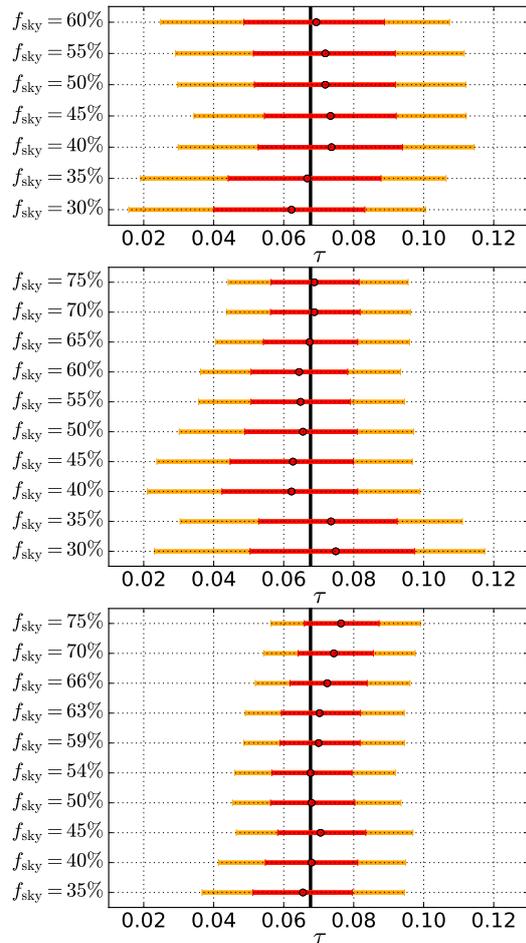


Figure 4.6: Estimates of τ from the LFI 70 GHz map (top), for the combination of WMAP Ka, Q and V bands (middle), and for the WMAP+LFI dataset (bottom) analysed in different masks. In this plots, the points represent the best-fit values and the red and yellow bars represent the 68% and 95% C.L., respectively. The vertical grey line represents the best-fit value for the WMAP+LFI baseline dataset, which uses 54% of the sky; see text for details.

all the τ constraints are nicely compatible with each other for LFI and WMAP. For the WMAP+LFI, the τ posteriors are still visually compatible with each other, but we observe a clear trend towards high values of τ for large sky fractions. It is worth mentioning that all the masks we are using for a given dataset are nested, and largely overlapped, so relying on a simple visual comparison can be misleading and we need a more accurate statistical test to assess consistency. Thus, for each dataset, we generated a Monte Carlo of 1000 CMB maps, with $\tau = 0.065$ and a set of realistic noise simulations extracted from the noise covariance matrix of the cleaned datasets (Eq. 4.5 and Eq. 4.15) through Cholesky decomposition. For every mask, we processed all those maps through a pixel-based likelihood algorithm implementing the function in Eq. (4.4), fitting τ and $\ln(10^{10} A_s)$ from a grid of models. All the other Λ CDM parameters were kept fixed to the best-fit of Pagano et al. (2019). Then we were able to build the statistics of the difference $\Delta\tau_{ij} \equiv |\tau_i - \tau_j|$ between the τ estimates for each pair of masks $\{i, j\}$, and, finally, compare this with the values of $\Delta\tau_{ij}$ obtained from the real data

In Table 4.9, we report the PTEs for the three datasets analysed, defined as the percentage of simulations that have an absolute parameter shift larger than the same quantity measured on the data. As explained above, the simulations used for this test contain only the CMB signal and noise drawn by cleaned map covariance matrix. We note that foreground residuals and, thus, chance correlations between such residuals and noise realisations are not included in the simulations; this makes the test conservative since it is more difficult to pass. The scatter we see on the data is perfectly compatible with the signal plus noise simulations, independently for LFI and WMAP, and for all the sky fractions considered. The WMAP+LFI dataset, instead, shows mild failures for sky fractions larger than 60%. This comes as some of a surprise since the corresponding results based on WMAP show excellent PTEs (see, e.g., columns 2 and 3 of Table 4.9). Again, we verified that the shift between the $\Delta\tau$ for WMAP and WMAP+LFI are compatible with what is seen in our signal plus noise simulations. We find that all the shifts are within $2\text{-}\sigma$. This indicates that as far as τ is concerned, all masks return consistent values and may be thus used in the analysis. However, we remark that there is a clear trend towards larger values of τ for masks with a sky fraction larger than $f_{\text{sky}}=63\%$ (again, see Fig. 4.6). Based on these considerations and on the fact that this is the one performing better in the ℓ -by- ℓ tests described in Sec. 4.5, we opt for a conservative choice and select the 54% mask as the baseline for WMAP+LFI. This dataset provides an error on τ that is 12% smaller than the one obtained from WMAP on the 75% mask.

For the baseline mask, in Table 4.10, we report the constraints on τ , $\ln(10^{10} A_s)$, both with $r = 0$ and variable r from the low-multipole dataset alone, having fixed the other Λ CDM parameters to the best-fit of Pagano et al. (2019). In the next section, we offer a detailed discussion on the τ constraints and its consequences for the cosmological scenario.

4.7 Reionization constraints

The CMB large-scale polarization data provide an almost direct measurement of the optical depth to reionization, being $C_\ell^{\text{EE}} \propto \tau^2$ and $C_\ell^{\text{TE}} \propto \tau$ for multipoles $\ell \lesssim 20$. In this section, we use the WMAP+LFI dataset in polarization, together with the Commander 2018 solution in temperature, to derive updated constraints on τ from CMB measurements at low frequencies.

For the cosmological parameter tests presented in this PhD thesis, we adopted the reionization model given in Lewis (2008). This is the default model in camb⁴ and it has been used for the Planck baseline cosmological results (TANH). In this model, the phase change in the intergalactic medium from the almost completely neutral state (up to a residual ionization fraction of 10^{-4} , remaining after recombination) to the ionized state is described as a sharp transition. The hydrogen reionization is assumed to happen simultaneously with the first reionization of helium, whereas the second reionization of helium is fixed at a redshift of $z = 3.5$ and is, again, described as a sharp transition. This choice is motivated by the expectations drawn from quasar spectra. Nevertheless, we expect the modeling of the helium double ionization to have a minor impact on the final results because varying the corresponding reionization redshift between 2.5 and 4.5 changes the total optical depth by less than 1% (Planck Collaboration Int. XLVII 2016). For the tests presented in this chapter, we did not explore different reionization models, however, it has been shown in Planck Collaboration VI (2018) that τ constraints from the latest Planck data have little sensitivity with regard to the actual details of the reionization history. Furthermore, earlier claims by Heinrich and Hu (2018) of a mild evidence, in Planck 2015 LFI data, for a more complex model of the ionization fraction, with hints of early reionization, have not been confirmed by alternate analyses of the same data set

⁴<https://camb.info>

Table 4.9: Consistency of the τ parameter values estimated on different masks. For each pair of masks defined in Sect. 4.2, we report the percentage of simulations with absolute parameter shift larger than the same quantity measured on the data. Each column corresponds to a different dataset, LFI (left), WMAP (center), and WMAP+LFI (right).

LFI PTE [%]		WMAP PTE [%]		WMAP+LFI PTE [%]	
Masks	$ \Delta\tau $	Masks	$ \Delta\tau $	Masks	$ \Delta\tau $
30%–35% ...	49.0	30%–35% ...	84.3	35%–40% ...	56.9
30%–40% ...	21.3	30%–40% ...	21.1	35%–45% ...	38.0
30%–45% ...	27.2	30%–45% ...	33.4	35%–50% ...	72.3
30%–50% ...	36.4	30%–50% ...	49.0	35%–54% ...	76.0
30%–55% ...	40.3	30%–55% ...	49.7	35%–59% ...	55.9
30%–60% ...	54.0	30%–60% ...	49.6	35%–63% ...	54.1
35%–40% ...	21.2	30%–65% ...	65.6	35%–66% ...	37.6
35%–45% ...	33.4	30%–70% ...	71.6	35%–70% ...	26.5
35%–50% ...	52.8	30%–75% ...	71.4	35%–75% ...	17.6
35%–55% ...	56.6	35%–40% ...	15.9	40%–45% ...	46.0
35%–60% ...	79.6	35%–45% ...	22.8	40%–50% ...	97.5
40%–45% ...	95.8	35%–50% ...	43.1	40%–54% ...	96.2
40%–50% ...	75.5	35%–55% ...	48.8	40%–59% ...	74.9
40%–55% ...	80.2	35%–60% ...	47.3	40%–63% ...	71.7
40%–60% ...	55.1	35%–65% ...	63.4	40%–66% ...	48.5
45%–50% ...	69.7	35%–70% ...	72.7	40%–70% ...	32.9
45%–55% ...	76.3	35%–75% ...	73.7	40%–75% ...	18.6
45%–60% ...	50.6	40%–45% ...	91.5	45%–50% ...	34.6
50%–55% ...	95.6	40%–50% ...	67.0	45%–54% ...	49.6
50%–60% ...	58.3	40%–55% ...	79.1	45%–59% ...	88.7
55%–60% ...	40.8	40%–60% ...	83.5	45%–63% ...	96.8
.....	–	40%–65% ...	64.5	45%–66% ...	68.7
.....	–	40%–70% ...	56.2	45%–70% ...	46.9
.....	–	40%–75% ...	58.1	45%–75% ...	29.7
.....	–	45%–50% ...	63.1	50%–54% ...	92.9
.....	–	45%–55% ...	78.1	50%–59% ...	58.5
.....	–	45%–60% ...	83.6	50%–63% ...	56.1
.....	–	45%–65% ...	59.2	50%–66% ...	29.1
.....	–	45%–70% ...	52.6	50%–70% ...	16.6
.....	–	45%–75% ...	55.1	50%–75% ...	7.7
.....	–	50%–55% ...	87.3	54%–59% ...	32.7
.....	–	50%–60% ...	83.3	54%–63% ...	37.8
.....	–	50%–65% ...	77.9	54%–66% ...	13.4
.....	–	50%–70% ...	66.5	54%–70% ...	6.5
.....	–	50%–75% ...	68.7	54%–75% ...	3.2
.....	–	55%–60% ...	87.9	59%–63% ...	78.6
.....	–	55%–65% ...	55.1	59%–66% ...	18.6
.....	–	55%–70% ...	45.0	59%–70% ...	8.0
.....	–	55%–75% ...	49.9	59%–75% ...	3.4
.....	–	60%–65% ...	33.4	63%–66% ...	7.2
.....	–	60%–70% ...	30.5	63%–70% ...	3.3
.....	–	60%–75% ...	36.6	63%–75% ...	1.5
.....	–	65%–70% ...	59.9	66%–70% ...	14.0
.....	–	65%–75% ...	66.5	66%–75% ...	4.1
.....	–	70%–75% ...	95.7	70%–75% ...	12.2

Table 4.10: Constraints on $\ln(10^{10}A_s)$, τ , and r from the WMAP+LFI likelihood. We show mean and 68% confidence levels. For r , the 95% upper limit is shown.

Parameter	Λ CDM	Λ CDM + r
$\ln(10^{10}A_s) \dots$	2.978 ± 0.050	$2.82^{+0.15}_{-0.08}$
$\tau \dots\dots\dots$	$0.069^{+0.011}_{-0.012}$	$0.067^{+0.011}_{-0.012}$
$r_{0.002} \dots\dots\dots$	\dots	≤ 0.79
$10^9 A_s e^{-2\tau} \dots$	$1.715^{+0.081}_{-0.092}$	$1.48^{+0.20}_{-0.14}$

(e.g., Villanueva-Domingo et al. (2018); Hazra and Smoot (2017); Dai et al. (2019); Hazra et al. (2019)). In particular, Millea and Bouchet (2018) have shown how the significance of those findings has been likely overestimated due to the choice of unphysical priors.

Having fixed the reionization model, first of all, we want to study the constraints from the large scales alone. Using the pixel-based likelihood framework of Sec. 4.6 (lowTEB), we only fit for τ , $\ln(10^{10}A_s)$, and r , while keeping all the other Λ CDM parameters fixed to the best-fit values given in Pagano et al. (2019). Our results are shown in Table 4.10, where the parameter, r , is estimated at a scale of $k = 0.002 \text{ Mpc}^{-1}$. The derived constraint on τ is

$$\tau = 0.069^{+0.011}_{-0.012} \quad (68\%, \text{lowTEB}), \quad (4.17)$$

which corresponds to a 5.8σ detection from the low-frequency CMB polarization data.

We then extended the analysis to include data from the small scales, specifically adding the *Planck* 2018 likelihood for TT, TE, EE angular power spectra (Planck Collaboration V 2019). This time, we let all the six base Λ CDM parameters vary, and we sampled from the space of possible cosmological parameters with an MCMC exploration using CosmoMC (Lewis and Bridle 2002). The reionization optical depth estimated in this case is⁵:

$$\tau = 0.074^{+0.010}_{-0.011} \quad (68\%, \text{TT,TE,EE}). \quad (4.18)$$

The parameter constraints we derived for pure Λ CDM are given in Table 4.11, where, for the purposes of comparison, we also report the *Planck* 2018 baseline results. The two compared datasets differ by the low- ℓ likelihoods. In one case, there is the pixel-based likelihood developed in this thesis (lowTEB), while in the other case, the low- ℓ likelihood is a combination of the Blackwell-Rao estimator for the Commander temperature solution and the E-mode power spectrum based *Planck* Legacy HFI likelihood (lowE). The latter likelihood provides a constraint on τ that is about 1.5 times tighter and 1.4σ lower in value than the one we obtain from the WMAP+LFI likelihood. Due to the well known degeneracy between A_s and τ , this also translates to a 33% tighter constraint on $\ln(10^{10}A_s)$ and 1.8σ lower in value. All the other cosmological parameters, rather, are in good agreement, differing by at most 36% of the σ . A similar tendency is also found when comparing Table 4.11 with an analogous analysis shown in Pagano et al. (2019).

Comparing the constraints from Table 4.10 and Table 4.11, we note that the values of $\ln(10^{10}A_s)$ and τ derived from the large scales alone are 1.9σ and 0.4σ lower, respectively. This behaviour was first noticed in Planck Collaboration XI (2016) and it is known to be induced by the low- ℓ anomaly, that is, the power deficit in the measured *TT* power spectrum with respect to the best-fit model at multipoles between $\ell = 20$ and 30. When limiting the analysis to the large scales, that is, to multipoles up to 30, the deficit has a high relative weight in the final solution, leading to a value of the overall amplitude of the spectrum that is lower than the one from the full analysis, which includes multipoles up

⁵In the following, the presence of the lowTEB dataset should be always understood.

Table 4.11: Parameter constraints for the Λ CDM cosmology (as defined in [Planck Collaboration XVI 2014](#)), illustrating the impact of replacing the low- ℓ baseline *Planck* 2018 likelihood (lowE) with the WMAP+LFI likelihood presented in this thesis (lowTEB). We also show the change when including the high- ℓ polarization likelihood in the analysis.

Parameter	TT+lowE 68 % limits	TT+lowTEB 68 % limits	TTTEEE+lowE 68 % limits	TTTEEE+lowTEB 68 % limits
$\Omega_b h^2$	0.02212 ± 0.00022	0.02218 ± 0.00022	0.02236 ± 0.00015	0.02241 ± 0.00015
$\Omega_c h^2$	0.1206 ± 0.0021	0.1200 ± 0.0021	0.1202 ± 0.0014	0.1197 ± 0.0014
$100\theta_{MC}$	1.04077 ± 0.00047	1.04086 ± 0.00047	1.04090 ± 0.00031	1.04096 ± 0.00031
τ	0.0522 ± 0.0080	$0.071^{+0.010}_{-0.011}$	$0.0544^{+0.0070}_{-0.0081}$	$0.074^{+0.010}_{-0.011}$
$\ln(10^{10} A_s)$	3.040 ± 0.016	3.076 ± 0.021	3.045 ± 0.016	3.082 ± 0.021
n_s	0.9626 ± 0.0057	0.9645 ± 0.0058	0.9649 ± 0.0044	0.9664 ± 0.0044
H_0	66.88 ± 0.92	67.12 ± 0.93	67.27 ± 0.60	67.51 ± 0.61
Ω_m	0.321 ± 0.013	0.317 ± 0.014	0.3166 ± 0.0084	0.3134 ± 0.0084
Ω_Λ	0.679 ± 0.013	0.683 ± 0.013	0.6834 ± 0.0084	0.6866 ± 0.0084
σ_8	0.8118 ± 0.0089	0.825 ± 0.010	0.8120 ± 0.0073	0.8259 ± 0.0091
z_{re}	7.50 ± 0.82	9.3 ± 1.0	7.68 ± 0.79	$9.51^{+0.98}_{-0.97}$
$10^9 A_s$	2.092 ± 0.034	$2.167^{+0.043}_{-0.049}$	$2.101^{+0.031}_{-0.034}$	$2.181^{+0.043}_{-0.049}$
$10^9 A_s e^{-2\tau}$	1.884 ± 0.014	1.882 ± 0.014	1.884 ± 0.012	1.882 ± 0.012
Age/Gyr	13.830 ± 0.037	13.819 ± 0.037	13.800 ± 0.024	13.791 ± 0.024

to $\ell = 2500$. Due to the aforementioned degeneracy, this also results in a lower value for τ .

Since one of the main results of this chapter is the τ constraint from the WMAP+LFI dataset, we want to further comment on the robustness of this result. In Fig. 4.7, we show the good agreement between the estimates of τ from LFI and WMAP separately. The two were derived using their own $f_{sky} = 50\%$ mask. The consistency between the two experiments is further confirmed by the null test that we performed, estimating τ from the half-difference map of the two data sets, LFI–WMAP. The posterior distribution for this case is reported in the same figure and it is compatible with noise, giving an upper limit of $\tau \leq 0.059$ at 95% CL.

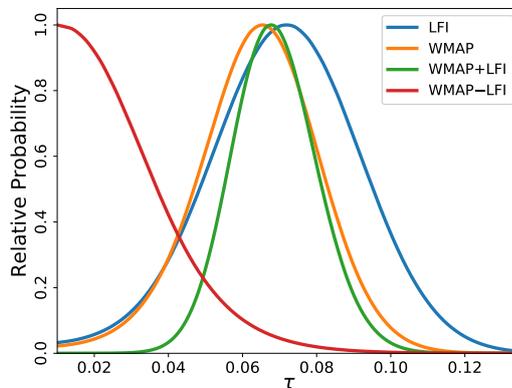


Figure 4.7: Posterior distributions of τ . LFI (blue) and WMAP (orange) are computed on their own 50% mask, WMAP+LFI (green) in the union of the two, retaining 54% of the sky, while WMAP–LFI (red) is computed in their intersection, retaining 46% of the sky.

Differently from the baseline low- ℓ *Planck* 2018 likelihood, which is based on the TT ,

EE , and BB power spectra, the pixel-based likelihood used in this thesis also includes the information contained in the TE cross power spectrum. In order to investigate the impact of this extra information, we build a polarization-only version of the pixel-based likelihood, which contains only the Q and U maps and the QQ, QU, and UU blocks of the covariance matrix, in an analogy of what was done in [Planck Collaboration XI \(2016\)](#). When we use this latter likelihood, we employ low- ℓ *TT Commander* likelihood based on the Blackwell-Rao estimator. The value of τ measured nulling the TE cross correlation is

$$\tau = 0.062 \pm 0.012 \quad (68\%, \text{Commander} + \text{lowP}), \quad (4.19)$$

which represents roughly a half- σ downward shift with respect to the full TEB likelihood, already seen on the LFI only likelihood in [Planck Collaboration XI \(2016\)](#). Such behavior is also shown by WMAP which, on 50% sky, yields $\tau = 0.055^{+0.019}_{-0.017}$ forcing $TE = 0$ and $\tau = 0.064^{+0.017}_{-0.015}$ with the full TEB likelihood. For WMAP, the same behavior is also present on larger masks; for example on the 75% sky fraction we measure $\tau = 0.065^{+0.013}_{-0.014}$ when $TE = 0$ and $\tau = 0.070^{+0.012}_{-0.013}$ when also TE is varied.

In all the previous cases, when TE is forced to zero, the τ posterior shifts closer to the HFI determination ([Planck Collaboration V 2019](#); [Pagano et al. 2019](#)), which is based only on EE estimates. Posteriors of the full pixel-based likelihood and the one without TE for WMAP+LFI are shown in [Fig. 4.8](#).

In order to verify if such behaviour is coherent with our error budget, we compare the shift in τ with a set of simulations. In the left panel of [Fig. 4.9](#) we show the histogram of $\Delta\tau$, defined as the difference between the τ estimated from the full TEB likelihood (“Full”) and the τ estimated forcing $TE = 0$ (“noTE”), for a Montecarlo of 1000 signal and noise simulations. This analysis shows that nullifying TE still provides an unbiased estimation of τ and also that the shift observed in data is not anomalous, representing a 1.7 σ fluctuation. We also show in the right panel of [Fig. 4.9](#), a similar plot for the ratio of 1- σ errors defined as $\sigma_{\tau_{\text{noTE}}}/\sigma_{\tau_{\text{Full}}}$; also, in this case, the value measured on data is compatible with the simulations. This test also suggests that for this dataset, removing TE degrades σ_{τ} by about 5% on average.

Adding to the CMB temperature and polarization data the *Planck* lensing likelihood ([Planck Collaboration VIII 2018](#)) and baryon acoustic oscillation (BAO) measurements ([Alam et al. 2017](#); [Beutler et al. 2011](#); [Ross et al. 2015](#)) breaks the degeneracy more efficiently with the amplitude of the scalar perturbations providing

$$\tau = 0.0714^{+0.0087}_{-0.0096} \quad (68\%, \text{TT,TE,EE+Lensing+BAO}). \quad (4.20)$$

Such bounds are slightly less constraining when compared with the existing *Planck* HFI based likelihood (see, e.g., [Planck Collaboration V 2019](#); [Pagano et al. 2019](#), and our [Fig. 4.10](#)).

Assuming the TANH model for the ionization fraction the τ constrain can be directly converted into a mid-point reionization redshift of

$$z_{\text{re}} = 9.28 \pm 0.84 \quad (68\%, \text{TT,TE,EE+Lensing+BAO}). \quad (4.21)$$

This value is higher but still compatible with analogous estimates that instead use the *Planck* HFI based large-scale polarization likelihood, $z_{\text{re}} = 7.82 \pm 0.71$ ([Planck Collaboration VI 2018](#)) and $z_{\text{re}} = 8.21 \pm 0.58$ ([Pagano et al. 2019](#)).

The WMAP+LFI CMB map and the corresponding covariance matrix are packaged in low- ℓ likelihood modules compatible with the `clik` infrastructure ([Planck Collaboration XV 2014](#); [Planck Collaboration ES 2013, 2015, 2018](#)) which are made publicly available⁶.

⁶The WMAP+LFI likelihood module is available on https://web.fe.infn.it/~pagano/low_ell_datasets/wmap_lfi_legacy

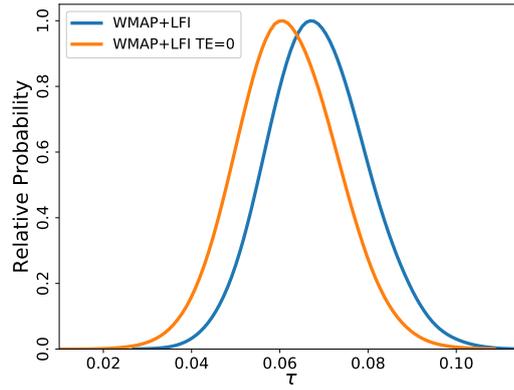


Figure 4.8: Comparison of τ posterior distributions for the WMAP+LFI using the full TQU likelihood (blue) and imposing $TE = 0$ in order to factorize the T and QU parts of the likelihood (orange).

We provide both a likelihood module that implements Eq. 4.4 inverting the full covariance matrix and one that implements the Sherman-Morrison-Woodbury (SMW) formula (Golub and Van Loan 1996) which allows us to speed up the computation by an order of magnitude (see Planck Collaboration XI 2016, Appendix B.1 for details) but does not include TB and EB. In both cases, in order to keep full compatibility with the codes of the `clik` framework, we do not treat the regularization noise as described in Sec. 4.3, but instead we sum a single realization. Such noise realization has been chosen in order to have a deviation for $\ln(10^{10}A_s)$ and τ with respect to the baseline case smaller than 1% in units of σ .

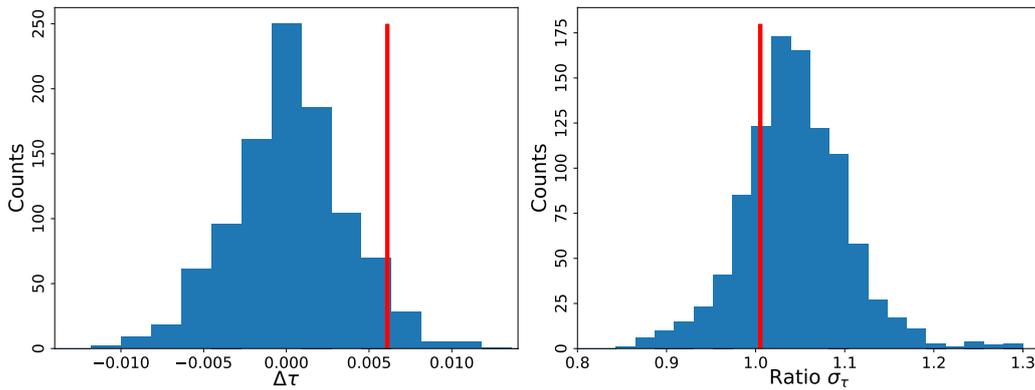


Figure 4.9: Left panel: Histogram of $\Delta\tau \equiv \tau_{\text{Full}} - \tau_{\text{noTE}}$ obtained analysing a Montecarlo of 1000 simulations. The red vertical bar shows the same quantity evaluated on the data. Right panel: Histogram of $\sigma_{\tau_{\text{noTE}}} / \sigma_{\tau_{\text{Full}}}$ obtained analysing a Montecarlo of 1000 simulations. The red vertical bar shows the same quantity evaluated on the data.

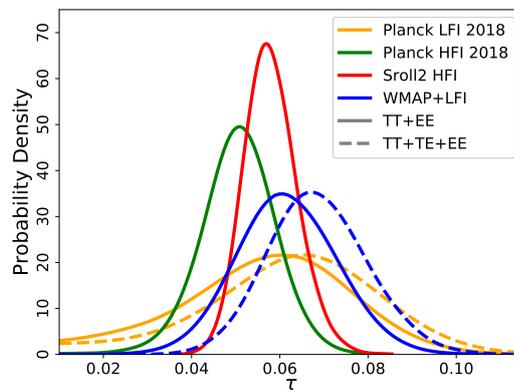


Figure 4.10: Posterior distributions for τ for various datasets. Solid lines represent τ constraints ignoring TE , dashed lines assume full TEB likelihood. Only τ and $\ln(10^{10}A_s)$ are sampled, whereas the remaining Λ CDM parameters are fixed to common fiducial values. The green and yellow lines show results obtained with the official *Planck* Legacy low- ℓ likelihoods ([Planck Collaboration V 2019](#)). The red line is obtained running the SRoll2 likelihood ([Pagano et al. 2019](#)). The blue lines represent the results of [Natale et al. \(2020\)](#).

Is the lack of power anomaly in the CMB correlated with the orientation of the Galactic plane?

In Sec. 3.2.1, we have highlighted the importance of rejection or confirmation of the null-hypothesis. This approach allows us to deeply investigate some properties of the chosen model. It does not require the computational effort needed to evaluate the Bayesian evidence, representing a flexible tool for data analysts. In the context of Λ CDM, cosmologists are continuously searching estimators to test departures from such a model. This interest brought to the finding several claims of unexpected statistical properties (or anomalies) of the CMB fluctuations.

In this chapter we analyse the low-variance anomaly, that is a feature of the CMB temperature anisotropy pattern present in both WMAP (Monteserin et al. 2008; Cruz et al. 2011; Gruppuso et al. 2013) and *Planck* data (Planck Collaboration VII 2018). It shows up at large angular scales, where the instrumental noise is negligible, with a statistical significance around 2-3 σ C.L. depending on the estimator employed. This effect is correlated with other CMB anomalies, see e.g. Copi et al. (2007, 2009, 2010); Schwarz et al. (2016), which are sensitive to the lack of power with respect to expectations of the Λ CDM model, see Muir et al. (2018) for further details. For this reason, we will use the expressions lack-of-power and low-variance as synonyms.

A statistical fluke is of course the simplest explanation for this phenomenon. However, in this case, one has to accept to live in a rare Λ CDM realisation. In any case, there are at least three reasons why this anomaly is worth of further investigations (Gruppuso and Sagnotti 2015):

1. it is unlikely that the effect is due to an unaccounted instrumental systematics: both WMAP and *Planck* observe it with similar significance despite being two separate experiments with different data gathering schemes and scanning strategies;
2. it is not natural to attribute this effect to foreground residuals: the latter are not expected to be correlated to the CMB, so a foreground residual should increase and not lower the total anisotropy power¹. A similar argument would also apply to possible extensions of the Λ CDM as long as their source is statistically independent from the primary CMB anisotropy (Gruppuso 2007; Bunn and Bourdon 2008).
3. it is suspiciously dependent on the Galactic mask: its statistical significance increases when only high Galactic regions are considered, which is usually a conservative choice

¹Note also that typically (and in particular at large scales where this chapter is focused) the foreground mitigation is performed at the map level (in the harmonic or pixel space) and not at the C_ℓ level.

in CMB data analysis (Gruppuso et al. 2016). It was also shown Gruppuso et al. (2018) that this effect was dominated by odd over even multipoles, see e.g. Kim and Naselsky (2010a,b); Gruppuso et al. (2011).

In this chapter we want to focus on the last item by estimating, from a statistical point of view, how likely is to find a CMB map of the Λ CDM model with such a behaviour between low- and high-Galactic latitudes. To perform this analysis we will use random rotations (see Appendix A.1) of simulated CMB maps in order to evaluate among all the possible orientations what is the probability of having most of the power at low-Galactic latitudes. The adopted estimator is the variance, V , of the temperature anisotropies, $\delta T(\hat{\mathbf{n}})$,

$$V \equiv \langle (\delta T(\hat{\mathbf{n}}))^2 \rangle, \quad (5.1)$$

where $\hat{\mathbf{n}}$ is the unit-vector pointing a given direction of observation. V is built through the angular power spectrum (APS), C_ℓ :

$$V = \sum_{\ell=2}^{\ell_{max}} \frac{2\ell+1}{4\pi} C_\ell, \quad (5.2)$$

where the maximum multipole, ℓ_{max} , is set to 29 in the following since we want to be consistent with the maximum multipole considered in the *Planck* pixel-based low- ℓ Likelihood functions (Planck Collaboration XI 2016). However, the dependence of V upon ℓ_{max} is very weak for $\ell_{max} \gtrsim 10$ and therefore such a choice does not impact significantly on our results.

5.1 Data set and simulations

5.1.1 CMB maps and masks

We use data products from the *Planck* 2018 data release, available in the *Planck* Legacy Archive². In particular we employ the temperature Commander 2018 map (Planck Collaboration IV 2018) downgraded to HEALPIX³ (Gorski et al. 2005) resolution $N_{side} = 16$ with a Gaussian beam with full width half maximum, FWHM, of 440 arcmin. The map is shown in the left panel of Fig. 5.1. As a consistency check we also employ the SMICA temperature map (Planck Collaboration IV 2018), also downgraded from high resolution to $N_{side} = 16$. These CMB maps have been delivered already with a constrained CMB realisation along the Galactic plane. We have added to those maps a regularisation noise realisation with 2 μ K rms, consistently considered in the extraction of the APS. This choice is consistent with the procedure adopted in Planck Collaboration V (2019). We checked that such a noise has a negligible impact on our results. The maps have been masked with several Galactic masks, shown in the right panel of Fig. 5.1 and whose sky fractions are listed in Table 5.1. More specifically, the considered masks are the $N_{side} = 16$ confidence mask provided with the 2018 Commander solution (Planck Collaboration IV 2018), named Std 2018, and other four masks built extending the edges of the Likelihood 2015 standard mask (Planck Collaboration XI 2016) by 12, 18, 24 and 30 degrees, called respectively Ext₁₂, Ext₁₈, Ext₂₄ and Ext₃₀. This choice is done in order to make contact with previous works, i.e. Gruppuso et al. (2016, 2018), and to compare the impact of the most recent *Planck* 2018 data with respect to that of the 2015 release, see Appendix A.2.

²<https://www.cosmos.esa.int/web/planck/pla>

³<http://healpix.sourceforge.net>

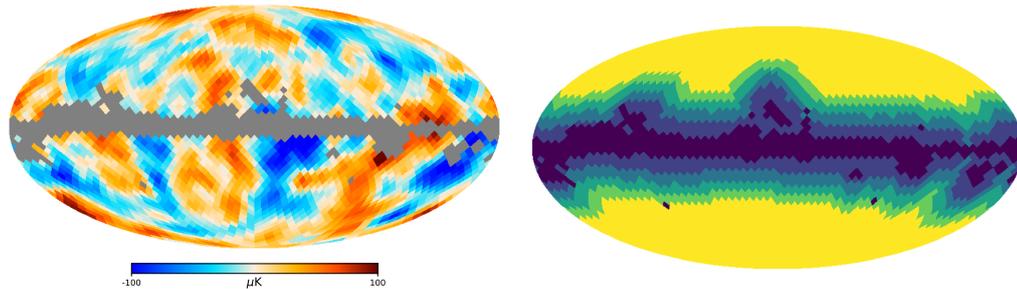


Figure 5.1: Left panel: **Commander** 2018 map smoothed at 440 arcmin, where the Std 2018 mask is applied. Right panel: Galactic temperature masks considered in this chapter. The dark blue region is for the 2018 standard case. The blue region is for the Ext₁₂ case. The light blue region is for the Ext₁₈ case. The dark green region is for the Ext₂₄ case. The green region is for the Ext₃₀ case.

Mask	Sky Fraction [%]
Std 2018	85.6
Ext ₁₂	70.8
Ext ₁₈	59.1
Ext ₂₄	48.7
Ext ₃₀	39.4

Table 5.1: Observed sky fractions for the masks shown in Figure 5.1.

5.1.2 Sets of simulations

We generate 10^5 CMB temperature maps at **HEALPix** resolution $N_{side} = 16$ randomly extracted from the *Planck* 2018 best-fit model through the **synfast** function of **healpy** (Gorski et al. 2005) with a Gaussian beam of 440 arcmin FWHM. To provide numerical regularisation, a different random noise realisation, with rms of $2 \mu\text{K}$, is added to each of the CMB simulations, as done for the observed **Commander** and **SMICA** 2018 maps. This set is used to estimate the statistical significance of the low-variance in a ΛCDM framework. A subset of 10^3 simulations of this set of ΛCDM realisations is referred to as *ensemble 0*. Another subset of 10^3 simulations constrained to have variance V close to the value observed by **Commander** 2018, $V_c = 2090.02 \mu\text{K}^2$ obtained with the Std 2018 mask, is called *ensemble 1*. More precisely a map \mathbf{m}_i with variance V_i belongs to *ensemble 1*, if $V_c - 20 \mu\text{K}^2 \leq V_i \leq V_c + 20 \mu\text{K}^2$. The analysis of the stability of our results with respect to the choice of the threshold of $20 \mu\text{K}^2$ is given in Appendix A.3. Note that in the case of **SMICA** the variance is also constrained in the same range which contains the value observed in the data ($V_s = 2085.57 \mu\text{K}^2$).

5.1.3 Angular power spectrum estimator

We use the variance V as estimator for the lack of power, built through Eq. (5.2). The C_ℓ are obtained with an optimal angular power spectrum estimator, namely **BolPo1** (Gruppuso et al. 2009), an implementation of the Quadratic Maximum Likelihood (QML) method (Tegmark 1997; Tegmark and de Oliveira-Costa 2001). The choice of the QML algorithm minimises the introduction of extra statistical uncertainty in our analysis with respect to other, suboptimal, APS estimators (Molinari et al. 2014). For each of the simulated maps and for the various masks defined above, we have used the estimates of **BolPo1** to build the variance, V . In Fig. 5.2 we show the APS of the **Commander** 2018 temperature map estimated with the five masks shown in Fig. 5.1 and whose sky fraction is reported in Table 5.1.

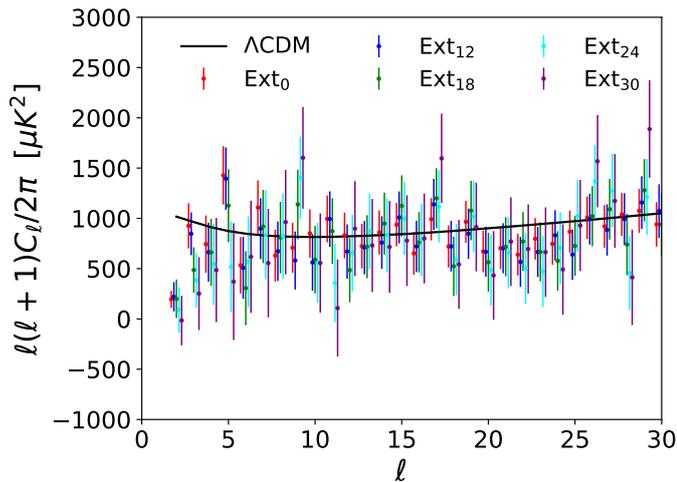


Figure 5.2: APS of the Commander 2018 temperature map estimated with the five masks shown in Fig. 5.1 and whose sky fraction is reported in Table 5.1. The uncertainty shown for each multipole do not include cosmic variance.

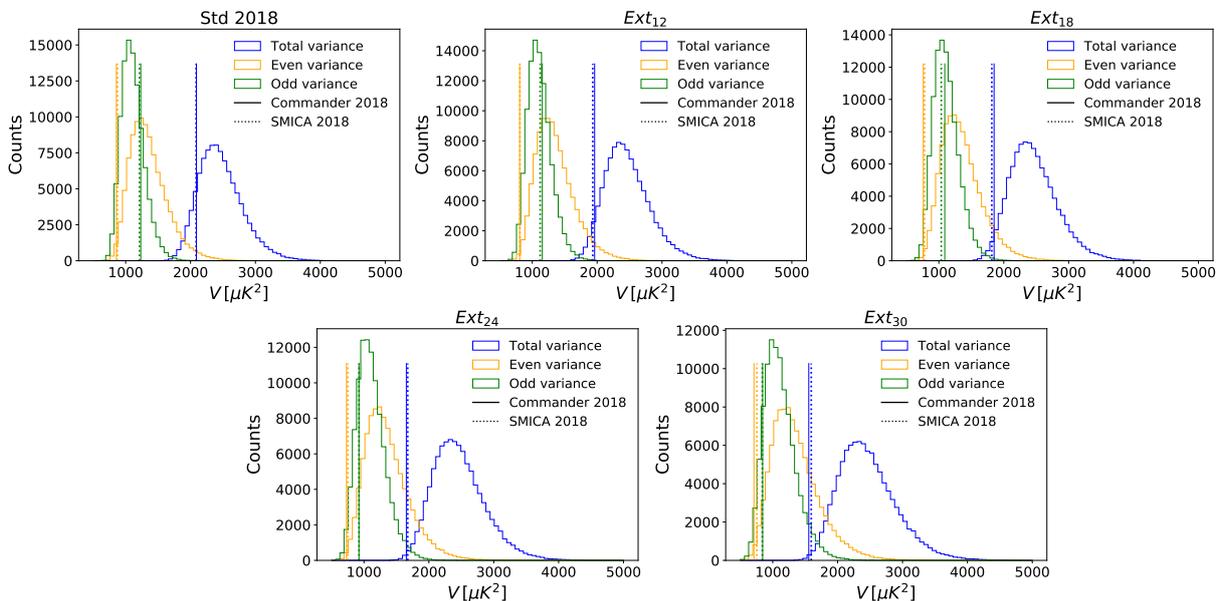


Figure 5.3: Each panel shows the empirical distribution of V in μK^2 expected in a Λ CDM model (blue) and computed through Eq. (5.2) for the masks listed in Table 5.1. We report also the even and odd splits of the variance, through Eq. 5.3 (orange and green, respectively). Vertical dashed and dotted bars correspond to the *Planck* 2018 *Commander* and *SMICA* CMB solutions, respectively.

5.2 Analysis in Λ CDM framework

As already known in the literature, the observed value of V is low and its statistical significance increases considering regions at high Galactic latitude, see e.g. [Monteserin et al. \(2008\)](#); [Cruz et al. \(2011\)](#); [Gruppuso et al. \(2013\)](#); [Planck Collaboration XXIII \(2014\)](#); [Planck Collaboration XVI \(2016\)](#); [Planck Collaboration VII \(2018\)](#). Employing the *Bolpol* code to extract the TT APS for each of the 10^5 Λ CDM simulations, we have built the probability distribution functions of V for each of the five masks shown in Fig. 5.1. The MC distributions are displayed in Fig. 5.3 where they are compared to the corresponding *Planck* 2018 observed values shown as vertical bars.

In the same panels we provide also V_+ (V_-), shown in orange (green), defined as V

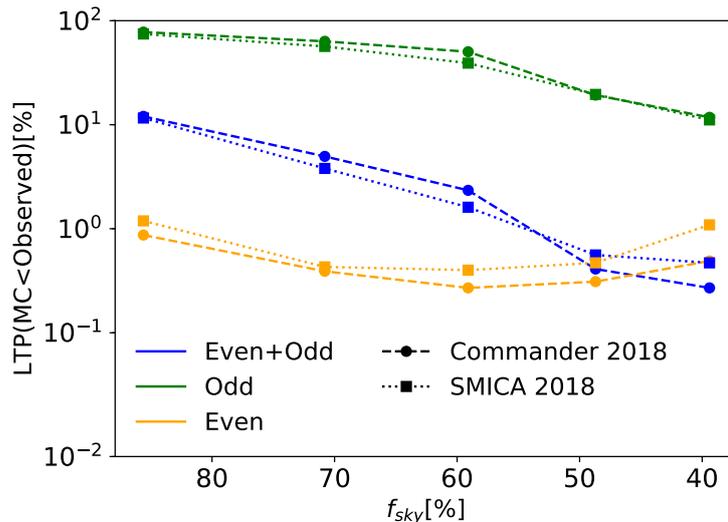


Figure 5.4: Lower tail probability of the *Planck* 2018 **Commander** and **SMICA** maps with respect to the 10^5 Λ CDM simulations as a function of the sky fraction.

but where the sum in Eq. (5.2) is performed only over the even (odd) multipoles, i.e.

$$V_{\pm} = \sum_{\ell=2}^{\ell_{max}} \left[\frac{1 \pm (-1)^{\ell}}{2} \right] \frac{2\ell + 1}{4\pi} C_{\ell}. \quad (5.3)$$

In addition, we display as vertical bars, with the same color convention, the corresponding observed *Planck* 2018 values for V_{\pm} .

Fig. 5.4 shows the three lower tail probabilities, henceforth LTP, for V_{+} (orange), V_{-} (green) and V (blue) against the observed sky fraction of the five cases of Fig. 5.3. V shows a monotonic behaviour: as one considers regions at higher and higher Galactic latitude the *Planck* observed values shift towards lower variances more rapidly than the increase of the width of the distribution due to sampling variance because of the smaller observed sky fraction considered. In other words, the observed values are more and more unlikely and for the extreme case, i.e. Ext₃₀ mask, we find a compatibility with Λ CDM model only at 0.3% C.L. for the **Commander** map and 0.5% for **SMICA**. This is dominated by V_{+} which is constantly low, independently on the considered sky fraction. Indeed, for **Commander**, its LTP varies around 0.3 – 0.5%, for all the considered sky fractions lower than the Std 2018 one. For **SMICA**, instead, its LTP varies in a slightly higher but still low range [0.5%, 1.1%]. On the other hand, V_{-} is more sensitive to the sky fraction, decreasing monotonically as one takes into account regions at higher and higher Galactic latitude. However, its LTP remains inside the 1σ dispersion of the MC’s, reaching $\sim 11\%$ in the Ext₃₀ mask, independently from the employed CMB solution.

The fact that the LTP of V decreases when using more aggressive masks suggests that the low power of the *Planck* data is somehow anisotropically distributed on the map. In other words, the increasing discrepancy of the data with respect to Λ CDM when we exclude from the analysis pixels around the Galactic plane, indicates a sort of “localisation” of most of the power around the Galactic plane itself.

Moreover, Fig. 5.3 and 5.4 show that, at large angular scales, such a low-Galactic-latitude power turns out to be dominated by the odd multipoles, see also Gruppuso et al. (2018).

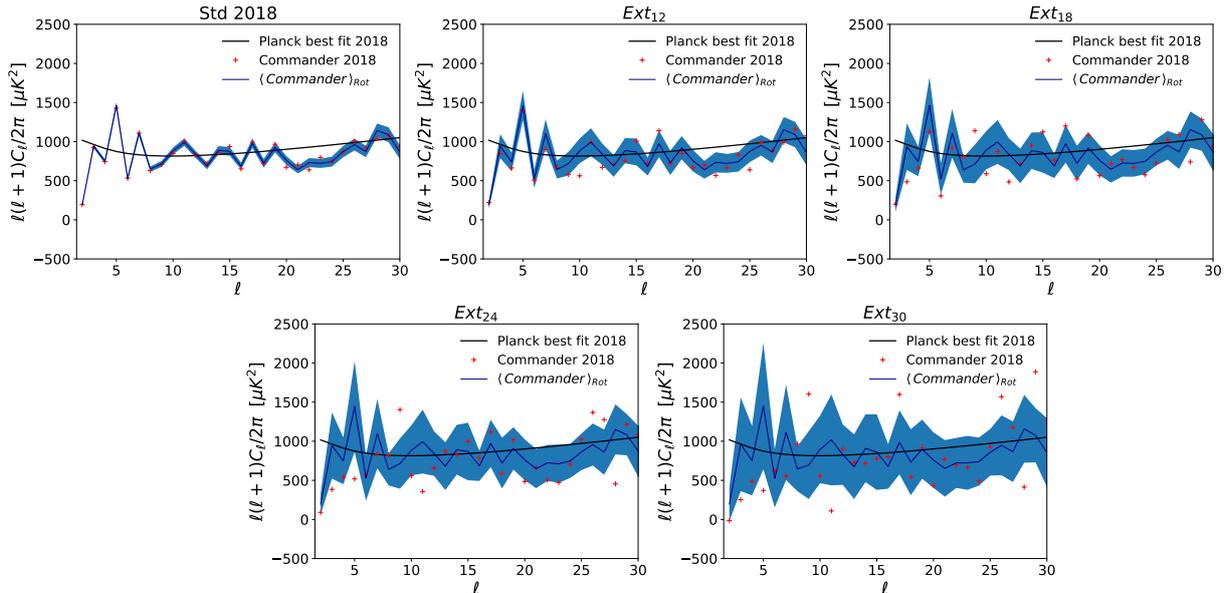


Figure 5.5: Each panel shows the TT APS of the *Commander* 2018 map estimated using different masks (red symbols). Blue line and blue region are respectively the average and the standard deviation of 10^3 random rotations of the *Commander* 2018 map. Note that in each mask the MC average is equal to the estimates obtained in the Std 2018 mask demonstrating that the variance is a mathematical object invariant under rotations only on average: the presence of a mask breaks the rotational symmetry for the single realization.

5.2.1 Variance analyses including rotations

We now further investigate the dependency of V with respect to the Galactic mask by implementing random rotations of the maps (see Appendix A.1 for details which include the validation). This is performed in order to evaluate among all the possible orientations what is the probability of having most of the power at low Galactic latitude. The above procedure can be seen as a sort of look-elsewhere effect on the orientation of the mask. For computational reasons we reduce the number of MC simulations by considering the *ensemble 0* made of 10^3 maps generated from the *Planck* 2018 best-fit model. Note that V is invariant under rotation of the input maps by construction only in the full sky case. In fact, when a mask is applied, the variance V is not conserved under rotation for a single realisation but invariance is restored only on ensemble average. This effect is nicely captured already at the angular power spectrum level: in Fig. 5.5 each panel shows the average and the statistical uncertainty at 1σ of the TT spectra of 10^3 random rotations of the *Commander* map for the various masks⁴. Notice that the APS estimates obtained with the Std 2018 mask are recovered only on average (blue lines) in the other masks. Moreover, as expected, the standard deviation (blue region) increases as the mask gets larger, allowing less observed sky for the analysis. In addition, still in Fig. 5.5 we show the TT spectrum of the *Commander* 2018 map without any rotation (red symbols).

We analyse random rotations of the *ensemble 0* and corresponding observed data building two estimators, the LTP-estimator (Section 5.2.1) and the r -estimator (Section 5.2.1). With the former we investigate separately for each mask how anomalous is the particular orientation of the Galactic plane. With the latter we quantify the statistical significance of the lowering trend of V with respect to its value in the Std 2018 mask with all the possible orientations.

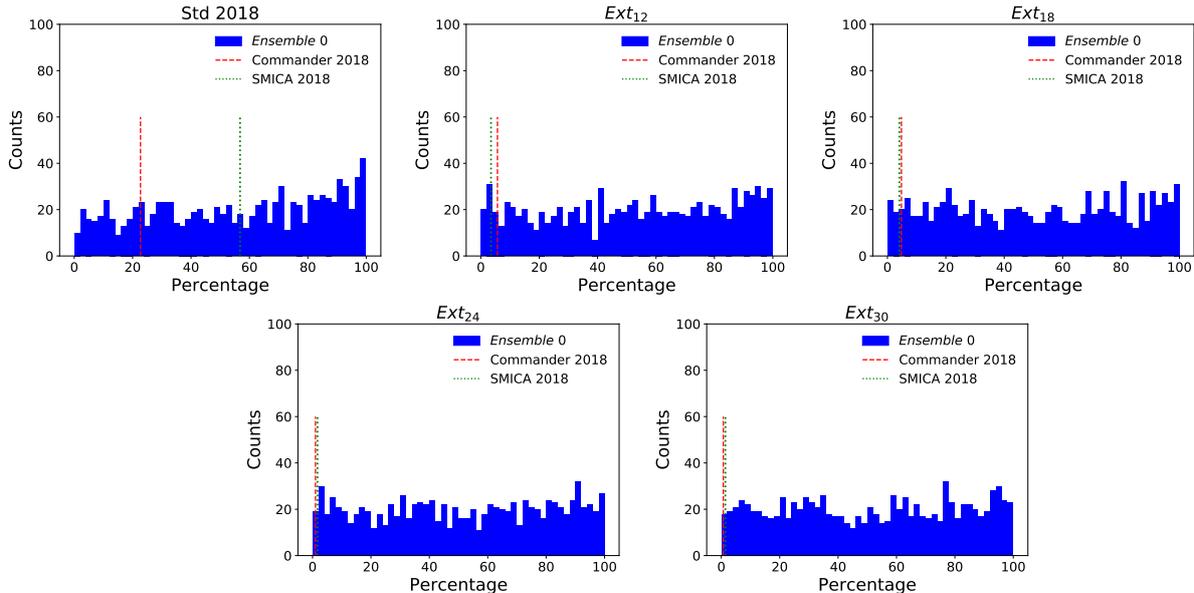


Figure 5.6: Histograms of the LTP of finding a rotated map of the *ensemble 0* with $V^{rot} < V$, where V is the variance of the corresponding unrotated map. Each panel shows the results obtained using a different mask. Red dashed and green dotted vertical bars are the LTP for **Commander** and **SMICA** respectively.

LTP estimator

For each map \mathbf{m}_i belonging to *ensemble 0* we build the histogram of V_i obtained through 10^3 random rotations of that map. Hence, we compute the LTP of that map \mathbf{m}_i , denoted with LTP_i , with respect to the corresponding set of rotations. This can be repeated for $i = 1, \dots, 10^3$, i.e. for all the maps of the *ensemble 0* and for all the considered masks. Thus, for each mask, we obtain a MC of 10^3 values of LTP representing the distribution of probabilities expected in a Λ CDM model. Since the variance does not depend on the orientation, the distribution of LTP is expected to be uniform, that is, each LTP is equiprobable. The empirical distribution of the LTP-estimator for each considered mask shown in Fig. 5.6 confirms our expectations. In the same Figure we also show the LTP obtained from *Planck* data as vertical bars, red for **Commander** and green for **SMICA**. The corresponding values are reported in left panel of Table 5.2. When we consider higher Galactic latitude, we find that the probability of observing a LTP with respect to its rotations lower than the corresponding LTP of **Commander** (**SMICA**) 2018 is anomalous at $\sim 2.8\sigma$ ($\sim 2.5\sigma$). Indeed, in the *Ext30* case, only 5 (13) out of 10^3 maps of the *ensemble 0* have a lower LTP than the **Commander** (**SMICA**) 2018 map, i.e. only in the 0.5% (1.3%) of the cases the anomaly associated to the power localisation around the Galactic plane is higher than data (see right panel of Tables 5.2).

r -estimator

We use here the r -estimator defined as

$$r \equiv \frac{V_{std} - V_{mask}}{\max_{j \in \text{rotations}} \{V_{std}^{(j)} - V_{mask}^{(j)}\}}, \quad (5.4)$$

where V_{std} is the variance computed in the Std 2018 mask, while V_{mask} is the variance computed in one of the other four extended masks. The numerator of Eq. (5.4) fixes the sign of the r -estimator as determined by the decrease ($r > 0$), or increase ($r < 0$),

⁴We obtain a similar behaviour for **SMICA** that is not shown here for sake of brevity.

Mask	LTP [%]		Mask	LTP [%]	
	$V_c^{(\text{rot})} < V_c$	$V_s^{(\text{rot})} < V_s$		$LTP_i < LTP_c$	$LTP_i < LTP_s$
Std 2018	22.7	56.8	Std 2018	18.8	49.1
Ext ₁₂	5.7	3.5	Ext ₁₂	6.7	4.5
Ext ₁₈	4.8	4.1	Ext ₁₈	4.7	4.4
Ext ₂₄	1.0	1.7	Ext ₂₄	1.0	1.5
Ext ₃₀	0.7	1.4	Ext ₃₀	0.5	1.3

Table 5.2: Left table: The probability of obtaining a value of the variance of the rotated **Commander** map (second row), $V_c^{(\text{rot})}$, and rotated **SMICA** map (third row), $V_s^{(\text{rot})}$, smaller than the unrotated one, V_c and V_s respectively. Right table: LTP of obtaining a simulation of the *ensemble 0* with LTP lower than the one obtained with the **Commander** map, LTP_c , or **SMICA** map, LTP_s .

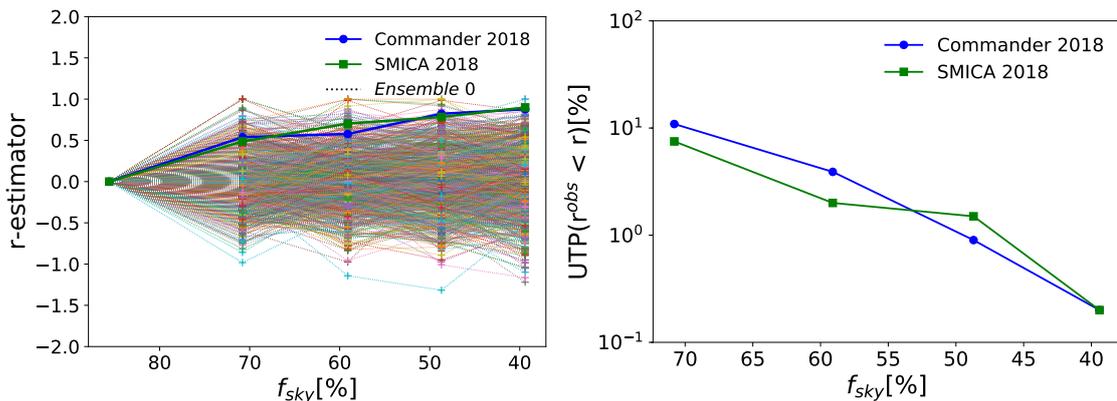


Figure 5.7: Left panel: r -estimator computed with Eq. (5.4) versus the sky fraction. The coloured dotted lines stand for the r value obtained from the *ensemble 0*. Blue and green solid lines stand for **Commander** and **SMICA** respectively. Right panel: UTP of obtaining a simulation with r larger than the one obtained with **Commander** (blue line) or **SMICA** (green line) as a function of the sky fraction.

of the variance as we widen the Galactic mask. This behaviour is normalised by the denominator, which picks up the maximum decrease among all the rotations⁵. The r -estimator is therefore upper bounded by 1, but it can become lower than -1. In other words, the r -estimator represents the fractional change of V , computed in an extended mask with respect to the Std 2018 mask value, relative to the maximum decrease across rotations. For example, $r = 0.5$ means that, we are dealing with a map which, in a given mask, has a variance difference with respect to the standard mask equal to exactly half of the maximum difference which can be found among all rotations. In the left panel of Fig. 5.7 we show the r -estimator for all the considered cases. Dotted lines connect the MC values of r represented with a plus symbol. Solid blue line connects the **Commander** 2018 values (dot symbols) and the solid green line connects the **SMICA** 2018 values (square symbols). For this estimator we consider the upper tail probability, UTP, defined as the fraction of simulations with larger values of r than the observed one. They are shown in the right panel of Fig. 5.7 and quoted in Table 5.3. Notice that both **Commander** and **SMICA** present an increase of r for higher and higher Galactic latitudes and in the Ext₃₀ case, they are close to 1, being $r^c = 0.88$ for **Commander** and $r^s = 0.90$ for **SMICA**. This means that the observed maps in the Ext₃₀ case are almost aligned to the direction which maximizes the lowering of V obtainable through rotations. The probability corresponding to this event is 0.2% for both **Commander** and **SMICA**. This leads to an anomalous value of r at a level of 3.1σ .

⁵In the denominator of r we include also the unrotated case, denoted here as the 0^{th} rotation.

Mask	UTP [%]	
	$r^C < r$	$r^S < r$
Ext ₁₂	10.9	7.5
Ext ₁₈	3.9	2.0
Ext ₂₄	0.9	1.5
Ext ₃₀	0.2	0.2

Table 5.3: UTP of obtaining a simulation of the *ensemble 0* with r larger than the one obtained from the data. Second column shows the UTP for **Commander**, third column the UTP for **SMICA**.

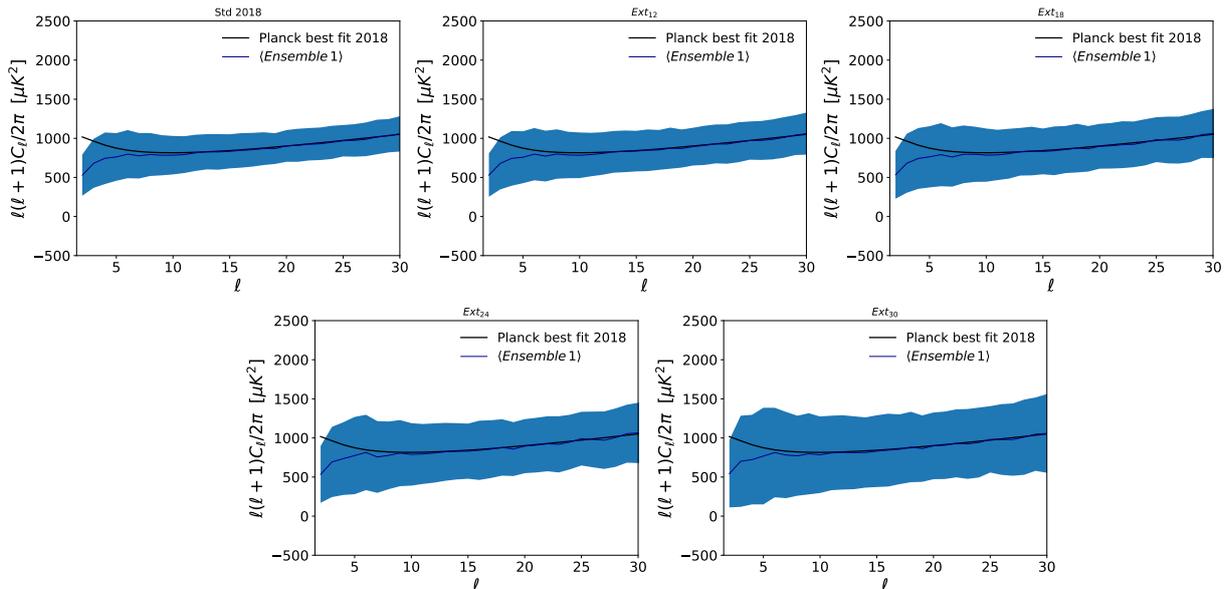


Figure 5.8: Each panel shows the *Planck* 2018 best-fit model (black solid line) and the average APS of *ensemble 1* (blue line), with its 1σ dispersion (blue region) for all the considered masks.

5.3 Analysis of Λ CDM simulations with low variance

In this section we repeat the analysis performed in Section 5.2 but now considering simulated maps which have almost the same variance V as the one observed by the CMB solutions (**Commander** and **SMICA**) of the *Planck* 2018 release. These are collected in the *ensemble 1*, as described in Section 5.1. The aim of this analysis is to check whether the previous results still hold when the variance is constrained to be low also across the simulations. In other words we would like to exclude the possibility that the observed trend of a lowering variance when extending the Galactic mask, is connected to the low value of the variance measured in the Standard mask. In Fig. 5.8 we display the *Planck* 2018 best-fit model (black solid line) and the average of *ensemble 1* (blue line), with its standard deviation (blue region) for all the considered masks. Notice the increase of the statistical uncertainty as the observed sky fraction decreases. This figure shows that *ensemble 1* behaves differently from the fiducial power spectrum only at low- l . In other words, selecting a subset of Λ CDM realisations with low variance is in fact equivalent to choosing maps with suppressed C_ℓ at low multipoles⁶.

We evaluate the variance V for each element of *ensemble 1* and for each of the considered masks. Results are shown in Fig. 5.9 where each panel provides the histogram of V for each mask. Dashed red line represents V as measured from **Commander**, and the dashed

⁶Note that we recover empirically the well-known correlation between low- V and low- C_2 anomalies Muir et al. (2018).

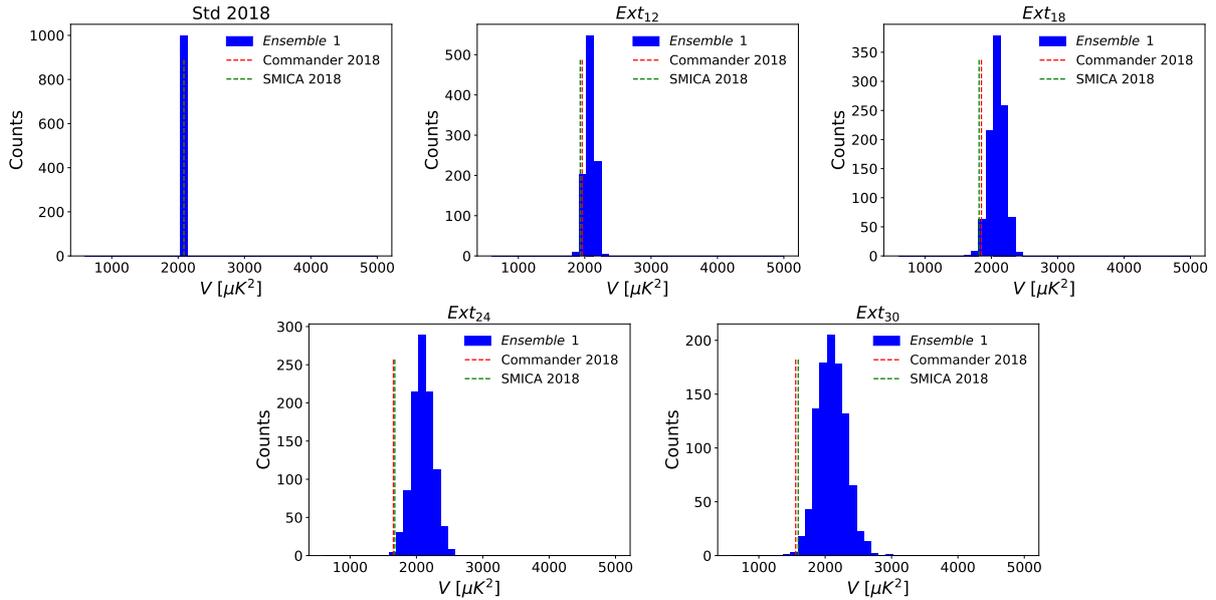


Figure 5.9: Histograms of the variances V of the maps belonging to the *ensemble 1* computed with the masks Std 2018, Ext₁₂, Ext₁₈, Ext₂₄ and Ext₃₀. The red dashed line identifies the variance of the **Commander** map, V_c . The green dashed line identifies the variance of the **SMICA** map, V_s .

Mask	LTP [%]	
	$V < V_c$	$V < V_s$
Std 2018	50.7	41.5
Ext ₁₂	4.1	1.9
Ext ₁₈	2.3	1.1
Ext ₂₄	0.2	0.2
Ext ₃₀	0.3	0.4

Table 5.4: The probability of obtaining a value for the variance V smaller than that of **Commander** (second column), V_c , or **SMICA** (third column), V_s , for a map of the *ensemble 1*. Note that the difference between the Ext₂₄ and Ext₃₀ case is of the order of the numerical sensitivity of the *ensemble 1*, since it is made of 10^3 simulations.

green line stands for V of **SMICA**. In the left panel of Fig. 5.10 we display the LTP of the *Planck* 2018 data in percentage as a function of the sky fraction. They are also reported in Table 5.4 for convenience. We find that the monotonic behaviour shown in Fig. 5.4 for the 10^5 Λ CDM simulations is almost⁷ recovered for the *ensemble 1*: V still decreases at high Galactic latitudes with a percentage of compatibility at the level of 0.3 – 0.4% in the Ext₃₀ case. This means that a “low variance” model (low as the one observed by *Planck*) is not enough to explain this behaviour at high Galactic latitude. Notice also that this effect is largely dominated by the quadrupole and the octupole. This is shown in the right panel of Fig. 5.10, where the LTP vs the observed sky fraction is shown when we exclude only the quadrupole (blue dashed lines) or both the quadrupole and the octupole (red dashed lines) in the computation of V .

5.3.1 Variance analyses including rotations

As for the *ensemble 0* we now include random rotations in the analysis of the *ensemble 1*. We still use the LTP-estimator and the r -estimator defined above.

⁷Note that for the **Commander** case the difference between the two last cases, i.e. Ext₂₄ and Ext₃₀ case, is of the order of the numerical sensitivity of the *ensemble 1*, since it composed of 10^3 simulations.

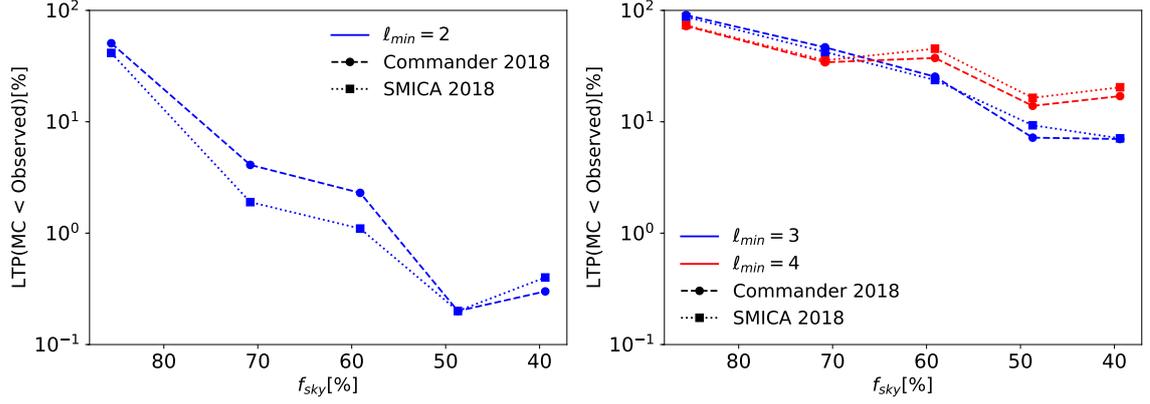


Figure 5.10: Right panel: LTP of the variance estimator for the *Planck* 2018 data in percentage as a function of the sky fraction. Left panel: the same as in right panel but with $\ell_{min} = 3$ (blue line) or $\ell_{min} = 4$ (red line).

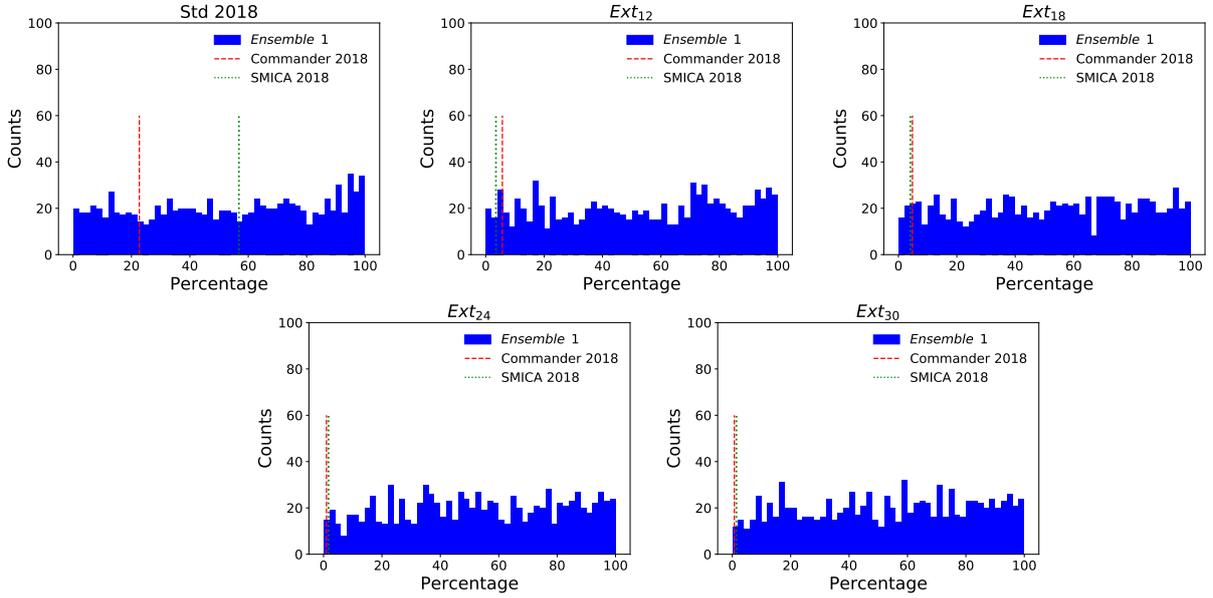


Figure 5.11: Histograms of the LTP of finding a rotated map of the *ensemble 1* with $V^{rot} < V$, where V is the variance of the corresponding unrotated map. Each panel shows the results obtained using a different mask. Red dashed and green dotted vertical bars are the LTP for *Commander* and *SMICA* respectively.

LTP estimator

For each map \mathbf{m}_i belonging to *ensemble 1* and its rotations we obtain the MC of 10^3 values of LTP_i . In Fig. 5.11 we show the histograms of such LTP_i for each considered mask. The observed LTP (i.e. those obtained from *Planck* data and shown in left panel of Table 5.2) are also shown in the same figure as vertical bars, red for *Commander* and green for *SMICA*. Notice that, by construction, even in a Λ CDM model constrained to have a low-variance as *ensemble 1*, the variance does not depend on the orientation. Therefore the distribution of LTP is still uniform as it is found in the histograms of Fig. 5.11. In this case we find for *ensemble 1* a very similar behaviour to *ensemble 0*. For *Commander* (*SMICA*) the LTP estimator gives a $\sim 2.8\sigma$ ($\sim 2.6\sigma$) anomaly at high Galactic latitude, see Table 5.5.

r -estimator

We apply here the r -estimator to the *ensemble 1* simulations. In Fig. 5.12 we show the results for all the considered cases. Dotted lines connect the MC values of r represented with a plus symbol. Solid blue line connects the *Commander* values (dot symbols) and the

Mask	LTP [%]	
	$LTP_i < LTP_c$	$LTP_i < LTP_s$
Std 2018	21.6	52.9
Ext ₁₂	5.7	3.2
Ext ₁₈	4.5	3.9
Ext ₂₄	0.7	1.3
Ext ₃₀	0.5	0.9

Table 5.5: LTP of obtaining a simulation of the *ensemble 1* with LTP lower than the one obtained with the **Commander** map, LTP_c , and **SMICA** map, LTP_s .

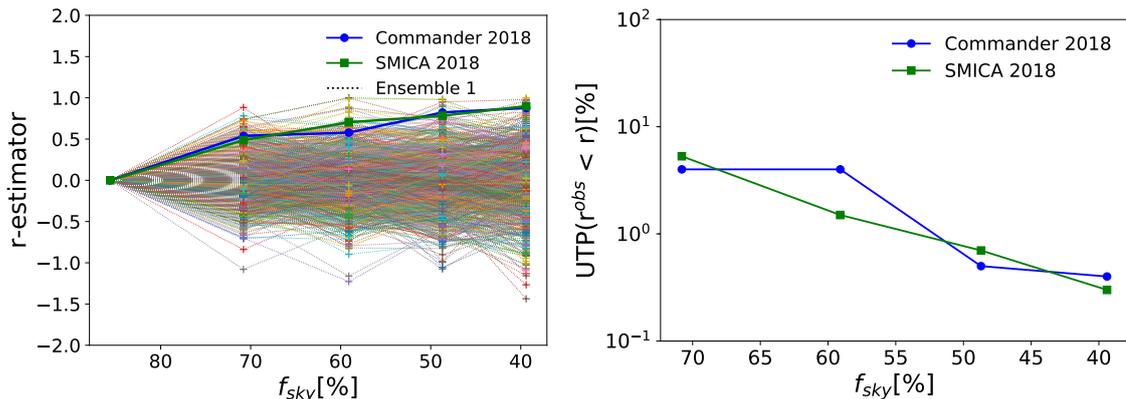


Figure 5.12: Left panel: r -estimator computed with Eq. (5.4) versus the sky fraction. The coloured dotted lines stand for the r value obtained from the *ensemble 1*. Blue and green solid lines stand for **Commander** and **SMICA** respectively. Right panel: UTP of obtaining a simulation with r larger than the one obtained with **Commander** (blue line) or **SMICA** (green line) as a function of the sky fraction.

solid green line connects the **SMICA** values (square symbols). The UTP are shown in the right panel of Fig. 5.12 and quoted in Table 5.6. At high Galactic latitude we find an anomalous value for r at the level of $\sim 2.9\sigma$ with a UTP of 0.4% for **Commander** and 0.3% for **SMICA**. In conclusions the results for the *ensemble 1* are similar to those of *ensemble 0* even when rotations are considered.

Mask	UTP [%]	
	$r^c < r$	$r^s < r$
Ext ₁₂	4.0	5.3
Ext ₁₈	4.0	1.5
Ext ₂₄	0.5	0.7
Ext ₃₀	0.4	0.3

Table 5.6: UTP of obtaining a simulation with r larger than the one obtained from the data. Second column shows the UTP for **Commander**, third column the UTP for **SMICA**.

A Appendix

A.1 Generating the rotations

Random rotations of temperature CMB maps are generated following an harmonic-based approach through a **Python** algorithm. We consider maps at **HEALPix** resolution $N_{side} = 16$ which are harmonic-expanded to obtain the initial $a_{\ell m}^{in}$ coefficients. These coefficients are

then rotated through the Wigner rotation matrices, $\mathbf{R}(\vartheta, \varphi, \psi)$, whose rotations angles $(\vartheta, \varphi, \psi)$ (also known as Euler angles), are randomly extracted from uniform distributions. Technically this is performed thanks to the `healpy` subroutine `rotate_alm`. After the rotation, the final map, or simply the rotated map, \mathbf{m}^R can be written as

$$\mathbf{m}^R = \sum_{\ell m} \left(\sum_{m'} \mathbf{R}_{mm'}(\vartheta, \varphi, \psi) a_{\ell m'}^{\text{in}} \right) Y_{\ell, m}(\theta, \phi). \quad (5)$$

To validate the procedure which implements random rotations, we consider a map which is zero except for a spot of 9° , see Fig. 13. This is done simply setting to 1, nine neighboring pixels and then smoothing⁸ the map with a Gaussian beam with a FWHM= 9° . For convenience we call \mathbf{m}_0 this initial map. Starting from \mathbf{m}_0 we perform N_{rot} rotations⁹ considering \mathbf{m}_{i-1} as the input for i^{th} rotation, with $i = 1, \dots, N_{rot}$. We then compute the

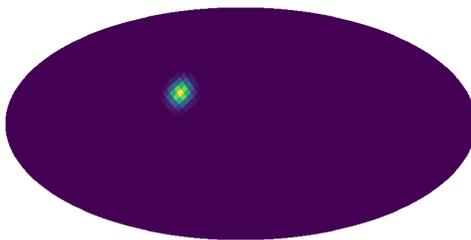


Figure 13: A test map at HEALPix resolution $N_{side} = 16$ with all pixels zero except for 9 pixels set to 1 and after convolution with a Gaussian beam of 9° .

following total map,

$$\mathbf{m}^{\text{tot}} = \sum_{i=0}^{N_{rot}} \mathbf{m}_i, \quad (6)$$

which is shown in Fig. 14, for $N_{rot} = 2, 50$ and 500 . The idea is to use \mathbf{m}^{tot} to test whether the set of considered rotations is able to “cover uniformly” all the possible directions. This is our requirement for validation which is quantified computing the APS of \mathbf{m}^{tot} and comparing the monopole with higher order multipoles: when the former dominates over the latter we can safely state that the set of rotations is sufficiently populated to have its isotropic part leading over accidental anisotropies. Note that, in turn, this procedure provides the minimum number of rotations which are needed to fulfill the requirement mentioned above. The left panel of Fig. 15 shows the behaviour of the lowest multipoles, namely the monopole C_0 , the dipole C_1 , the quadrupole C_2 , and the octupole C_3 , against the number of rotations. The monopole component increases its magnitude quadratically versus the number of rotations whereas low- ℓ components oscillate around a very slowly monotonic growth. We repeat this procedure 50 times and compute the mean distribution of the same first low- ℓ components, see right panel of Fig. 15. The mean behaviour of the different components, and the hierarchy among the multipoles, is substantially unchanged with respect to what obtained with the single realisation. In particular the hierarchy among low- ℓ multipole components seems to become stable for $N_{rot} > 900$. Most importantly, we find that the magnitude of the ratio C_0/C_1 at $N_{rot} = 1000$ is of the order 10^3 : therefore we choose this threshold to define the minimal number of rotations needed to cover sufficiently homogeneously the whole sky.

⁸The smoothing is applied in order to minimise aliasing effects when going from real to harmonic space and vice versa.

⁹In other words, we apply N_{rot} times Eq. (5).

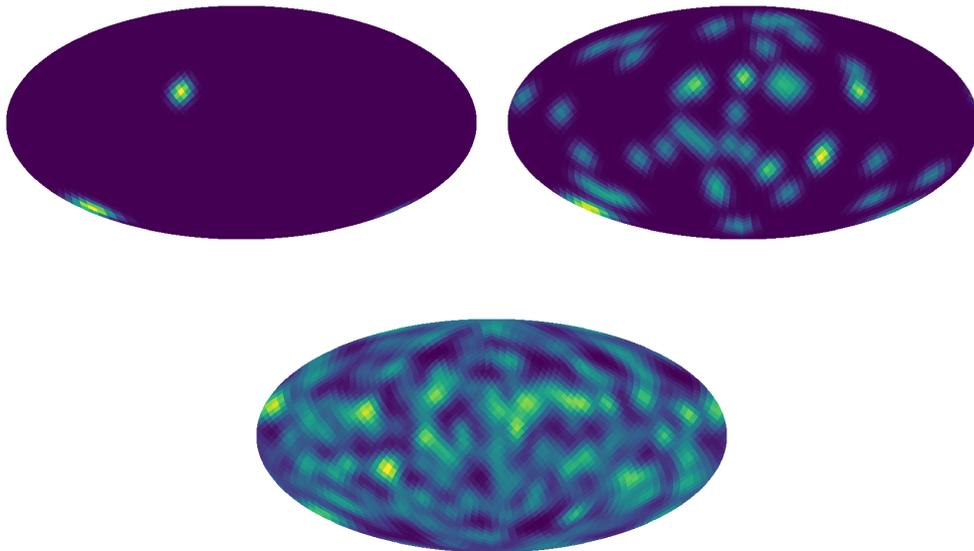


Figure 14: Total map \mathbf{m}^{tot} computed through Eq. 6 at HEALPix resolution $N_{side} = 16$ for $N_{rot} = 2$ (top left panel), $N_{rot} = 50$ (top right panel) and for $N_{rot} = 500$ (bottom panel).

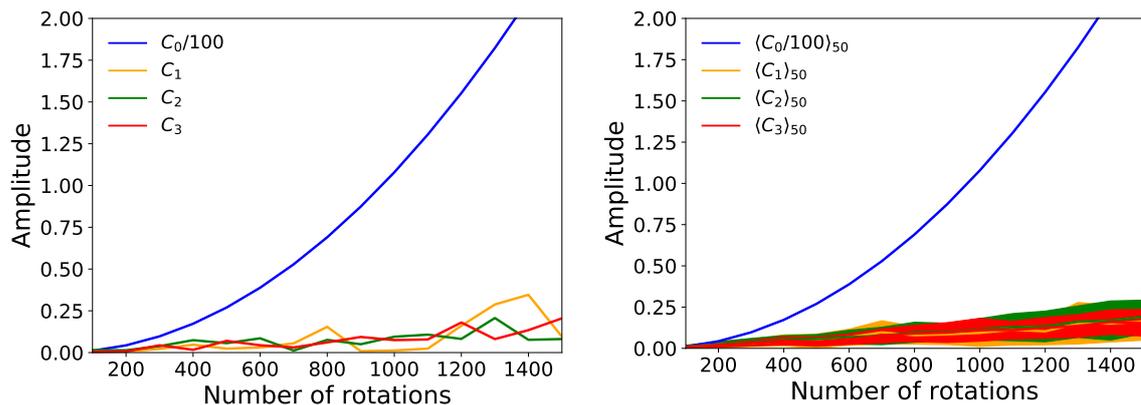


Figure 15: Left panel: amplitude of the first low- ℓ components of the APS of the test map after each rotation. Right panel: the average over 50 repetitions of the machinery described in Sec. A.1. The filled regions correspond to the 1σ dispersion of C_ℓ components.

A.2 Comparison between 2018 and 2015 *Planck* release

In this section we consider the 2015 *Planck* data. This analysis is performed mainly because the *Planck* 2015 standard mask (Planck Collaboration XI 2016), henceforth called Std 2015, is smaller than the 2018 one. Its observed sky fraction is 93.6%, see Fig. 16, versus 85.6% of the Std 2018, see Fig. 5.1 and Table 5.1. Hence we employ here the Commander 2015 map used in Planck Collaboration XI (2016) still at HEALPix resolution $N_{side} = 16$ and FWHM of 440 arcmin and consistently to what performed for the 2018 case, we added to this map a regularisation noise of $2 \mu\text{K rms}$. The masks used during this analysis are the same listed in Table 5.1, with the exception of the Std 2018, which has been replaced with the Std 2015. Similarly to what performed in Section 5.1 for the generation of the *ensemble 0*, we build here a MC of 10^4 maps using the *Planck* 2015 best-fit model. From these maps, we select a subset of 10^3 maps with variance V within $20 \mu\text{K}^2$ from the value computed with the Commander 2015 map, i.e. $V_c = 2060.09 \mu\text{K}^2$. This set of simulations is called *ensemble 1-2015*. The behaviour of V as a function of the masks obtained with the *ensemble 1-2015* is shown in Fig. 17 and the corresponding LTP

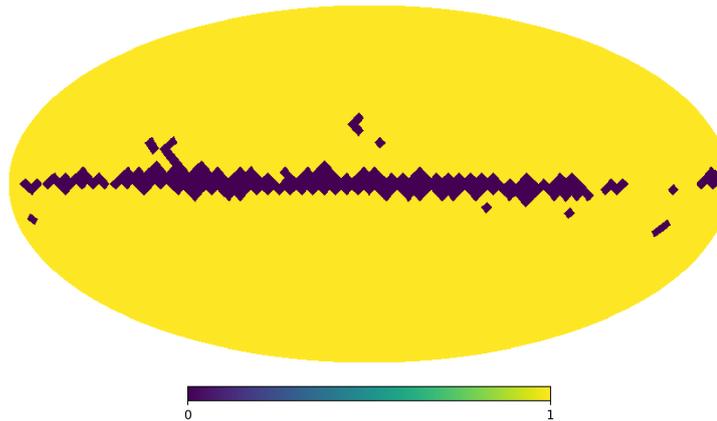
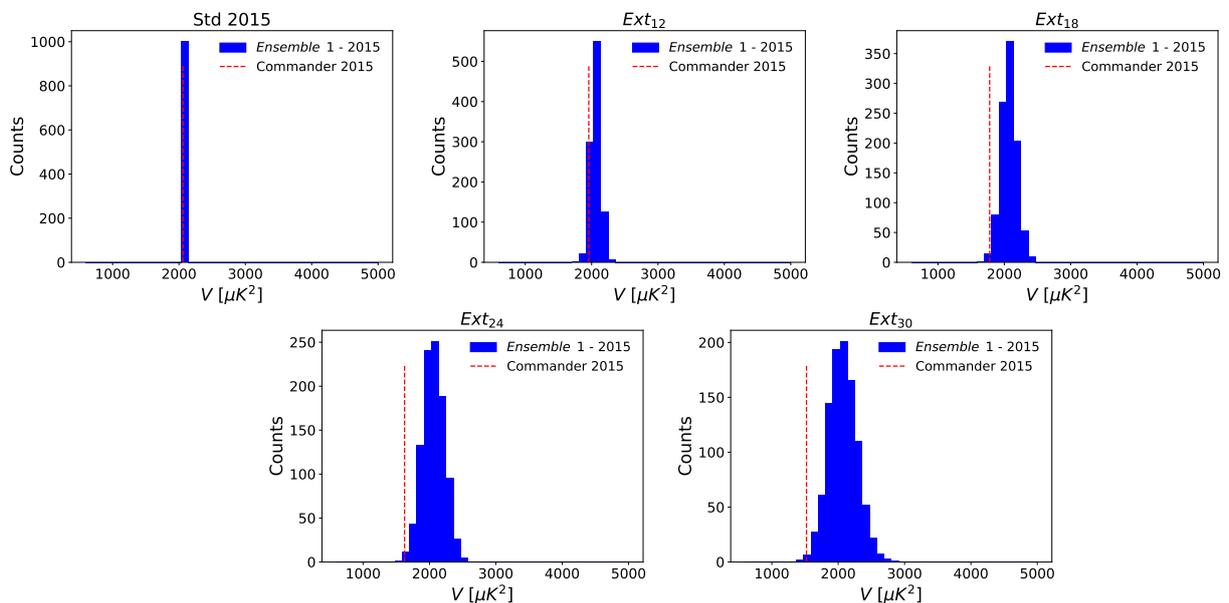


Figure 16: Std 2015 temperature mask.

Figure 17: Histograms of the variance V of the maps belonging to *ensemble 1 - 2015* computed for the masks Std 2015, Ext₁₂, Ext₁₈, Ext₂₄ and Ext₃₀. The red dashed line identifies the variance of the *Commander 2015* map, V_c .

are reported in the first column of Table 7. We recover a similar monotonic behaviour as for the 2018 case. That is, in the Ext₃₀ case, the behaviour of the 2015 data is anomalous at $\sim 2.9\sigma$.

We take into account now the rotations applied to *ensemble 1-2015*. The results for the LTP-estimator and r -estimator are shown in Fig. 18 and Fig. 19. For the LTP-estimator we find in the Ext₃₀ mask a LTP of 0.5% which is in line with the 2018 analysis. All the LTP for this estimator are reported in Table 7. On the other hand, the r -estimator gives $r^c = 0.80$ for the Ext₃₀ mask with a p-value of 1.2%. While the general behaviour of r across the mask is recovered here, the probability at high Galactic latitude is slightly higher.

A.3 Dependence on threshold

In this section we study the impact on our results of the threshold of V we choose to select the maps of the *ensemble 1* from the 10^5 Λ CDM simulations. Specifically, in addition to the threshold of $20 \mu\text{K}^2$ used in Section 5.1, we choose two other thresholds at $10 \mu\text{K}^2$

Mask	LTP [%]		
	$V < V_c$	$V_c^{(\text{rot})} < V_c$	$LTP_i < LTP_c$
Std 2015	47.2	64.2	58.3
Ext ₁₂	7.2	11.1	10.0
Ext ₁₈	0.8	2.9	2.4
Ext ₂₄	0.4	1.6	1.6
Ext ₃₀	0.3	0.5	0.5

Table 7: The probability of obtaining a value for the variance V smaller than that of **Commander** 2015 for a map of the *ensemble* 1-2015 (first column). The probability of obtaining a value of the variance of the rotated **Commander** 2015 map, $V_c^{(\text{rotated})}$, smaller than the unrotated one (second column). LTP of obtaining a simulation with LTP_i lower than the one obtained with the **Commander** 2015 map, LTP_c (third column).

Mask	UTP [%]
	$r^c < r$
Ext ₁₂	12.1
Ext ₁₈	2.3
Ext ₂₄	1.4
Ext ₃₀	1.2

Table 8: UTP of obtaining a simulation of the *ensemble* 1 - 2015 with r larger than the one obtained with the **Commander** 2015 map.

and $30 \mu\text{K}^2$. These will define two new subsets of 10^3 CMB temperature maps which have a variance V close to the value observed by **Commander** 2018. We refer to these two additional subsets as *ensemble* 2 (E2) and *ensemble* 3 (E3), respectively.

Therefore, we repeat on E2 and E3, the same analysis previously performed on *ensemble* 1 for both the considered estimators, focusing on the mask Ext₃₀. We start building the distribution of V , see Fig. 20, where the left panel refers to E2 while the right one to E3. The LTP of **Commander** 2018 are $LTP_{E2}(V_c < V_i)=0.2\%$ and $LTP_{E3}(V_c < V_i)=0.7\%$, which are consistent with what obtained with *ensemble* 1.

As done for the *ensemble* 1, we can apply random rotations to E2 and E3 and build the LTP-estimator and the r -estimator in the Ext₃₀ case. The distributions of the former are shown in the left panels of Fig. 21 and Fig. 22 for the E2 and E3 case respectively. The vertical dashed bars stand for the **Commander** 2018 values of the estimator, see again Table 5.5. The LTP of the LTP-estimator, turn out to be 0.3% and 0.2% for E2 and E3 respectively. In the right panels of Fig. 21 and Fig. 22 we show the r -estimator for the E2 and E3. We find $UTP_{E2}(r_c < r_i)=0.4\%$ and $UTP_{E3}(r_c < r_i)=0.5\%$ for E2 and E3 respectively. We conclude that our results are stable with respect to the choice of the threshold which defines the set of constrained realisations.

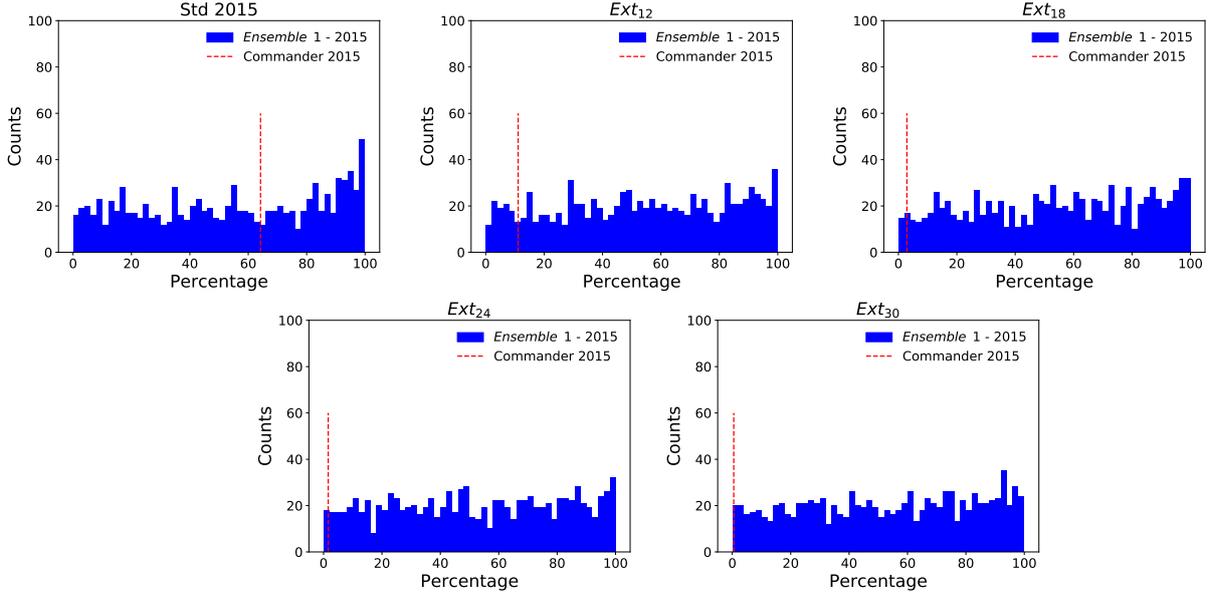


Figure 18: LTP of finding a rotated map of the *ensemble 1 - 2015* with $V^{rot} < V$, where V is the variance of the corresponding unrotated map. Each panel shows the results obtained using a different mask. The dashed vertical bars are the LTP of *Commander 2015*.

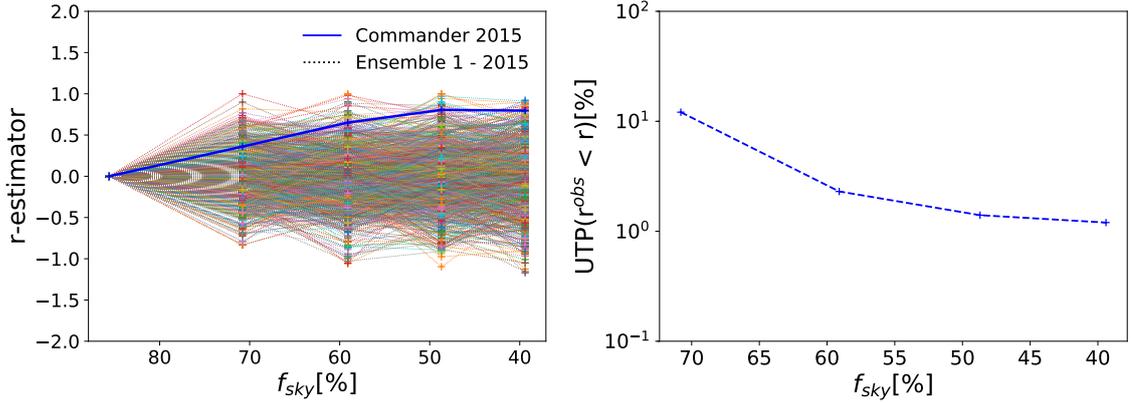


Figure 19: Left panel: r -estimator computed with Eq. (5.4) versus sky fraction. Right panel: UTP of obtaining a simulation with r larger than the one obtained with *Commander 2015* as a function of the sky fraction.

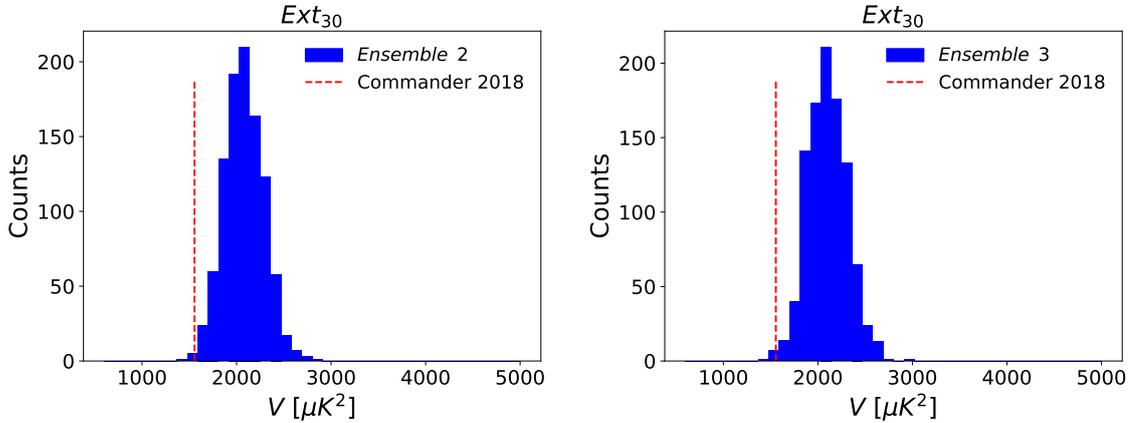


Figure 20: Variance distribution of the *ensemble 2* (left panel) and *ensemble 3* (right panel) for the *Ext30* mask. Red dashed line corresponds to the variance of the *Commander 2018* map.

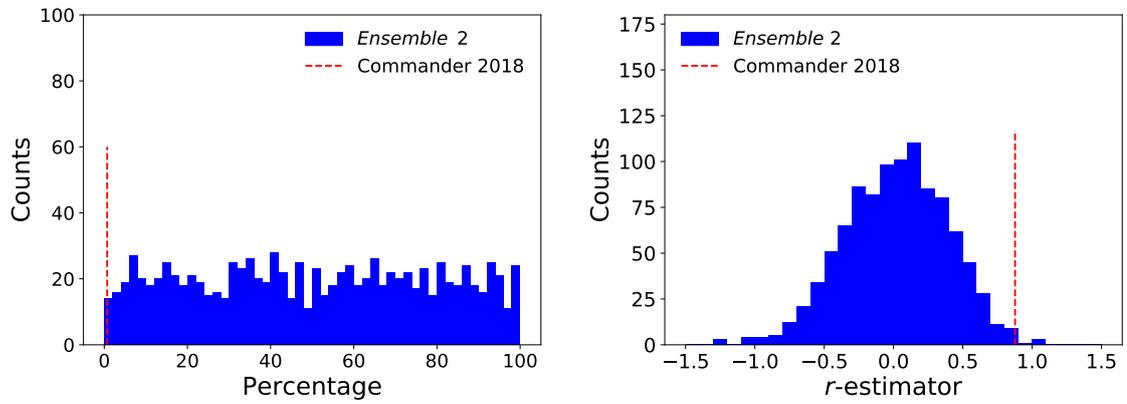


Figure 21: Left panel: distribution of probability of observing, in a Λ CDM model with low variance, a lower value with respect to V_C due to random rotations of *ensemble 2*. Right panel: r -estimator computed with Eq. (5.4) for the *ensemble 2*. Both the results have been obtained using the Ext₃₀ mask.

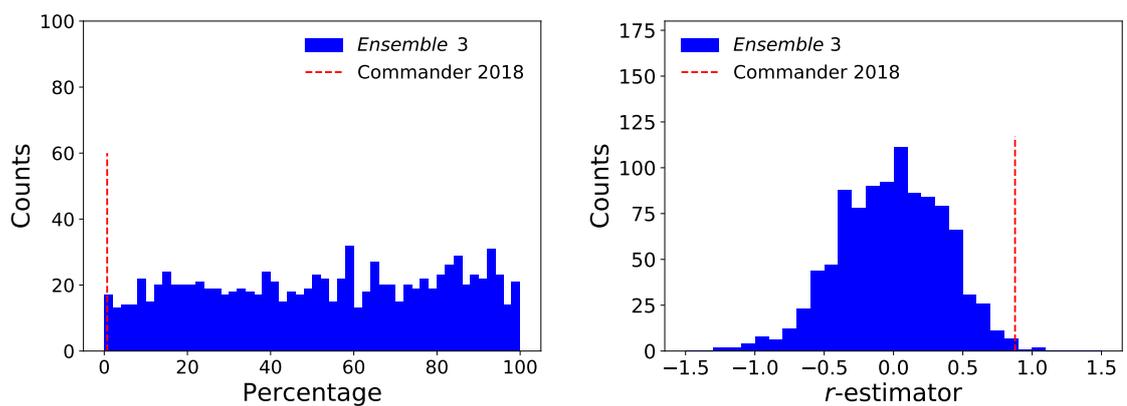


Figure 22: The same as Fig.21 but for *ensemble 3*.

A small-scale estimate of the reionization optical depth and its interplay with Λ CDM extensions

As pointed out in [Planck Collaboration VI \(2018\)](#), there is no strong evidence in favour of any of the extensions to the Λ CDM model. Furthermore, it has also shown the robustness of the constraints on the base- Λ CDM parameters to these extensions. The results are stable both when are used *Planck* data alone and when are added data from external sources. However, *Planck* measurements are far from the cosmic variance limit, which is crucial to discriminate between different cosmological models. It demands the need for new experiments specifically designed to overcome what has done so far.

Despite the community interest in the improvement of polarization and lensing measurements, there are no studies in the literature devoted to evaluating the information content carried individually by these two probes. Nonetheless, the gain achievable on a single parameter by improving their measurements is not explicitly quantified. Both probes are crucial for reducing the volume of parameter space, not only for the Λ CDM model but also for its extensions. On one hand, lensing can help to improve the neutrino masses or the effective number of relativistic species constraints. Thus, relatively interesting extensions, in this optic, can be both Λ CDM+ m_ν and Λ CDM+ N_{eff} models. On the other hand, future large-scale polarization measurements are promising stringent constraints on the reionization optical depth. Such an achievement can easily break the Δ - τ degeneracy. Here Δ represents a characteristic scale which modify the power spectra damping both temperature and polarization power at low multipoles ([Gruppuso and Sagnotti 2015](#)). It constitutes one possible solution to the lack of power anomaly at large angular scales in the anisotropy power spectrum since it breaks the power-law at very low wavenumbers. In this Λ CDM+ Δ model the scalar power spectrum takes the form of

$$P_\Delta(k) = A_s \frac{(k/k_*)^3}{\left[(k/k_*)^2 + (\Delta/k_*)^2\right]^{2-\frac{n_s}{2}}}. \quad (6.1)$$

In this chapter, we are going to apply some method seen in [Sec. 3.4](#). The aim is to quantify the effects of lensing and large-scale polarization measurements induced on Λ CDM parameters. As extended models, we consider the possibility of varying the neutrino mass, the effective number of relativistic species, ad Δ . All these three parameters have a characteristic effect on both CMB power spectra and lensing potential. Thus, separating the two probes, we study how a different experimental setup will affect their determination. The idea beyond that is the possibility of breaking the $A_s e^{-2\tau}$ degeneracy in two ways. The first involves large-scale polarization measurement to constraint the value of τ since the EE

spectrum is the most sensitive to a change of this parameter. The second employs large-scale temperature and lensing likelihood. This choice allows us to break the degeneracy by determining the value of A_s .

6.1 Dataset and forecasts

We employ three different large-scale polarization datasets. Precisely, we use the low-frequency dataset described in Chapter 4, hereafter WMAP+LFI; the spectrum based dataset described in Pagano et al. (2019), hereafter `Sroll2`; the spectrum based dataset described in Planck Collaboration V (2019), hereafter `SimAll`. All these three datasets employ the `Commander` 2018 map (Planck Collaboration V 2019) in temperature. We further combine these with the `plik_TTTEEE` *Planck* high- ℓ (Planck Collaboration V 2019) and the *Planck* lensing likelihoods (Planck Collaboration VIII 2018). We will improperly refer to these joint likelihoods with the name of the corresponding low- ℓ datasets.

Furthermore, we consider forecasts for future CMB experiments by performing a Bayesian likelihood analysis using a Monte Carlo Markov Chains (MCMC) method, following the approach described in Perotto et al. (2006). In particular, having set the Λ CDM parameters to the bestfit values given in Pagano et al. (2019), we produce three mock datasets. We refer to the first as `stdSO`. It includes the publicly available “goal” noise curves for the 93 GHz channel at small scales, $30 \leq \ell \leq 4000$, over 40% of the sky. These are released by the SO collaboration and fully described in Sec. 2.2 of Ade et al. (2019). For the largest angular scales not probed by SO, we include TT and EE from *Planck* over 80% of the sky in the range $2 \leq \ell \leq 29$. It is worth noting that, this choice is different from the one adopted by the SO collaboration which use only the *Planck* intensity data with a prior on the optical depth of $\tau = 0.06 \pm 0.01$. This gives us lower constraints on τ -estimate. We also include an additional 20% of the sky from *Planck* in the range $30 \leq \ell \leq 2500$. This is done, accordingly with Ade et al. (2019), in order to produce an overall sky area of 60%, which is compatible with the area used by *Planck* after masking the Galaxy. We use a simple white noise model, with an amplitude of 30 μK arcmin in temperature and 60 μK arcmin in polarization, to account for the *Planck* noise characteristics. The full width half maximum (FWHM) of the beam is set to 7° . We refer to the second as `CV+SO`, that differs from the previous one by the inclusion of cosmic variance limited resolution on the EE power spectrum over 100% of the sky in the range $2 \leq \ell \leq 29$. The last one, hereafter `CV`, is an ideal experiment in which we have cosmic variance limited resolution on power spectra in all the available multipole range ($2 \leq \ell \leq 4000$). The lensing noise curves are built through the quadratic estimator described in Okamoto and Hu (2003).

For these three different experimental configurations, we also produce a set of extended Λ CDM mock data. Precisely, these include Λ CDM+ Δ model with $\Delta = [0.5, 1.5, 2.5, 3.5] \times 10^{-4} \text{ Mpc}^{-1}$; a Λ CDM+ m_ν model with $m_\nu = [0.1, 0.15] \text{ eV}$; a Λ CDM+ N_{eff} model with $N_{eff} = [2.9, 3.11]$.

Finally, for each of the datasets described, we build two sub-datasets. The former combines the EE power spectrum in the low-multipole range ($2 \leq \ell \leq 29$) with the TT , TE and EE power spectra in the high-multipole range ($30 \leq \ell \leq 4000$). This choice guarantees the possibility of break the $A_s e^{-2\tau}$ degeneracy by constraining the reionization optical depth through the large-scale polarization measurement. The latter combines the TT power spectrum in the low-multipole range with the TT , TE , EE power spectra plus lensing potential in the high-multipole range. As for the large-scale polarization probe, also this combination allows us to break the $A_s e^{-2\tau}$ degeneracy, this time by constraining the scalar amplitude A_s through lensing observation. In what follows, we refer to these two likelihoods as `lEhTE` (lowE_highTTTEEE) and `lThTEphi` (lowT_highTTTEEE_lensing) respectively. These two datasets are used to quantify the information content of lensing

potential and large-scale polarization measurements in terms of information theory.

6.2 Methods

In this section, we are going to explore the methods used to quantify the effects induced by the two cosmological probes on both Λ CDM parameters and some of its extensions. We firstly report the constraints on Δ , m_ν and N_{eff} . We focus the attention primarily on Δ since there are no-updated constraints on it. Subsequently, we forecast which experimental setup is needed to have a more than 3σ detection of this parameter. We further quantify the information content of lensing and large-scale polarization measurement using concepts of information theory.

We use the Shannon-entropy defined in Sec. 3.4. If we denote the posterior distribution as $p(\boldsymbol{\theta})$, given an opportune choice of parameters $\boldsymbol{\theta}$, it reads

$$S = - \int d\boldsymbol{\theta} p(\boldsymbol{\theta}) \ln p(\boldsymbol{\theta}). \quad (6.2)$$

The value of S depends on both the dimensionality of the problem and the logarithmic volume of parameter space bounded by 1σ contour: better is the constraints, and higher is its absolute value. In other words, S assumes small numbers for a peaked likelihood. However, the exact computation of this object requires the evaluation of Bayesian evidence, that is in general an expensive computational challenge. One can give up his calculation and ask himself what distribution maximises it, by obtaining a lower bound for the information content. For this propose, we observe that among all distributions with a fixed variance, the Gaussian distribution is the one that maximise the Shannon-entropy, see Sec. 3.4.1. As it maximises entropy, this solution represents the distribution that assumes the least information given the constraints on the variance. This result provides a huge simplification, since it allows us to compute the S analytically once we know the covariances Σ of the chosen parameters. The Shannon entropy, in this case, reads (cfr. Eq. (3.88))

$$S = \frac{1}{2} \ln \left[(2\pi)^d |\Sigma| e^d \right], \quad (6.3)$$

where d is the dimensionality of the distribution. We can compare the lThTE ϕ and lEhTE datasets with their union to estimate the information content carried by these two probes by comparing probability densities functions with statistical non-independence. That is, one can define relative entropies between two distributions. The relative entropy in the Shannon entropy case is the Kullback-Leibler divergence (Kullback and Leibler 1951)

$$\Delta S = \int d\boldsymbol{\theta} p(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}. \quad (6.4)$$

Notice that, in general, also the evaluation of this object is computationally expensive. However, since we are comparing the two normally-distributed posteriors, $p(\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\theta}; \mu_1, \Sigma_1)$ and $q(\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\theta}; \mu_2, \Sigma_2)$ (the one that maximise the Shannon-entropy), we can use the analytical expression given by (cfr. Eq. (3.91))

$$\begin{aligned} \Delta S [p(\boldsymbol{\theta})|q(\boldsymbol{\theta})] &= \frac{1}{2} (\mu_1 - \mu_2)^\top \Sigma_2^{-1} (\mu_1 - \mu_2) + \\ &+ \frac{1}{2} \left[\text{Tr} \left(\Sigma_1 \Sigma_2^{-1} \right) - d - \log \left(\frac{\det \Sigma_1}{\det \Sigma_2} \right) \right]. \end{aligned} \quad (6.5)$$

An interesting interpretation can be given by setting $p(\boldsymbol{\theta})$ equal to the Gaussian approximation of the full dataset posterior and $q(\boldsymbol{\theta})$ equal to the Gaussian approximation of the

lThTE ϕ or lEhTE posterior. In this case the entropy divergence (ΔS_{FT} or ΔS_{FE}) quantifies by how much the entropy will decrease by adding polarization lE (lowE) or lensing lTh ϕ (lowT_high ϕ) measurement to the data. Thus, we expect that more information is gained adding a complementary dataset, higher is the value of the Kullback-Leibler divergence. This is well captured in Fig. 6.1, where it is shown the relative entropies for the full likelihood of a Λ CDM for a successive and cumulative probe combination adding CMB-lensing, galaxy clustering and weak lensing to the primary CMB (Pinho et al. 2020). Here we can observe a reduction in uncertainty achieved through the combination of cosmological probes reflected by smaller absolute information entropies. The same is repeated for five parameters individually.

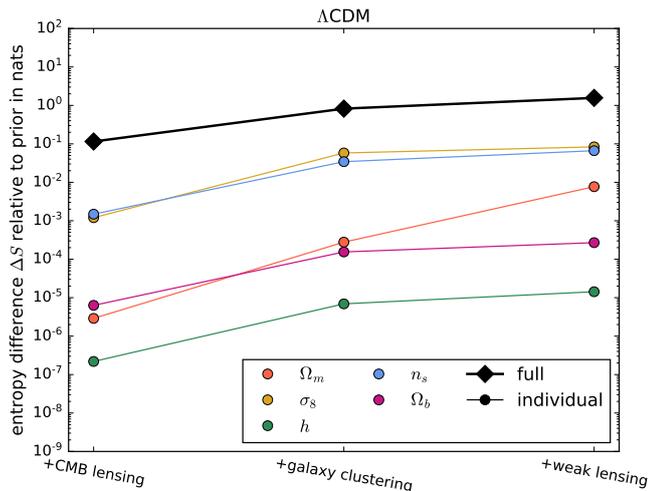


Figure 6.1: Relative entropies ΔS for the full likelihood of a Λ CDM, and for $n = 5$ parameters individually, both marginalised and conditionalised, for a successive and cumulative probe combination adding CMB-lensing, galaxy clustering and weak lensing to the primary CMB (Pinho et al. 2020).

Notice that the relative entropy is sensitive to differences between parameter values derived with different probes, but in addition there is a dependence on the difference between the errors. As an example, we can consider Eq.(6.5) where it is clear that each mismatch in parameter estimates can be reabsorbed by varying the magnitude of the covariance matrix. Indeed, if we assume that $p(\boldsymbol{\theta})$ is derived combining the posterior $q(\boldsymbol{\theta})$ obtained using old dataset with some addition of complementary data, it can be shown that (see, e.g. Seehars et al. 2014)

$$\langle \Delta S [p(\boldsymbol{\theta})|q(\boldsymbol{\theta})] \rangle = -\frac{1}{2} \log \frac{|\Sigma_1|}{|\Sigma_2|} \quad (6.6)$$

$$\sigma^2(\Delta S) = \frac{1}{2} \text{Tr} \left[\left(\Sigma_1 \Sigma_2^{-1} - \mathbb{I} \right)^2 \right]. \quad (6.7)$$

Thus, if there is no improvement by adding some complementary data in constraining parameters, we expect $\langle \Delta S \rangle \sim 0$ and $\sigma^2(\Delta S) \sim 0$. Then, each mismatch in parameter estimates will increase the “surprise” ($\Delta S - \langle \Delta S \rangle$) of obtaining a ΔS different from zero. This is not true anymore if there is also a change in the magnitude of the covariance matrix. This property can be used to quantify the magnitude of systematical errors and tensions between data sets (Amara and Refregier 2013; Seehars et al. 2014, 2016; Nicola et al. 2019; Pinho et al. 2020). However, in this thesis, we aim to quantify only the information gain when we add polarization or lensing measurement to the complementary dataset. This kind of analysis is left for future work.

6.3 Information content of different CMB probes

The most recent constraint up to date on Δ is given in Gruppuso et al. (2018), where they used the Commander 2015 solution (Planck Collaboration XI 2016) in temperature and a noise weighted combination of WMAP and *Planck* in polarization. They shown that the value of this scale parameter changes with the amount of excluded Galactic latitude, capturing the enhancement of the lack-of-power anomaly with the reduction of the temperature sky fraction (Gruppuso and Sagnotti 2015; Monteserin et al. 2008; Cruz et al. 2011; Gruppuso et al. 2013; Planck Collaboration VII 2018). The value that they quoted using the Commander 2015 standard mask (Planck Collaboration XI 2016) is $\Delta = (1.7 \pm 0.9) \times 10^{-4} \text{ Mpc}^{-1}$ (68% CL). With only an available sky fraction of $f_{\text{sky}} = 39\%$, they found a value of $\Delta = (3.5 \pm 1.1) \times 10^{-4} \text{ Mpc}^{-1}$ (68% CL), which correspond to a $\sim 3 \sigma$ detection. The corresponding dependency from the sky fraction in polarization is not yet analyzed.

Thus, the first step is to extract updated constraints on Δ . Here we employ the three different datasets described in Sec. 4.1, namely WMAP+LFI, SimAll and Sroll2, using the MCMC CosmoMC (Lewis and Bridle 2002) sampler to estimate the parameters posterior distributions. All these three dataset employ the Commander 2018 mask (Planck Collaboration V 2019) in temperature. The values that we find, at 68% CL, are shown in Tab. 6.1. As we can see, all the values are consistent with the previous quoted in Gruppuso et al. (2018), leading, in the most constraining case, to a 1.6σ detection. Note that, since the HFI based datasets better constrain the value of the optical depth τ , the value of Δ tends to decrease due to the mild degeneracy between the two parameters, see left panel of Fig. 6.2. The corresponding marginal posterior probabilities for Δ in the three different cases are shown in the right panel of Fig. 6.2.

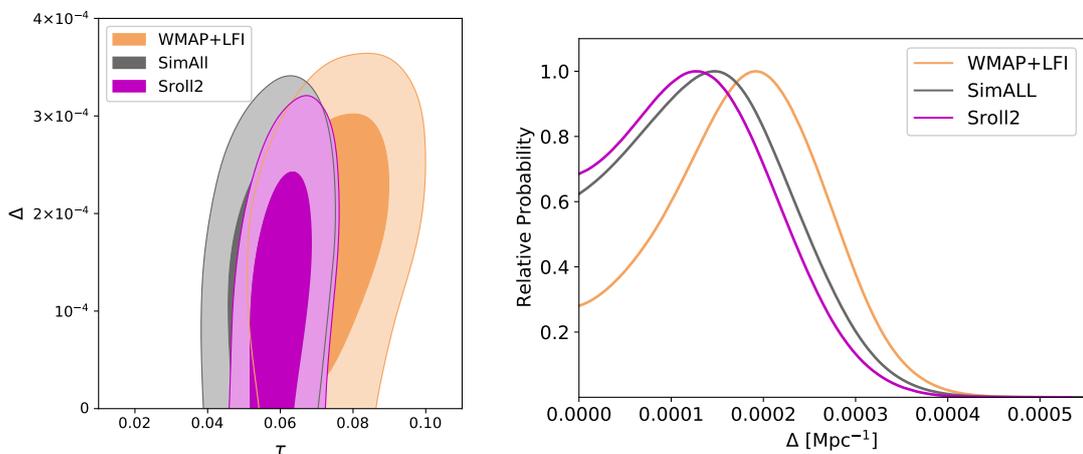
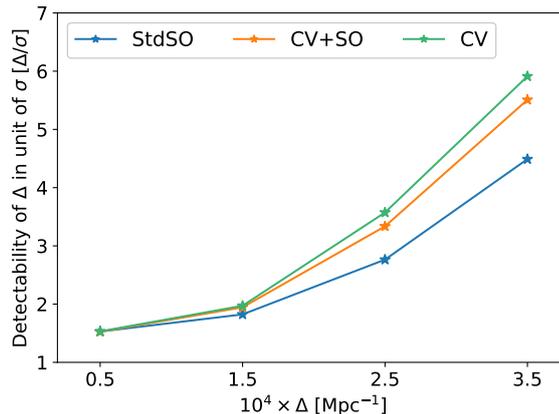


Figure 6.2: Left Panel: Joint posterior probability for scale parameter Δ and optical depth τ , computed from WMAP+LFI, SimAll, and Sroll2 datasets. Right Panel: Marginal posterior probability for scale parameter Δ , computed from WMAP+LFI, SimAll, and Sroll2 datasets.

Table 6.1: Constraints on Δ at 68% CL for the different dataset employed.

Dataset	$10^4 \times \Delta [\text{Mpc}^{-1}]$
WMAP+LFI	1.79 ± 0.83
SimAll...	1.43 ± 0.82
Sroll2...	1.32 ± 0.78

It is worth noting that, due to the increasing statistical significance of the lack-of-

Figure 6.3: Improvement on the detectability of Δ in unit of σ as a function of Δ .

power anomaly with the decreasing of the available sky fraction, we are not able to fix a fiducial value for this scale parameter. Thus, it is interesting to forecast the detectability of Δ due to an improvement of the experimental setup as a function of Δ itself. In fact, increasing the value of this scale parameter means more suppressing the power on a large scale. As a consequence, there is a decreasing of cosmic-variance. This undoubtedly affects the capability of constraining Δ . Here we use the three mock datasets, StdSO, CV+SO, and CV, built in a Λ CDM+ Δ framework with $\Delta = [0.5, 1.5, 2.5, 3.5] \times 10^{-4} \text{ Mpc}^{-1}$. This range of values is consistent with the findings reported in Tab. 6.1. Precisely, the lowest value, $\Delta = 0.5 \times 10^{-4} \text{ Mpc}^{-1}$, corresponds to a 1.6σ shift with respect to the WMAP+LFI value. The highest value, $\Delta = 3.5 \times 10^{-4} \text{ Mpc}^{-1}$, as well as being consistent with the finding in the extended mask (Gruppuso et al. 2018), correspond, in the worst case, to a 2.8σ shift with respect to the Sroll12 value.

The choice, in our opinion, the only reasonable one, to select a range of Δ 's leads to answer the question: *Which is the lowest detectable value of Δ once the experimental setup is improved?* We found that the lowest value of Δ allowing for a $\sim 2 \sigma$ detection is $\Delta = 1.5 \times 10^{-4} \text{ Mpc}^{-1}$, with an improvement of the $\sim 15\%$ on the error between the StdSO and the CV cases. The value of $\Delta = 2.5 \times 10^{-4} \text{ Mpc}^{-1}$, that correspond to a 0.8σ shift with respect to WMAP+LFI and to a 1.5σ shift with respect to Sroll12, can potentially be detected at more than 3σ for experiment like CV+SO. Fixing $\Delta = 3.5 \times 10^{-4} \text{ Mpc}^{-1}$, its detectability can reach $\sim 6 \sigma$, with an improvement of 24% on the error between the StdSO and the CV cases. These results are shown in Fig. 6.3.

Table 6.2: Constraints on m_ν (at 95% CL) and N_{eff} (at 68% CL) for the different dataset employed.

Dataset	m_ν (eV)	N_{eff}
WMAP+LFI	< 0.38	2.92 ± 0.19
SimAll...	< 0.25	2.89 ± 0.19
Sroll12...	< 0.27	2.89 ± 0.19

We report also the values of m_ν and N_{eff} obtained from the WMAP+LFI dataset. This choice is due to the absence of such an estimate in literature because of the recent delivery of this dataset. The results are shown in Tab. 6.2, where we report also the values obtained from SimAll and Sroll12 for completeness.

At this point, we can proceed in the analyses and quantify the information carried by the two discussed probe. The left panel of Fig. 6.4 shows the values of the Shannon entropy computed as in Eq. (6.3) using the posteriors of WMAP+LFI, SimAll and

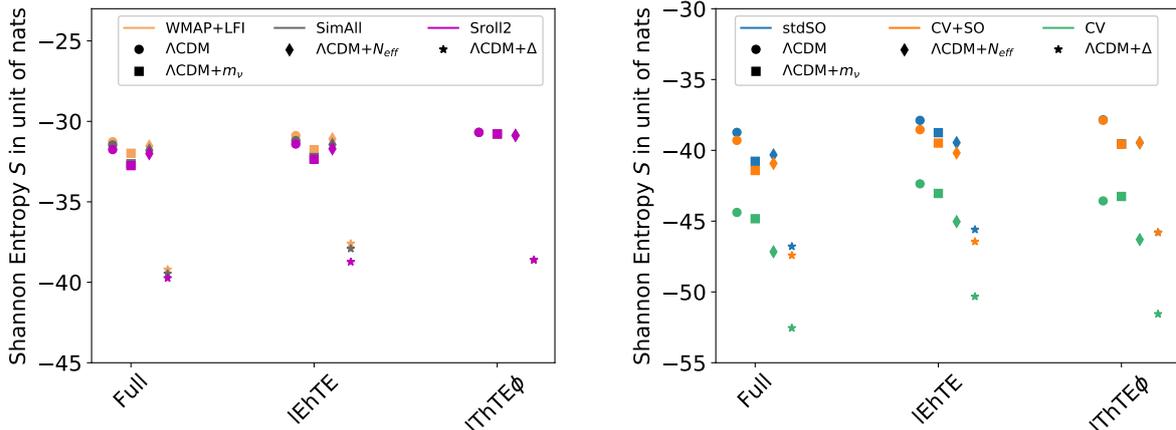


Figure 6.4: Left Panel: Behaviour of the Shannon entropy computed by using the current available CMB data. Right Panel: Behaviour of the Shannon entropy computed by using the forecasts datasets.

Sroll12 datasets. Here we report the information for the full, lEhTE and lThTE ϕ likelihoods. The Sroll12 dataset, as expected, is the one with the higher constraining power. This result shows how the reduction in the volume of parameter space, due to a better breakdown of the $A_s e^{-2\tau}$ degeneracy, strongly depends on the low- ℓ polarization measurement. However, for the Λ CDM model, there is no significant difference in the overall information content between the two probes. This effect is basically due to the choice of the base parameters. Indeed, only three over six of them strongly depend from the large scale measurements, one of which is also an internally derived parameter. Instead, the current precision level of large-scale polarization measurement, compared with the lensing likelihood, emphasize its global higher information content in both Λ CDM+ m_ν and Λ CDM+ N_{eff} frameworks. The Λ CDM+ Δ model is strongly dependent from the temperature measurement by construction. This effect makes the lThTE ϕ likelihood favoured respect the large-scale polarization in disentangling the mild degeneracy between τ and Δ . Finally, note that all three datasets share the same large-scale temperature, small-scale data, and lensing likelihood. This explains why the Shannon entropy of the full dataset is driven by the behaviour of the lEhTE dataset. Furthermore, the Shannon entropy results, in general, lower for the extended models. This result reflects the change in both the dimensionality and the volume of parameters space. Note that, this reduction is drastically enhanced for the Λ CDM+ Δ model, since there is also a reduction of the cosmic variance in the first multipoles. This confirms that the comparison can be done only between values computed assuming the same fiducial model.

The right panel of Fig. 6.4 shows the forecast values of the Shannon entropy computed using the mock datasets. As expected, there is a reduction of the S with respect to the current values reported in the left panel of Fig. 6.4. Furthermore, the importance of lensing measurement became more evident. Indeed, for an SO-like experiment the global information content is comparable for the two probes if the underlying model is Λ CDM. For a Λ CDM+ m_ν and Λ CDM+ N_{eff} models the information carried by lThTE ϕ is greater than the one obtained by lEhTE. Notice that, a CV+SO like experiment largely improve the constraint on τ and thus on A_s . Despite that, by adding cosmic variance limited large-scale polarization measurements there is only a slight improvement on the overall information content carried by lEhTE dataset. This result is true for all the cosmological models considered. Finally, a CV ideal CMB experiment tends to enhance the discrepancies between information content of lensing and polarization measurement in favour of lThTE ϕ dataset.

What we have achieved so far does not give us any information on single param-

ter. It infers only global behaviour. To quantify the effects induced by adding the two probes we use Kullback-Leibler divergence as explained in Sec. 6.2. Precisely, to compute the gain induced by adding large-scale polarization measurement we use the divergence $\Delta S_{FT} \equiv \Delta S[p_F|p_T]$, where $p_F(\boldsymbol{\theta})$ and $p_T(\boldsymbol{\theta})$ are the full and lThTE ϕ posterior distribution, respectively. Similarly, to compute the gain induced by adding the lensing measurement we use the divergence $\Delta S_{FE} \equiv \Delta S[p_F|p_E]$, where $p_E(\boldsymbol{\theta})$ is the lThTE ϕ posterior distribution. As previously stated, we use the Gaussian approximation in Eq. (3.91).

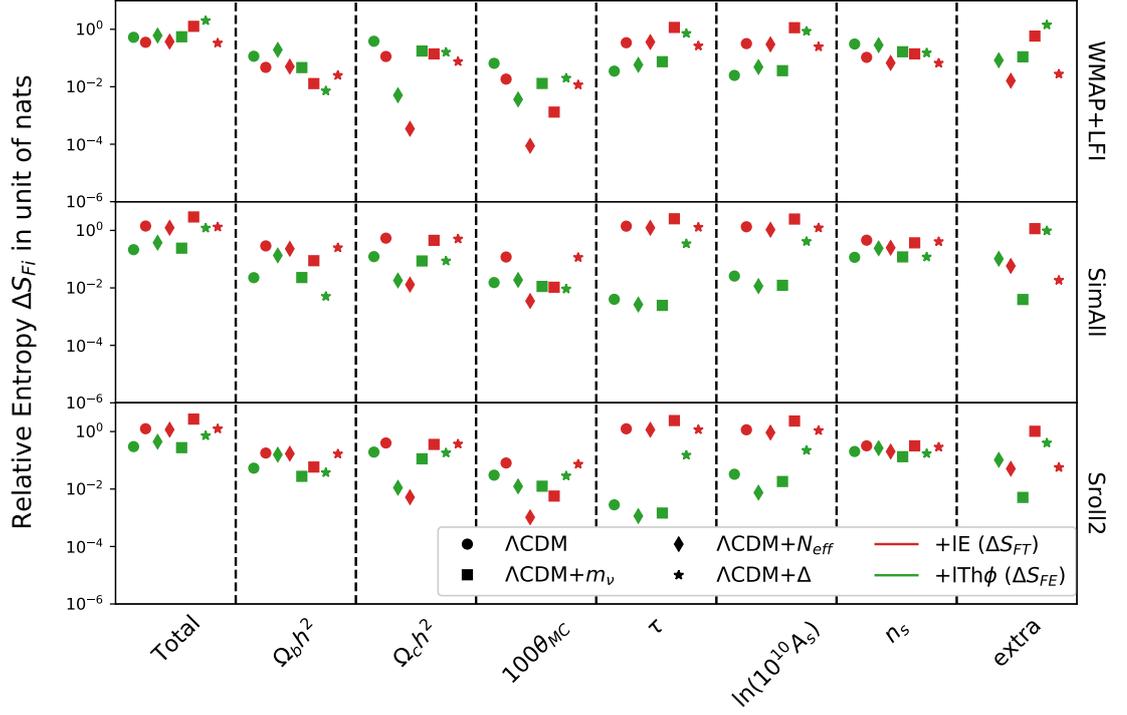


Figure 6.5: Behaviour of the Kullback-Leibler divergences computed by using the current available CMB data. The first column shows the total behaviour, while the others show the gain for each parameter obtained using the corresponding marginal distributions. The last column refers to the extra parameter of the theory.

Fig. 6.5 shows the results for WMAP+LFI, SimAll and Sroll12 datasets. The first column represents the total behaviour, while the others show the gain for each parameter obtained using the corresponding marginal distributions. The last column refers to the extra parameter of the theory. Focussing the attention on the first column, we can see that the global behaviour founded with the basic computation of the Shannon entropy has recovered. There are only small differences due to the more complex structure of the Kullback-Leibler divergence, that depends on shifts between parameters, change in degeneracy directions, and overall change in the volume of parameter space. The total behaviour of ΔS is quite similar for SimAll and Sroll12. The results for each considered model, except for $\Lambda\text{CDM}+\Delta$, highlight the highest gain obtained when IE data has added. This effect is mainly driven by the constraints on τ given by large-scale polarization data. It is known that the small-scale data prefers a higher A_s , thus using only the lThTE ϕ likelihood push the value of τ towards higher value widening also the respective posterior. Furthermore, since the optical depth affects the relative power between large-scales and intermediate and small-scales (that have their power suppressed by $e^{-2\tau}$), there is a partial degeneracy with n_s . A variation of n_s can affects the relative heights of the first few peaks making this parameter partially degenerate with $\Omega_b h^2$. Thus IE measurements are of crucial importance to break all these internal degeneracies of the model. The $\Lambda\text{CDM}+\Delta$

model shows a similar behaviour. However, due to the partial degeneracy between τ and Δ , the constraint on the optical depth becomes weaker. The overall result is a shrinking of the difference between τ from lEhTE and lThTE ϕ . It is worth notice that, as pointed out in [Planck Collaboration VI \(2018\)](#), the stringent constraint on τ coming from large-scale polarization measurements pushes the scalar amplitude, A_s , through lower values to match the high- ℓ value. This reduction of parameter spaces allows only lowest value for the masses of neutrino, giving, as a result, more stringent constraints using only the lEhTE likelihood. The same behaviour is not seen by forecast since there is no mismatch between high- ℓ and low- ℓ measurements. The impact of the two probes on the estimation of extra parameter for WMAP+LFI remains the same. However, the overall situation reflects both the weakest constraint on the reionization optical depth and the preference for this dataset towards higher value of τ . This effect makes the lTh ϕ dataset more suited for giving constraints on cosmological parameters. The only exception remains the Λ CDM+ m_ν model, in which the neutrino mass constraint plays a crucial role.

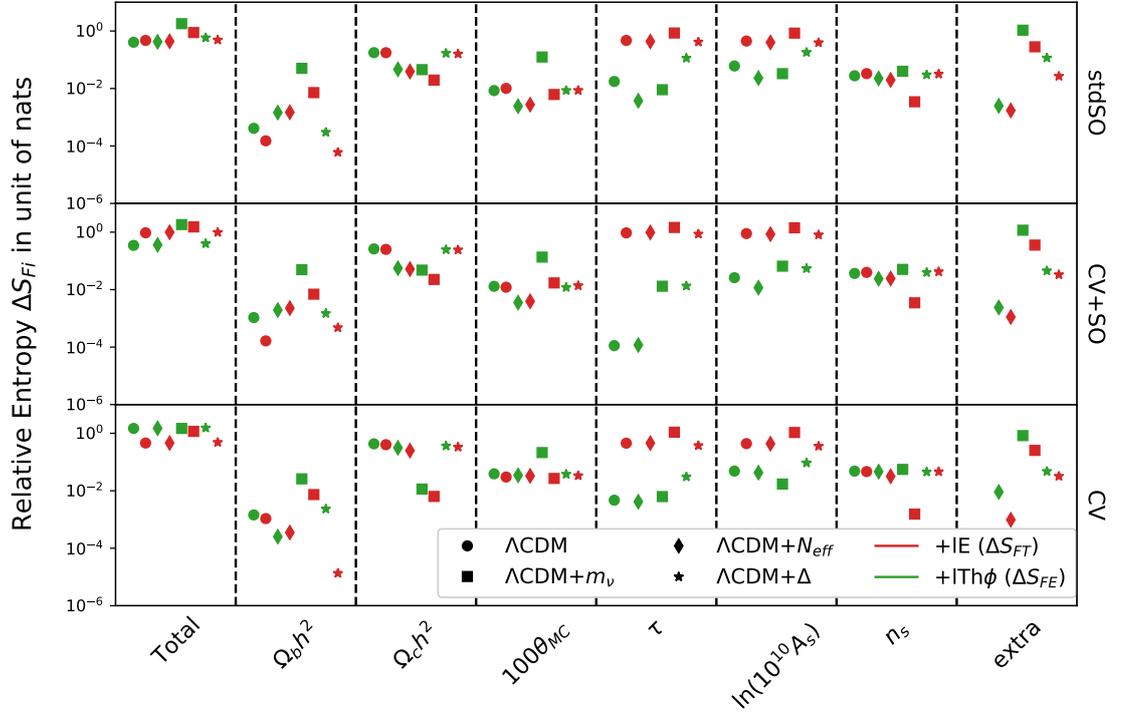


Figure 6.6: Behaviour of the Kullback-Leibler divergences computed by using the forecasts datasets. The first column shows the total behaviour, while the others show the gain for each parameter obtained using the corresponding marginal distributions. The last column refers to the extra parameter of the theory.

Fig. 6.6 shows the results for the mock datasets. As we can see from the last column, the lTh ϕ data helps us to better constraint the extended parameters. The results in the Λ CDM+ m_ν framework show how lensing is crucial in disentangling the degeneracy between m_ν and the angular size at the time of last-scattering, θ_* . Once θ_* is fixed, the parameter space with larger neutrino masses is reduced, giving tighter constraints on the m_ν . The large-scale polarization power spectrum, instead, is not very sensitive to a change of θ_* , making this effect unchanged for all the configuration setup employed. This result allows for an higher tail in the posterior of m_ν estimated through the lE dataset. As compensation, since this dataset knows nothing about a change of the relative heights of the even and odd acoustic peaks, there is a mild preference toward lower values of $\Omega_b h^2$. The mild degeneration with n_s also push the scalar spectral index toward lower

values. This highlights in which way the lensing measurements are affecting the neutrino mass estimation. For the other models, there is no evidence of effects induced on single parameter employing the two different probes. The total behaviour has then driven by the change in the degeneracies directions, jointly with the information on the extra parameter. The only two exceptions are the $\Omega_b h^2$ and n_s estimates in the $\Lambda\text{CDM}+\Delta$ and $\Lambda\text{CDM}+N_{eff}$ models, respectively. When Δ is included in the model, the value of n_s remains stable when estimated using lThTE ϕ or lEhTE. This effect follows from the fact that Δ change the position of the reionization peak, as well as dumping it. Varying n_s , instead, has as a consequence a change in the tilt. Thus the two effects are mostly uncorrelated. However, the shift of the reionization peak makes the large-scale polarization estimate partially degenerate with $\Omega_b h^2$, since, as stated before, the low- ℓ is poorly affected by a change of the relative heights of the even and odd acoustic peaks. This effect is not seen in the lThTE ϕ dataset due to the presence of lensing. In the $\Lambda\text{CDM}+N_{eff}$ framework, the degeneracy between N_{eff} and n_s makes the lensing more informative in disentangling this effect.

6.3.1 The effect induced by the presence of a missing parameter

In this section, we are going to explore the effect induced by the presence of a missing parameter. This analysis completes the information extrapolated previously by estimating how the extra parameter affects the base- ΛCDM estimates and which is the responsible probe of it. We use the mock datasets built in the ΛCDM , $\Lambda\text{CDM}+\Delta$, $\Lambda\text{CDM}+m_\nu$, and

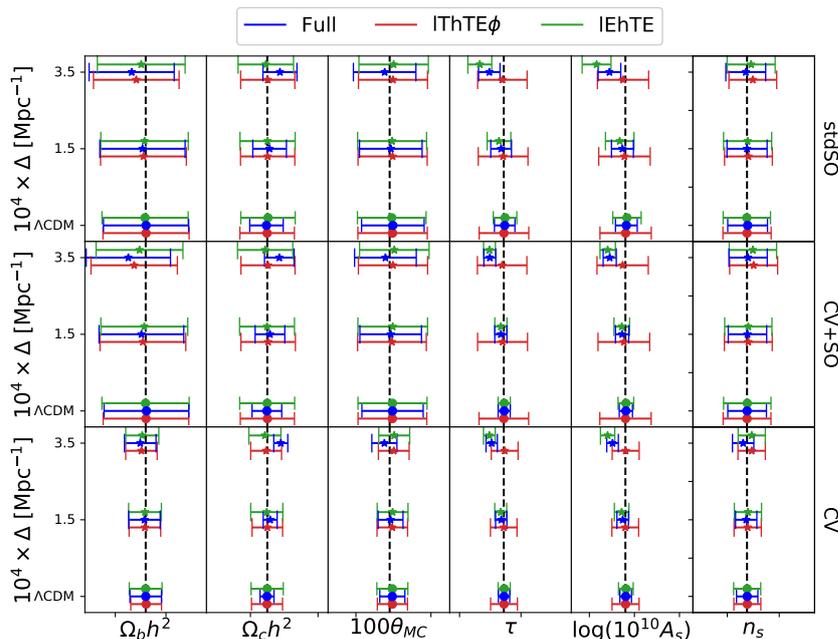


Figure 6.7: Behaviour of the base- ΛCDM parameters estimated imposing $\Lambda\text{CDM}+\Delta$ as theory and ΛCDM as model. The black dotted lines represent the fiducial values of ΛCDM parameters.

$\Lambda\text{CDM}+N_{eff}$ frameworks and described in Sec. 6.1. For each theory we estimate the base- ΛCDM parameters, assuming ΛCDM as a model. Again, here we consider the full, lThTE ϕ , and lEhTE likelihoods. Note that, each probe (lensing or polarization) is affected in a different way, depending on the extended model considered. Since other parameters can mimic the same effect, this method allow us to spot which observable has to be better measured to break these degeneracies.

Figures 6.7-6.9 show the results of these estimates. Each parameter has a peculiar fingerprint. The presence of an unaccounted Δ -like parameter in the model, for example,

does not affect the values of τ and A_s estimated when the lensing likelihood is used to break the $A_s e^{-2\tau}$ degeneracy, see Fig. 6.7. However, there is a mild preference towards higher values of n_s to compensate the extra-tilt induced by Δ . This effect pushes θ_* and $\Omega_b h^2$ toward higher and lower values, respectively. When the full likelihood has considered, the A_s estimate has dragged by lensing measure. To compensate for this effect, a higher value of $\Omega_c h^2$ has preferred. As a consequence, θ_* results lower than the best-fit value.

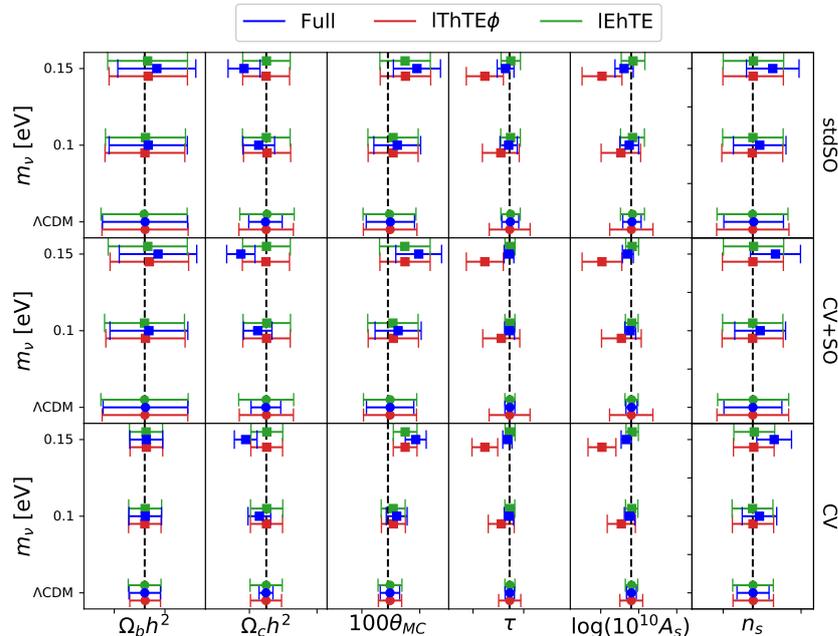


Figure 6.8: Behaviour of the base- Λ CDM parameters estimated imposing Λ CDM+ m_ν as theory and Λ CDM as model. The black dotted lines represent the fiducial values of Λ CDM parameters.

The presence of an unaccounted m_ν -like parameter in the model, instead, has a different effect on the base- Λ CDM parameters, see Fig. 6.8. The main effect is to compensate for the presence of $m_\nu > 0.06$ eV by lowering of A_s estimated using the lThTE ϕ likelihood. The large-scale polarization dataset is mostly insensitive to the presence of m_ν leaving the τ estimate unaltered. Both the likelihood employed show a preference toward a higher value of θ_* . When the full likelihood has considered, the information on τ coming from large-scale polarization measurement dominates, dragging both τ and A_s toward their fiducial values. This effect has compensated by a lowering of $\Omega_c h^2$ and n_s . To remain in a high probability region of parameter space, the value of θ_* increase further.

Assuming Λ CDM as the model with Λ CDM+ N_{eff} as the underlying theory has the most dramatic impact on the six base parameters, see Fig. 6.9. The presence of an unaccounted N_{eff} -like parameter changes the relative height of the acoustic peaks. This effect makes N_{eff} partially degenerate with $\Omega_b h^2$ and $\Omega_c h^2$. It is worth to notice that the constraining power in terms of $\Omega_c h^2$ estimate of the lensing likelihood is only poorly higher than the lE dataset. The value of $\Omega_b h^2$, instead, results slightly more affected by the presence of N_{eff} in the lThTE ϕ dataset. This effect is driven by the degeneracy with n_s , and also in this case the lensing estimate prefer values slightly far away from the fiducial value. The major difference between the two datasets relies on the effect induced on A_s , and then τ .

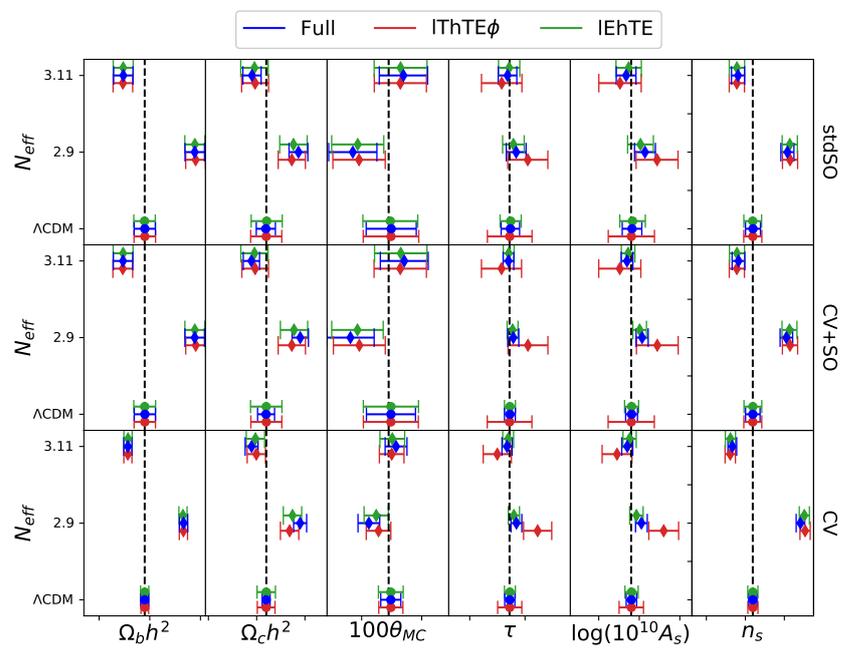


Figure 6.9: Behaviour of the base- Λ CDM parameters estimated imposing Λ CDM+ N_{eff} as theory and Λ CDM as model. The black dotted lines represent the fiducial values of Λ CDM parameters.

Conclusions

In this thesis, we have studied three application of statistical analyses techniques to CMB data. We have used Bayesian, frequentist, and information theory approaches.

We start describing the basics of the currently-accepted Standard Model of Cosmology in Chapter 1. We describe the homogeneous background parametrization, reviewing the thermal history of the Universe and highlighting the connections between inflation and the CMB radiation. In Chapter 2, we move our attention to the CMB, which is the main observable used in this thesis. We characterize both temperature and polarization anisotropies, briefly reviewing the fingerprint of reionization on CMB. Before presenting the original part of the work, in Chapter 3, we review some concepts on data analysis techniques, focussing the attention on three different frameworks in which interpreting probability.

We reserve the last three chapters for the original work of this thesis. In Chapter 4, we present a novel CMB pixel-space likelihood focussed on polarization at large angular scales, whose main cosmological target is the optical depth to reionization, τ . The underlying dataset combines foreground-mitigated WMAP Ka, Q, and V bands with *Planck* LFI 70 GHz channel in an optimally weighted CMB map. In the foreground cleaning of WMAP bands, we adopt the *Planck* 353 GHz channel as a dust template, instead of the WMAP dust model based on starlight-derived polarization directions. As a synchrotron template, the K band is used for WMAP channels, while *Planck* 30 GHz is used for the 70 GHz map. The corresponding covariance matrix is computed coherently and fed, together with the cleaned CMB map, into a pixel space likelihood, made publicly available. We produced a set of masks with increasing sky fraction and used them to test the performance of the component separation, the quality of polarization power spectra, and the overall stability of τ constraints, showing a remarkable stability among sky fractions.

For the baseline dataset, which retains 54% of the sky, the ℓ -by- ℓ probability to exceed the χ^2 of the measured angular power spectra (PTE) does not show any major outlier, with only $\ell = 18$ and 23 of BB and $\ell = 23$ of EB spectra at more than $2.5\text{-}\sigma$. Consequently, the integrated PTEs are perfectly consistent with simulations both on the reionization peak only (i.e., $\ell = 2 \div 10$) and on the full multipole range (i.e., $\ell = 2 \div 29$).

Regarding the reionization optical depth estimation, we compared the variation of τ estimated on different sky fractions with a Montecarlo of signal plus noise, finding no significant deviations for the baseline dataset compared with other sky fractions, up to $f_{\text{sky}} \sim 70\%$.

Sampling the parameter space with our low- ℓ likelihood only, we find $\tau = 0.069_{-0.011}^{+0.012}$. When CMB small scales, BAO observations, and *Planck* lensing likelihood are included, we shrink optical depth constraint down to $\tau = 0.0714_{-0.0096}^{+0.0087}$. Such bounds are slightly less constraining when compared with the existing *Planck* HFI based likelihood (see, e.g., [Planck Collaboration V 2019](#); [Pagano et al. 2019](#), and our Fig. 4.10), yet they represent a novel measurement obtained with an independent pipeline that adopts different data and likelihood approximation and includes TE correlations, while the *Planck* HFI estimates are currently restricted to EE information. The τ estimates obtained with the likelihood package discussed in this chapter is, in general, very compatible with the *Planck* HFI based

constraints, with a preference for slightly higher τ values, probably driven by the inclusion of TE. This bound is also in perfect agreement with the *Planck* LFI Legacy likelihood (Planck Collaboration V 2019). Within the Λ CDM model, τ can be also constrained without the use of polarization data to break the degeneracy between A_s and τ , combining temperature and weak lensing data. In this kind of analysis, the *Planck* collaboration found $\tau = 0.080 \pm 0.025$ from temperature and lensing data and $\tau = 0.078 \pm 0.016$ when BAO is added (Planck Collaboration ES 2018). Those values, despite being slightly higher than previous findings (Weiland et al. 2018; Planck Collaboration XIII 2016), are still in agreement with our constraints.

The likelihood package we provided is built on a real-space estimator and does not assume rotational invariance, but, rather, only the Gaussianity of the fields, which secures several advantages. For instance, it can also be easily used for constraining non-rotationally invariant cosmologies, including, naturally, the TB and EB spectra in the parameter exploration. Furthermore, it allows to obtain an independent estimate of the reionization optical depth that can be used for different purpose. It can be used to test extended reionization model since it provides an optimal dataset for robustness test due to the combination of all the channels, TT, TE, EE, and BB (see, e.g. Paoletti et al. 2020). For the same reason, it can be used to add complementary information in the study of some CMB anomalies (see, e.g. Chiocchetta et al. 2020).

In Chapter 5, we analyse the lack-of-power anomaly, a well known characteristic of the CMB temperature anisotropy pattern showing up at large angular scales. In particular, we focus on the intriguing fact that this feature is statistically more significant (at a $\sim 3\sigma$) when only high Galactic latitude data are taken into account. The latter observations suggests that most of the large scale anisotropy power happens to be mainly localised around the Galactic plane. This might sound bizarre because the early universe should not know anything about the “direction” of the disk of our Galaxy. To tackle the issue, we evaluated how often a Λ CDM realisation happens to have most of its power localised at low Galactic latitude.

To support the analysis, we generate a Λ CDM Monte Carlo set of 10^5 CMB maps from the *Planck* 2018 best-fit model. By analysing this set, we first show that the *Planck* 2018 data exhibits the same trend of decreasing CMB field variance while increasing the Galactic mask, which was found previously in the literature. We then proceed to randomly rotate the simulated maps (denoted as *ensemble 0*), as well as the data, 10^3 times. The rotated maps are employed to compute the empirical distribution function of two estimators, based on the CMB field variance (Section 5.2). With the LTP-estimator (Section 5.2.1) we test to what extent the low CMB anisotropy power in the data depends on the orientation of the Galactic plane. With the r -estimator (Section 5.2.1) we assess instead the behaviour against rotation of the decreasing trend of the CMB variance at increasing Galactic latitude. The introduction of random rotations is a key-element to evaluate whether the lack of power anomaly is indeed correlated with Galactic latitude.

To further investigate this behaviour we also select from the 10^5 Λ CDM simulations set a smaller set, of 10^3 maps, which exhibits the same low-variance as the one observed in the *Commander* and *SMICA* 2018 maps. We call this set *ensemble 1* and repeat the analyses performed on the *ensemble 0*.

We find that even when performing random rotations, our CMB sky is anomalous in power at about $2.8 - 2.5\sigma$ depending on the considered component separation method when employing the LTP estimator. Specifically, only 5 maps out of 10^5 have a LTP at high Galactic latitude (in the *Ext₃₀* mask) smaller than the *Planck Commander* data. For the r -estimator we evaluate that only the 0.2% of the maps show a larger value of r between *Std 2018* and *Ext₃₀* masks, again with respect to *Commander*. Results are substantially stable if we employ *SMICA* in place of *Commander*. Finally, using the low-variance constrained

simulation of *ensemble 1* yields similar results, showing that having a low-variance field in the first place is not enough to justify the observed trend with Galactic latitude.

In conclusion, the introduction of rotations do not spoil the lack of power anomaly at high Galactic latitude which turns out to be quite stable against the “look-elsewhere effect” spawned by random rotations of the reference frame.

In Chapter 6, we focus on the comparison between the information content carried separately by lensing and large-scale polarization measurements. We establish the effects induced by an improvement of the experimental setup on Λ CDM, Λ CDM+ Δ , Λ CDM+ m_ν , and Λ CDM+ N_{eff} parameters. For each dataset employed, we build two different likelihoods. The first combines the EE power spectrum in the low-multipole range with the TT, TE and EE power spectra in the high-multipole range (lEhTE). The second combines the TT power spectrum in the low-multipole range with the TT, TE, EE power spectra plus lensing potential in the high-multipole range (lThTE ϕ). Both these likelihoods are capable of break the internal degeneracy between A_s and τ , then providing two independent datasets that can be used to quantify the information content carried by the main CMB observables. We compare them using an information theory-based approach as the main framework.

We find that, in a Λ CDM+ Δ framework, even for a CV-limited experiment on all scales, the minimum value of Δ detectable at more than 3σ is $\Delta = 2.5 \times 10^{-4} \text{ Mpc}^{-1}$. This effect is an intrinsic problem of the model. Indeed, due to the cosmic-variance, low values of Δ can be reabsorbed varying both τ and A_s . Furthermore, in this model there are no strong degeneracies between Δ and the other Λ CDM parameters, with the exception of τ . We show that, even if the lThTE ϕ likelihood carried the highest information content, this is due only to the changes in degeneracies directions. Indeed, the measurement of τ coming from large-scale EE spectrum helps to break the degeneracy with Δ . The difference between the two probes in constraining this parameter is partially cancelled when a CV setup is employed in the low-multipole range. This effect is highlighted when we do not include the presence of Δ in the model. That is, by assuming a Λ CDM model with an underlying theory which includes Δ , we find that the presence of this characteristic scale does not affect the Λ CDM parameter estimated using the lThTE ϕ likelihood.

In a Λ CDM+ m_ν framework, we show how the lensing measurement results crucial in disentangling the degeneracy between θ_* and m_ν . Indeed, the low- ℓ polarization power spectrum is slightly sensitive to a shift of the acoustic peak. This allows the model to remain in a high probability region of the parameter space even for lower values of $\Omega_b h^2$ and n_s . This is a characteristic fingerprint of the neutrino mass. If we have a theory with a massive neutrino, but we assume a Λ CDM as model, both lThTE ϕ and lEhTE datasets prefer higher value for θ_* . This result highlight the importance of lensing measurement to fix the value of the angular size at the time of last-scattering.

The presence of N_{eff} has a non-trivial effect on all Λ CDM parameters. Even in this case, as for m_ν , the most informative probe is lensing. We show that this behaviour is driven by the degeneracy with n_s , which is partially degenerate with $\Omega_b h^2$. This effect has remarked when we have the unaccounted presence of N_{eff} in the model. Indeed, in this case, the estimate of n_s and $\Omega_b h^2$ result more shifted from the fiducial value, showing a larger bias with respect the lEhTE dataset estimates.

In conclusion, our results spoil the induced effects on parameter estimation by considering both lensing and large-scale polarization measurements separately. In particular, we quantify the importance of improving lensing measurement to constrain some characteristic features of some of the most debated cosmological models, such as Λ CDM + m_ν and Λ CDM + N_{eff} . We also show how future large-scale polarization measurement can increase our understanding of more exotic scenarios, such as Λ CDM+ Δ . This leads to a more competitive need to improve both the observables at our disposal.

Acknowledgements

I am greatly indebted to my supervisor Paolo Natoli for his support and guidance. I can only express admiration for him, both as an advisor and as a man. He always knew how to find the right words, even in the difficult moments of this journey. This thesis, without him, it would not exist.

I also thank Alessandro Gruppuso. He helped me to move the first steps in the world of research.

I would like to reserve a special thanks to Luca Pagano. He taught me a lot by guiding me, step by step, into the intriguing world of data analysis.

I consider both of them “co-supervisors ad honorem”. Thanks for providing guidance and feedback throughout this project, and on each step I made during these three years.

Thank you to Massimiliano Lattanzi and Martina Gerbino. It was always helpful to talk with them about different aspects of physics.

All of them have always advised me on every aspect of my professional growth. Thanks for never stopped motivating me with new questions and problems. I am very glad to have the opportunity to work with them.

Thank you to Margherita for its moral support, for its ability to bear with me, and for making this experience so much fun.

Thank you to my father for always believe in me and for its support during these years.

Finally, this dissertation would not have been possible without the love, encouragement, patience and tolerance of Mariapia. She was always present during the happiest and hardest moments of life.

Bibliography

- Adachi, S. et al. (2020). A measurement of the CMB E-mode angular power spectrum at subdegree scales from 670 square degrees of POLARBEAR data. *Astrophys. J.*, 904(1):65.
- Ade, P. et al. (2017). A Measurement of the Cosmic Microwave Background *B*-Mode Polarization Power Spectrum at Sub-Degree Scales from 2 years of POLARBEAR Data. *Astrophys. J.*, 848(2):121.
- Ade, P. et al. (2018a). BICEP2 / Keck Array x: Constraints on Primordial Gravitational Waves using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season. *Phys. Rev. Lett.*, 121:221301.
- Ade, P. et al. (2019). The Simons Observatory: Science goals and forecasts. *JCAP*, 1902:056.
- Ade, P. A. R., Ahmed, Z., Aikin, R. W., Alexander, K. D., Barkats, D., Benton, S. J., Bischoff, C. A., Bock, J. J., Bowens-Rubin, R., Brevik, J. A., Buder, I., Bullock, E., Buza, V., Connors, J., Cornelison, J., Crill, B. P., Crumrine, M., Dierickx, M., Duband, L., Dvorkin, C., Filippini, J. P., Fliescher, S., Grayson, J., Hall, G., Halpern, M., Harrison, S., Hildebrandt, S. R., Hilton, G. C., Hui, H., Irwin, K. D., Kang, J., Karkare, K. S., Karpel, E., Kaufman, J. P., Keating, B. G., Kefeli, S., Kernasovskiy, S. A., Kovac, J. M., Kuo, C. L., Larsen, N. A., Lau, K., Leitch, E. M., Lueker, M., Megerian, K. G., Moncelsi, L., Namikawa, T., Netterfield, C. B., Nguyen, H. T., O'Brient, R., Ogburn, R. W., Palladino, S., Pryke, C., Racine, B., Richter, S., Schillaci, A., Schwarz, R., Sheehy, C. D., Soliman, A., St. Germaine, T., Staniszewski, Z. K., Steinbach, B., Sudiwala, R. V., Teply, G. P., Thompson, K. L., Tolán, J. E., Tucker, C., Turner, A. D., Umiltà, C., Vieregg, A. G., Wandui, A., Weber, A. C., Wiebe, D. V., Willmert, J., Wong, C. L., Wu, W. L. K., Yang, H., Yoon, K. W., and Zhang, C. (2018b). Constraints on primordial gravitational waves using *planck*, *wmap*, and new *bicep2/keck* observations through the 2015 season. *Phys. Rev. Lett.*, 121:221301.
- Alam, S. et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *Mon. Not. Roy. Astron. Soc.*, 470(3):2617–2652.
- Amara, A. and Refregier, A. (2013). Model breaking measure for cosmological surveys. *Physical Review D*, 89.
- Baumann, D. (2011). Inflation. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small*, pages 523–686.
- Benabed, K., Cardoso, J. F., Prunet, S., and Hivon, E. (2009). TEASING: a fast and accurate approximation for the low multipole likelihood of the Cosmic Microwave Background temperature. *Mon. Not. Roy. Astron. Soc.*, 400:219.

- Bennett, C. L., Larson, D., Weiland, J. L., Jarosik, N., Hinshaw, G., Odegard, N., Smith, K. M., Hill, R. S., Gold, B., Halpern, M., Komatsu, E., Nolte, M. R., Page, L., Spergel, D. N., Wollack, E., Dunkley, J., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., and Wright, E. L. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *ApJ Supp.*, 208:20.
- Beutler, F., Blake, C., Colless, M., Jones, D. H., Staveley-Smith, L., Campbell, L., Parker, Q., Saunders, W., and Watson, F. (2011). The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant. *Mon. Not. Roy. Astron. Soc.*, 416:3017–3032.
- Bunn, E. F. and Bourdon, A. (2008). Contamination cannot explain the lack of large-scale power in the cosmic microwave background radiation. *Phys. Rev. D*, 78:123509.
- Carroll, S. M. (2019). *Spacetime and Geometry: An Introduction to General Relativity*. Cambridge University Press.
- Carron, J., Amara, A., and Lilly, S. J. (2011). Probe combination in large galaxy surveys: application of Fisher information and Shannon entropy to weak lensing. *Monthly Notices of the Royal Astronomical Society*, 417(3):1938–1951.
- Caticha, A. (2008). Lectures on Probability, Entropy, and Statistical Physics.
- Chen, X. (2010). Primordial Non-Gaussianities from Inflation Models. *Adv. Astron.*, 2010:638979.
- Chiocchetta, C., Gruppuso, A., Lattanzi, M., Natoli, P., and Pagano, L. (2020). Lack-of-correlation anomaly in CMB large scale polarisation maps.
- Choi, S. K. et al. (2020). The Atacama Cosmology Telescope: A Measurement of the Cosmic Microwave Background Power Spectra at 98 and 150 GHz.
- Copi, C., Huterer, D., Schwarz, D., and Starkman, G. (2007). The Uncorrelated Universe: Statistical Anisotropy and the Vanishing Angular Correlation Function in WMAP Years 1-3. *Phys. Rev. D*, 75:023507.
- Copi, C. J., Huterer, D., Schwarz, D. J., and Starkman, G. D. (2009). No large-angle correlations on the non-Galactic microwave sky. *Mon. Not. Roy. Astron. Soc.*, 399:295–303.
- Copi, C. J., Huterer, D., Schwarz, D. J., and Starkman, G. D. (2010). Large angle anomalies in the CMB. *Adv. Astron.*, 2010:847541.
- Cox, R. T. (1946). Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14(2):1–13.
- Cruz, M., Vielva, P., Martinez-Gonzalez, E., and Barreiro, R. (2011). Anomalous variance in the WMAP data and Galactic Foreground residuals. *Mon. Not. Roy. Astron. Soc.*, 412:2383.
- Dai, W.-M., Ma, Y.-Z., Guo, Z.-K., and Cai, R.-G. (2019). Constraining the reionization history with CMB and spectroscopic observations. *Phys. Rev.*, D99(4):043524.
- Delouis, J. M., Pagano, L., Mottet, S., Puget, J. L., and Vibert, L. (2019). SRoll2: an improved mapmaking approach to reduce large-scale systematic effects in the Planck High Frequency Instrument legacy maps. *Astron. Astrophys.*, 629:A38.

- Dicke, R., Peebles, P., Roll, P., and Wilkinson, D. (1965). Cosmic Black-Body Radiation. *Astrophys. J.*, 142:414–419.
- Dodelson, S. (2003). *Modern Cosmology*. Academic Press, Amsterdam.
- Durrer, R. (2008). *The Cosmic Microwave Background*. Cambridge University Press.
- Ellis, G. F. R., Maartens, R., and MacCallum, M. A. H. (2012). *Relativistic Cosmology*. Cambridge University Press.
- Fixsen, D. (2009). The Temperature of the Cosmic Microwave Background. *Astrophys. J.*, 707:916–920.
- Fixsen, D., Cheng, E., Gales, J., Mather, J. C., Shafer, R., and Wright, E. (1996). The Cosmic Microwave Background spectrum from the full COBE FIRAS data set. *Astrophys. J.*, 473:576.
- Fixsen, D. et al. (1994). Cosmic microwave background dipole spectrum measured by the COBE FIRAS. *Astrophys. J.*, 420:445.
- Frieden, R. and Gatenby, R. A. (2006). *Exploratory Data Analysis Using Fisher Information*. Springer-Verlag, Berlin, Heidelberg.
- Fuskeland, U., Wehus, I. K., Eriksen, H. K., and Næss, S. K. (2014). Spatial variations in the spectral index of polarized synchrotron emission in the 9-yr WMAP sky maps. *Astrophys. J.*, 790:104.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.
- George, E. et al. (2015). A measurement of secondary cosmic microwave background anisotropies from the 2500-square-degree SPT-SZ survey. *Astrophys. J.*, 799(2):177.
- Gerbino, M., Lattanzi, M., Migliaccio, M., Pagano, L., Salvati, L., Colombo, L., Gruppuso, A., Natoli, P., and Polenta, G. (2020). Likelihood methods for CMB experiments. *Front. in Phys.*, 8:15.
- Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations*. The Johns Hopkins University Press, third edition.
- Gorski, K., Hivon, E., Banday, A., Wandelt, B., Hansen, F., Reinecke, M., and Bartelman, M. (2005). HEALPix - A Framework for high resolution discretization, and fast analysis of data distributed on the sphere. *Astrophys. J.*, 622:759–771.
- Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *ApJ*, 622:759–771.
- Gregory, P. (2005). *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support*. Cambridge University Press.
- Gruppuso, A. (2007). A Complete Statistical Analysis for the Quadrupole Amplitude in an Ellipsoidal Universe. *Phys. Rev. D*, 76:083010.
- Gruppuso, A., De Rosa, A., Cabella, P., Paci, F., Finelli, F., Natoli, P., de Gasperis, G., and Mandolesi, N. (2009). New estimates of the CMB angular power spectra from the WMAP 5 yrs low resolution data. *Mon. Not. Roy. Astron. Soc.*, 400:463–469.

- Gruppuso, A., Finelli, F., Natoli, P., Paci, F., Cabella, P., De Rosa, A., and Mandolesi, N. (2011). New constraints on Parity Symmetry from a re-analysis of the WMAP-7 low resolution power spectra. *Mon. Not. Roy. Astron. Soc.*, 411:1445–1452.
- Gruppuso, A., Kitazawa, N., Lattanzi, M., Mandolesi, N., Natoli, P., and Sagnotti, A. (2018). The Evens and Odds of CMB Anomalies. *Phys. Dark Univ.*, 20:49–64.
- Gruppuso, A., Kitazawa, N., Mandolesi, N., Natoli, P., and Sagnotti, A. (2016). Pre-Inflationary Relics in the CMB? *Phys. Dark Univ.*, 11:68–73.
- Gruppuso, A., Natoli, P., Paci, F., Finelli, F., Molinari, D., De Rosa, A., and Mandolesi, N. (2013). Low Variance at large scales of WMAP 9 year data. *JCAP*, 07:047.
- Gruppuso, A. and Sagnotti, A. (2015). Observational Hints of a Pre-Inflationary Scale? *Int. J. Mod. Phys. D*, 24(12):1544008.
- Guth, A. H. (1987). The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. *Adv. Ser. Astrophys. Cosmol.*, 3:139–148.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.
- Hazra, D. K., Paoletti, D., Finelli, F., and Smoot, G. F. (2019). Joining bits and pieces of reionization history.
- Hazra, D. K. and Smoot, G. F. (2017). Witnessing the reionization history using Cosmic Microwave Background observation from Planck. *JCAP*, 1711(11):028.
- Heinrich, C. and Hu, W. (2018). Does *Planck* 2015 polarization data favor high redshift reionization? *Phys. Rev.*, D98(6):063514.
- Henning, J. et al. (2018a). Measurements of the Temperature and E-Mode Polarization of the CMB from 500 Square Degrees of SPTpol Data. *Astrophys. J.*, 852(2):97.
- Henning, J. W., Sayre, J. T., Reichardt, C. L., Ade, P. A. R., Anderson, A. J., Austermann, J. E., Beall, J. A., Bender, A. N., Benson, B. A., Bleem, L. E., Carlstrom, J. E., Chang, C. L., Chiang, H. C., Cho, H.-M., Citron, R., Moran, C. C., Crawford, T. M., Crites, A. T., de Haan, T., Dobbs, M. A., Everett, W., Gallicchio, J., George, E. M., Gilbert, A., Halverson, N. W., Harrington, N., Hilton, G. C., Holder, G. P., Holzappel, W. L., Hoover, S., Hou, Z., Hrubes, J. D., Huang, N., Hubmayr, J., Irwin, K. D., Keisler, R., Knox, L., Lee, A. T., Leitch, E. M., Li, D., Lowitz, A., Manzotti, A., McMahon, J. J., Meyer, S. S., Mocuano, L., Montgomery, J., Nadolski, A., Natoli, T., Nibarger, J. P., Novosad, V., Padin, S., Pryke, C., Ruhl, J. E., Saliwanchik, B. R., Schaffer, K. K., Sievers, C., Smecher, G., Stark, A. A., Story, K. T., Tucker, C., Vanderlinde, K., Veach, T., Vieira, J. D., Wang, G., Whitehorn, N., Wu, W. L. K., and Yefremenko, V. (2018b). Measurements of the temperature and e-mode polarization of the CMB from 500 square degrees of SPTpol data. *The Astrophysical Journal*, 852(2):97.
- Hinshaw, G. et al. (2013). Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. *Astrophys. J. Suppl.*, 208:19.
- Hinshaw, G., Larson, D., Komatsu, E., Spergel, D. N., Bennett, C. L., Dunkley, J., Nolte, M. R., Halpern, M., Hill, R. S., Odegard, N., Page, L., Smith, K. M., Weiland, J. L., Gold, B., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., Wollack, E., and Wright, E. L. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. *ApJ Suppl.*, 208:19.

- Hogg, D. W., Eisenstein, D. J., Blanton, M. R., Bahcall, N. A., Brinkmann, J., Gunn, J. E., and Schneider, D. P. (2005). Cosmic homogeneity demonstrated with luminous red galaxies. *Astrophys. J.*, 624:54–58.
- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.
- Kamionkowski, M., Kosowsky, A., and Stebbins, A. (1997). Statistics of cosmic microwave background polarization. *Phys. Rev. D*, 55:7368–7388.
- Kim, J. and Naselsky, P. (2010a). Anomalous parity asymmetry of the Wilkinson Microwave Anisotropy Probe power spectrum data at low multipoles. *Astrophys. J. Lett.*, 714:L265–L267.
- Kim, J. and Naselsky, P. (2010b). Anomalous parity asymmetry of WMAP power spectrum data at low multipoles: is it cosmological or systematics? *Phys. Rev. D*, 82:063002.
- Kosowsky, A., Milosavljevic, M., and Jimenez, R. (2002). Efficient cosmological parameter estimation from microwave background anisotropies. *Phys. Rev. D*, 66:063007.
- Krachmalnicoff, N., Baccigalupi, C., Aumont, J., Bersanelli, M., and Mennella, A. (2016). Characterization of foreground emission on degree angular scales for CMB B-mode observations - Thermal dust and synchrotron signal from Planck and WMAP data. *Astron. Astrophys.*, 588:A65.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86.
- Lattanzi, M., Burigana, C., Gerbino, M., Gruppuso, A., Mandolesi, N., Natoli, P., Polenta, G., Salvati, L., and Trombetti, T. (2017). On the impact of large angle CMB polarization data on cosmological parameters. *JCAP*, 1702(02):041.
- Lewis, A. (2008). Cosmological parameters from WMAP 5-year temperature maps. *Phys. Rev. D*, 78(2):023002.
- Lewis, A. and Bridle, S. (2002). Cosmological parameters from CMB and other data: A Monte Carlo approach. *Phys. Rev. D*, 66(10):103511.
- Lewis, A. and Bridle, S. (2002). Cosmological parameters from CMB and other data: A Monte Carlo approach. *Phys. Rev.*, D66:103511.
- Louis, T. et al. (2017). The Atacama Cosmology Telescope: Two-Season ACTPol Spectra and Parameters. *JCAP*, 06:031.
- Lyth, D. H. and Liddle, A. R. (2009). *The primordial density perturbation: Cosmology, inflation and the origin of structure*.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092.
- Millea, M. and Bouchet, F. (2018). Cosmic microwave background constraints in light of priors over reionization histories. *A&A*, 617:A96.
- Molinari, D., Gruppuso, A., Polenta, G., Burigana, C., De Rosa, A., Natoli, P., Finelli, F., and Paci, F. (2014). A comparison of CMB Angular Power Spectrum Estimators at Large Scales: the TT case. *Mon. Not. Roy. Astron. Soc.*, 440(2):957–964.

- Monteserin, C., Barreiro, R., Vielva, P., Martinez-Gonzalez, E., Hobson, M., and Lasenby, A. (2008). A low CMB variance in the WMAP data. *Mon. Not. Roy. Astron. Soc.*, 387:209–219.
- Muir, J., Adhikari, S., and Huterer, D. (2018). Covariance of CMB anomalies. *Phys. Rev. D*, 98(2):023521.
- Mukhanov, V. (2005). *Physical Foundations of Cosmology*. Cambridge University Press, Oxford.
- Natale, U., Gruppuso, A., Molinari, D., and Natoli, P. (2019). Is the lack of power anomaly in the CMB correlated with the orientation of the Galactic plane? *JCAP*, 12:052.
- Natale, U., Pagano, L., Lattanzi, M., Migliaccio, M., Colombo, L., Gruppuso, A., Natoli, P., and Polenta, G. (2020). A novel CMB polarization likelihood package for large angular scales built from combined WMAP and Planck LFI legacy maps. *accepted by Astron. Astrophys.*
- Nicola, A., Amara, A., and Refregier, A. (2019). Consistency tests in cosmology using relative entropy. *JCAP*, 01:011.
- Okamoto, T. and Hu, W. (2003). CMB lensing reconstruction on the full sky. *Phys. Rev.*, D67:083002.
- Pagano, L., Delouis, J. M., Mottet, S., Puget, J. L., and Vibert, L. (2019). Reionization optical depth determination from Planck HFI data with ten percent accuracy.
- Page, L., Hinshaw, G., Komatsu, E., Nolta, M. R., Spergel, D. N., Bennett, C. L., Barnes, C., Bean, R., Doré, O., Dunkley, J., Halpern, M., Hill, R. S., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Peiris, H. V., Tucker, G. S., Verde, L., Weiland, J. L., Wollack, E., and Wright, E. L. (2007). Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Polarization Analysis. *ApJ Supp.*, 170:335–376.
- Pandey, B. and Sarkar, S. (2015). Testing homogeneity in the Sloan Digital Sky Survey Data Release Twelve with Shannon entropy. *Mon. Not. Roy. Astron. Soc.*, 454(3):2647–2656.
- Paoletti, D., Hazra, D. K., Finelli, F., and Smoot, G. F. (2020). Extended reionization in models beyond Λ CDM with Planck 2018 data. *JCAP*, 09:005.
- Penzias, A. A. and Wilson, R. W. (1965). A Measurement of excess antenna temperature at 4080-Mc/s. *Astrophys. J.*, 142:419–421.
- Perotto, L., Lesgourgues, J., Hannestad, S., Tu, H., and Wong, Y. Y. Y. (2006). Probing cosmological parameters with the CMB: Forecasts from full Monte Carlo simulations. *JCAP*, 0610:013.
- Pinho, A. M., Reischke, R., Teich, M., and Schäfer, B. M. (2020). Information entropy in cosmological inference problems.
- Planck Collaboration ES (2013). *The Explanatory Supplement to the Planck 2013 results*, <https://www.cosmos.esa.int/web/planck/pla>. ESA.
- Planck Collaboration ES (2015). *The Explanatory Supplement to the Planck 2015 results*, <https://www.cosmos.esa.int/web/planck/pla>. ESA.

- Planck Collaboration ES (2018). *The Legacy Explanatory Supplement*, <https://www.cosmos.esa.int/web/planck/pla>. ESA.
- Planck Collaboration XV (2014). *Planck* 2013 results. XV. CMB power spectra and likelihood. *A&A*, 571:A15.
- Planck Collaboration XVI (2014). *Planck* 2013 results. XVI. Cosmological parameters. *A&A*, 571:A16.
- Planck Collaboration XXIII (2014). *Planck* 2013 results. XXIII. Isotropy and statistics of the CMB. *A&A*, 571:A23.
- Planck Collaboration XI (2016). *Planck* 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters. *A&A*, 594:A11.
- Planck Collaboration XIII (2016). *Planck* 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13.
- Planck Collaboration XVI (2016). *Planck* 2015 results. XVI. Isotropy and statistics of the CMB. *A&A*, 594:A16.
- Planck Collaboration I (2018). *Planck* 2018 results. I. Overview, and the cosmological legacy of *Planck*. *A&A*, *submitted*.
- Planck Collaboration II (2018). *Planck* 2018 results. II. Low Frequency Instrument data processing. *A&A*, *in press*.
- Planck Collaboration III (2018). *Planck* 2018 results. III. High Frequency Instrument data processing. *A&A*, *in press*.
- Planck Collaboration IV (2018). *Planck* 2018 results. IV. Diffuse component separation. *A&A*, *in press*.
- Planck Collaboration V (2019). *Planck* 2018 results. V. Power spectra and likelihoods. *A&A*, *submitted*.
- Planck Collaboration VI (2018). *Planck* 2018 results. VI. Cosmological parameters. *A&A*, *submitted*.
- Planck Collaboration VII (2018). *Planck* 2018 results. VII. Isotropy and statistics. *A&A*, *in press*.
- Planck Collaboration VIII (2018). *Planck* 2018 results. VIII. Gravitational lensing. *A&A*, *in press*.
- Planck Collaboration Int. XLVII (2016). *Planck* intermediate results. XLVII. Constraints on reionization history. *A&A*, 596:A108.
- Reichardt, C. et al. (2020). An Improved Measurement of the Secondary Cosmic Microwave Background Anisotropies from the SPT-SZ + SPTpol Surveys.
- Ross, A. J., Samushia, L., Howlett, C., Percival, W. J., Burden, A., and Manera, M. (2015). The clustering of the SDSS DR7 main Galaxy sample – I. A 4 per cent distance measure at $z = 0.15$. *Mon. Not. Roy. Astron. Soc.*, 449(1):835–847.
- Ross, S. (1998). *A First Course in Probability*. Prentice Hall International Editions. Prentice Hall.

- Sayre, J. et al. (2020). Measurements of B-mode Polarization of the Cosmic Microwave Background from 500 Square Degrees of SPTpol Data. *Phys. Rev. D*, 101(12):122003.
- Schwarz, D. J., Copi, C. J., Huterer, D., and Starkman, G. D. (2016). CMB Anomalies after Planck. *Class. Quant. Grav.*, 33(18):184001.
- Seehars, S., Amara, A., Refregier, A., Paranjape, A., and Akeret, J. (2014). Information Gains from Cosmic Microwave Background Experiments. *Phys. Rev. D*, 90(2):023533.
- Seehars, S., Grandis, S., Amara, A., and Refregier, A. (2016). Quantifying Concordance in Cosmology. *Phys. Rev. D*, 93(10):103507.
- Seljak, U. and Zaldarriaga, M. (1996). A Line of sight integration approach to cosmic microwave background anisotropies. *Astrophys. J.*, 469:437–444.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423.
- Tegmark, M. (1996). A method for extracting maximum resolution power spectra from microwave sky maps. *MNRAS*, 280:299–308.
- Tegmark, M. (1997). How to measure CMB power spectra without losing information. *Phys. Rev. D*, 55:5895–5907.
- Tegmark, M. and de Oliveira-Costa, A. (2001). How to measure CMB polarization power spectra without losing information. *Phys. Rev. D*, 64:063001.
- Trotta, R. (2008). Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemp. Phys.*, 49:71–104.
- Trotta, R. (2017). Bayesian Methods in Cosmology.
- Villanueva-Domingo, P., Gariazzo, S., Gnedin, N. Y., and Mena, O. (2018). Was there an early reionization component in our universe? *JCAP*, 1804(04):024.
- Weiland, J., Osumi, K., Addison, G., Bennett, C., Watts, D., Halpern, M., and Hinshaw, G. (2018). Effect of Template Uncertainties on the WMAP and Planck Measures of the Optical Depth Due To Reionization. *Astrophys. J.*, 863:161.
- Zaldarriaga, M. (1997). Polarization of the microwave background in reionized models. *Phys. Rev. D*, 55:1822–1829.
- Zaldarriaga, M. and Harari, D. D. (1995). Analytic approach to the polarization of the cosmic microwave background in flat and open universes. *Phys. Rev. D*, 52:3276–3287.