SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI

---

dissertation submitted for the degree of
“doctor philosophiæ”
in astrophysics and cosmology

Analyzing large-scale polarization data for next-generation CMB experiments

candidate:
Kevin Wolz

supervisors:
Dr. Nicoletta Krachmalnicoff
Prof. Carlo Baccigalupi

Academic Year 2022-2023

Kevin Wolz: *Analyzing large-scale polarization data for next-generation CMB experiments*

E-mail: kevin.wolz93@gmail.com

TABLE OF CONTENTS

LIST OF ACRONYMS

ACT: Atacama Cosmology Telescope (experiment)

ADCNL: Analog-to-Digital Converter Nonlinearity

AIC: Akaike Information Criterion

BAO: Baryonic Acoustic Oscillations

BIC: Bayesian Information Criterion

BICEP: Background Imaging of Cosmic Extragalactic Polarization (experiment)

CAMB: Code for Anisotropies in the Cosmic Microwave Background (software)

CDM: Cold Dark Matter

CIB: Cosmic Infrared Background

CL: Credible Level

CLASS: Cosmic Linear Anisotropy Solving System (software)

CMB: Cosmic Microwave Background

CNN: Convolutional Neural Network

COBE: Cosmic Background Explorer (experiment)

CRHWP: Cryogenic Rotating Half-Wave Plate

CTIO: Cerro Tololo Inter-American Observatory (experiment)

DES: Dark Energy Survey (experiment)

DESI: Dark Energy Spectroscopic Survey (experiment)

FFP: Full Focal Plane (algorithm)

FGBuster: ForeGroundBuster (software)

FLRW: *Friedmann-Lemaître-Robertson-Walker* (spacetime metric)

FWHM: Full Width at Half Maximum

GMF: Galactic Magnetic Field

GR: General Relativity

HEALPix: Hierarchical Equal Area isoLatitude Pixelisation of a 2-sphere (algorithm)

HFI: High Frequency Instrument (*Planck* experiment)

HL likelihood: *Hamimeche-Lewis* likelihood

HSC: Hyper Suprime-Cam (experiment)

HWPSS: Half-Wave Plate Synchronous Signal

ISW: Integrated Sachs-Wolfe effect

KiDS: Kilo-Degree Survey (experiment)

kSZ: kinetic Sunyaev-Zel'dovich effect

Λ**CDM:** Λ Cold Dark Matter

LAT: Large Aperture Telescope

LFI: Low Frequency Instrument (*Planck* experiment)

LiteBIRD: Lite (Light) satellite for the studies of B-mode polarization and Inflation from cosmic background Radiation Detection (experiment)

LSS: (cosmic) Large-Scale Structure

LSST: Legacy Survey of Space and Time (experiment)

MCMC: Markov-Chain Monte Carlo (algorithm)

NERSC: National Energy Research Scientific Computing Center

NILC: Needlet Internal Linear Combination (algorithm)

NN: Neural Network

SW: (ordinary) Sachs-Wolfe effect

PCL: Pseudo-C_ℓ (power spectrum estimator)

POLARBEAR: POLARization of the Background Radiation experiment

PySM: Python Sky Model (software)

PTE: Probability To Exceed

QML: Quadratic Maximum Likelihood (power spectrum estimator)

ReLU: Rectified Linear Unit

SAT: Small Aperture Telescope

SED: Spectral Energy Distribution

SEVEM: Spectral Estimation Via Expectation Maximisation (algorithm)

SMICA: Spectral Matching Independent Component Analysis (algorithm)

SO: Simons Observatory (experiment)

S-PASS: S-Band Polarization All Sky Survey (experiment)

SPT: South Pole Telescope (experiment)

WMAP: Wilkinson Microwave Anisotropy Probe (experiment)

ABSTRACT

Current and near-future precision measurements give us the unique opportunity to refine the standard model of cosmology. One of the next milestones is the experimental proof of cosmic inflation, which would resolve outstanding problems related to the standard model and shed light on the origin of cosmic structure. Upcoming cosmic microwave background (CMB) experiments aim at a first detection of the curl-like polarization B-modes, a signal that is sourced by cosmological gravitational waves and considered a "smoking gun" of inflation. A key question, not only relevant for B-mode searches but for precision cosmology in general, is how to find the best analysis strategy when systematic effects match or dominate the signal of interest. In this Thesis, we present three original works that address different facets of this question, and apply them to three complementary problems in cosmological data analysis.

The main focus of this Thesis is the robust measurement of cosmological B-modes in the presence of Galactic foreground emission, considered to be a major source of systematic bias. We developed, tested, and validated the power-spectrum-based B-mode analysis pipeline for the Simons Observatory, an upcoming experiment that is expected to set new constraints within the next five years. The second work aims at constraining the optical depth to reionization from current large-scale CMB polarization data, which are known to contain non-Gaussian instrumental systematic effects that are difficult to write down in a likelihood model. We developed a novel likelihood-free estimator based on neural networks (NNs) and applied it to real *Planck* data to retrieve the optical depth directly from maps. This represents, to our awareness, the first cosmological parameter estimation on CMB polarization maps that is performed entirely by NNs. In the third work, we address the difficulty to efficiently marginalize over many astrophysical "nuisance parameters" that commonly arise in two-point analyses of the cosmic large-scale structure (LSS). We present a novel analytical likelihood approximation based on Laplace's method and apply it to real LSS data, achieving a speedup of a factor 3-5 compared to the standard "brute-force" approach.

The rising significance of systematic effects, exemplified by the search for cosmological B-modes, requires efficient problem-tailored analysis methods that are easy to interpret and to employ. It is essential that such methods are robust, meaning that we can reliably exclude potential sources of bias. We anticipate that novel methods, such as hybrid analysis pipelines or NN-based estimators, will prove highly beneficial to this endeavor.

Introduction

At the beginning of the 21st century, humans can measure and understand the physical processes at the largest scales in our Cosmos – about one trillion trillion times larger than their average body size – to sub-percent precision. This remarkable gain in knowledge initiated what scientists call the era of “precision cosmology”. This was possible due to the collaborative effort of experimental scientists to build instruments that can precisely measure tiny anisotropies in the cosmic microwave background (CMB) radiation that was released when the Universe was merely 380,000 years old, accurately matching the predictions that theoretical scientists made half a century ago. Today we know that the Universe is expanding (at an accelerated rate), and we can accurately trace back its evolution in time, thanks to the standard model of cosmology, which also predicts that the Cosmos itself, and with it said background radiation, originated from a hot, dense state known as the “big bang”.

This “cosmic symphony” between theory and experiments has cracks – and they run deep. In fact, we do not know what has initially given rise to those tiny fluctuations in the cosmic fabric, nor why they seem to permeate the Universe at its largest scales. The next generation of CMB experiments is preparing to measure a minuscule pattern in the polarization signal of the CMB, called B-modes, indicating the existence of primordial gravitational waves. Those, in turn, would confirm a popular theory according to which the very young Universe underwent rapid expansion – “cosmic inflation” – solving large parts of the cosmic puzzle. This signal has not been detected so far, but many theories of the early Universe could be constrained – or ruled out – by high-sensitivity measurements that either lead to a detection or a tighter upper limit of the B-mode amplitude.

Measuring the primordial B-mode signal from the first split seconds of our Universe is challenging: first, we are talking about a fluctuation of a few tens of nK in a 2.7 K background, which requires excellent control of instrumental noise. Second, we measure a much larger foreground signal from our own Galaxy, which represents a complex, poorly characterized contaminant that can cause spurious detection of primordial B-modes. Third, despite using cutting-edge technology, measuring such a small, polarized signal requires scientists to be aware and in control of all possible kinds of instrumental systematic effects that may bias the primordial signal.

My own role in this endeavor, as part of the Simons Observatory (SO) Collaboration, has been the development and validation of the B-mode power spectrum analysis pipeline, which will assess SO's first-year data and is designed to either measure, or set new constraints, on primordial gravitational waves. A major part of the work described in this Thesis culminated in a pipeline comparison and validation project, which featured different algorithms, designed to independently extract and measure primordial B-modes from SO's polarization data. With SO expecting to see its first light in late 2023, we used simulated maps of the microwave sky in order to probe the pipelines' ability to robustly constrain a primordial signal in the presence of complex Galactic foregrounds. The results indicate robust agreement between all pipelines if foregrounds are simple, while biases caused by more complex foregrounds can be successfully mitigated by using specific algorithmic extensions that are presented in this work.

Apart from my work within a collaboration, I took the opportunity to independently develop a novel algorithm to estimate cosmological parameters on the sky, based on artificial neural networks. Motivated by the fact that complex residuals that contaminate the sky are often hard to model analytically, this work takes a likelihood-free approach that relies only on having access to realistic simulations. We estimate the optical depth to reionization τ from real polarized *Planck* maps and obtain results compatible with literature methods, albeit with 30% larger error bars. The second conclusion of this work is that applying a neural network to real data required a large effort to thoroughly validate the network on different types of simulations, and experiment with different training setups. This allows for a promising outlook with regards to using neural networks as a complementary method for future B-mode analyses.

This Thesis is structured as follows.

In Chpt. 1, I introduce the modern picture of cosmology based on Einstein's General Relativity theory (GR), starting with the remarkably simple hypotheses that our Universe is homogeneous and isotropic at supergalactic scales. This leads to the Λ cold dark matter (CDM) model that implies that our Universe is made of dark energy, dark matter, and standard model particles. I then add linear perturbations to this theoretical framework, describing small fluctuations that are widely believed to originate from primordial quantum fluctuations, before they are stretched out to cosmic scales in a phase of early expansion known as inflation.

Chapter 2 takes this basic description and enriches it with the physical interactions of photons in the early Universe, leading up to the decoupling of photons from standard-model particles and giving rise to the CMB. I explain how we can understand, predict, and formalize angular fluctuations in the CMB total intensity and polarization, introducing the power spectrum statistic. After that, I move to current measurements of the CMB power spectrum, highlighting the *Planck* experiment and the Background Imaging of Cosmic Extragalactic Polarization (BICEP) and *Keck* experiment, before characterizing two major present-day challenges centered around CMB polarization, those being Galactic foregrounds and instrumental systematics. I close this chapter by presenting three next-generation experiments, the Simons Observatory, the Lite (Light) satellite for the studies of B-mode polarization and Inflation from cosmic background Radiation Detection (LiteBIRD), and CMB-S4.

In Chpt. 3, I present my contribution to constraining large-scale B-modes with SO, a power spectrum inference pipeline that I co-developed, tested and validated on simulations as a key part of the work presented in this Thesis. I start by outlining the power spectrum estimation algorithm and explaining how it solves practical issues related to observing the partial sky, after which I present the likelihood and theoretical model of the B-mode power spectrum. Lastly, I detail the parameter inference process, stressing the validation and code development work I carried out.

Chapter 4 puts the power spectrum pipeline I co-developed into the prespective of a pipeline validation project, featuring three different B-mode inference algorithms that will be used to constrain the amplitude of primordial gravitational waves with SO. My role in this was to coordinate and develop the code for the pipeline comparison. The goal of this project was to assess the robustness of all three methods if the sky signal contains non-Gaussian realistic foregrounds. Here, I describe the three algorithms with two extensions and the sky simulation suite that were used. The results agree between all pipelines for simple foregrounds, while more complex foregrounds can be successfully treated with said algorithmic extensions.

In Chpt. 5, I present a collaboration-independent work, in which I introduce a neural network (NN)-based parameter estimation algorithm suitable for polarized CMB maps, and use it to infer the optical depth to reionization τ directly from *Planck* maps – as a “first” in the field of CMB data analysis. I motivate this likelihood-free approach by the need of modeling complex map-level residuals, which is often hard to do analytically, before introducing neural networks and our specific choice for the architecture, and the sky simulations that were used. I detail why and how we went to great lengths to meticulously validate the NN training process in various steps using simulations, before finally presenting the results on real *Planck* data.

Chapter 6 completes this Thesis by reporting a side project that I was involved in, related to the problem of how to efficiently marginalize over many ($\gtrsim 10$) systematic parameters in the context of inferring cosmology from the cosmic large-scale structure (LSS). My role in this project was to co-develop a code that approximates marginal likelihoods based on Laplace's method, speeding up inference by a factor 3-10. I introduce and motivate the need for such a code in LSS analysis, explain its mathematics and take this opportunity to highlight its connection with volume effects, an issue that is of relevance also to CMB analysis. I close the chapter by presenting results on real and simulated LSS data and offering future prospects for this method.

Finally, in the Conclusions, I summarize the original scientific work presented in this Thesis, described in Chpts. 3 to 6. I highlight the main results and limitations of our methods, and single out potential future improvements and perspectives for each of the works presented.

CHAPTER 1

Λ cold dark matter cosmology

Cosmology is the science that studies the Universe in its entirety and its physical properties at scales larger than the typical size of a galaxy. At its basis is Einstein's theory of General Relativity (GR). It offers a description of the Universe as a whole at the largest scales, where gravity is the only relevant interaction. In addition to GR, most cosmological theories assume that the Cosmos is spatially homogeneous and isotropic, meaning that at all times, any observer at rest with respect to the cosmic expansion sees the same sky in all directions. This is clearly wrong, given that we observe anisotropic structures at large scales, such as the clustering of galaxies. However, when averaged over sufficiently large scales, these quantities are indeed evenly distributed and form a homogeneous background, justifying our assumption. In this context, we obtain a remarkably simple picture of the Universe that can globally expand or contract, depending on its exact energy and matter composition. Knowing the precise content of the Universe at a given time, cosmologists can determine its background expansion throughout cosmic history. Experimental data (e.g., Planck Collaboration I, 2020) have shown that the present Universe is expanding in an accelerated manner, driven by four main constituents that dominate the energy-matter budget: the elusive dark energy (about 70%), symbolized by the Greek letter Λ, which, as we shall see, is responsible for the cosmic acceleration, the equally poorly understood cold dark matter (CDM, about 25%), matter from massive particles of the standard model of particle physics (about 5%), and electromagnetic radiation (about 0.01%). This model, called "ΛCDM", can be refined to describe small fluctuations about the homogeneous background which, as we shall see, are the key to understanding and probing the history of our early Universe.

In this chapter, we introduce the basic ingredients of homogeneous cosmology and present the ΛCDM model, which represents the current standard description of our Universe at supergalactic scales. We start by deriving the theory of the cosmic background expansion in Sect. 1.1, before refining this picture by studying linear cosmological perturbations in Sect. 1.2.

1.1 Background expansion

In this section, we discuss homogeneous and isotropic cosmology, describing the components that make up the Universe, and explaining conceptual problems that may be solved by a hypothetical inflationary era at very early cosmic times.

1.1.1 Homogeneous isotropic universe

In GR, space and time lose their absolute meaning from Newtonian physics. Instead they depend on the observer's state of motion and must be combined in the concept of a four-dimensional manifold called "spacetime". We can describe the geometry of spacetime in terms of a metric, a generalized distance measure between points. Two neighboring points in spacetime with a local coordinate distance $\mathrm{d}x^\mu$, $\mu = 0, .., 3$ (where $\mu = 0$ labels time and $\mu = 1, 2, 3$ labels the space coordinates) have a coordinate-independent spacetime distance $\mathrm{d}s$ given by

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \,, \tag{1.1}$$

where $g_{\mu\nu}$ denotes the spacetime metric, and we sum over repeated indices. The importance of the metric is the ability to measure distances between any two points in spacetime and thereby describing its entire geometry.

Gravity results from the interplay between mass, or energy, and the geometry of spacetime, encoded in the spacetime metric. More specifically, gravity is equivalent to the curvature of spacetime and its sole effect is to make masses move along curved paths. Meanwhile, the presence of mass and energy itself causes the curvature of spacetime, explaining the apparent attractive force that is gravity. This interplay is mathematically described by the Einstein equations

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} - g_{\mu\nu}\Lambda \,. \tag{1.2}$$

The left-hand side of (1.2) represents the local curvature of spacetime via the Einstein tensor $G_{\mu\nu}$, while the right-hand side describes the local energy-mass distribution, represented by the energy-momentum tensor $T_{\mu\nu}$, multiplied by Newton's gravitational constant G. We refer to Wald (1984) for a complete definition of the tensors used in Eq. (1.2) starting from the metric tensor. The constant Λ accounts for the energy density of the vacuum, also known as the cosmological constant, which is a key quantity for the dynamics of the Universe, as we shall see later in this chapter.

The assumption of spatial homogeneity and isotropy leads to an enormous simplification of the geometry of the Universe. We can subdivide spacetime into a family of spatially homogeneous and isotropic slices at constant time, giving us back the notion of an absolute time in the Newtonian sense, called “cosmic time” t. We can therefore measure infinitesimal time intervals $\mathrm{d}t$ independently of our location in space, while spatial distances assume a time-dependent prefactor $a(t)$ called the “cosmic scale factor”. This leads to a spacetime metric of the form ($i, j \in \{1, 2, 3\}$)

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{d}t^2 + a^2(t)\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j \,, \tag{1.3}$$

where we can write the the spatial metric γ_{ij} in spherical coordinates (r, θ, φ)

$$\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j = \mathrm{d}r^2 + \chi^2(r)\,(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\varphi^2)\,. \tag{1.4}$$

We have one remaining geometrical degree of freedom in the curvature parameter K, which can be negative, zero, or positive, corresponding to a hyperbolic, flat, or spherical spatial geometry, respectively. The radial coordinate $\chi(r)$, called the “comoving radius”, depends on the spatial geometry as

$$\chi(r) \equiv \begin{cases} |K|^{-1/2}\sinh(\sqrt{|K|}r)\,, & K < 0 \quad \text{(hyperbolic)} \\ r\,, & K = 0 \quad \text{(flat)} \\ K^{-1/2}\sin(\sqrt{K}r)\,, & K > 0 \quad \text{(spherical)} \end{cases} . \tag{1.5}$$

Equation (1.3) is the so-called *Friedmann-Lemaître-Robertson-Walker* metric of cosmology (FLRW, Friedmann, 1922; Lemaître, 1931; Robertson, 1935; Walker, 1937). It describes the background geometry of the Universe in terms of its scale factor $a(t)$, which allows for the Universe to expand or contract in time. In the late 1920s, Georges Lemaître and Edwin Hubble first showed that the Universe is expanding at a rate $H(t) \equiv \mathrm{d}a/\mathrm{d}t$ today known as the “Hubble rate”. Evaluated at the present, $t = t_0$, the Hubble rate is known as the “Hubble constant” H_0 and amounts to about 70 km/s/Mpc.[1] As an important consequence of measuring light across cosmological distances, a photon traveling in an expanding FLRW universe for some finite time Δt will have its physical wavelength stretched, or “redshifted”. This happens because during the photon’s journey, the cosmic scale factor a increases, and with it the physical length of any object. This relative increase in the physical wavelength is known as the “redshift” z, and amounts to

$$z \equiv -1 + \frac{a(t_0)}{a(t_0 - \Delta t)}\,, \tag{1.6}$$

where t_0 is the cosmic time at the end of the photon’s journey.

This extraordinary simplification of the geometry of the Universe puts equally strong constrains on the allowed shape of the energy and matter distribution. Spatial isotropy prevents any energy-momentum fluxes or anisotropic stresses, leaving us with a perfect fluid that is parameterized only

---

[1] One Megaparsec (Mpc) equals about 3.09×10^{22} m, or 3.26 million light years.

by its global energy density $\rho(t)$ and pressure $P(t)$. Inserting the energy-momentum tensor of this "cosmic fluid" intro the Einstein equations (1.2) leads to the well-known Friedmann equations, which relate cosmic geometry (via the scale factor a and its time derivatives a' and a'') to the cosmic energy and matter content, given by the fluid parameters P and ρ, and the cosmological constant Λ. Defining units in which the speed of light is one, the Friedmann equations read

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{\rho} - \frac{K}{a^2} + \frac{\Lambda}{3} \tag{1.7}$$

$$\dot{H} + H^2 = \frac{\ddot{a}}{a} = \frac{-4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3}\,. \tag{1.8}$$

These are the key equations describing the expansion in an homogeneous and isotropic cosmology. They relate the time-dependent scale factor to the time-dependent pressure and energy density of the cosmic fluid, allowing us to know the scale of the Universe at all times, assuming we know the constants K and Λ (Peebles, 1993). If the cosmic fluid, including the cosmological constant, happens to have an energy density equal to the "critical density" ρ_{crit},

$$\rho_{\text{crit}}(t) \equiv \frac{3H^2(t)}{8\pi G}\,, \tag{1.9}$$

then the curvature is forced to be zero. The term Λ results as viable general solution of the Einstein equations (1.2) and was originally introduced in Einstein (1917) to ensure that the Universe was static. This interpretation turned out to be wrong, since the effect of a cosmological constant is to drive the exponential expansion of the Universe. A more modern interpretation of the cosmological constant, more generally known as "dark energy", is that of a quantum-mechanical vacuum energy that permeates the Cosmos. Unlike the other constituents, whose energy densities get diluted as the Universe expands, the cosmological constant does not vary, and inevitably ends up dominating the cosmic energy density. In the following, we discuss the different components in this model in more detail.

1.1.2 Constituents of the Universe

In order to solve the Friedmann equations and determine the cosmic expansion history, we need to specify the properties of the cosmic fluid in an FLRW universe. Considering the ratio between pressure and energy density, we can distinguish between three components: matter, radiation, and the cosmological constant. Nonrelativistic matter has a negligible kinetic energy as compared to its mass, and therefore behaves like a pressureless fluid, $P = 0$. Radiation, consisting of relativistic particles, exhibits radiation pressure given by one third of its energy density, $P = \rho/3$. A cosmological constant is a component with a constant energy density, implying a negative pressure that exactly outweighs its energy density, $P = -\rho$. To see this, we can use the covariant energy-momentum conservation from GR to find (Peebles, 1993)

$$\dot{\rho} = -3H\ (\rho + P)\ , \tag{1.10}$$

which is known as the "continuity equation". By inserting $P = -\rho$, we find indeed that the energy density is constant. Equation (1.10) allows us to infer how the individual components' energy densities evolve in time. If, in turn, we insert a given component's energy density into the first Friedmann equation (1.7), we learn how a universe dominated by each single component would scale in time. This defines four cosmological eras that a general FLRW universe can undergo, which are listed in the following.

- **Radiation-dominated era:** This era is dominated by relativistic particles with an energy density that scales like $\rho \propto a^{-4}$, accounting for the dilution of the particle number density in an expanding volume ($n \propto a^{-3}$) and the redshifting of wavelength ($1/\lambda \propto a^{-1}$). The cosmic scale factor behaves like $a \propto t^{1/2}$, resulting in a decaying energy density as $\rho \propto t^{-2}$.

- **Matter-dominated era:** This era is dominated by nonrelativistic matter with an energy density that scales like $\rho \propto a^{-3}$, accounting for the dilution of the particle number density in an expanding universe. The universe scales like $a \propto t^{2/3}$, resulting in a decaying energy density as $\rho \propto t^{-2}$.

- **Curvature-dominated era:** If the universe is dominated by curvature, the first Friedmann equation yields $H^2 = K/a^2$, meaning that the universe expands linearly in time. If we were to associate an energy density to curvature, it would decay quadratically in both a and t.

- **Λ-dominated era:** This era is dominated by vacuum energy, or a cosmological constant, with a constant energy density. The cosmic scale factor increases exponentially like $a \propto e^{\sqrt{\Lambda/3}t}$.

In a real universe, these components generally form a mixed fluid, and the total energy density and the scale factor behaves in a more complicated way that depends on the relative abundances of each of them. However, we can see that by the relative scaling of the energy densities with a, the components form a hierarchy: radiation decays faster than matter, which in turn decays faster than curvature, while the cosmological constant does not decay at all. For any multicomponent universe, at sufficiently early times, there must have therefore been a radiation-dominated era, followed by a matter-dominated phase, and so on. Knowing the abundances of each of these components at present would therefore allow us to evolve them back in time and making predictions for the future. We define the fractional density parameter of a component X as the ratio between its energy density and the critical energy density, which was defined in Eq. 1.9,

$$\Omega_X(t) = \frac{\rho_X(t)}{\rho_{\mathrm{crit}}(t)} \,. \tag{1.11}$$

Using this definition, the first Friedmann equation (1.7) can be rewritten as

$$H^2 = H_0^2 \left(\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_K a^{-2} + \Omega_\Lambda \right) \,, \tag{1.12}$$

where we defined $\Omega_K = -K/H_0^2$, and H_0 is the Hubble function evaluated at present. Current measurements from *Planck* (Planck Collaboration VI, 2020) constrain the Hubble and density parameters to

$$\begin{aligned} H_0 &= (67.66 \pm 0.42)\,\mathrm{km/s/Mpc} \\ \Omega_\Lambda &= 0.6889 \pm 0.0056 \\ \Omega_m &= 0.3111 \pm 0.0056 \\ \Omega_K &= 0.001 \pm 0.002 \end{aligned}$$

at 68% credible level (CL).

Besides the unknown nature of Λ, which completely lacks a fundamental physical explanation so far, an almost equally puzzling component is hidden behind the matter sector. The visible ordinary matter, such as stars, galaxies, planets, intergalactic dust, and black holes, cannot account for the apparent gravitational mass observed dynamically in galaxies and galaxy clusters. Baryons, as astrophysicists name all matter composed by particles from the standard model of particle physics, can therefore only account for about one sixth of the total nonrelativistic mass. The remaining unknown component, dubbed "cold dark matter", behaves like a nonrelativistic ("cold") fluid that – since the moment when it decoupled from all the other particles – has interacted only through its own gravity. As we shall see in the next chapter, anisotropies in the background radiation and abundances of light nuclei (Planck Collaboration VI, 2020) constrain the baryonic energy density to $\Omega_b h^2 = 0.02233 \pm 0.00015$ and the cold dark matter density to $\Omega_c h^2 = 0.1198 \pm 0.0012$, where $h \equiv H_0/(100\,\mathrm{km/s/Mpc})$.

The finding that the observable Universe is dominated by two unknown components, the cosmological constant Λ and cold dark matter, has given the current cosmological model its name "Λ cold dark matter" (ΛCDM). In the following section, we complement this scenario with the current hypothesis about the very early Universe, which predates and leads into the radiation-dominated era.

1.1.3 The early Universe

The cosmological standard model ΛCDM, discussed previously, fits data exceptionally well. Observations of the relative temperature anisotropies in the CMB (see Chpt. 2) confirm the underlying assumption of the approximate isotropy of the Universe at its largest scales. While the observational data from the CMB seem to paint a coherent picture of an expanding FLRW universe that started with a big bang about 13.7 billion years ago, the observed isotropy raises a conceptual problem related to causality. Information can only travel at the speed of light, which defines a causal horizon beyond which no two points in space could have talked to each other since the big bang. This implies that no process in the early Universe could homogenize the CMB at scales larger than the causal horizon at photon decoupling, which corresponds to an angular scale of about one degree today. The fact that we still observe an isotropic CMB at scales well beyond degree scales raises the question as to whether our picture of the early Universe is complete. This is known as the **"horizon problem."**

This question comes along with another conceptual issue: as we have seen in the previous section, we observe a Universe with a spatially flat geometry. One can show that in a ΛCDM universe, the curvature grows with time, meaning that it had to be much smaller in the past, to the point where its smallness cannot be associated to a coincidence. This is called the **"curvature problem."**

Cosmic inflation is a theoretical paradigm proposed to solve these problems through a modification of our ΛCDM Universe at very early times (Guth, 1981). Inflation is defined as a phase of exponential expansion at $t \lesssim 10^{-32}$s after the big bang, which made the Universe grow by about 60 orders of magnitude within an extremely short amount of time. The typical length scale of physical interactions before inflation was stretched out to scales far beyond the causal horizon, far enough to even contain the causal horizon at the present day. This makes the present Universe look homogeneous at its largest observable scales, simply by the fact that the primordial fluid could interact over those scales and thermalize before inflation started. This solves the horizon problem. It is possible to show that furthermore, inflation suppresses spatial curvature Ω_K like $1/a^2$, allowing for the possibility that the curvature may have well been large at the beginning of inflation (Linde, 2008). This solves the curvature problem.

A broad variety of inflationary models has been studied (Linde, 2008; Martin et al., 2014). They all have in common that at very early times, one or several quantum fields called "inflatons" with negative pressure dominate the energy density of the Universe. This negative pressure leads to an exponential expansion of the Universe, similar to what a cosmological constant does at late times. The end of inflation is usually marked by the inflaton reaching a potential minimum and decaying into the particles of the standard model of particle physics (Tsujikawa, 2003). The simplest family of inflationary models is "single-field slow-roll inflation", based on the assumption that inflation is caused by the inflaton slowly rolling down a potential well. This potential dominates the inflaton energy until it reaches a minimum and starts decaying. One of the very first inflationary models proposed, known as the Starobinsky R^2 model, was derived from the interaction of a classical gravitational field background with several massless quantum fields of different spins (Starobinsky, 1980).

The most intriguing consequence of inflation is that it explains the origin of primordial fluctuations, which we shall discuss in the following section. While the physical interactions were in the microscopic regime at the beginning of inflation, they reached macroscopic scales by its end. This suggests that inflation stretches primordial quantum fluctuations to macroscopic scales, turning them into classical fluctuations. This connects quantum field theory with large-scale cosmology and introduces the picture of vacuum fluctuations giving rise to the primordial seeds of the CMB anisotropies. The fundamental physics of inflation can be described by a single or several inflaton fields that dominate the dynamics of the Universe at high energies and eventually decay into the standard model particles at the end of the inflationary phase.

An experimental confirmation of inflation would be nothing short of a revolution in cosmology, as it would allow us to probe energy scales of about 10^{15} GeV, far beyond the reaches of any human-made experiment. A proof of inflation would also ascertain us that cosmological perturbations at the largest observable scales today are a pristine probe of the physical processes within the first 10^{-32} seconds after the big bang.

1.2 Linear perturbations

We now move away from the strictly homogeneous description of the Universe to introduce small fluctuations in the FLRW metric. This is the aim of linear cosmological perturbation theory, and sets the foundations of our current understanding of the early Universe.

As we have seen in the last chapter, the geometry of the Universe is intrinsically related to its energy-matter content. To take into account this relation, we must study the linear perturbations of the Einstein tensor, representing the geometry of the Universe, and the energy-momentum tensor of the cosmic fluid, representing its particle content:

$$G_{\mu_\nu} \longrightarrow \bar{G}_{\mu\nu} + \delta G_{\mu\nu} \ , \quad T_{\mu_\nu} \longrightarrow \bar{T}_{\mu\nu} + \delta T_{\mu\nu} \ . \tag{1.13}$$

We refer to Hu (2003); Durrer (2008) for a general treatment of cosmological perturbations. We do not derive the full set of evolution equations and the resulting dynamics for those. Instead, we limit ourselves to a description of the different kinds of perturbations and their key statistical properties, which are needed for the discussion in the following chapters.

A conceptual difficulty arises from the equivalence principle of GR: a gravitational field and a corresponding acceleration of the reference system are physically equivalent. To mathematically satisfy this statement, the Einstein equations must, under general coordinate transformations, maintain their mathematical form (1.2). If we decompose $G_{\mu\nu}$ and $T_{\mu\nu}$ into a background and a perturbation as above, we must ensure that the equivalence principle holds for the background and the perturbed quantities separately. This can be addressed by choosing "gauge-invariant" perturbation variables that do not depend on the choice of coordinates. This way, we ensure that the solutions describe physical effects and are not artifacts of an unfavorable choice of coordinates.

In the following, we write the linearly perturbed Einstein and energy-momentum tensors in a mathematically convenient form, called scalar-vector-tensor decomposition. As we shall see latetr, this description provides a complete set of physically relevant quantities as long as the perturbations are small. This allows us to independently consider each one of these three types of perturbations, and study them individually as they evolve in time.

1.2.1 Scalar and tensor perturbations

Based on the Helmholtz decomposition, we can express any field in a given spacetime as a sum of three components – scalars, vectors, and tensors – that transform differently under spatial rotations. Physically, scalar fields represent fluctuations in the local spatial curvature, the gravitational potential, and the energy density, while tensor perturbations represent gravitational waves. Vector perturbations are usually associated to rotational velocity fields, which are not generated in most inflationary models, and, moreover, decay in an expanding universe (Hu, 2003). Therefore, vector perturbations can be safely neglected at this stage. Scalar and tensor perturbations are independent in linear theory, which allows us to study their evolution separately.

As an example, we apply this decomposition to the perturbations of the spacetime metric, $g_{\mu\nu}$. In contrast to the homogeneous FLRW background metric, the time-time component g_{00} is locally affected by a scalar metric perturbation Φ, while vector (V_i) and scalar (B) perturbations give rise to nonzero time-space components g_{0i}. Finally, the space-space metric components contain all three types of perturbations (scalars Ψ and E, vectors C_i, and tensors h_{ij}, see Hu, 2003):

$$\begin{aligned} g_{00} &= -a^2(1+2\Phi)\,, \\ g_{0i} &= a^2(V_i - B_{,i})\,, \\ g_{ij} &= a^2[(1-2\Psi)\delta^{\mathcal{K}}_{ij} + D_{ij}E - \frac{1}{2}(C_{i,j} + C_{i,j}) + h_{ij}]\,, \end{aligned} \tag{1.14}$$

where $D_{ij} \equiv (\partial_i \partial_j - \frac{1}{3}\delta^{\mathcal{K}}_{ij} \nabla^2)$, $\delta^{\mathcal{K}}$ is the Kronecker symbol, ∇^2 is the Laplace operator, and the subscript $,i$ denotes the partial derivative with respect to the spatial coordinate x^i.

As explained in the last section, the need for a coordinate-independent theory makes it necessary to define gauge invariant quantities. The gravitational wave perturbation h_{ij} already has this property, while vector perturbations, as argued above, can be neglected. The scalar metric perturbations can be combined with the perturbation of the fluid density $\delta\rho$, the time derivative of the density $\dot{\rho}$, and the Hubble rate H, to give the gauge-invariant quantity (Riotto, 2003)

$$\zeta \equiv \Psi + \frac{\delta\rho}{\dot{\rho}} \,. \tag{1.15}$$

This is the gravitational potential on slices of uniform energy density. In the most common scenario, where perturbations are adiabatic and we can neglect anisotropic stress in the cosmic fluid, ζ completely describes all scalar perturbations (Hu, 2003). We speak of an adiabatic perturbation if the energy density and the pressure are related by

$$\frac{\delta\rho}{\dot{\rho}} = \frac{\delta P}{\dot{P}} \,, \tag{1.16}$$

which, in particular, implies that no energy is transferred between the perturbation and the background environment.

This leaves us with a minimal set of two variables, the scalar gravitational potential at uniform energy density ζ, and the tensorial amplitude of gravitational waves h_{ij}, that quantify the linear perturbations of the Universe about a homogeneous FLRW background. Now, we may ask how we can optimally compress these two fields into numbers that we can measure in an experiment. To retrieve meaningful information from cosmic fluctuations, we need to know how to compare the amplitude of perturbations between different points in spacetime. One elegant way of doing so is to assume that perturbations are isotropic and homogeneous in the statistical sense, that is, they follow the same statistical distribution across the entire sky as observed by any observer in spacetime. We can regard this as the generalization of the cosmological principle for perturbations. Adopting this assumption, we first describe any given perturbation field $f(t, \boldsymbol{x})$ in terms of its spatial Fourier transform:

$$f(t, \boldsymbol{k}) \equiv \int \frac{\mathrm{d}^3 x}{(2\pi)^3} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \,, \tag{1.17}$$

where we fixed cosmic time t. Statistical isotropy then implies that the two-point correlation in Fourier space, $\langle f(t, \boldsymbol{k})[f(t, \boldsymbol{k}')]^* \rangle$, is nonzero only if $\boldsymbol{k} = \boldsymbol{k}'$ and only depends on the absolute magnitude of the Fourier wave vector $\boldsymbol{k}$:

$$\langle f(t, \boldsymbol{k})\, f(t, \boldsymbol{k}')^* \rangle = (2\pi)^3 \delta^3_{\mathcal{D}}(\boldsymbol{k} - \boldsymbol{k}')\, P_f(t, k) \,, \tag{1.18}$$

where $\delta^3_{\mathcal{D}}$ is the three-dimensional Dirac delta distribution, $k \equiv |\boldsymbol{k}|$, and $P_f(t, k)$ is called the **power spectrum** of the field f. The power spectrum $P_f(k)$ encodes the two-point statistics of the perturbation field. Assuming statistical homogeneity and isotropy thus allows to drastically reduce the number of degrees of freedom of the perturbation fields from initially four dimensions $(t, \boldsymbol{k})$ to two dimensions (t, k). If the field f is a Gaussian random variable, then its power spectrum characterizes the full probability distribution of f. Both of these properties are predicted by most models of the very early Universe, including a large class of inflationary models.

In the following section, we present the standard power law parameterization of the power spectrum of primordial perturbations, starting with scalar perturbations.

1.2.2 Power spectrum of scalar perturbations

As we saw earlier, scalar perturbations can, under fairly general conditions, be entirely described by the Fourier-space power spectrum of the perturbation ζ of the gravitational potential. While this significantly reduces the complexity of our theoretical model, we can do even better. Most models of the very early Universe, such as inflation, predict primordial scalar perturbations as being produced in an almost scale-independent way, which would imply $P_\zeta(k) = \text{const.}$. Indeed, since the inflationary phase needs to end at some point and the inflaton needs to keep track of this “progress”, the power spectrum cannot be perfectly scale invariant. A simple and common parameterization of the primordial curvature power spectrum is the power law,

$$P_\zeta(k) = A_s \left(\frac{k}{k_*}\right)^{n_s} , \tag{1.19}$$

measured at a conventional scale $k_* = 0.05\ \text{Mpc}^{-1}$. Here, we introduced A_s, the power spectrum amplitude of scalar fluctuations, and n_s, the scalar power spectral index. Equation (1.19) quantifies the statistics of the seeds of the cosmic scalar perturbations, which give rise to the scalar CMB anisotropies at decoupling and the LSS of the matter density and galaxy distributions in the late Universe (Planck Collaboration X, 2020).

The parameters A_s and n_s are the two early-Universe parameters among the six parameters of the cosmological standard model ΛCDM and represent the initial conditions for the cosmological perturbations. Current measurements by *Planck* give the following constraints (Planck Collaboration VI, 2020)

$$\begin{aligned} \ln(10^{10} A_s) &= 3.043 \pm 0.014 \\ n_s &= 0.965 \pm 0.004 \,, \end{aligned}$$

where the uncertainties indicate the corresponding central 68% credible intervals.

1.2.3 Power spectrum of tensor perturbations

The power spectrum of primordial tensor perturbations h_{ij} is commonly written as

$$P_h(k) = A_t \left(\frac{k}{k_*}\right)^{1-n_t}, \tag{1.20}$$

which is measured at a pivot scale of $k_* = 0.05$ Mpc. Here, A_t is the power spectrum amplitude of primordial tensor perturbations, and n_t is called the tensor spectral index. Note that scale invariance would imply $n_t = 1$, not 0.

We parameterize the abundance of primordial gravitational waves by the **tensor-to-scalar ratio**, the (positive) ratio of the power spectrum amplitudes between tensor and scalar perturbations:

$$r \equiv \left.\frac{A_t}{A_s}\right|_{k=k_*}, \tag{1.21}$$

where $k_* = 0.05$ Mpc, as usual. At the time of writing this Thesis (2023), the BICEP and *Keck* array (BICEP2/Keck Collaboration, 2021) place the tightest experimental constraint on the tensor-to-scalar ratio, with

$$r < 0.036 \quad \text{(BICEP/\textit{Keck} 2021)} \tag{1.22}$$

at 95% CL. Near-future experiments, such as SO (SO Collaboration, 2019), LiteBIRD (LiteBIRD Collaboration, 2023), and CMB-S4 (Abazajian et al., 2019), are preparing ground-based and space missions to tighten this constraint by a factor of 3-10. This may potentially allow for the first detection of primordial gravitational waves if r is at a level of $r \gtrsim 0.003$-0.01, and would otherwise allow us to exclude a large class of theoretical models, such as Starobinsky R^2 inflation, which predicts a value of $r \sim 0.0046$ (see e.g., LiteBIRD Collaboration, 2023).

In the following chapter, we describe the CMB anisotropies in detail and relate the primordial perturbations to the observable features in the CMB temperature and polarization power spectra. We devote a large part of the following chapter discussing the experimental side of CMB analysis. In that context, we present past measurements, discuss systematic effects, and summarize future CMB surveys.

CHAPTER 2

Cosmic microwave background

The FLRW cosmological model, founded on three essential assumptions – General Relativity, spatial homogeneity, and isotropy – does not only lead to a stunningly simple, yet equally successful picture of the homogeneous background Universe. By introducing linear perturbations, we obtain a cosmological model allowing us to associate primordial quantum fluctuations with their later manifestations at cosmic scales: as tiny gravity wells in the photon-baryon plasma, then as ripples in the cosmic background radiation, and, finally, as the filamentary structure of dark matter, galaxies, and intergalactic dust that permeates the largest scales of our observable Universe. Before Arno Penzias and Robert W. Wilson discovered the cosmic microwave radiation in 1964 (Penzias & Wilson, 1965), the idea of its existence had merely been the result of an elegant theoretical hypothesis. Afterwards, we had the CMB – an almost uniform blackbody radiation with an average temperature of 2.7 K – as evidence and visible afterglow of the big bang. Since then, the CMB has served cosmologists as a direct probe of the physics around the time of the last scattering between photons and baryons, about 380,000 years after the big bang.

The primordial fluctuations, conjectured to be the seeds of cosmic structure, eventually imprinted themselves into the CMB radiation. In order to understand how this was possible, we need to study the high-energy interactions of photons in the primordial plasma, which left their own, distinctive marks in the photons' spatial distribution. After the formation of the CMB, the photons kept free-streaming for nearly 13.4 billion years during which they entered the reionized Universe, and were slightly deflected in the gravitational potential of the late-time matter distribution. This series of events has left predictable imprints in the power spectra of the CMB anisotropies, both in temperature and in polarization, which makes the CMB the most powerful cosmological probe to date.

Following the detection of the CMB, many experiments ensued that measured not only its background temperature, but also its angular fluctuations. Thanks to ever more precise experiments, such as the Cosmic Background Explorer (COBE)[1], the Wilkinson Microwave Anisotropy Probe

---

[1] URL: lambda.gsfc.nasa.gov/product/cobe

(WMAP)[2], and *Planck*[3], we have been able to confirm theoretical predictions at an astonishing accuracy. While some aspects of the so-called “standard model of cosmology” ΛCDM are in perfect agreement with experimental data, other aspects pose fundamental conceptual problems, such as the horizon and curvature problem (see Sect. 1.1.2), which might be solved by cosmic inflation. One of the key science goals of current and future CMB experiments is the first measurement of polarization B-modes, the curl-like pattern in the CMB polarization. B-modes of primordial origin, as we discuss in this chapter, would directly probe the amplitude of gravitational waves in the very early Universe and potentially rule out or corroborate inflationary cosmology. As groundbreaking of an achievement this would be for theoretical physics, it is challenging to realize from the experimental perspective. Future experiments must be especially thorough in characterizing Galactic polarized foregrounds and instrumental contamination, which may both lead to significant bias in B-mode measurements.

In this chapter, we start by introducing the physics of photon decoupling that led to the formation of the CMB in Sect. 2.1. We then formulate the mathematical description of the CMB and polarization angular power spectra in Sect. 2.2, before discussing the predicted features from physical effects before and after photon decoupling in Sect. 2.3. Afterwards, in Sect. 2.4, we move to observations of the CMB anisotropies with a focus on the polarization B-mode power spectrum. In Sects. 2.6 and 2.7, we introduce Galactic foregrounds and instrumental contamination in real experiments, before closing with an overview on next-generation CMB polarization experiments in Sect. 2.5.

2.1 Decoupling

From the theory of the homogeneous cosmology presented in the previous chapter, the CMB is predicted to be a isotropic blackbody radiation with a temperature of 2.7 K. Primordial perturbations introduce small variations that are imprinted in this background radiation. To understand how this takes place in detail, we must study the physical interactions of photons through the history of the early Universe until decoupling, the era when photons started free-streaming as the cosmic background radiation. After the conjectured inflationary phase within the first 10^{-32} seconds after the big bang, or at an equivalent energy scale of 10^{15} GeV, the standard model particles formed a dense, tightly coupled primordial plasma. These particles interacted frequently via the strong and electroweak forces, creating a thermal equilibrium. As the Universe kept expanding, their average kinetic energy - known as the plasma’s temperature - kept decreasing accordingly. Some of the particle interactions got inhibited due to the lower temperatures, marking “phase transitions” that

---

[2] URL: map.gsfc.nasa.gov

[3] URL: cosmos.esa.int/web/planck

left behind temporarily stable reaction products. Examples are the first formation of protons and neutrons from quarks and antiquarks after the quantum chromodynamic phase transition around 160 MeV, the decoupling of neutrinos around 1 MeV, and the phase of electron-positron annihilation that resulted in a net number of electrons in the Universe. At about 100 keV, or a temperature of 10^9 K, nuclear fusion started to occur in a process known as "primordial nucleosynthesis", which produced the cosmic abundances of light nuclei, a mixture of ionized hydrogen, helium, and traces of other elements. This plasma was still dominated by photons, tightly coupled with baryonic matter via elastic Compton scattering, and cooling as the Universe expanded.

When the Universe had cooled to about 10^4 K, matter had become largely nonrelativistic and started to dominate the global energy density. The mass density of both baryonic and cold matter was no longer negligible with respect to their kinetic energy density and, as a result, tiny gravitational potential wells from primordial curvature perturbations started to attract matter. This produced small overdensities that followed the curvature perturbations. Meanwhile, photons and baryons remained tightly coupled via elastic Compton scattering. At that time, two competing forces dominated the plasma dynamics: while gravity tried to concentrate the mass inside the potential wells, radiation pressure pushed outward, trying to disperse the plasma. This resulted in a driven harmonic oscillation, during which the fluid would alternately cluster inside the potential wells and "run over". This phenomenon added an important imprint on the CMB anisotropies called "baryonic acoustic oscillations" (BAOs), which we present in the following section.

As the acoustic oscillations continued, the large Compton scattering cross section kept the plasma tightly coupled. During that era, electrons were immersed in a sea of photons and prevented photons from freely traveling over any cosmologically relevant distance. By the time the Universe had cooled down to energies well below the hydrogen ionization energy of 13.6 eV (about 1.6×10^4 K), electrons had slowly started to combine with protons to form temporarily stable neutral hydrogen atoms. This process, called "recombination", continued until the bulk of electrons was locked away in hydrogen atoms, making it much less likely for a photon to encounter an electron, to the point where photons could practically free-stream over the length of the observable Universe. This process, called "photon decoupling", took place at a temperature of about 3000 K and defines the birth of the CMB.

To see how primordial perturbations of a given comoving Fourier mode k (corresponding to a comoving length scale $\lambda = 2\pi/k$ in real space) are imprinted in the CMB temperature anisotropies, we must consider how they evolve in time. The linearized Einstein equations lead to solutions for perturbations in the scalar gravitational potential that are approximately constant in time regardless of their spatial scale, while for density and gravitational wave perturbations, we obtain different solutions for scales much larger than the Hubble horizon ($k \ll aH$) and scales much smaller than the Hubble horizon ($k \gg aH$). Density perturbations remain approximately constant at superhorizon

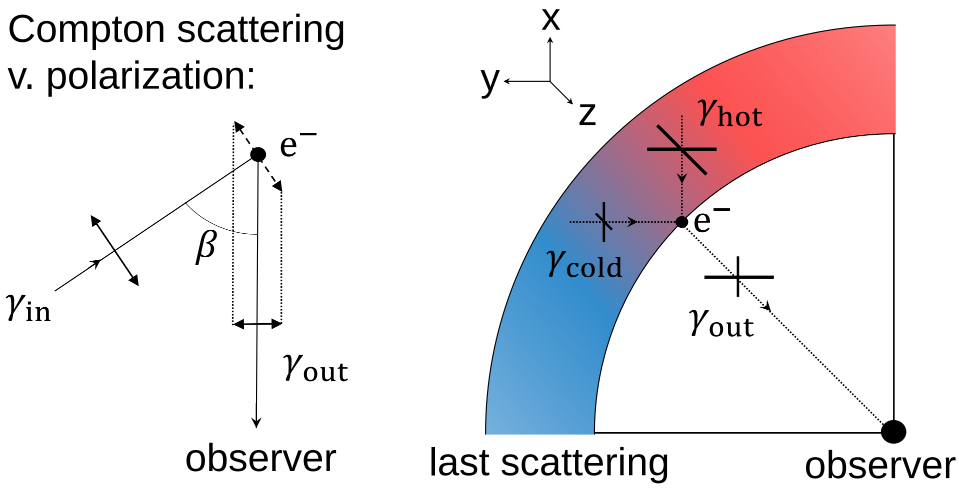

Figure 2.1: Illustration of Compton scattering, highlighting its polarization dependence (*left panel*) and the generation of polarization anisotropies from local temperature quadrupole pattern at last scattering (*right panel*). Figures adopted and modified from Hu & White (1997).

scales, while subhorizon scales oscillate in a radiation-dominated universe and grow linearly with a in a matter-dominated universe. Gravitational waves, too, maintain a constant amplitude at superhorizon scales and perform a damped oscillation at subhorizon scales (Hu, 2003). This means that the comoving Hubble horizon at decoupling, $1/k_*$, marks a characteristic scale, which leads to three regimes:

- **$k \gg k_*$**: perturbations of comoving size much smaller than $1/k_*$ had started decaying before the CMB formed, resulting in a decreased amplitude for the CMB temperature fluctuations.
- **$k \sim k_*$**: perturbations of comoving size of order $1/k_*$ imprinted themselves in the CMB at their maximum amplitude, leading to a maximum in the CMB temperature power spectrum known as the "recombination peak".
- **$k \ll k_*$**: perturbations of comoving size much larger than $1/k_*$ had not entered the horizon by the time the CMB was formed, since no causal mechanism could have imprinted them into the CMB. Therefore, they are not present in the CMB.

The comoving Hubble horizon at decoupling can be translated into a physical scale in the present-day Universe, where it subtends an angle of about one degree on the sky.

Compton scattering plays another important role, that is, it generates the polarization anisotropies of the CMB. Compton scattering changes the polarization of the incoming photon in the scattering plane, depending on the scattering angle. Figure 2.1 in the left panel shows the scattering plane in the rest frame of the electron. An incoming photon makes the electron oscillate along the incident polarization direction, but only the projection of this oscillation amplitude perpendicular to the outgoing photon direction can translate into the final polarization amplitude. If this process keeps happening in an isotropic fluid, no net polarization will be emitted on average. If, instead, we look at a baryon-photon plasma that features a quadrupole pattern of cold and hot spots, as shown in the right panel of Fig. 2.1, then photons that are Compton-scattered in the observer's direction is linearly polarized on average. Thus, soon-to-be CMB photons around the time of decoupling adopted anisotropies in linear polarization that traced the local quadrupole pattern in the temperature. The photons had only a narrow time window to pick up this polarization pattern. This had to happen late enough so that quadrupole patterns did not immediately get washed out via thermalization, and early enough so that electrons were still available for Compton scattering. As a result, polarization perturbations occur only at small scales, resulting in a sharp decline in the power spectrum of the polarization at large angles.

To summarize, the anisotropies of the CMB in temperature and polarization contain information about initial perturbations, and acoustic oscillations between baryons and photons at the time of decoupling. Linear polarization, generated from temperature quadrupole anisotropies through Compton scattering, is not only an important experimental cross check of our theory but carries information independent of the temperature field. As we show in the next section, the parity-odd polarization B-modes represent a uniquely clean window to primordial gravitational waves. In the following, we give a formal characterization of the CMB temperature and polarization amplitudes.

2.2 CMB anisotropies

To test the theoretical explanation for the CMB anisotropies presented previously, we need an efficient and robust characterization of temperature and polarization fluctuations on the 2-sphere. Temperature anisotropies can be described by a scalar field $\delta T/T(\hat{\boldsymbol{n}})$ which does not depend on our choice of coordinates and expresses, as is intuitively clear from its definition, the relative fluctuations of the thermodynamical temperature of the CMB. Linear polarization reaching us from a given angle in the sky can be described by an electric field vector in the plane perpendicular to the propagation direction. We can parameterize it by two numbers known as the linear Stokes parameters Q and U. Choosing a fixed coordinate frame where the electric field lies in the x-y plane, Q quantifies the difference between the squared vertical and horizontal polarization amplitudes, whereas U is the difference between the diagonally oriented squared polarization amplitudes,

$$\begin{aligned} Q &= |E_x|^2 - |E_y|^2 \,, \\ U &= 2|E_x||E_y| \cos 2\chi \,. \end{aligned} \tag{2.1}$$

Here, $|E_x|$ and $|E_y|$ are the absolute values of the x- and y-components of the electric field vector, respectively, and χ is the angle that the electric field vector spans with the x-axis, known as the polarization angle (Rybicki & Lightman, 1985).

In contrast to T, the linear polarization parameters Q and U depend on the how we choose to rotate our coordinate system, calling for an alternative parameterization. As we show in the following, this can be achieved geometrically by recognizing parity-even and parity-odd polarization modes on the 2-sphere, the so-called E- and B-modes. The point of this exercise is partially to be able to robustly compare different experiments, but more importantly, it allows us to single out primordial gravitational waves, which are the only source of polarization B-modes from the early Universe. The T-, E-, and B-modes can then be efficiently compressed into different angular power spectra (TT, TE, EE, BB, etc.), which represent the angular projection of the Fourier power spectrum. For effects imprinted at decoupling, the angular power spectra allow us to translate the three-dimensional comoving scales $1/k$ into angular scales θ as $\theta = 1/(k\chi_*)$, where χ_* is the comoving distance between us and the surface of last scattering (see Sect. 1.1).

Let us now sketch the essential mathematical transformations needed to arrive at the standard characterization of CMB anisotropies on the sphere. The temperature fluctuations $\delta T/T(\hat{\boldsymbol{n}})$ can be re-written in harmonic space by means of the scalar coefficients $a^T_{\ell m}$,

$$a^T_{\ell m} = \int \mathrm{d}\hat{\boldsymbol{n}}\, \frac{\Delta T}{T}(\hat{\boldsymbol{n}}) Y_{\ell m}(\hat{\boldsymbol{n}})\,, \tag{2.2}$$

where $Y_{\ell m}(\hat{\boldsymbol{n}})$ is the spherical harmonic function of degree ℓ at order m. For a linearly polarized field represented by the complex number $P = Q + iU$, where Q and U are the Stokes parameters, we can analogously define the spin-weighted harmonic coefficients ${}_{\pm2}a^P_{\ell m}$ as

$${}_{\pm2}a^P_{\ell m} = \int \mathrm{d}\hat{\boldsymbol{n}}\, P(\hat{\boldsymbol{n}})\,{}_{\pm2}Y_{\ell m}(\hat{\boldsymbol{n}})\,, \tag{2.3}$$

where ${}_{\pm2}Y_{\ell m}(\hat{\boldsymbol{n}})$ are the spin-weighted spherical harmonics. The Stokes parameters Q and U depend on the orientation of the coordinate frame in which we measure polarization on the sphere, which makes them difficult to compare among experiments. To solve this dilemma, Kamionkowski et al. (1997) and Seljak & Zaldarriaga (1997) introduced an alternative, coordinate-independent parameterization that divides the two polarization degrees of freedom into irrotational E-modes and divergence-free B-modes,

$$\begin{aligned} a^E_{\ell m} &= -\frac{1}{2}\left({}_2a_{\ell m} + {}_{-2}a_{\ell m}\right) \\ a^B_{\ell m} &= -\frac{1}{2i}\left({}_2a_{\ell m} - {}_{-2}a_{\ell m}\right) . \end{aligned} \tag{2.4}$$

While E- and B-modes do not change under rotations of the 2-sphere, it can be shown that they have opposite behavior under parity inversion $\hat{\boldsymbol{n}} \mapsto -\hat{\boldsymbol{n}}$: E-modes are not affected while B-modes change their sign under parity inversion.

Most crucially for CMB data analysis, the E- and B-mode decomposition of polarization anisotropies on the sphere has a consequential physical interpretation. It can be mathematically demonstrated (Kamionkowski et al., 1997) that scalar perturbations only source temperature and the E-mode polarization patterns in the CMB, while tensor perturbations lead to temperature, E-, and B-mode anisotropies. This makes polarization B-modes in the CMB a direct tracer of primordial gravitational waves, provided other, late-time sources can be isolated. For this reason, current searches for cosmological gravitational waves focus on primordial B-modes in the CMB polarization. We will explore the expected shape of this primordial signal in the following section.

Our cosmological theory of linear anisotropies is based on the assumption of statistical isotropy, meaning that rotations on the sky do not change the large-scale statistical properties of the CMB. We can make best use of this symmetry by casting our observations into angular power spectra. The angular power spectrum C^{XY}_ℓ is defined as the correlation between spherical harmonic coefficients on the isotropic sky,

$$\langle a^X_{\ell m} (a^Y_{\ell' m'})^* \rangle = \delta^{\mathcal{K}}_{\ell\ell'} \delta^{\mathcal{K}}_{mm'} C^{XY}_\ell , \tag{2.5}$$

where $X, Y \in \{T, E, B\}$ and $\delta^{\mathcal{K}}$ is the Kronecker symbol. We note that isotropy implies that different multipole pairs (ℓ, m) are statistically uncorrelated. In the important special case that X and Y are Gaussian fields with zero mean, Eq. (2.5) implies that their spherical harmonic coefficients are statistically completely described by their variance, given by C^{XY}_ℓ.

Estimating power spectra from a field on the 2-sphere usually relies on the assumption that ensemble averaging $\langle \cdot \rangle$ can be replaced by spatial averaging. It can be shown that an unbiased estimator of the power spectrum is simply given by the average quadratic harmonic coefficient over m:

$$\tilde{C}_\ell \equiv \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2 . \tag{2.6}$$

This estimator is a quadratic form of Gaussian random variables, and as such a random variable itself. The fact that we construct it from a finite set of data points (the $2\ell + 1$ measured harmonic coefficients) implies that this estimate must have a statistical variance itself, known as "cosmic variance". The cosmic variance represents a fundamental statistical limitation of our spherical harmonic basis and cannot be reduced by improving one's experiment. It is worth stressing that angular power spectra are strictly defined on the full sky. If we wish to estimate the temperature and polarization power spectra on the partial sky, as is the case for ground-based experiments, we need to account for the statistical correlation between different pairs of (ℓ, m) and the mixing of E- and B-modes. In particular, we need to know the polarization field on the entire sphere to unequivocally distinguish a B-mode from an E-mode, and vice versa. We address these practical complications when discussing power spectrum estimators in 3.1.

With the power spectrum, we have at hand an efficient summary statistic of the CMB anisotropies. To compare observations with theoretical predictions, we need to be able to compute the expected shape of the CMB temperature and power spectra from cosmological parameters. This can be done by solving the relativistic Boltzmann equation, which describes the time evolution of photon momenta and positions (or, equivalently, Fourier wave vectors) in a perturbed FLRW metric, in the presence of a tight coupling between photons and baryons. This coupling term, quantifying elastic Compton scattering between photons and electrons, generates a hierarchy of coupled equations that can be solved iteratively. Standard numerical Boltzmann solvers, such as the Code for Anisotropies in the Microwave Background (`CAMB`, Lewis et al., 2000; Howlett et al., 2012) and the Cosmic Linear Anisotropy Solving System (`CLASS`, Lesgourgues, 2011), are easy-to-use, well-maintained algorithms that allow fast and accurate computation of angular power spectra given an input set of cosmological parameters. In the next section, we explore the shape of the CMB power spectra and describe how early- and late-time cosmology is encoded in the various features.

2.3 Imprints in the CMB power spectra

Having described the photons' view on cosmic history until decoupling, let us now change perspective and describe how the CMB angular power spectra look when we observe them 13.4 billion years after decoupling. Causality, as expressed through the Hubble horizon, serves as a translator between cosmic time and angular scales, allowing us to view the CMB power spectra almost like a screenplay of cosmic history. As we argued in the previous section, this includes the very early eras such as inflation as a possible origin of primordial perturbations, and later events, such as baryonic acoustic oscillations and photon decoupling, which marks the birth of the CMB. Although in the post-decoupling era, photons streamed through the Universe almost undisturbed, they kept interacting with their environment, recording late-time events such as cosmic reionization. By using numerical Boltzmann solvers such as `CAMB`, we can predict the exact shapes of the CMB power spectra as a function of the cosmological parameters.

In this section, we have a closer look at the features in the CMB angular power spectra and

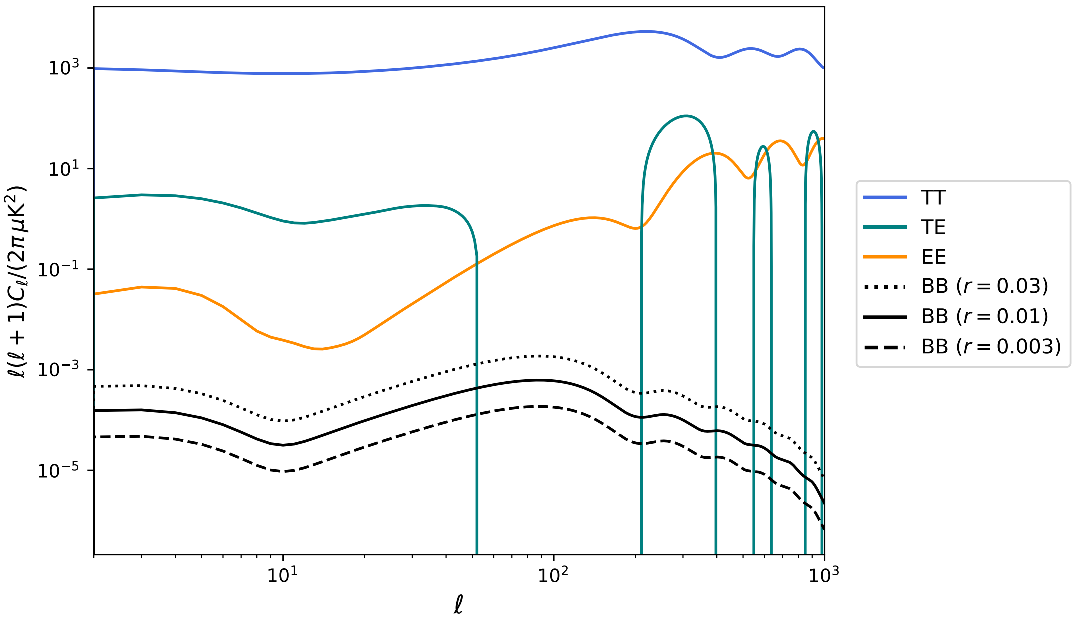

Figure 2.2: Temperature and polarization angular power spectra expected from a standard ΛCDM cosmology. We show the TT, TE, and EE spectra resulting from primordial scalar perturbations, with their expected imprints from recombination and late-time physics. For comparison, we show the tensor-only BB power spectrum (no lensing B-modes) for three different values of the tensor-to-scalar ratio r.

discuss how they vary along with the cosmological parameters of the standard ΛCDM model, considering an optional nonzero gravitational wave amplitude. We can categorize these imprints into polarization-independent effects, such as baryonic acoustic oscillations, as well as polarizing effects, such as the large-scale reionization bump. We start with the larger temperature and E-mode power spectra, before investigating the much smaller B-modes, which are a main focus of this Thesis.

2.3.1 General effects

Figure 2.2 shows the theoretical predictions of the TT, TE, and EE angular power spectra, calculated using `CAMB` assuming the *Planck* 2018 best-fit ΛCDM cosmology. For comparison, we also show the expected BB power spectrum from primordial perturbations assuming different values of the tensor-to-scalar ratio r. We can see that due to the low polarization fraction of the CMB of about 10%, the TT power spectrum is the largest one, while the TE and EE powers are smaller by 1-2 and 2-4 orders of magnitude, respectively. We start by describing a set of general, non-polarizing effects related to the era of decoupling and later times, which leave characteristic, calculable imprints in all

CMB temperature and polarization power spectra. Later, we describe polarizing effects responsible for additional features in the BB, EE, and TE power spectra.

- **Sachs-Wolfe (SW) effect:** At angular scales larger than the horizon at decoupling, $\ell \lesssim 100$, the temperature perturbations could not have causally evolved before decoupling. Instead, the temperature anisotropies observed today are equal to the temperature anisotropies at decoupling minus a gravitational redshift contribution. The latter effect expresses the photon energy lost by climbing out of the local gravity well at decoupling, while the temperature anisotropies are themselves related to local overdensities at decoupling. For a matter-dominated Universe with adiabatic density perturbations, one can show that both of these terms combine to an observed temperature anisotropy of $\delta T/T = -\Phi/3$, where Φ is the local gravitational potential at decoupling (Sachs & Wolfe, 1967; White & Hu, 1997). This effect is visible as a "valley" of nearly constant amplitude in the TT autospectrum at multipoles $\ell \lesssim 100$.

- **Integrated Sachs-Wolfe effect (ISW):** The largest observed angular scales are linked to the late-time Universe, where the Hubble horizon was close to the size of the observable Universe. At those late times, the Universe started to be dominated by dark energy, which makes gravitational potentials grow faster than the surrounding Universe (Rees & Sciama, 1968). As a result, CMB photons traversing the cosmic LSS lose some net energy beyond the term $-\Phi/3$ from the ordinary SW effect, creating a slight upturn in the CMB at the largest multipoles of the TT autospectrum.

- **Baryonic acoustic oscillations (BAOs):** In the presence of potential wells generated from primordial perturbations, the baryon-photon plasma performed acoustic oscillations driven by the competition between gravity and pressure. The spatial scale of the gravity wells determined the oscillation frequency ω via $k = \omega c_s$, relating the perturbation scale k to the phases of the acoustic peaks. In particular, the first peak in the angular power spectrum at $\ell \sim 220$ corresponds to the acoustic mode that had just reached the first peak by the time of decoupling (Planck Collaboration I, 2016). The second peak at $\ell \sim 538$ corresponds to the acoustic mode reaching its first minimum at decoupling, and so on. The scale associated with the peaks is the sound horizon at last scattering, one of the six cosmological parameters of the standard ΛCDM model, defined as

$$\theta_* \equiv \frac{r_*}{\chi(t_*)}\,, \tag{2.7}$$

  where r_* is the comoving sound horizon at decoupling and $\chi(t_*)$ is the comoving radius, defined in Eq. (1.5), evaluated at the time of decoupling, t_*. The position of the acoustic peaks probes the background expansion parameters H_0, Ω_m and the spatial curvature Ω_K to high precision, while their amplitude is sensitive to the primordial power spectrum amplitude A_s and the fractional energy density of baryons, Ω_b. The acoustic peaks are the dominant feature at intermediate to small scales ($200 \lesssim \ell \lesssim 2000$) in TT, TE, and EE.

- **Reionization:** The ignition of the first ionizing sources about 200 million years after decoupling led to the reionization of the intergalactic medium. The newly freed electrons started undergoing Compton scattering with CMB photons, partially erasing anisotropies at scales smaller than the Hubble horizon at the beginning of reionization. This leads to a decrease in the CMB TT, TE, and EE power spectra at all but the very large angular scales ($\ell \gtrsim 20$) by a factor $e^{-2\tau}$, where τ denotes the optical depth to reionization

  $$\tau \equiv \int_{t_*}^{t_0} \sigma_T \, n_e(t) \, \mathrm{d}t \, . \tag{2.8}$$

  Here, t_0 is the present time, σ_T is the Thomson scattering cross section, and $n_e(t)$ is the electron number density. The optical depth τ represents a late-time parameter of the standard ΛCDM model of cosmology.

- **Damping:** The decoupling process is not instantaneous, and therefore leads to a suppression of TT, TE, and EE power at small angular scales $\ell \gtrsim 2000$ that correspond to the thickness of the surface of last scattering. Alternatively, we can interpret this suppression of the acoustic peaks as the effect of imperfect coupling between photons and baryons, leading to a diffusion damping of the small-scale acoustic peaks, which have undergone more oscillations than the larger-scale peaks.

2.3.2 Polarization-specific effects

Let us now focus on polarization specific effects that do not effect the temperature anisotropies but generate features in the BB, EE, and TE power spectra. As can be seen from Fig. 2.2, the expected BB power spectrum from primordial perturbations is smaller than EE by about two orders of magnitude. This is due to the fact that B-modes cannot be generated from density perturbations, as we saw in Sect. 2.2. Hence, unlike E-modes, they do not inherit the prominent small-scale acoustic peak structure that is generated at decoupling. Instead, there are three more subtle effects that dominate the features of the BB angular power spectrum, which we discuss here.

- **Recombination peak:** Primordial tensor perturbations, or gravitational waves, are commonly assumed to originate from an inflationary phase in the very early Universe, producing an almost scale-invariant power spectrum (see Sect. 1.1.3), with an amplitude that scales linearly with the tensor-to-scalar ratio r. They are imprinted in the CMB polarization B-modes at the time of decoupling, which leads to the "recombination peak" in the BB power spectrum around angular scales $\ell \sim 90$, corresponding to the horizon scale at the time of decoupling (Zaldarriaga & Seljak, 1997). Similarly, scalar perturbations are imprinted into polarization E-modes at decoupling, causing a recombination peak in the EE and in TE power spectra.

- **Reionization bump:** The second time that primordial perturbations would be able to generate E- and B-mode polarization patterns is during the time of reionization. When the Universe stopped being electrically neutral, the largest causally interacting scales contributed to today's observed angular scales of $\ell \sim 20$, forming a large-scale "reionization bump" in the BB, EE, and TE power spectra. Smaller-scale perturbations had already entered causal interactions before, so they are suppressed. Larger scales have been scattering with the CMB in the later Universe ever since, so they all contribute to the reionization bump. The amplitude of the reionzation bump is approximately proportional to τ^2 (Reichardt, 2016; Kaplinghat et al., 2002).

- **Gravitational lensing:** Nonlinear distortions of the CMB polarization field from the gravitational lensing effect caused by the late-time matter distribution convert E-modes into B-modes. The lensing B-mode signal can be viewed as an additional source of stochastic noise, with an almost white power spectrum of about 5 μK-arcmin amplitude on the largest angular scales (Okamoto & Hu, 2003), which is expected to be the dominant contribution at multipole scales $30 \lesssim \ell \lesssim 300$ probed by the SO SATs. We can artificially scale this "lensing template" by a factor $A_{\rm lens}$, mimicking the practice of subtracting the lensing contribution on the map level ("delensing", Kesden et al., 2002; Smith et al., 2012).

Figure 2.3 shows how variations in the tensor-to-scalar ratio r, the optical depth τ, and the amplitude of gravitational lensing, $A_{\rm lens}$, affect the features in the BB power spectrum.

2.4 Current measurements of power spectra

Having established a theoretical understanding that allows us to connect CMB power spectra with cosmological parameters, we can now shift our attention to experimental constraints. In this section, after briefly reviewing the highlights in past CMB observations, we discuss today's state-of-the-art power spectrum measurements. We specifically focus on the 2018 legacy results from the *Planck* satellite and the latest constraints on polarization B-modes from the BICEP/*Keck* experiment.

Since the detection of the CMB in 1964, experimental cosmology has arguably seen its largest success story so far. The CMB had been predicted to be essentially a blackbody of a few Kelvin

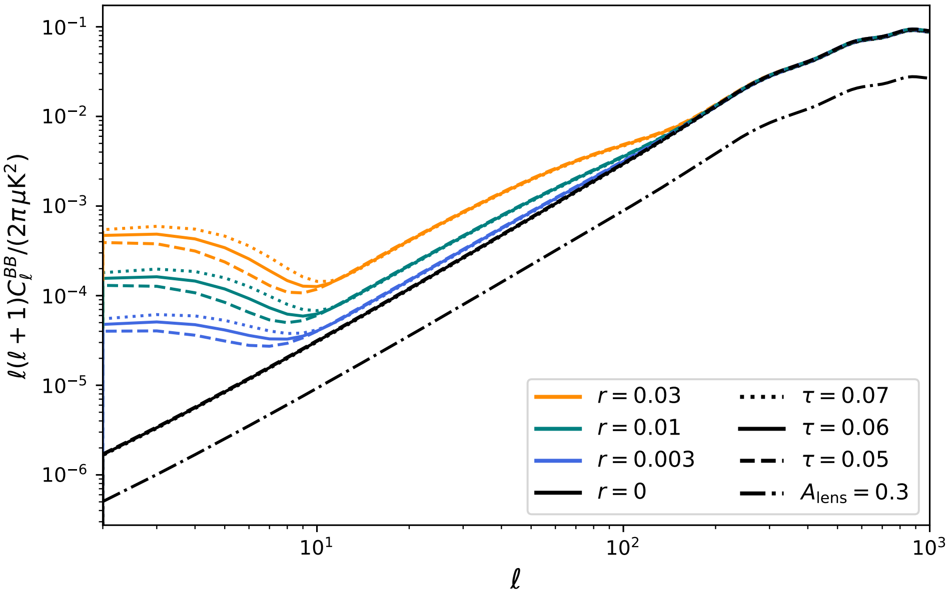

Figure 2.3: BB angular power spectrum for different values of the the primordial tensor-to-scalar ratio (color variation) and different values of the reionization optical depth, (line style variation). We also plot a tensor-free spectrum which has the lensing B-mode contribution reduced ("delensed") to 30%.

temperature, with small anisotropies coming from primordial perturbations. With the remnant radiation of the big bang at their grasp, scientists started to design the first experiments to map the CMB anisotropies and determine its spectral distribution, but it took 25 years to realize the first survey that could accomplish both of these tasks with the necessary precision. In 1989, the COBE satellite was launched on a two-year mission to measure the spectral distribution of the CMB between 60 GHz and 3 THz on the full sky, at seven degrees angular resolution. COBE showed that the CMB monopole is perfectly fit a blackbody spectral distribution with a temperature of 2.725 ± 0.002 K and achieved the first measurement of small, 100μK-level anisotropies in the CMB angular distribution (Mather et al., 1999). These findings confirmed theoretical expectations and encouraged the formation of larger science collaborations that would later build instruments able to exploit the cosmological information within the CMB anisotropies. In 2001, the WMAP satellite was launched to map the full sky of CMB radiation anisotropies in five frequency bands between 23 and 90 GHz at an angular resolution of 0.2 degrees. The mission was a groundbreaking success, establishing the spatially flat ΛCDM model as the standard cosmological model and measuring the five early-Universe parameters to better than five percent accuracy (Hinshaw et al., 2013). WMAP also performed the first full-sky measurements of the polarized CMB and Galactic foregrounds at 23 GHz. In 2009 followed the latest CMB survey in space, the European Space Agency's

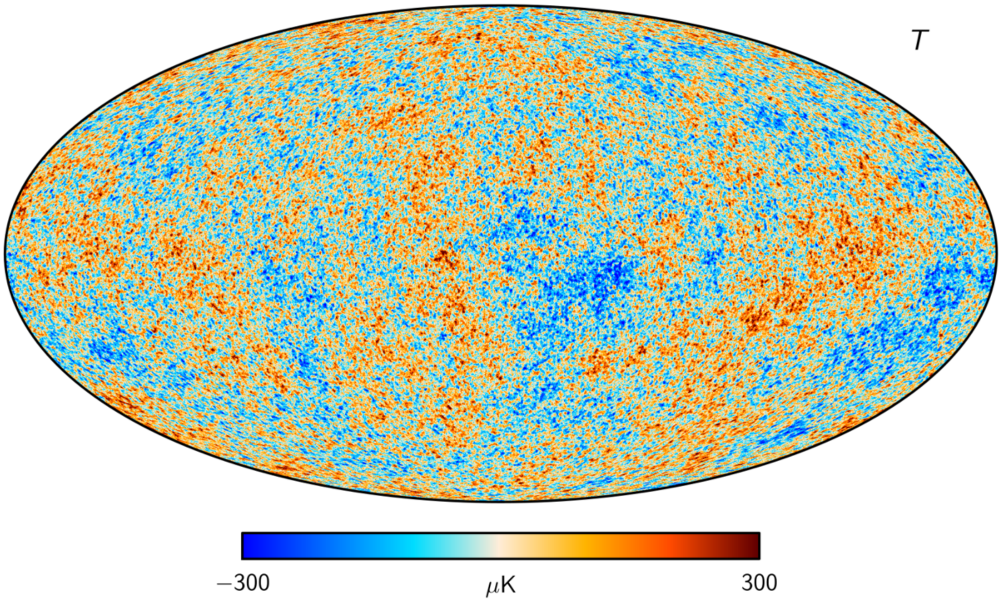

Figure 2.4: The *Planck* CMB temperature map as produced from multifrequency maps using the `Commander` component separation algorithm. Image is taken from Planck Collaboration I (2020).

Planck satellite, to further refine WMAP's results and provide more sensitive, higher-resolution measurements in temperature and polarization, which we describe in the following.

2.4.1 *Planck* temperature and polarization

The European Space Agency's *Planck* satellite mission (Planck Collaboration I, 2020), aimed at further constraining the parameters of the ΛCDM model beyond what WMAP found, by measuring the maps of CMB anisotropies at higher sensitivity and down to smaller scales. *Planck* aimed at full-sky polarization measurements that had been pioneered by WMAP, probing Galactic foreground emission as a non-negligible contaminant to the CMB. *Planck* carried two instruments, the Low Frequency Instrument (LFI) and the High Frequency Instrument (HFI) that measured the sky at nine different frequency bands from 23 GHz to 857 GHz, among which the lower seven frequency channels from 23 GHz to 353 GHz were sensitive to polarization. Both instruments scanned the full sky in a four-year mission, having wto main science goal: constraining cosmology with the CMB, and characterizing Galactic and extragalactic emission for studying astrophysics and improving CMB foreground subtraction.

Planck pioneered component separation, the use of multifrequency maps to distinguish sky components by means of their spectral differences and produce cleaned component maps. The *Planck* science team developed several component separation algorithms, based on different methods to enable robust comparison:

- `Commander` (Planck Collaboration X, 2016; Eriksen et al., 2008) is a Bayesian parametric fitting algorithm using templates in pixel space.

- **Needlet Internal Linear Combination** (`NILC`, Basak & Delabrouille, 2012, 2013) is a model-independent ("blind") cleaning algorithm that aims at removing all sky components with spectral emission laws different from the CMB.

- **Spectral Estimation Via Expectation Maximisation** (`SEVEM`, Leach et al., 2008; Fernández-Cobos et al., 2016) is an internal template-cleaning approach in pixel space that produces clean CMB maps typically constructed as difference maps between two neighboring *Planck* channels.

- **Spectral Matching Independent Component Analysis** (`SMICA`, Cardoso et al., 2008) uses an independent power-spectrum-based algorithm to synthesize clean CMB maps.

Figure 2.4 shows the map of *Planck*'s CMB temperature anisotropies produced by the `Commander` algorithm. Component-separated maps from *Planck* include polarized CMB maps at about 5 arcmin angular resolution, and maps of polarized Galactic foregrounds at an angular resolution of 80 arcmin or better, most prominently thermal dust and synchrotron emission, which are the most detailed full-sky maps of Galactic foregrounds to date. We address these in more detail in Sect. 2.6.

The *Planck* 2018 legacy data include angular power spectra of the CMB temperature and polarization. Figure 2.5 shows the TT, EE, and TE power spectra with error bars, which include the uncertainty from foreground subtraction and the expected sampling variance associated to observing a limited number of modes on the sky. The lower right panel of Fig. 2.5 shows the power spectrum of the gravitational lensing potential, which represents the angle-averaged line-of-sight integral of the gravitational potential of the matter distribution in the late Universe, causing a distortion in the CMB power spectrum at small scales. The blue lines show the theoretical prediction from the best-fit flat ΛCDM model which, as we can see, match the observed data to an astonishing precision and accuracy. The intermediate scales of these power spectra are measured at high S/N which allowed the detection of eight acoustic peaks. The TT and TE power spectra are dominated by sample variance at most scales $\ell \lesssim 1800$, essentially exhausting all cosmological information available, while the error bars of the EE power spectrum is dominated by instrumental noise. In addition, *Planck* measured a null low-ℓ BB polarization power consistent with instrumental noise.

The *Planck* power spectra directly constrain six primary parameters in the flat ΛCDM scenario. Adopting the convention that the Hubble parameter at present is $H_0 = 100h$ km/s/Mpc, these parameters and their values (with 68% central CL) as measured by *Planck* are: the density of cold dark matter, $\omega_c \equiv \Omega_c h^2 = 0.1198 \pm 0.0012$; the density of baryons, $\omega_b \equiv \Omega_b h^2 = 0.02233 \pm 0.00015$; the amplitude, $\ln(10^{10} A_s) = 3.043 \pm 0.014$, and spectral index, $n_s = 0.965 \pm 0.004$, of

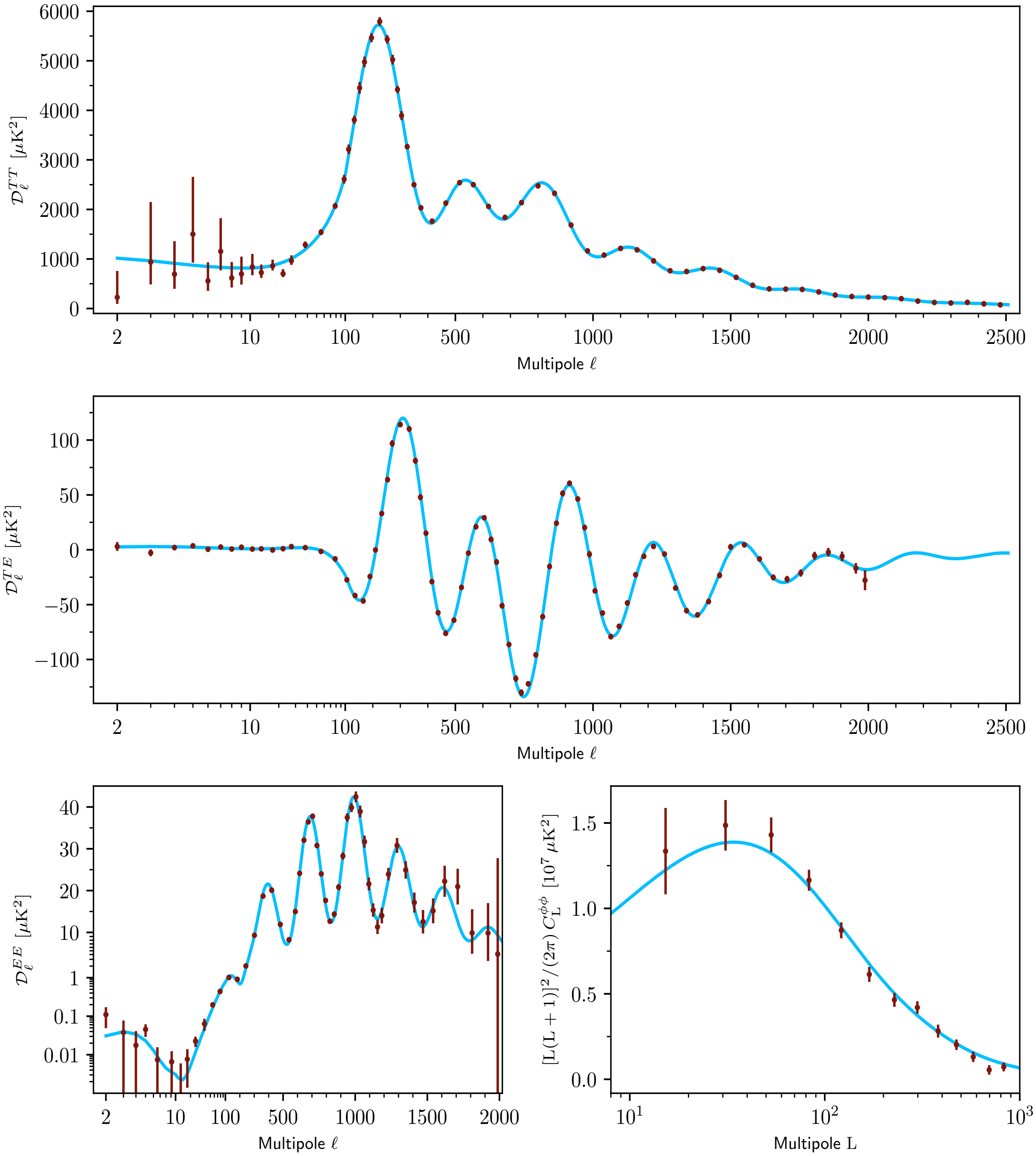

Figure 2.5: The *Planck* CMB TT, TE, and EE angular power spectra as published in the 2018 legacy release (Planck Collaboration I, 2020). These are foreground-subtracted and frequency-averaged C_ℓs representing the temperature autocorrelation (*top*), the temperature-polarization cross-correlation (*middle*), the E-mode polarization autocorrelation (*bottom left*), and the gravitational lensing potential (*bottom right*). The anisotropy power spectra are plotted with a multipole axis that goes smoothly from logarithmic at low ℓ to linear at high ℓ. In all panels, the blue line is the best-fit Planck 2018 model, based on the combination of TT, TE, and EE.

a power-law spectrum of adiabatic perturbations; the angular scale of the acoustic oscillations, $100\theta_* = 1.0411 \pm 0.0003$; and the optical depth to reionization, $\tau = 0.054 \pm 0.007$ (Planck Collaboration VI, 2020). While the combinations ω_c, ω_b are primarily constrained by the relative height of the acoustic peaks, the sound horizon θ_* determines the absolute peak position. The overall normalization of all power spectra at post-reionization scales $\ell \gtrsim 20$ is sensitive to the primordial power spectrum amplitude A_s, in combination with the optical depth as $A_s e^{-2\tau}$. This degeneracy can be broken by measuring the reionization bump in EE at $\ell \lesssim 20$, as we shall address in more detail in Chpt. 5.

In summary, the measurement of the *Planck* temperature and E-mode power spectra has established the flat ΛCDM model as the simplest and best-fitting "standard" model of cosmology, although several fundamental questions on its validity remain open (see Sect. 1.1.3). We might obtain answers to some of these questions from experimentally constraining primordial polarization B-modes, enabling us to crucially narrow down the space of viable models of the very early Universe, including inflation. This is the ultimate goal of several future experiments that we introduce in Sect. 2.5. In the following, we briefly present the BICEP/*Keck* experiment, a ground-based telescope array that has achieved the tightest constraints on primordial B-modes at the time of writing (2023).

2.4.2 Degree-scale polarization B-modes

While low-noise TT, TE, and EE power spectra from Planck have corroborated the ΛCDM model, recent and ongoing ground-based experiments such as BICEP/*Keck* have since been further constraining the amplitude of the CMB BB power spectrum.

The BICEP and *Keck* array (hereinafter BK) are two CMB telescopes located at the South Pole at 2800 m altitude. Since 2006, several surveys have been conducted, each including a number of instrument upgrades. The 2021 science results included measurements of the polarized microwave sky in three frequency bands between 95 and 220 GHz on about 2% of the sky in a patch centered around the south celestial pole. To constrain polarization B-modes, BK measures the low-resolution polarized sky at angular scales $50 < \ell < 200$, which are only weakly contaminated by lensing B-modes, and combine data with foreground-dominated channels from WMAP and *Planck* to characterize the amplitude of polarized Galactic foregrounds. BK uses a multifrequency power-spectrum-based analysis pipeline to model and marginalize over foreground signals in a parametric Bayesian approach.

The CMB community's progress on constraining low-ℓ B-modes is tightly connected with Galactic science via the characterization of degree-scale polarized Galactic foregrounds. In 2014, the BICEP2 Collaboration had reported a detection of primordial B-modes at a significance of $> 5\sigma$ (BICEP2 Collaboration, 2014), a signal that, as later analyses of *Planck* data in the same sky region revealed, can be entirely ascribed to Galactic thermal dust emission (Planck Collaboration Int. XXX, 2016; BICEP2/Keck Collaboration & Planck Collaboration, 2015). In recent years, thermal dust and synchrotron emission (Krachmalnicoff et al., 2018) have shown to be a ubiquitous and significant contaminant to degree-scale B-modes, a topic we discuss in more detail in Sect. 2.6. The current state-of-the-art constraints on the tensor-to-scalar ratio, based on the 2018 BK data combined with foreground-sensitive archival data from *Planck*, give $r \lesssim 0.032$ at 95% CL (Tristram et al., 2022). The successor experiment of BICEP/*Keck*, called BICEP Array, is projected to improve the statistical uncertainty of currently $\sigma(r) = 0.009$ (68% central CL) to $\sigma(r) \lesssim 0.003$ by 2027 (BICEP2/Keck Collaboration, 2022).

At small to intermediate angular scales, the predominant effect in the BB power spectrum of the CMB is caused by gravitational lensing by the late-time matter distribution (see Sect. 2.3). This effect was first observed in 2013 by the POLARization of the Background Radiation experiment (POLARBEAR, POLARBEAR Collaboration, 2014)[4], which estimated the lensing signal from the four-point correlation function of E- and B-modes observed at a 30-square-degree patch of 2 arcsec map resolution in the southern sky. More recent *Planck* data were able to detect this signal at a significance of about 40σ (Planck Collaboration VIII, 2020).

Figure 2.6 shows the most recent constraints on the angular power spectra of the CMB polarization B-modes in comparison to the theoretically expected lensing signal and the conjectured signal from primordial gravitational waves. The figure highlights measurements by BICEP/*Keck* for intermediate to large scales, as well as recent constraints of the lensing signal by POLARBEAR (POLARBEAR Collaboration, 2017), the South Pole Telescope (SPT, Sayre et al., 2020) and the Atacama Cosmology Telescope (ACT, Choi et al., 2020).

To summarize, the ΛCDM parameters are well constrained from temperature and E-mode polarization data, while measuring degree-scale primordial B-modes requires significant efforts in constraining foregrounds at multiple frequencies, and building experiments that achieve high sensitivity. Moreover, reducing the statistical uncertainty on the BB power spectrum warrants subtracting the intermediate-scale lensing B-modes on the map level, by a method called “delensing”. This requires a synergy between experiments that measure the large scales (typically SAT-like telescopes on the ground or in space) and LAT-like telescopes that measure the lensing B-modes at sub-degree scales in the same area on the sky. This is especially relevant for high-sensitivity future

---

[4] URL: lambda.gsfc.nasa.gov/product/polarbear

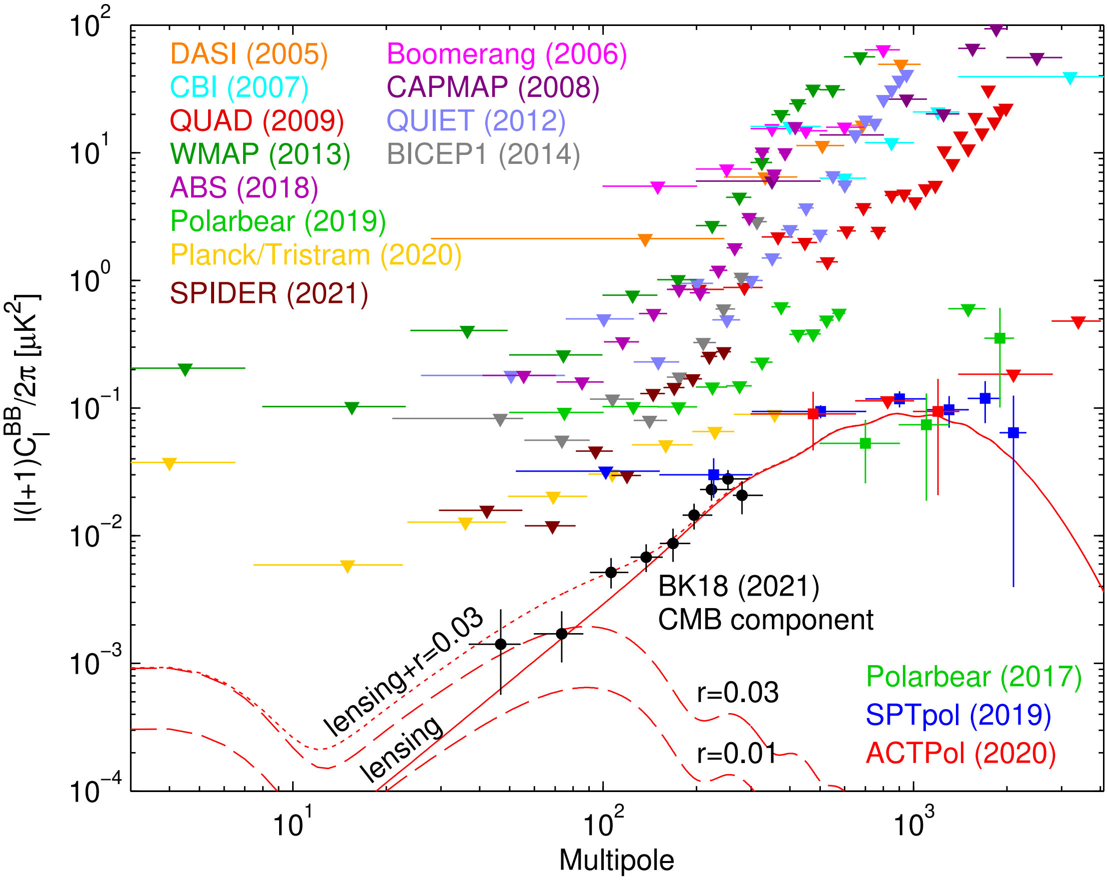

Figure 2.6: Current constraints on B-mode power spectra at large and intermediate scales (markers with error bars; upper limits are given by downward pointing triangles), compared to the theoretical scenarios of lensing B-modes combined with different levels of primordial gravitational waves ($r = 0.03$ or $r = 0$). The current upper limit on primordial gravitational waves is given by $r < 0.032$ at 95% CL (Tristram et al., 2022). This figure is adopted from BICEP2/Keck Collaboration (2021), for the original publications of the data sets see references therein.

experiments, as we shall see in the next section.

2.5 Future experiments

In this section, we outline three representative next-generation CMB experiments to carry on the legacy of WMAP and *Planck* in constraining our cosmological model. As we shall see in the following, future CMB experiments are heavily investing in the prospect of gaining a new understanding of the very early Universe through the search for primordial gravitational waves. Being the main focus of Chpts. 3 and 4, we start by describing the SO experiment, a fully funded ground-based CMB survey expected to receive its first light by 2024. Looking just a few years ahead, we then continue with the LiteBIRD and CMB-S4 experiments, a next-generation satellite mission and the first "Stage 4" ground-based experiment.

2.5.1 Simons Observatory

The Simons Observatory (SO) is a ground-based CMB telescope array located in the Atacama desert, Chile, at about 5200 meters altitude. Its experimental setup consists of three 42cm-diameter SATs, to measure intermediate to large scales on one northern and one southern sky patch covering a total of about 10% of the sky, and one 6m-diameter LAT to measure small scales on about 40% of the sky, at six frequency bands between 27 and 280 GHz. Each SAT contains a CRHWP to demodulate the polarized signal and thus avoid large-scale noise related to the atmosphere. SO aims at measuring primordial B-modes and non-Gaussianity, probing late-time cosmology and astrophysics through small-scale observations of gravitational lensing and maps of CMB spectral distortions caused by high-energy electrons in galaxy clusters. Further science goals of SO are mapping the Galaxy to unprecedented precision, and using high-sensitivity polarization as an independent cross-check of the ΛCDM cosmological model. In particular, the nominal instrument design of SO is predicted to constrain primordial gravitational waves at a precision of $\sigma(r) \lesssim 0.003$ after five survey years (SO Collaboration, 2019). This would allow us to detect, or rule out, a power of primordial gravitational waves corresponding to $r = 0.01$ at 2-3σ significance, thereby crucially constraining the models of the primordial Universe. Figure 2.7 shows SO's predicted constraints on primordial parameters in the r-n_s plane in comparison with an exemplary selection of inflationary models that could be constrained, or ruled out, assuming a vanishingly small tensor-to-scalar ratio ($r = 0$) or primordial gravitational waves with $r = 0.01$.

The LAT will observe a patch covering about 40% of the sky at arcminute resolution, chosen to optimize the overlap with the two SAT patches and the target regions of near-future galaxy surveys, such as the Dark Energy Spectroscopic Instrument (DESI, DESI Collaboration, 2016) and the Rubin

Observatory/Legacy Survey of Space and Time (LSST, Ivezić et al., 2019), enabling combining and cross-correlating probes. The SAT sky patches cover about 10% of the sky while avoiding the highly foreground-contaminated Galactic plane, and are wide enough to constrain multipoles down to $\ell \sim 30$ while keeping the variation in elevation, and the resulting ground pickup, to a minimum. The frequency channels include two CMB-sensitive channels centered at 93 GHz and 145 GHz, two low-frequency channels at 27 GHz and 39 GHz to constrain Galactic synchrotron emission, and two high-frequency channels at 225 GHz and 280 GHz, designed to characterize the Galactic thermal dust component.

Using three highly sensitive low-resolution SATs serves the key science purpose of constraining degree-scale B-modes at the required low noise level, while the LAT measures the small-scale lensing B-modes in a fully overlapping sky patch in order to delens the SAT B-modes. Apart from this, the LAT enables the pursuit of many more small-scale science goals that we address below. The polarized SATs and LAT sky signals are both contaminated by various foregrounds at different scales, such as diffuse Galactic synchrotron and dust at the SAT large scales, and extragalactic radio point sources and diffuse late-time signals such as the cosmic infrared background (CIB) at the smaller LAT scales. Those contaminants must be subtracted using different methods according to the required sensitivity. Degree-scale Galactic foregrounds for the SATs require component separation algorithms that make use of multifrequency observations, as described in Chpt. 4. Depending on the scientific goals, foregrounds on the LAT are removed either through blind component separation methods such as `ILC` (Bennett et al., 2003; Eriksen et al., 2004), and/or deprojection of the sky contaminants (see Sect. 2.5 of SO Collaboration, 2019).

The Simons Observatory Collaboration has identified six key science goals for cosmology (SO Collaboration, 2019), which we briefly present here.

1. **Primordial perturbations.** The SO SATs target the measurement of primordial gravitational waves through degree-scale B-modes at a baseline precision of $\sigma(r) = 0.003$ (68% central CL), which would allow to constrain models of the very early Universe. Moreover, measurements of the total intensity and E-mode power spectra with the SO LAT will help tighten the constraints on scalar primordial perturbations, with a predicted 0.5% precision on the parameter combination $e^{-2\tau} A_s$. Finally, SO aims at constraining the local non-Gaussianity of primordial perturbations to $\sigma(f_{\mathrm{NL}}) = 2$ at 68% central CL, by cross-correlating specific effects in the LAT CMB anisotropies (the kinetic Sunyaev-Zel'dovich (kSZ) effect tracing the bulk motion of high-energy electrons in galaxy clusters (Coulton et al., 2020; Sunyaev & Zeldovich, 1980) and the gravitational lensing effect on the CMB, respectively) with external galaxy survey maps.

2. **Effective number of relativistic species.** Through their effects on the expansion rate, primordial element abundances and radiation perturbations affect the damping and the position of the acoustic peaks in the temperature, polarization E-modes, and TE cross-power spectra. The SO LAT can constrain these effects and help improve current measurements of the effective number of relativistic species in the early Universe, with a projected baseline precision of $\sigma(N_{\text{eff}}) = 0.07$ at 68% central CL.

3. **Neutrino mass.** SO aims at constraining the upper limit of the total mass of neutrinos to $\sigma(\sum m_\nu) = 40$ meV (SO baseline, 68% central CL), which leaves an imprint in the small-scale lensing distortions and the acoustic peak positions in the CMB power spectra through the total matter density, as well as galaxy clustering properties. These three effects can be exploited in different ways via precision measurements by SO LAT, in combination with external galaxy clustering data.

4. **Deviation from Λ.** The matter power spectrum holds information on the late-time expansion history and possible deviations from the cosmological constant scenario. The SO LAT can constrain the root mean square of matter fluctuations on an 8 Mpc/h scale, σ_8, through several means that involve galaxy clustering and gravitational lensing of the CMB, measured in combination with external data sets. In addition, better indirect constraints on the Hubble constant, achieved through precision measurements of the acoustic peak positions, can indicate deviations from Λ. The SO baseline scenario predicts a 2% precision measurement of σ_8 at redshift 1-2, and a precision on the Hubble constant of $\sigma(H_0) = 0.4$ km/s/Mpc at 68% central CL, respectively.

5. **Galaxy evolution.** The kinetic and thermal Sunyaev-Zel'dovich effects are well-characterized local spectral distortions of the CMB caused by elastic scattering between CMB photons and high-energy electrons present in galaxy clusters (Sunyaev & Zeldovich, 1980). By measuring the Sunyaev-Zel'dovich effects in massive halos, SO aims at informing and refining models of galaxy evolution, measuring the efficiency of energy injection by feedback to 2% precision and the fraction of non-thermal pressure to 6% precision (68% central CL; for more details, see Battaglia et al., 2017). These constraints will be derived from the LAT temperature maps, combined with galaxy positions measured by the DESI spectroscopic galaxy survey.

6. **Duration of reionization.** The kSZ effect provides a way to detect additional variance in the temperature CMB attributed to the patchy reionization paradigm, and thus constrain the duration of the epoch of reionization, Δz_{re} (Calabrese et al., 2014). This is possible through measurements of the CMB temperature and E-mode polarization anisotropies by the SO LAT.

SO baseline constraints predict a measurement precision of $\sigma(\Delta z_{\mathrm{re}}) = 0.6$ at 68% central CL.

Secondary science targets of SO include constraining big bang nucleosynthesis, the nature and interactions of dark matter, and the nonlinear growth of structure. Moreover, SO will be able to map extragalactic sources, detect transient sources, and constrain Galactic science (Hensley et al., 2022).

To summarize, SO is a ground-based CMB experiment in the Chilean Atacama desert which will measure CMB polarization and temperature anisotropies with the main goals of constraining primordial gravitational waves and the early Universe, the neutrino mass, the reionization epoch, and late-time physics such as dark energy and galaxy evolution. With two complementary types of telescopes, three SATs and one LAT, and in six frequency bands, SO will delens the degree-scale B-mode signal and characterize degree-scale Galactic foregrounds to be able to achieve a robust detection of $r = 0.01$ at 95% CL, by establishing a statistical uncertainty of $\sigma(r) = 0.003$ at 68% central CL, improving on current constraints from Tristram et al. (2022) by a factor of three. With the planned start of full science observations in 2024, SO is expected to reach its nominal science goals after five years of observations. Subsequently, SO will obtain two additional SATs from UK funds and one extra SAT from Japanese funds, together with a US-funded doubling of the number of optics tubes in the LAT. These upgrades combined will increase the observatory's total number of detectors from nominally 60,000 to 100,000, which will subsequently be renamed the Advanced Simons Observatory. In the next section, we present two other future CMB experiments, the LiteBIRD satellite, expected to launch in the late 2020s, and the ground-based CMB-S4 experiment, designed as the first "Stage 4" CMB experiment in the early 2030s.

2.5.2 LiteBIRD and CMB-S4

The late 2020s and early 2030 will see the arrival of two ambitious future CMB experiments, LiteBIRD and CMB-S4. While the former is a satellite mission with a narrow focus on constraining primordial B-modes on the full sky, the latter is poised to constrain a wide range of cosmological parameters in an ambitious, multi-site ground-based CMB experiment.

LiteBIRD (Hazumi et al., 2020) is a satellite commissioned by the Japan Aerospace Exploration Agency scheduled to launch in the late 2020s and perform a three-year full-sky survey. Its primary goal is to constrain large-scale B-modes in order to measure the tensor-to-scalar ratio at a precision of $\sigma(r) \sim 0.001$ (60% central CL), and either detect gravitational waves associated with specific inflationary models, or set upper limits to rule out a large class of models of the very early Universe. In particular, LiteBIRD's indispensable contribution to CMB science will be a full-sky

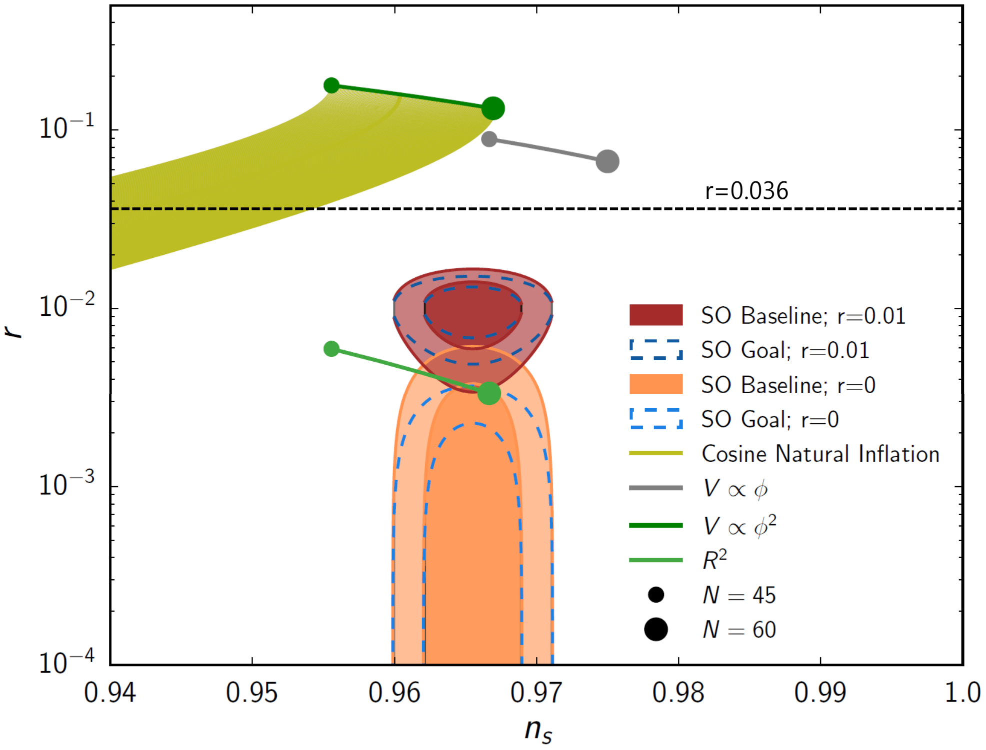

Figure 2.7: Predicted constraints on the primordial tensor-to-scalar ratio r and the scalar spectral index n_s achieved by a five-year SO nominal survey, assuming input models with $n_s = 0.965$, and either $r = 0$ or $r = 0.01$. The contours correspond to 68% and 95% central CL for SO baseline (solid contours) and SO goal constraints (dashed contours). The figure also shows predictions by some models of cosmic inflation assuming that the cosmic scale factor grew by a factor e^N, with $N \in [45, 60]$, in line with current expectations. Inflationary models considered are: Cosine Natural inflation (Freese et al., 1990), Starobinsky (R^2) inflation (Starobinsky, 1980; Starobinskij, 1992), and ϕ^n inflation (Linde, 1983; McAllister et al., 2010). The horizontal dashed black line indicates the 95% CL upper limit on primordial gravitational waves from BICEP2/Keck Collaboration (2021). This figure was taken from SO Collaboration (2019).

survey of the polarized microwave sky, including a high-sensitivity measurement of the large-scale reionzation bump. Upcoming ground-based experiments, such as SO, BICEP Array, and CMB-S4, will provide complementary small-scale measurements necessary to delens large-scale B-modes measured by LiteBIRD. LiteBIRD's polarization-sensitive transition-edge sensor (TES) array will observe in 15 frequency bands between 34 and 448 GHz, and is expected to reach a typical angular resolution of 0.5 degrees at 100 GHz. LiteBIRD will be the first full-sky CMB experiment since *Planck*, and its increased sensitivity and better control of systematics will allow to probe large-scale polarization effects such as the optical depth to reionization τ and the sum of neutrino masses, cosmic birefringence, primordial magnetic fields, spectral distortions in the CMB, and large-scale anomalies observed in *Planck* temperature maps.

CMB-S4 (Carlstrom et al., 2019) is a next-generation ground-based experiment of ambitious dimension that is designed to set new limits on primordial physics and the cosmic thermal history, and provide a "centimeter to millimeter legacy data set" for astrophysics and cosmology. Scheduled for two seven-year surveys in the 2030s, CMB-S4 will consist of 21 telescopes located at the South Pole and in the Chilean Atacama desert. One one hand, a low-resolution B-mode survey will observe three percent of the sky through one LAT and up to 18 SATs in nine frequency bands between 20 and 270 GHz, targeting a noise level below 1μK-arcmin. On the other hand, a wide and deep survey set out to scan 70% of the sky through two additional 6m LATs will produce a 1.5 arcmin-resolution legacy data set to probe small-scale temperature and polarization anisotropies in the CMB at unprecedented sensitivity. While upcoming surveys and science goals are yet to be defined in detail, two of CMB-S4's main science goals are to constrain the primordial tensor-to-scalar ratio at a precision of $\sigma(r) \sim 5 \times 10^{-4}$ and pin down the uncertainty in effective number of relativistic particles to $\Delta N_{\text{eff}} \leq 0.06$ at 95% CL (Abazajian et al., 2019).

In Table 2.1, we show an overview of the predicted constraints on a selection of key cosmological parameters with SO, LiteBIRD and CMB-S4.

2.6 Polarized Galactic foregrounds

The challenge of measuring the tiny B-mode signal in the CMB is significantly aggravated by the large polarized emission from our own Galaxy. The two main sources, thermal emission from interstellar dust and synchrotron emission from cosmic ray electrons, are the dominant signals on the microwave sky at degree angular scales. Current data cannot constrain sky emission models well enough to allow for a foreground-bias-free measurement of r, no matter the true amplitude of the primordial signal. Characterizing Galactic foregrounds is therefore a priority for near-future B-mode experiments. Galactic synchrotron and dust emission, although of different physical

Table 2.1: Selection of uncertainties on cosmological parameters at 68% central CL, according to current measurements and future predictions for SO (SO Collaboration, 2019), LiteBIRD (LiteBIRD Collaboration, 2023) and CMB-S4 (Abazajian et al., 2019). Parameters that represent extensions to the ΛCDM model are: the tensor-to-scalar ratio r, the local primordial non-Gaussianity parameter $f_{\rm NL}^{\rm local}$, the sum of neutrino masses $\sum m_\nu$ and the effective number of relativistic species $N_{\rm eff}$. Current results are taken from (Planck Collaboration VI, 2020), unless indicated differently.

Predicted 68% CL uncertainties (ΛCDM plus extensions)

Parameter	Current	SO (baseline)	LiteBIRD	CMB-S4
n_s	0.0038	0.0020	-	0.0020
τ	$0.0060^\dagger$	-	0.0020	0.0025
r	$0.009^\ddagger$	0.003	0.001	0.0005
$f_{\rm NL}^{\rm local}$	5.1^*	2.0	-	0.57
$\sum m_\nu$ [meV]	60	40	12	15
$N_{\rm eff}$	0.15	0.07	0.15	0.03

† Pagano et al. (2020) ‡ BICEP2/Keck Collaboration (2021) * Planck Collaboration IX (2020)

origin, ultimately trace the morphology of the Galactic magnetic field (GMF). In the following, we briefly discuss both components in terms of their phenomenology, current models and experimental limitations, and their impact on CMB B-mode measurements.

2.6.1 Synchrotron emission

Synchrotron emission is caused by the accelerated motion of cosmic ray electrons in the presence of the GMF, with an amplitude that depends on the magnetic field strength. The degree-scale morphology of Galactic synchrotron emission is not limited to the Galactic disk, but extends far beyond, following the magnetic field lines. Measurements show that linearly polarized Galactic synchrotron emission can be approximated by a power law (Planck Collaboration X, 2016),

$$f_s^\nu = A_s \left(\frac{\nu}{\nu_{0,s}} \right)^{\beta_s} \tag{2.9}$$

with a spectral index $\beta_s \sim -3$, which also follows if we assume a power law spectral energy distribution (SED) for the cosmic ray electrons. Synchrotron emission is intrinsically polarized, with a polarization fraction of $\lesssim$15% at intermediate and high Galactic latitude (Planck Collaboration XXV, 2016; Page et al., 2007). We therefore expect no substantial difference between the polarized and unpolarized synchrotron SED. While *Planck* data alone are not sensitive enough to measure the spatial variability of the synchrotron spectral index, recent ground-based experiments show that at degree scales, β_s varies at the percent level when observed in the southern hemisphere (Harper et al., 2022; Krachmalnicoff et al., 2018).

The BB power spectrum can be roughly described by a power law, $D_\ell \equiv \ell(\ell+1)C_\ell/2\pi = A_s(\ell/80)^{\alpha_s}$, with amplitude A_s and spectral index α_s. This simple model assumes that the synchrotron morphology is the same for all frequencies and therefore does not account for variations in the SED spectral index. Synchrotron emission dominates CMB B-modes at frequencies below 70 GHz. Krachmalnicoff et al. (2016) find that there is no region in the sky, nor frequency below 100 GHz, where synchrotron emission would be subdominant with respect to CMB B-modes. Recent analyses of data from the S-Band Polarization All Sky Survey (S-PASS) combined with Planck, WMAP data show that the minimum synchrotron contamination at degree angular scales at 90 GHz is at the level of an equivalent tensor-to-scalar $r_{\text{synch}} \sim 10^{-3}$ (Krachmalnicoff et al., 2018).

The Q-U-I JOint Tenerife Experiment (QUIJOTE)[5] (Rubiño-Martín et al., 2010) is a ground-based experiment located at the Teide Observatory in Tenerife, Canary Islands, with the scientific goal of observing the polarized microwave sky at degree angular resolution at frequencies between 10 and 40 GHz. During a wide Galactic survey between 2012 to 2018, QUIJOTE's Mid Frequency Instrument (MFI) observed the northern sky, including the Galactic plane, at degrees between 10 and 20 GHz. Recent results from this survey (de la Hoz et al., 2023) find a mean synchrotron spectral index of $\beta_s = -3.08$ and with a degree-scale spatial dispersion of 0.13. Furthermore, the power-law emission model describes data sufficiently well outside the Galactic plane but fails for data inside this region. These results corroborate that characterizing synchrotron emission at a sensitivity relevant for future B-mode experiments requires data models beyond a simple uniform power-law SED.

2.6.2 Thermal dust emission

Thermal dust emission is caused by elongated dust grains, which align with the Galactic magnetic field lines and radiate thermally with a nonzero fraction of linear polarization. While the exact physical properties of these interstellar dust grains are unknown, some models consider multiple dust populations (Guillet, V. et al., 2018), while others assume a single population (Draine & Hensley, 2021). The dust SED in polarization is commonly described by a modified blackbody law,

$$f_d^\nu = A_d \left(\frac{\nu}{\nu_{0,d}}\right)^{\beta_d - 2} \frac{B_\nu(T_d)}{B_{\nu_{0,d}}(T_d)} \tag{2.10}$$

with temperature $T_d \sim 20$ K, a spectral index $\beta_d \sim 1.5$, and an arbitrary pivot frequency that we fix to $\nu_{0,d} = 353$ GHz. Observations at large scales find that the dust spectral indices in total intensity

---

[5] URL: research.iac.es/proyecto/quijote/pages/en/home.php

and in polarization agree within 10% (Planck Collaboration XI, 2020; Ashton et al., 2018).

At subdegree scales, the morphology of polarized Galactic dust emission is poorly known due to the lack of higher-resolution observations. Degree-scale observations are consistent with a BB power spectrum model with a negative spectral index and significant variation among different regions of the sky (Planck Collaboration XI, 2020). Moreover, when considering different dust populations with different spectral indices that emit into the same line of sight, their average SED is expected to differ from a modified blackbody law (Chluba et al., 2017). In recent years, alternative models have been designed that account for spatial variation of the spectral emission laws of Galactic foregrounds. Those include parameterizing frequency decorrelation, or accounting for linear fluctuations in the dust SED via the so-called "moment expansion" (Tegmark, 1998; Chluba et al., 2017; Azzoni et al., 2021). We address this latter model in Chpt. 3, when talking about the C_ℓ-based component separation pipeline for SO.

Synchrotron and dust emission may, to some degree, exhibit a statistical correlation, considering their connection with the GMF. A common assumption in power-spectrum-based foreground models is to include a scale-independent correlation coefficient, which current data constrain to be lower than 20% (Choi & Page, 2015; Krachmalnicoff et al., 2018).

2.6.3 CMB contamination

Figure 2.8 shows current constraints on the BB power spectrum amplitudes of diffuse polarized Galactic synchrotron and thermal dust foregrounds, estimated from from S-PASS, WMAP, and *Planck* (Krachmalnicoff et al., 2018; Planck Collaboration Int. XXX, 2016). The power spectra are measured on 184 patches at intermediate and large Galactic latitudes ($|b| > 20$ deg), covering a sky fraction of about 1% each, and assuming power-law power spectra with fixed spectral indices $\alpha_s = -1$ and $\alpha_d = -0.42$. Synchrotron and dust maps are evaluated at 2.3 and 353 GHz, respectively, and multifrequency emissions are extrapolated using a power-law SED for synchrotron and a modified-blackbody SED for dust, assuming $\beta_s = -3.22 \pm 0.08$, $T_d = 19.6$ K, and $\beta_d = 1.59$. The figure shows that degree-scale Galactic foreground emission surpasses the CMB emission at all frequencies considered. Krachmalnicoff et al. (2018) further conclude that not accounting for the morphological complexity of Galactic B-mode emission would lead to biases on the tensor-to-scalar ratio at the level $\delta r \sim 0.01$-0.001 at intermediate to high Galactic latitudes. Therefore, future B-mode searches rely on multifrequency observations to be able to characterize and subtract foregrounds at the necessary accuracy.

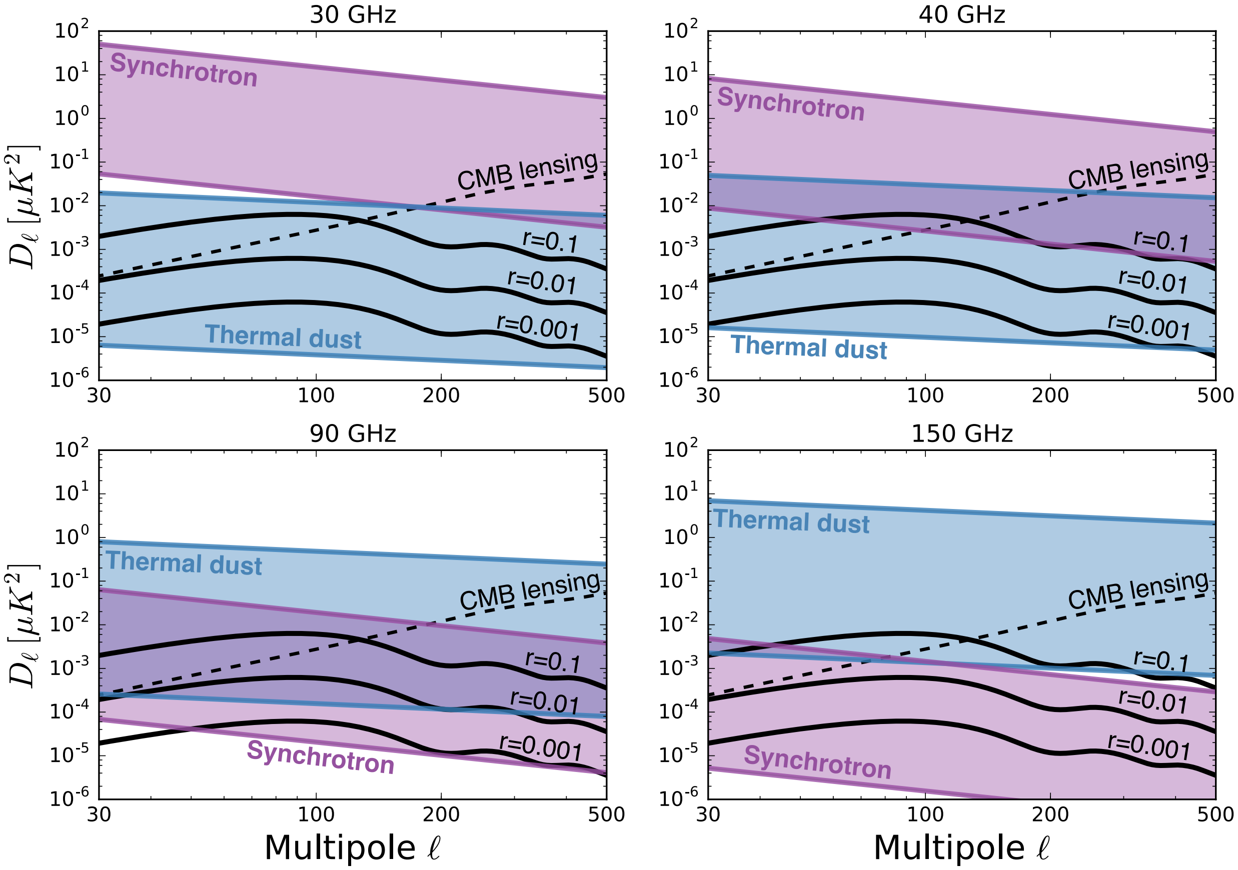

Figure 2.8: Constraints on the rescaled BB power spectrum amplitudes, $D_\ell \equiv \ell(\ell+1)C_\ell/2\pi$, inferred from real Galactic synchrotron and dust maps at 2.3 GHz and 353 GHz, respectively, compared with theoretical predictions of cosmological and lensing B-modes. Emissions are interpolated to different frequencies using isotropic SEDs of a power-law and a modified blackbody shape, respectively. Foreground power spectra are evaluated in 184 circular sky regions at intermediate and high Galactic latitudes, each with a sky fraction of about 1%, assuming isotropic, fixed power-law power spectral indices. Original work and figure: (Krachmalnicoff et al., 2018).

2.7 Instrumental systematic effects

Instrumental systematic effects, as well as Galactic foregrounds, are the dominant potential sources of bias for the next generation of CMB experiments, presented in Sect. 2.5. In this section, we briefly address some of the most relevant sources of systematic uncertainty in near-future ground-based CMB experiments, such as SO, which is the focus of Chpts. 3 and 4. While the detailed inclusion of instrumental systematic effects are beyond the scope of the data analysis techniques discussed there, the tight control of all possible sources of bias is essential for the robust detection of primordial B-modes.

For ground-based experiments, two important contaminants are **atmospheric emission and ground pickup**. These effects determine the choice of the telescope site and the design of the optical elements, and are therefore closely related to the instrumental systematics discussed below. In practice, the bulk of atmospheric systematic effects, mostly caused by the abundance of precipitable water vapor, as well as emission from the terrain surrounding the telescope, can be mitigated by applying a set of filters to the telescope timestream data. Timestream filtering affects mainly low sampling frequencies, which predominantly translate to large angular scales on the sky, and thus leads to a reduced number of measured modes at low to intermediate ℓ. This contributes to a statistical large-scale noise known as **$1/f$ noise** that decreases the statistical sensitivity from the ground. As we shall see later, employing an optical component called a cryogenic rotating half-wave plate (CRHWP) can improve on this by trading in some of the $1/f$ noise for another, potentially less harmful type of systematic effect.

In the following, we give a brief but general overview on instrumental contaminants expected to be important for SO, although those can be considered relevant for most other next-generation CMB experiments, too. With the scope of this Thesis being CMB polarization, we focus on polarization effects. We do not attempt to give a complete list of all possible instrumental effects that SO may encounter, and refer to the literature works cited in this section, and references therein, for an extended discussion.

We begin by discussing three categories of instrumental effects: optical effects, detector-related effects, and effects connected to polarization modulation. We then focus on a subset of instrumental effects that specifically impact degree-scale BB measurements and explain how they translate to a bias the tensor-to-scalar ratio.

Optical systematic effects. The first category of instrumental effects are those related to optical elements of the telescope, as described in Gallardo et al. (2018). This concerns the instrument beam, orthogonal detector pairs, and the telescope pointing.

- The instrument beam, which is the optical response of the instrument to an incoming signal in the field of view, can vary with the detector position within the optical array, the weather conditions on site, or the observed frequency. To avoid bias in the signal amplitude, the beam shape must be well characterized through calibration from external polarized sources, such as planets like Jupiter, or flying drones. The instrument beam contains a central main beam, near sidelobes adjacent to the main beam, and far sidelobes, which can be up to 180 degrees away from the main beam. Near sidelobes can produce ghost images resulting from reflection, scattering, or diffraction by optical elements, while far sidelobes can contain large-amplitude stray light which degrades the detector performance. Sidelobe pickup can be mitigated by reflective baffles that guide light rays away from the receiver.

- Pair-differencing systems such as the SO Large Aperture Telescope (LAT) relies on estimating the Stokes Q and U parameters by subtracting the signals from co-pointed detectors sensitive to orthogonal polarization states. Instrumental temperature-to-polarization (T-to-P) and E-to-B leakage can be caused through differential transmission or reflection by optical elements. These two types of leakage are especially important since total intensity is generally larger than the polarized intensity, and CMB E-modes are much larger than B-modes. Both types of leakage can also arise from imperfect detector pair orthogonality. The latter effect can be mitigated by employing a CRHWP as a polarization modulator as implemented in the SO SATs, which we discuss below.

- Another possible systematic effect are telescope pointing uncertainties that can arise either from mechanical jitter whilst rotating the instrument, or long-term distortions in the optical elements. The first effect causes a small-scale degradation of the beam and can be reduced through careful calibration, while the latter effect creates beam distortions that are more difficult to model and can be minimized, for instance, by mechanical reinforcements.

Detector-related systematic effects. The second category are systematics related to polarization-sensitive detector arrays. Following Crowley et al. (2018), we focus on nonlinear bolometer response, gain variations, and readout crosstalk.

- The polarization-sensitive detectors, called bolometers, are known to have a nonlinear response to the input signal that can vary with environmental conditions such as the Earth's atmosphere along the telescope's line of sight. These generally cause a global variation in T, Q, and U which translate to $1/f$ noise in timestream data, and must be estimated from timestream simulations. In the case of SO LAT, where polarization is measured through pair-differencing, this also causes T-to-P and E-to-B leakage.

- Detector gain drifts are caused by the long-timescale instability of the thermal bath temperature during observations of a scanning telescope. This signal-independent effect depends on the type of detectors used, and generally causes large-scale noise and multiplicative bias in the power spectra.

- Lastly, detector crosstalk is induced by the detector readout architecture, and refers to any spurious coupling of one detector's signal into the measured signal of another detector. In a pair-differencing experiment such as SO LAT, detector crosstalk leads to T-to-P leakage and E-to-B leakage, while in CRHWP experiments such at the SATs, detector crosstalk does not contaminate the polarization channels with the brighter temperature signal, as we shall see next.

Systematic effects from polarization modulation. A cryogenic, continuously-rotating half-wave plate (CRHWP), as used in the SO SATs, is an optical system that modulates the incoming polarization signal to alleviate systematic effects arising from slowly varying noise and detector pair-differencing. This mitigates T-to-P leakage but introduces other systematic effects, like the half-wave plate synchronous signal (HWPSS) which causes a non-linearity in the detector response, introducing another, subdominant, source of T-to-P leakage. Salatino et al. (2018) studied the CRHWP systematic effects for a telescope with the SO SATs' specifications, and find a total leakage coefficient well below one percent.

Effects on degree-scale B-modes. Having explored general instrumental systematic effects relevant for a ground-based CMB experiment like SO, we present here an example of a specific forecast for SO. In a recent work, Abitbol et al. (2021) quantify requirements on instrumental systematics calibration to guarantee a bias of the tensor-to-scalar ratio r at or below $\delta r \sim 10^{-3}$. The authors consider gain variations, uncertainties in the frequency band centers, and uncertainties in the polarization angle. Errors in the frequency bandpasses may be caused by time-variable atmospheric effects, which influence the Galactic foreground parameters that leak into the cleaned CMB. A frequency-dependence of the polarization angle, which may be caused by material properties of the CRHWP, leads to E-to-B leakage and a corresponding bias on r. The authors find that gain calibration factors and bandpass center frequencies must be known to percent level or better, while polarization angles must be calibrated to the level of a few tenths of a degree to guarantee a robust measurement of r.

These forecasts show how scientific goals translate from cosmological constraints into instrumental requirements, and demonstrate the synergy that is needed between instrumentation, calibration, and science analysis working groups within a large collaboration such as SO.

CHAPTER 3

Power spectrum pipeline to constrain primordial B-modes

The Simons Observatory aims at measuring or tightening the experimental constraints on primordial gravitational waves through degree-scale polarization B-modes in the CMB. This would have a significant impact on theories of the very early Universe, such as cosmic inflation. We know that this signal on the sky would be dominated by Galactic thermal dust and synchrotron emission and B-modes from gravitational lensing due to the cosmic LSS. Moreover, instrumental correlated noise and systematic effects, which typically depend on the details of the experimental setup, may significantly contaminate the sky signal (see Sect. 2.7). Measuring this extremely faint gravitational wave background thus requires using algorithms that accurately characterize and clean each of these contaminants on the partial sky. Subtracting the B-mode lensing contribution, also called "delensing", requires E-mode observations at intermediate scales or tracers of the LSS (Lewis & Challinor, 2006; Namikawa et al., 2022), while noise mitigation requires excellent knowledge of instrumental properties. Removing Galactic foreground emission relies on observations at multiple frequencies to separate components based on the spectral difference compared to the CMB (Delabrouille & Cardoso, 2007; Leach et al., 2008). To achieve this, the SO SATs measure the degree-scale polarized sky in six frequency bands from 27 to 280 GHz (SO Collaboration, 2019; Ali et al., 2020).

Estimating angular power spectra (C_ℓs) of a polarized field on the partial sky brings two main challenges. Firstly, observing only a fraction of the sphere breaks isotropy by cutting off harmonic modes on the sky, which introduces uncertainty about the angular scales that we measure. This leads to a coupling between different ℓ-modes estimated on the partial sky (Hivon et al., 2002). Secondly, since E- and B-modes are geometrical patterns defined on the full sky, observing the partial sky makes them intrinsically hard to disentangle and leads to "leakage" from E- to B-modes, and vice versa. Given that the E-modes of the CMB are much larger in amplitude than CMB B-modes, the main practical challenge for C_ℓ-based B-mode analyses is thus to distinguish leaked, ambiguous modes from the true B-mode signal, a technique known as "B-mode purification" (Smith, 2006; Smith & Zaldarriaga, 2007; Bunn, 2011). In the first part of this chapter (Sect. 3.1), we address the question of how to practically solve these issues when estimating power spectra.

The remaining part of this chapter addresses the question of how to infer cosmological parameters from degree-scale multifrequency B-mode power spectra, introducing power spectrum likelihoods (Sect. 3.2) and data models (Sect. 3.3). We then introduce the core work of this Thesis, the development and validation of a power-spectrum-based component separation pipeline for analyzing degree-scale B-modes with SO (Sect. 3.4). The power spectrum pipeline, based on an approach used by the BICEP/*Keck* Collaboration (BICEP2/Keck Collaboration, 2016b, 2018), is one of three algorithms developed for SO with the aim of robustly inferring r on partial sky maps in the presence of Galactic foregrounds and instrumental noise. This pipeline computes BB power spectra from four data split maps at the six SO frequencies, considering the partial sky as observed by the SO SATs. The data splits are assumed to contain noise that is uncorrelated at the map level. We compute "cross-split" power spectra by correlating unequal data splits, thereby avoiding bias in the power spectra coming from the map-level noise autocorrelation. The resulting noise-unbiased multifrequency power spectrum estimates are then compared against a theoretical multi-component power spectrum model in a likelihood that is sampled within a Bayesian MCMC framework. The pipeline infers cosmological parameters associated with the CMB as well as astrophysical parameters associated with Galactic foregrounds, and can therefore retrieve the marginal posterior of the tensor-to-scalar ratio by integrating ("marginalizing") over foreground parameters. In order to validate the SO component separation pipelines in their ability to perform this task, they were tested against models of Galactic foregrounds of varying complexity in a pipeline comparison project (Wolz et al., 2023a), which we describe in Chpt. 4.

This chapter is structured as follows. We start by reviewing present power spectrum estimation and the associated challenges in Sect. 3.1. We then discuss the ingredients needed to perform statistical inference with CMB power spectra, starting with the C_ℓ likelihood in Sect. 3.2 and moving to theoretical models of degree-scale B-modes on the microwave sky in Sect. 3.3. Finally, in Sect. 3.4, we describe the SO C_ℓ pipeline, focusing on the implementation of the power spectrum estimation, the covariance computation, and the Bayesian inference framework, before outlining the numerical code in Sect. 3.4.4. We summarize our findings in Sect. 3.5.

3.1 Estimating power spectra on the partial sky

The angular power spectrum, the harmonic transform of the two-point correlation function, is an established summary statistic with a legacy worth of reliable estimators and inference methods suited to analyze CMB polarization. It has important practical and theoretical benefits: Compared to maps, the power spectrum provides efficient data compression while maintaining a physical meaning that is easy to interpret. Current *Planck* data support the hypothesis of a Gaussian CMB (Planck Collaboration VII, 2020), in which case the statistical information is fully encoded in the power spectrum. If the data contain non-Gaussian components, such as Galactic foreground emission, the power spectrum still represents a key statistical quantity. In this section, we explore the practical challenge of estimating unbiased and near-optimal power spectra at the partial sky, which is aggravated by the coupling of harmonic modes and E-to-B leakage. We then present the pseudo-C_ℓ estimator, which is the algorithm used in the SO C_ℓ pipeline, and explain how it addresses the aforementioned problems. We also briefly outline the quadratic maximum likelihood (QML) estimator.

3.1.1 Bandpower coupling and E/B leakage

Analyzing power spectra on the partial sky is a nontrivial task. On the full sky, computing the angular power spectrum of an isotropic field involves an integration over all points of the sphere, which is no longer possible on the incomplete sky. Observing only a part of the sky breaks isotropy and causes a coupling between different ℓ-modes. This is because on the partial sky, spherical harmonics do not form an orthogonal basis, hence they no longer provide a unique mapping between angular scales and ℓ-modes. By observing only a part of the sky, we cut off a subset of harmonic modes, which effectively introduces an intrinsic ambiguity as to what angular scales we measure, called "mode coupling". In real data analysis, our goal is therefore to capture the expected effect that the observation mask has on our measurement, which allows us to retrieve the estimated C_ℓs as they would appear on the full sky. This is achieved through the so-called "mode coupling matrix", which we discuss in detail in the context of the pseudo-C_ℓ estimator.

While mode coupling affects the estimation of power spectra of both scalar and polarization fields, an additional complication in the polarized case is E/B leakage. On the full sky, all E- and B-modes are uniquely defined, while on the partial sky, some of the polarization modes are ambiguous and cannot be assigned to one type or the other. Analogous to harmonic modes, E- and B-modes couple on the incomplete sky as a result of information loss from cutting off parts of the sky, and with it geometric patterns in Q and U. This causes part of the (unobserved) full-sky E-modes to become ambiguous modes, which bear the risk of being misinterpreted as B-modes, and vice versa. Due to the relative smallness of CMB B-modes with respect to E-modes

($C_\ell^{EE}/C_\ell^{BB} \sim 10^3$ for $\ell \sim 90$), E-to-B leakage can lead to a significant positive bias in degree-scale B-mode measurements, making this a high-risk contaminant in the search for primordial B-modes. Besides said bias, E-to-B leakage increases the B-mode power spectrum variance, adding to the statistical uncertainty associated with the primordial tensor-to-scalar ratio. Fortunately, reliable and efficient methods exist that tightly control and mitigate these effects, which we detail in the following section.

Mode coupling and E/B leakage can be addressed by different algorithms that estimate angular power spectra on the partial sky. The SO C_ℓ pipeline employs the pseudo-C_ℓ estimator, which is a versatile and efficient power spectrum estimator commonly applied to analyze the small angular scales of the CMB. We present the pseudo-C_ℓ algorithm in detail in the following section.

3.1.2 The pseudo-C_ℓ (PCL) estimator

The pseudo-C_ℓ (PCL) estimator (Hivon et al., 2002) is a semi-analytical algorithm designed to estimate unbiased, full-sky angular power spectra from scalar or polarized fields observed on the partial sky. It is based on the analytical computation of the mode coupling matrix requiring only the sky mask and not the map itself, making it extremely efficient. The PCL estimator is always unbiased, meaning it correctly reproduces the full-sky C_ℓ on average. Moreover, the PCL with B-mode purification is an optimal estimator of B-modes on the partial sky, meaning that it reaches the lowest possible variance, assuming that pixels on the sky are statistically uncorrelated. For ground-based experiments such as SO, this is often a sufficiently accurate assumption as long as correlated systematic effects, such as large-scale "stripes" from the atmosphere, or detector gain drift, can be neglected. Otherwise, the PCL estimator is no longer optimal but stays unbiased. We briefly address optimal quadratic estimators below.

Let us have a look at the PCL estimator in more detail. For two polarized fields labeled X and Y with measured harmonic coefficients $a_{\ell m}^{X/Y,p}$ and $p, q, p', q' \in \{E, B\}$, the PCL estimator can be written as

$$\hat{C}_\ell^{(X,p)(Y,q)} \equiv \sum_{p',q'} \sum_{\ell'=0}^{\infty} (\mathsf{M}^{-1})_{\ell\ell'}^{pq,\,p'q'} \tilde{C}_{\ell'}^{(X,p')(Y,q')} \,, \tag{3.1}$$

where

$$\tilde{C}_\ell^{(X,p)(Y,q)} \equiv \sum_{m=-\ell}^{\ell} \frac{a_{\ell m}^{X,p} (a_{\ell m}^{Y,q})^*}{2\ell + 1} \tag{3.2}$$

is the naïve, full-sky power spectrum estimator. The matrix M in Eq. (3.1) is the mode coupling matrix mentioned above. It is defined as the linear map between the true, full-sky polarization power spectrum $C_\ell^{(X,p)(Y,q)}$ and the expectation value of the full-sky estimator,

$$\sum_{p',q'} \sum_{\ell'=0}^{\infty} M_{\ell\ell'}^{pq,\,p'q'}\, C_{\ell'}^{(X,p')(Y,q')} = \left\langle \tilde{C}_\ell^{(X,p)(Y,q)} \right\rangle . \tag{3.3}$$

If M^{-1} exists, using it to linearly transform the naïve estimator as in Eq. (3.1) defines an unbiased estimator. Binning the multipoles into ℓ-bins is a common practical way to ensure the invertibility of the mode coupling matrix (Alonso et al., 2019).

For a polarized field, the mode coupling matrix contains 4x4 blocks that each map from the pair of polarization modes (p', q') to the pair (p, q). Using the basic assumption that the pixels on the sky are statistically uncorrelated, the mode coupling matrix can be written analytically as (Kogut et al., 2003; Hansen & Gorski, 2003)

$$\mathsf{M}_{\ell\ell'} = \begin{pmatrix} M^+_{\ell\ell'} & 0 & 0 & M^-_{\ell\ell'} \\ 0 & M^+_{\ell\ell'} & M^-_{\ell\ell'} & 0 \\ 0 & M^-_{\ell\ell'} & M^+_{\ell\ell'} & 0 \\ M^-_{\ell\ell'} & 0 & 0 & M^+_{\ell\ell'} \end{pmatrix} , \tag{3.4}$$

where

$$M^\pm_{\ell\ell'} \equiv \frac{2\ell'+1}{4\pi} \sum_{\ell''=0}^{\infty} P^w_{\ell''} \begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix}^2 \frac{1 \pm (-1)^{\ell+\ell'+\ell''}}{2} . \tag{3.5}$$

In the last step, we introduced the Wigner-3j symbol, and

$$P^w_\ell \equiv \sum_{m=-\ell}^{\ell} |a^w_{\ell m}|^2 , \tag{3.6}$$

where $a^w_{\ell m}$ is the scalar spherical harmonic transform of the sky mask $w(\hat{\boldsymbol{n}})$. Evaluating the PCL estimator is extremely fast, since M need only be computed once for each mask, and the Wigner-6j matrix elements can be read from precomputed lists. The numerical PCL estimation typically scales like the cube of the maximum multipole probed, ℓ^3_{max}). Together with its being unbiased, this makes the PCL estimator a standard choice in CMB analysis.

Let us discuss how the PCL estimator solves the issues of mode coupling and E-to-B leakage. Mode coupling manifests in the offdiagonal elements in the matrices $(M^\pm_{\ell\ell'})$, which depend directly on the mask's angular power spectrum defined in Eq. (3.6). Likewise, the leakage between E- and B-modes becomes evident from the nonzero offdiagonal blocks in the matrix M, which all happen to be proportional to $(M^-_{\ell\ell'})$, as can be seen from Eq. (3.4). If the mask is isotropic, meaning that we apply no sky cuts, it can be algebraically shown that the total coupling matrix M becomes the unity matrix, in which case the different ℓs decouple, and the PCL estimator reduces to the full-sky estimator. If the mask covers the incomplete sky, we can analytically solve the mode coupling issue by multiplying the full-sky estimate by the inverse mode coupling matrix as done in Eq. (3.1). This decouples the multipoles and allows us to obtain unbiased power spectrum estimates on the partial sky.

The standard PCL estimator, although correcting for mode coupling and power spectrum bias due to E/B leakage, does not solve the issue of power spectrum variance leakage between E- and B-modes. While the mode coupling matrix formalism is designed to retrieve an unbiased estimate of the full-sky power spectra of E- and B-modes (their two-point statistics), it does not provide unbiased estimates of the power spectrum variance (the four-point statistics). Since the cosmological B-mode signal targeted by SO is much smaller than the E-mode signal, the B-mode power spectrum variance is expected to be systematically amplified by the presence of ambiguous modes, unless we isolate and subtract the ambiguous signal on the map level before estimating the B-mode power spectrum. Such procedure is called "B-mode purification". The associated algorithm is described in Smith (2006); Bunn (2011) and a recent software implementation is contained in the `NaMaster` code framework (Alonso et al., 2019). To gain some insight into the purification algorithm, let us first acknowledge that on the (full) 2-sphere, we can define a differential operator $\mathcal{B}$ that turns a scalar field into B-modes (see Appendix A1 of Alonso et al., 2019). The standard procedure of estimating B-modes on the full sky can be described by three consecutive operations:

1. We construct a "full-sky B-mode basis" by applying $\mathcal{B}$ to the well-known scalar spherical harmonics $Y_{\ell m}$.
2. We project the polarization field that we want to analyze onto this full-sky B-mode basis.
3. We average over the full 2-sphere to obtain the final, unbiased B-mode harmonic coefficients.

If we observe the incomplete sky, then step 3 picks up ambiguous E-modes because of information loss due to the unobserved modes. To prevent this from happening, we can modify step 1 to construct a "cut-sky B-mode basis" by first multiplying the scalar spherical harmonics $Y_{\ell m}$ by the observation mask, and then applying $\mathcal{B}$ to the product of both. This new basis projects out a smaller set of modes called "pure" B-modes, which are orthogonal to the ambiguous modes, thus allowing us to safely perform a full-sky average in step 3. As a result, we obtain partial-sky estimates of pure B-mode harmonic coefficients, which are free from spurious E-modes. In particular, pure B-modes allow us to estimate unbiased BB power spectra with no variance leakage. To combine this procedure with the PCL algorithm, we need to make small modifications in the mode coupling matrix $M^{\pm}_{\ell\ell'}$ in Eq. (3.5) (see, e.g., Eq. (35) in Alonso et al., 2019). We also need to ensure that our observation mask is differentiable everywhere, which can be achieved by tapering the mask boundaries to avoid sharp edges, also known as "apodization" (see Grain et al., 2009, for a thorough discussion of these methods).

The SO C_ℓ pipeline uses the `NaMaster` code (Alonso et al., 2019)[1] to calculate the PCL estimator. The code is written in C and wrapped into `python` to facilitate combination with other astronomy software libraries. For increased efficiency, `NaMaster` is structured around two main classes (`NmtField` and `NmtWorkspace`) that incorporate field-level operations, such as spherical harmonic transformation and B-mode purification, and correlation-level operations, such as computing the PCL estimate and the mode coupling matrix. A `NmtField` object can be initialized with a sky mask, a pair of Stokes Q and U maps that represent the input data, a beam transfer function, and a flag to enable or disable B-mode purification. A `NmtWorkspace` object can compute the mode coupling matrix from a pair of fields and a given ℓ-binning scheme, and decoupled PCL estimates. Numerical stability and regularization of the mode coupling matrix warrants a binning into bandpower windows, which is included as a convenience function in `NaMaster`. We note that binning of C_ℓs is only equivalent to a simple averaging over ℓs if we observe the full sky, while on the partial sky, exchanging the order of binning and mode decoupling leads to different results. Therefore, binned bandpower windows do not retain a rectangular shape but are smoothed out depending on the details of the shape and borders of the mask. More details can be found in Alonso et al. (2019).

3.1.3 The quadratic maximum likelihood (QML) estimator

The PCL estimator ignores any statistical correlation between different pixels in both signal and noise. This is a good approximation if a) the noise bias can be avoided (e.g., via the cross-split technique which we discuss later in Sect. 3.4), and b) correlated modes (e.g., atmospheric "stripes" as a result of timestream filtering) can be neglected. In that case, the PCL estimator retrieves the true full-sky B-mode power spectrum without bias, but obtaining a minimum-variance estimator requires purification, as we saw in the previous section. It is possible to construct an unbiased, minimum-variance estimator of polarization E- and B-modes on the incomplete sky that also accounts for correlations between different pixels. This generalized estimator is known as the quadratic maximum likelihood (QML) algorithm. The QML estimator accounts for real-space correlation by weighting sky pixels with the full inverse pixel-pixel covariance matrix, instead of just weighting every pixel with its inverse variance, as done in the PCL case.[2] The QML algorithm

---

[1] Source code: github.com/LSSTDESC/NaMaster, documentation: namaster.readthedocs.io/

[2] The PCL estimator assigns infinite variance to unobserved pixels.

is computationally expensive, demanding $O(N_{\rm pix}^3)$ operations for a map of $N_{\rm pix}$ pixels, whereas PCL is generally much faster in comparison, requiring only $O(N_{\rm pix}^{3/2})$ operations (Tegmark & de Oliveira-Costa, 2001). Therefore, analysis pipelines of CMB polarization data usually apply QML only to low-resolution maps that capture multipoles up to $\ell = 200$ and resort to the faster PCL method for smaller scales. The QML estimator represents a generalization of the PCL estimator, and since we do not use it in the SO C_ℓ pipeline, we do not lay out the mathematical details here. We refer to Efstathiou (2004, 2006) for more details.

3.2 Power spectrum likelihoods

Inferring cosmological parameters from measured angular power spectra of a field on the sky requires, besides a theoretical data model to compare with, a likelihood that allows the assessment of statistical uncertainty. As we saw previously, the power spectrum statistic represents the two-point information of a given field, and describes its full statistical properties if this field is Gaussian. If non-Gaussian effects on the field level are sufficiently small, the power spectrum has a well understood probability distribution that allows the construction of fast-to-evaluate analytical likelihoods. Once we have a theoretical model, parameterized by a set of physical parameters as we shall see later in Sect. 3.3, we can evaluate the power spectrum likelihood at different values in parameter space. The final parameter distribution, called the posterior, combines the information from the data likelihood with a parameter prior distribution, and can be estimated by numerical sampling methods such as Monte Carlo Markov chain (MCMC).

In this section, we introduce several analytical likelihoods for power spectrum data that are commonly used in CMB polarization analysis. We focus on the Gaussian likelihood approximation that, as we shall see later, represents an accurate and numerically efficient choice for the SO C_ℓ pipeline. We briefly discuss alternative likelihood expressions that accurately describe the statistics of the power spectrum estimate in more general cases.

3.2.1 Gaussian approximation

For the SO C_ℓ pipeline, we employ the Gaussian likelihood approximation, which assumes that the binned C_ℓ multipoles can be described as Gaussian random variables. To see that this is a good approximation, we first observe that binned power spectrum estimates are generally given by the sum of $M = f_{\rm sky}\Delta\ell(2\ell + 1)$ harmonic modes, where $f_{\rm sky} \in (0, 1]$ is the observed sky fraction and $\Delta\ell$ is the width of the multipole bin. By virtue of the central limit theorem (see e.g., Eicker, 1966), the true power spectrum likelihood converges to a Gaussian distribution as $M \to \infty$, so the distribution of C_ℓ estimates is close to Gaussian as long as ℓ-modes and ℓ-bin widths are sufficiently large.

CMB polarization experiments such as SO measure B-mode power spectra across different frequency channels and at different angular scales, which are arranged in a data vector. The power spectrum likelihood quantifies the statistical distribution of this data vector, including correlations among pairs of channels and pairs of angular modes. The Gaussian likelihood approximation $P(\hat{\boldsymbol{C}}|\boldsymbol{\theta})$, evaluated at a given data vector $\hat{\boldsymbol{C}}$ and conditional on the model parameter vector $\boldsymbol{\theta}$, is given by

$$-2\ln P(\hat{\boldsymbol{C}}|\boldsymbol{\theta}) = \sum_{(\nu_1,\,\nu_1')} \sum_{(\nu_2,\,\nu_2')} \sum_{l,l'=1}^{N_{\rm bin}} \left(\hat{C}_l^{\nu_1\nu_2} - C_l^{\nu_1\nu_2}(\boldsymbol{\theta})\right)^T \left(\Sigma^{-1}\right)_{ll'}^{\nu_1\nu_2,\,\nu_1'\nu_2'} \left(\hat{C}_{l'}^{\nu_1'\nu_2'} - C_{l'}^{\nu_1'\nu_2'}(\boldsymbol{\theta})\right), \quad (3.7)$$

where $(\nu_i,\ \nu_i')$ label the frequency channel pairs and l, l' label the multipole bins. In the case of the SO survey with six frequency channels, the sums over channel pairs go over $N_{\rm cross} = 21$ terms corresponding to the number of unique cross-channel power spectra, and the sums over l, l' contain $N_{\rm bin} = 27$ terms corresponding to the number of ℓ-bins. Finally, Σ is the $(N_{\rm cross} \cdot N_{\rm bin}) \times (N_{\rm cross} \cdot N_{\rm bin})$ covariance matrix. For the SO case, ignoring covariance terms that couple different multipole bins, $l \neq l'$, is a good approximation, as we shall see later in Sect. 3.4.2.

The data vector that enters the likelihood contains B-mode purified BB pseudo-C_ℓ estimates obtained by the `NaMaster` code (see Sect. 3.4.1). Therefore, at the power spectrum likelihood level, we do not need to explicitly include polarization, and instead treat the B-modes in the six SO frequency channels as correlated scalar fields. Moreover, the power spectrum data do not contain any noise bias, since we estimate them using the cross-split technique, which we discuss later in Sect. 3.4.

We estimate the covariance Σ from simulations. This procedure can generate numerical noise that may result in a noninvertible covariance matrix. Binning the ℓ-modes into bandpowers helps mitigate numerical noise and decrease the number of nonzero subdiagonals related to masking the sky (see Sect. 3.4). At the same time, binning increases the number of modes that are summarized in the bandpower statistic, which significantly improves the accuracy of the Gaussian approximation. Although Eq. (3.7) lacks the determinant term $\propto \ln\det\Sigma$ that would make this an exact Gaussian distribution in the data vector $\hat{\boldsymbol{C}}$, we can assume that this omission is of little practical consequence. As shown in Kodwani et al. (2019), the parameter dependence of the covariance is generally weak, and treating the covariance as a constant affects the parameter uncertainties by less than about 10%. This is a subdominant contribution compared to other B-mode variance-inducing effects discussed in this chapter, such as E-to-B leakage, gravitational lensing, or the presence of Galactic foregrounds. We therefore ignore parameter dependence in the covariance in the context of this Thesis.

3.2.2 Exact likelihood

To put the Gaussian likelihood approximation into perspective, let us introduce the exact power spectrum likelihood on the full sky. Ignoring statistical moments of higher than second order, the spherical harmonic coefficients $a_{\ell m}$ of an isotropic random field on the 2-sphere are independent, identically distributed Gaussian random variables with mean zero and variance $\left\langle |a_{\ell m}|^2 \right\rangle = C_\ell$. The joint probability density of all spherical harmonics with constant ℓ is thus given by a product of $2\ell + 1$ Gaussian distributions,

$$
\begin{aligned}
-2\,\ln P(\{a_{\ell m}\}|C_\ell) &= \sum_{m=-\ell}^{\ell} \frac{|a_{\ell m}|^2}{C_\ell} \\
&= (2\ell+1)\left(\frac{\tilde{C}_\ell}{C_\ell} + \ln(C_\ell)\right) + \text{const.}\,,
\end{aligned} \tag{3.8}
$$

where we introduced the unbiased full-sky estimator $\tilde{C}_\ell \equiv \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2/(2\ell+1)$. From the second line in Eq. (3.8), we see that $\tilde{C}_\ell$ is a sufficient statistic of all $a_{\ell m}$ at fixed ℓ if, as we stated before, higher-order moments are negligible, which is equivalent to assuming that the field on the 2-sphere is a Gaussian random field. To obtain the posterior distribution $P(C_\ell(\boldsymbol{\theta})|\tilde{C}_\ell)$ of the model $C_\ell(\boldsymbol{\theta})$ given the data $\tilde{C}_\ell$ from the data likelihood $P(\tilde{C}_\ell|C_\ell(\boldsymbol{\theta}))$ at fixed ℓ, we need to apply two operations: first, we normalize Eq. (3.8) with respect to the random variable $\tilde{C}_\ell$ to obtain the correctly normalized (proper) data likelihood. Then, we use Bayes' theorem to obtain the proper posterior distribution, which means we normalize the probability density with respect to the random variable C_ℓ. The result is a (rescaled) χ^2 distribution with $(2\ell + 1)$ degrees of freedom:

$$
-2\,\ln P(C_\ell|\tilde{C}_\ell) = (2\ell+1)\left[\frac{\tilde{C}_\ell}{C_\ell} - \ln\left(\frac{\tilde{C}_\ell}{C_\ell}\right) - 1\right], \tag{3.9}
$$

which, due to the central limit theorem, asymptotically approaches the Gaussian approximation for $\ell \longrightarrow \infty$, Eq. (3.7) (Cramér, 1946; Eicker, 1966). For individual multipoles $\ell \gtrsim 30$ measured on the full sky, every C_ℓ estimate is a sum of $M \approx (2\ell + 1) \gtrsim 61$ harmonic modes, making it virtually indistinguishable from a Gaussian random variable. In practice, we measure binned power spectra on the partial sky using PCL estimates (see Sect. 3.1.2), replacing the sum of $a_{\ell m}$s by a linear combination, which modifies the effective number of modes per bandpower to approximately $M = f_{\rm sky}\Delta\ell(2\ell + 1)$ (Knox, 1997). For the SO C_ℓ pipeline, we have $f_{\rm sky}\Delta\ell \approx 1$, justifying our choice of using the Gaussian likelihood approximation for SAT B-modes.

On the partial sky, the full-sky power spectrum estimator $\tilde{C}_\ell$ is biased, and we can no longer use Eq. (3.9). In order to obtain the exact distribution of a Gaussian map on the incomplete sky, we would in principle have to resort to likelihoods that are defined in pixel space. Evaluating general pixel-based likelihoods scales like $\ell_{\max}^6$ which is computationally prohibitive for $\ell \gtrsim 200$ (Mortlock et al., 2002; Lewis et al., 2002). Pixel-based likelihoods are used in low-ℓ CMB data analysis, provided the bias from non-Gaussian systematic effects can be neglected (Planck Collaboration V, 2020).

An alternative to expensive pixel-based likelihoods is the C_ℓ-based *Hamimeche-Lewis* likelihood (HL likelihood, Hamimeche & Lewis, 2008). The analytical expression of the HL likelihood is based on a variable transformation of the term $\tilde{C}_\ell / C_\ell$ that allows one to rewrite Eq. (3.9) in the form of a Gaussian likelihood. This has an important numerical advantage: as with the Gaussian likelihood approximation described in Sect. 3.2.1, we can precompute the covariance matrix from simulations, which makes the likelihood evaluation considerably less expensive.

3.3 Modeling degree-scale B-modes on the microwave sky

In this section, we present theory models of B-mode emission in power spectrum space, which we compare to real multifrequency observations within the framework of the SO C_ℓ pipeline. All models include the lensed CMB and Galactic foregrounds, which are the dominant sky components at angular scales $30 \lesssim \ell \lesssim 300$ relevant for SO SAT measurements. In this context, we introduce the "C_ℓ-fiducial" and "C_ℓ-moments" models, which are extensively used in the SO pipeline comparison project described in Chpt. 4.

3.3.1 Cosmic microwave background

In its two main variants, our CMB BB power spectrum model consists of two components: the late-time gravitational lensing signal and the assumed primordial gravitational wave signal. The lensing-induced contribution to the B-mode power spectrum features the lensing peak at small scales around multipoles $\ell \sim 1000$ (see Sect. 2.3). At smaller multipoles $30 \lesssim \ell \lesssim 300$ probed by the SO SATs, the amplitude of the lensing B-modes are expected to be at least as large as the primordial signal, assuming the latest constraints $r \lesssim 0.036$ from BICEP2/Keck Collaboration (2021). The conjectured primordial B-mode contribution features a large-scale reionization bump at $\ell \sim 4$ and an intermediate-scale recombination bump at $\ell \sim 90$ (see Sect. 2.3 and Fig. 2.3). We parameterize this CMB model by the primordial tensor-to-scalar ratio r and the phenomenological $A_{\rm lens}$ parameter that normalizes the power of lensing B-modes,

$$D_\ell^c = r\, D_\ell^{\rm tens}(r=1) + A_{\rm lens}\, D_\ell^{\rm lens}\,, \tag{3.10}$$

where we introduced the scaled power spectrum $D_\ell \equiv \ell(\ell+1)C_\ell/2\pi$. In this simple model, we use $A_{\rm lens}$ to parameterize the uncertainty about the residual level of lensing B-modes after delensing. The tensor template $D_\ell^{\rm tens}$ is the theoretically predicted CMB B-mode contribution from primordial gravitational waves. The lensing B-mode template $D_\ell^{\rm lens}$ is a convolution of the predicted CMB E-mode power spectrum with the predicted lensing kernel (Lewis & Challinor, 2006). Both functions can be efficiently calculated by numerical Boltzmann solvers like `CAMB` (Lewis et al., 2000; Howlett et al., 2012). In state-of-the-art B-mode data analysis, a common choice is to directly estimate maps of lensing B-modes from observed E-modes and external data sets using established methods, such as the the quadratic estimator by Okamoto & Hu (2003). This lensing map is then subtracted from the raw B-mode map before estimating the power spectrum. This method, called "delensing", is usually the preferred approach since it reduces the total B-mode power, leading to less cosmic variance and, in consequence, tighter constraints on the cosmological parameters. Near-future CMB polarization experiments, such as SO, are commonly designed to use SATs for B-mode measurements at large angular scales and LATs to measure the small-scale gravitational lensing signal on an overlapping sky patch, allowing for map-based delensing. Simulation-based predictions using delensing pipelines show that a reduction of the lensing signal by well more than 50% (corresponding to $A_{\rm lens} < 0.5$) is realistic for SO (Namikawa et al., 2022).

3.3.2 Galactic foregrounds

Galactic foregrounds, specifically thermal emission from synchrotron radiation and Galactic dust grains, dominate the CMB lensing B-modes and any possible primordial component at most frequencies and scales (see Sect. 2.6). Even at the relative maximum of the CMB lensing signal at 90 GHz and $\ell \sim 100$, foregrounds significantly contaminate the CMB and must therefore be included in CMB B-mode analyses (Planck Collaboration XXX, 2016; Krachmalnicoff et al., 2016). Studies of Galactic foregrounds indicate that a potential primordial signal, currently constrained to $r < 0.0036$ at 95% CL by BICEP2/Keck Collaboration (2021), would be contaminated by foregrounds at an equivalent tensor-to-scalar ratio of at least 10^{-3} at all frequencies or scales (Krachmalnicoff et al., 2018). It is therefore crucial to make full use of multifrequency observations in order to constrain the SEDs of Galactic synchrotron and thermal dust emission to disentangle it from the CMB.

In the following, we present the theoretical models used to parameterize Galactic foregrounds in the SO C_ℓ pipeline. Thanks to past analyses of WMAP and *Planck* data, as well as recent complementary studies at low frequencies, such as the C-Band All Sky Survey (C-BASS, Harper et al., 2022) and S-PASS (Krachmalnicoff et al., 2018), we are able to constrain simple Galactic foreground emission models at a quantitative level. However, these constraints still leave a lot of freedom as to the details of the power spectrum model; we do not know yet which model will be the best one. The conservative approach to this issue is to consider a range of complexity in both our theory models and in our set of sky simulations, and to compare the model performance in different mock data scenarios. This is the premise of the pipeline comparison project described in Chpt. 4.

Let us now give an overview of the different foreground theory models implemented in the SO C_ℓ pipeline. We start from a simple model of uncorrelated synchrotron and thermal dust model without variations in their spectral indices. We then present two generalizations, the first of which considers synchrotron-dust correlation, and the second one of which accounts for small spectral index variations via the so-called moment expansion. While we consider only foreground emission in the BB autospectrum here, we note that EE and EB foreground emission is usually parameterized in the same way, but can differ considerably in their best-fit parameter values (see, e.g., Planck Collaboration XI, 2020).

Simple dust and synchrotron model. To lowest order, the SED can be described by a modified blackbody law $f_d^{\nu'}(\beta_d, T_d)$ (see Eq. (2.10)), parameterized by the dust temperature $T_d \sim 20$ K and the dust frequency spectral index $\beta_d \sim 1.5$ (Planck Collaboration XI, 2020). Similarly, the simplest theory model of the synchrotron spectral distribution is a power law SED $f_s^\nu(\beta_s)$ (see Eq. (2.9)), parameterized by the synchrotron spectral index $\beta_s \sim -3$ (Planck Collaboration XXV, 2016). The power spectra of synchrotron and dust, measured at the pivot frequencies of 23 GHz and 353 GHz, respectively, and a pivot scale $\ell_0 = 80$, are commonly modeled as power laws of the form

$$D_\ell^{\nu\nu'} = D_\ell^{d,\nu\nu'} + D_\ell^{s,\nu\nu'} \tag{3.11}$$

where $D_\ell \equiv \ell(\ell+1)C_\ell/2\pi$ as above, and

$$D_\ell^{d,\nu\nu'} \equiv f_d^\nu(\beta_d, T_d)\, f_d^{\nu'}(\beta_d, T_d)\, A_d \left(\frac{\ell}{\ell_0}\right)^{\alpha_d} , \quad D_\ell^{s,\nu\nu'} \equiv f_s^\nu(\beta_s)\, f_s^{\nu'}(\beta_d)\, A_s \left(\frac{\ell}{\ell_0}\right)^{\alpha_s} . \tag{3.12}$$

Both the SED and power spectrum models used here are only rough approximations and do not represent the complexity of the real sky. Based on *Planck* and WMAP measurements, the foreground power law amplitudes evaluated in the SO SAT sky region (which covers a sky fraction of $f_{\rm sky} \sim 0.1$ that avoids high Galactic emission, see Fig. 4.2) take typical mean values of $A_s \sim 1.6\mu\mathrm{K}^2$, $A_d \sim 30\mu\mathrm{K}^2$, while the power spectral indices amount to $\alpha_s \sim -1$, $\alpha_d \sim -0.2$ when evaluated in the same region (Wolz et al., 2023a). Foreground amplitude parameters and spectral indices are known to vary significantly between frequencies and regions on the sky (Krachmalnicoff et al., 2018; Planck Collaboration XI, 2020), giving rise to frequency decorrelation (see Sect. 2.6). One way to address this is by means of the moment expansion described below.

Dust-synchrotron correlation. One might expect Galactic dust and synchrotron emission to be statistically correlated at the power spectrum level, as both sources of emission are linked with the morphology of the GMF (Planck Collaboration XXV, 2016; Planck Collaboration Int. XXXII, 2016). Assuming a simple one-parameter model of synchrotron-dust correlation, the power spectrum obtains an additional term

$$D_\ell^{s\times d,\,\nu\nu'} \equiv \epsilon_{ds}\,(f_d^\nu f_s^{\nu'} + f_s^\nu f_d^{\nu'})\,\sqrt{A_d A_s \left(\frac{\ell}{\ell_0}\right)^{\alpha_s+\alpha_d}}\,, \tag{3.13}$$

where $\epsilon_{ds} \in [-1,\,1]$ quantifies the level of correlation between both components. Adding this correlation term to the simple dust and synchrotron model described in Eqs. (3.11) and (3.12), we obtain the $\boldsymbol{C_\ell}$**-fiducial model** that was adopted in the SO pipeline comparison project (see Chpt. 4).

Moment expansion. The simple C_ℓ-fiducial model might not be sufficient to describe Galactic foregrounds in real data. The spectral indices of thermal dust and synchrotron emission vary across sky regions and along the line of sight, which cause the observed emission laws to differ from the fiducial model defined in Eq. (3.11). As we show in Chpt. 4, neglecting plausible levels of spectral variation in the model can lead to biases of more than 2σ in the inferred value of the tensor-to-scalar ratio. This motivated extensions of the spectral emission model (Tegmark, 1998; Chluba et al., 2017) that parameterize the statistical fluctuations of the spectral index $\delta\beta$ across the sky around the mean value $\bar{\beta}$. If those variations are small compared to the mean SED $\bar{f}^\nu = f^\nu(\bar{\beta})$, Azzoni et al. (2021) suggest to perform a Taylor expansion of a sky component's foreground SED up to second order

$$f^\nu(\hat{\boldsymbol{n}}) \approx \bar{f}^\nu + \delta\beta(\hat{\boldsymbol{n}})\partial_\beta \bar{f}^\nu + \frac{1}{2!}[\delta\beta(\hat{\boldsymbol{n}})]^2\partial_\beta^2 \bar{f}^\nu\,, \tag{3.14}$$

where we used short notations $\partial_\beta \bar{f}^\nu \equiv \partial f^\nu/\partial\beta|_{\beta=\bar{\beta}}$ and $\partial_\beta^2 \bar{f}^\nu \equiv \partial^2 f^\nu/\partial\beta^2|_{\beta=\bar{\beta}}$. In the simplest version of this so-called "moment expansion", spectral index fluctuations $\delta\beta_{d,s}(\hat{\boldsymbol{n}})$ of thermal dust and synchrotron emission are assumed to be Gaussian random fields that are uncorrelated with any other field. We can then parameterize their power spectra by power laws

$$D_\ell^{\beta_{d,s}} = B_d \left(\frac{\ell}{\ell_0}\right)^{\gamma_{d,s}}\,. \tag{3.15}$$

Here, we introduced the four moment parameters $\{B_d,\,\gamma_d,\,B_s,\,\gamma_s\}$, which are the power spectrum amplitudes ($B_{d/s}$) and tilts ($\gamma_{d/s}$) corresponding to dust and synchrotron emission, respectively. Inserting Eqs. (3.14) and (3.15) into Eq. (3.11), we obtain the full, lowest-order moment expansion

$$D_\ell^{\nu\nu'} = D_\ell^{\nu\nu'}(\bar{\beta}_d, \bar{\beta}_s) + D_\ell^{\nu\nu'}\big|_{1\times1} + D_\ell^{\nu\nu'}\big|_{0\times2}\,, \tag{3.16}$$

where $D_\ell\big|_{1\times 1}$ denotes the correlation between two first-order terms in the moment expansion, and $D_\ell\big|_{0\times 2}$ denotes the correlation between a zeroth-order and a second-order term. Both terms have the following explicit forms (Azzoni et al., 2021):

$$
\begin{aligned}
D_\ell^{\nu\nu'}\Big|_{1\times 1} &\equiv \sum_{c\in\{d,s\}} \partial_{\beta_c}\bar{f}_\nu^c\,\partial_{\beta_c}\bar{f}_{\nu'}^c \sum_{\ell_1\ell_2} \frac{(2\ell_1+1)(2\ell_2+1)}{4\pi} \begin{pmatrix} \ell & \ell_1 & \ell_2 \\ 0 & 0 & 0 \end{pmatrix}^2 D_\ell^{\nu\nu'}(\bar{\beta}_d,\bar{\beta}_s) D_{\ell_2}^{\beta_c}\,, \\
D_\ell^{\nu\nu'}\Big|_{0\times 2} &\equiv \sum_{c\in\{d,s\}} \frac{1}{2}\left[\bar{f}_c^\nu\,\partial_{\beta_c}^2\bar{f}_c^{\nu'} + \bar{f}_c^{\nu'}\,\partial_{\beta_c}^2\bar{f}_c^{\nu}\right] D_\ell^{\nu\nu'}(\bar{\beta}_d,\bar{\beta}_s) \sum_{\ell_1} \frac{2\ell_1+1}{4\pi} D_{\ell_1}^{\beta_c}\,,
\end{aligned}
\tag{3.17}
$$

where we used the Wigner-3j symbol in the first line. Equation (3.16), in addition to the synchrotron-dust correlation term (3.13), defines the ***C_ℓ*-moments model** used in the context of the SO pipeline comparison project (see Chpt. 4).

3.4 Simons Observatory C_ℓ pipeline

In this section, we present the architecture of the power-spectrum-based C_ℓ pipeline that we co-developed and tested as part of the main work of this Thesis, with the goal of robustly constraining degree-scale CMB B-modes with SO. Figure 3.1 displays the main stages of the SO C_ℓ component separation pipeline.

We start from simulated multifrequency polarization Q and U maps with coadded CMB, foreground sky signals (which are convolved with a Gaussian beam), and instrumental noise, as detailed later in Sect. 4.2. The simulations can either serve as samples for the empirical covariance estimation, or as mock data for pipeline validation. In the first case, we employ coadded Gaussian map realizations of the fiducial power spectra of CMB, noise, and Galactic foregrounds using the `synfast` function of the `HEALPix` code (Górski et al., 2005). In the second case, we use dust and synchrotron templates from the Python Sky Model (`PySM`, Thorne et al., 2017; Zonca et al., 2021), which are more realistic and thus generally non-Gaussian. `PySM` is a numerical code to generate artificial maps of Galactic emission in intensity and polarization, and is based on publicly available data from *Planck* and WMAP. The first two stages estimate angular BB power spectra from maps, while the final two stages construct the covariance and infer the cosmological and foreground parameters.

The map-level noise has an autocorrelation that biases power spectrum estimates. Ideally, we can model and subtract the noise autocorrelation on the power spectrum level. While this is the most efficient approach, it requires an accurate characterization of the noise properties, which may not be available for real data, at least during the initial phase of an experiment. For the SO C_ℓ pipeline,

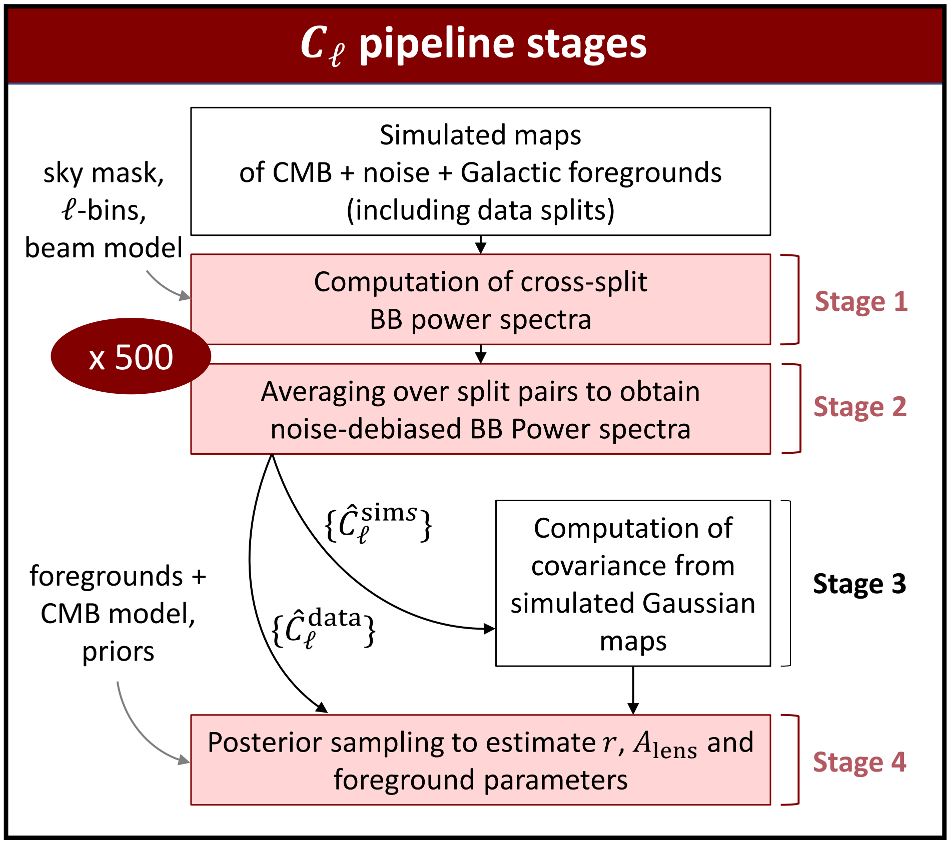

Figure 3.1: Schematic of the C_ℓ component separation pipeline, starting from reading multifrequency polarization maps that include data splits, estimating the sample covariance and noise-unbiased power spectra, and finally inferring cosmological and foreground parameters. All stages but the covariance calculation are repeatedly run on simulations. See text for details.

we therefore adopt the more conservative "cross-split" technique, which relies on calculating power spectra across unequal data splits. Averaging over this set of cross-split power spectra avoids the noise bias in a robust and model-independent way, although slightly increasing the computational cost. It all relies on the assumption that our experimental design allows for different data splits with largely uncorrelated noise. Real CMB experiments, such as BICEP/*Keck* or *Planck*, usually provide splits as a result of dividing data with regard to different criteria such as observation periods, sets of detectors, or scanning strategies. The computation of the power spectrum covariance in the SO pipeline relies on simulated split maps, which are simply realized by coadding independent noise realizations to a single sky signal.

3.4.1 Power spectrum estimation

In **Stage 1,** we compute the pseudo-C_ℓ estimates of all cross-split spectra $\{\hat{C}_\ell^{(\nu_1,s_1)(\nu_2,s_2)}\}$ of a given cross-frequency power spectrum, where ν_i and s_i denote the frequency band and the split, respectively. This procedure depends on three auxiliary inputs: the sky mask, the ℓ-binning scheme, and the instrument beam model. As explained in Sect. 3.1.2, we use the `NaMaster` implementation of the PCL algorithm, which analytically decouples ℓ-modes and mitigates E-to-B leakage caused by observing the partial sky through a sky mask. To facilitate B-mode purification, meaning the subtraction of ambiguous polarization modes from pure B-modes (see Sect. 3.1.2), we apodize the edges of the mask with a C^1-type kernel of 10 degree radius (see Grain et al., 2009). The multipole binning scheme spans a range between $\ell_{\rm min} = 30$ and $\ell_{\rm max} = 300$ with a bin width $\Delta\ell = 10$, resulting in $N_{\rm bin} = 27$ ℓ-bins. We deconvolve the C_ℓs considering a set of Gaussian beam transfer functions $b_\ell^{\nu_i}$ corresponding to the individual frequency channels:

$$\hat{C}_\ell^{\nu_1\nu_2} \rightarrow \hat{C}_\ell^{\nu_1\nu_2}/(b_\ell^{\nu_1} b_\ell^{\nu_2})\,. \tag{3.18}$$

In **Stage 2,** we average over the cross-split spectra, considering only unequal splits $s_1 \neq s_2$, to estimate the noise-unbiased power spectra. Note that if we included spectra with $s_1 = s_2$, we would pick up the noise auto-correlation that we want to avoid. Considering four splits, this means we sum over $4(4-1)/2 = 6$ unique cross-split spectra for frequency autocorrelations ($\nu_1 = \nu_2$) and $4^2 = 16$ unique cross-split spectra for frequency cross-correlations ($\nu_1 \neq \nu_2$). Given the six SO frequency channels, we end up with a final set of $6(6+1)/2 = 21$ noise-unbiased power spectrum estimates.

3.4.2 Covariance

As explained in Sect. 3.2, the key advantage of using the Gaussian likelihood approximation (or, alternatively, the HL likelihood) is the fact that the covariance matrix can be precomputed from simulations, thus removing the computational bottleneck of having to evaluate the covariance at every sampling step during the likelihood evaluation. In **Stage 3** as displayed in Fig. 3.1, we compute a set of binned BB power spectra $\{C_l^{\nu_1,\nu_2}\}$ from simulations and use them to calculate the BB power spectrum covariance Σ as

$$\Sigma_{ll'}^{\nu_1\nu_2,\nu_1'\nu_2'} = \delta_{ll'}^{\mathcal{K}}\, \hat{\mathrm{Cov}}(\{C_l^{\nu_1\nu_2}\},\, \{C_{l'}^{\nu_1',\nu_2'}\})\,, \tag{3.19}$$

where $\delta^{\mathcal{K}}$ denotes the Kronecker symbol, $(l,\, l')$ and $(\nu,\, \nu')$ label the ℓ-bins and the frequency channels, respectively, and $\hat{\mathrm{Cov}}(\boldsymbol{x},\, \boldsymbol{y})$ denotes the sample covariance between two data vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of length N, defined as

$$\hat{\mathrm{Cov}}(\boldsymbol{x},\, \boldsymbol{y}) \equiv \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})\,(y_i - \bar{y})\ , \tag{3.20}$$

where $\bar{x} \equiv \sum_i x_i/N$ and $\bar{y} \equiv \sum_i y_i/N$. We included the Kronecker symbol to imply that we neglect any coupling between unequal ℓ-bins $l \neq l'$, which we find to be a good approximation in our case, as we show below.

Covariance matrices that are calculated from simulations may be subject to two types of issues: numerical instability and incompatibility with data in the likelihood. The first one can arise from numerical noise that leads to a noninvertible covariance matrix, while the second one can arise from poorly chosen covariance simulations that result in bad estimates of the power spectrum variance. Here, we present a series of robustness checks as performed in the context of the SO pipeline comparison project (see Chpt. 4) to detect these potential issues regarding the covariance.

- **Number of simulations:** The finite number of simulations N_{sims} that are used to calculate the sample covariance (3.20) introduces numerical noise. If we approximate the binned power spectra as Gaussian random variables (which is a reasonable assumption given that the lowest ℓ-bin is effectively a sum of $M \approx f_{\mathrm{sky}}\Delta\ell(2\ell+1) = 61$ harmonic modes), their sample covariance follows a rescaled Wishart distribution (Wishart, 1928). We can then estimate the element-wise standard deviation as

  $$\sigma(\Sigma_{ij}) \approx \sqrt{(\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj})/N_{\mathrm{sims}}}\,, \tag{3.21}$$

  which gives us an expected S/N for the C_ℓ standard deviation of about 20% for $N_{\mathrm{sims}} = 100$ and about 9% for $N_{\mathrm{sims}} = 500$. We run the entire pipeline in both cases and find negligible bias on the parameter uncertainties. For the main analysis cases in the SO pipeline comparison project, we conservatively choose $N_{\rm sims} = 500$.

- **Truncation of subdiagonals[3]:** As shown in Sect. 3.1.1, C_ℓ estimates on the partial sky are correlated among different multipoles due to the anisotropy of the mask. As the correlation quickly reduces with the distance $|\ell - \ell'|$ (see Fig. 12 of Alonso et al., 2019), we can mitigate this effect by introducing ℓ-binning. As a result, we do not expect significant mode coupling between unequal ℓ-bins. However, numerical noise may affect the elements in the covariance matrix that have $l \neq l'$, causing it to be noninvertible. A straightforward solution to this is to set all subdiagonals to zero, provided this truncation does not erase actual physical information. We test the robustness of this choice by a) comparing the first and second subdiagonals ($|l - l'| = 1, 2$) with the expected numerical noise and b) comparing the minimum χ^2 statistic (see Appendix A) calculated from 100 Gaussian simulations and zero, one, or two included subdiagonals with the theoretically expected distribution. Figure 3.2 shows the results on the left panel. We find that all subdiagonals are consistent with numerical noise and a diagonal covariance is in excellent statistical agreement with the expectation.

- **Non-Gaussian foregrounds:** Polarized Galactic foregrounds are anisotropic and non-Gaussian, but their statistical properties beyond the two-point correlation are largely unknown (Delouis et al., 2022; Coulton & Spergel, 2019). In the SO C_ℓ pipeline, we ignore foreground-related non-Gaussianity at the level of the power spectrum covariance, for two main reasons: first, the considered non-Gaussian foreground models do not provide stochastic samples, and second, drawing Gaussian map realizations from a fiducial C_ℓ model is both numerically efficient, and easy to implement and understand from a statistical viewpoint. We may ask whether the C_ℓ covariance obtained in this way are a sufficiently good representation of a data set that contains a realistic level of foreground-induced non-Gaussianity. To investigate this for SO, we generated 100 sky maps with coadded CMB, noise, and foregrounds, each for different non-Gaussian dust and synchrotron templates from the `PySM` model (Zonca et al., 2021), and assessed the empirical distributions of the minimum χ^2 values in each case. In the right panel of Fig. 3.2, we show a comparison of the more complex `dmsm` model, which includes spatially varying frequency spectral indices, and the simpler `d0s0` model with constant spectral indices (see Sect. 4.2 for more details). As can be seen from the figure, the `d0s0` model is compatible with the theoretical expectation, which confirms that the Gaussian simulation-based covariance is an adequate choice for this foreground scenario. The `dmsm` model leads to sightly larger χ^2 values, indicating that the Gaussian covariance is still a rea-

---

[3] We refer to "subdiagonals", meaning all matrix entries adjacent to the diagonal (first sub- and superdiagonal), and then iteratively the respective adjacent elements further out from the diagonal (sub- and superdiagonals of order > 1).

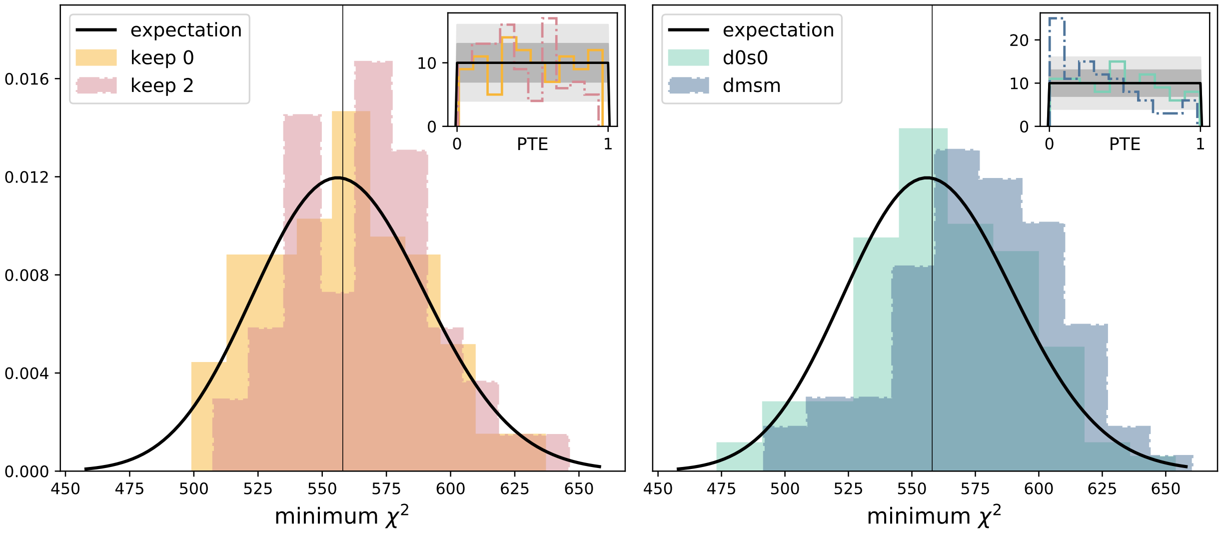

Figure 3.2: Statistical robustness tests of the empirical power spectrum covariance used with the SO C_ℓ pipeline. We show results from checking the truncation of subdiagonals (*left panel*) and the adequate representation of artificial non-Gaussian foreground data (*right panel*), considering 100 simulated maps of CMB, foregrounds, and noise in the SO "goal-optimistic" scenario (see Sect. 4.2) using the C_ℓ-fiducial model. The black solid line represents the expected distribution of the minimum χ^2 with $N - P = 558$ degrees of freedom. The small inlays show the corresponding probabilities to exceed (PTE) from simulations and the two gray-shaded ares represent the 1σ and 2σ standard error expected from uniform sampling. The left panel considers Gaussian foreground simulations and the right panel considers a diagonal C_ℓ covariance.

sonable, but not perfect description of this foreground model. While exploring non-Gaussian covariances compatible with complex, non-Gaussian foreground models is beyond the scope of this work, this should be the topic of future investigations. In the context of this work, however, we use Gaussian covariances, keeping in mind that our quantification of statistical uncertainty may be slightly inaccurate in the case of complex foregrounds scenarios, such as `dmsm` and `d10s5` (see Chpt. 4).

In real B-mode data analysis, significant mode coupling and non-Gaussianity can have other sources, such as small-scale leakage from delensing templates or instrumental systematic effects. These topics are highly relevant for the analysis of real data with the SO C_ℓ pipeline and we address them in the Conclusions.

3.4.3 Inference

Two of the C_ℓ pipeline's main purposes are estimating noise-unbiased power spectra from data splits and computing the covariance. In **Stage 4** as shown in Fig. 3.1, we construct the likelihood from the covariance and perform MCMC sampling to obtain the parameter's posterior distribution. The data likelihood combines the noise-unbiased power spectrum data $\hat{C}_\ell$, the theoretical model $C_\ell(\boldsymbol{\theta})$ that depends on the theory parameters $\boldsymbol{\theta}$, and the covariance matrix Σ:

$$P(\{\hat{C}_\ell\}|\{C_\ell(\boldsymbol{\theta})\}, \Sigma) \tag{3.22}$$

Each of these ingredients must be carefully chosen before the inference can commence. We previously discussed sky models and likelihoods in a general fashion. Now, let us address the sky models chosen for the SO pipeline comparison project.

As the nominal likelihood, we chose the Gaussian approximation (Sect. 3.2.1). Although, in principle, the HL likelihood (Hamimeche & Lewis, 2008) might be slightly more accurate at low ℓ, we verified through simulations that using the Gaussian approximation makes no practical difference for the inferred parameter posterior distribution.

As for the theoretical C_ℓ-based sky models, we implemented four alternatives:

- **C_ℓ-fiducial model.** This model, as introduced in Eq. (3.11), includes the lensed CMB, a possible primordial tensor contribution, and Galactic thermal dust and synchrotron emission. The latter two are modeled in frequency space by a modified blackbody law and a power law, respectively, and by power laws in harmonic space. Thermal dust and synchrotron may be statistically correlated among each other, which is accounted for by the model, but their spectral laws are assumed to be constant across the sky, with no frequency decorrelation. The model contains nine free parameters: the CMB parameters r and $A_{\rm lens}$, foreground amplitudes A_s and A_d for synchrotron and dust (evaluated at 23 GHz and 353 GHz, respectively), power spectral indices α_s and α_d, frequency spectral indices β_s and β_d, and the synchrotron-dust correlation ϵ_{ds}. In thermodynamic temperature units, this model can be written as

  $$D_\ell^{\nu\nu'}(\text{fiducial}) = D_\ell^c + D_\ell^{s,\,\nu\nu'} + D_\ell^{d,\,\nu\nu'} + D_\ell^{s\times d,\,\nu\nu'}\,, \tag{3.23}$$

  where CMB (c), dust (d), and synchrotron (s) components are given by Eqs. (3.10) and (3.12), and $D_\ell \equiv \ell(\ell+1)C_\ell/2\pi$.

- **C_ℓ-moments model.** This model performs a first-order harmonic expansion in the SED spectral indices of thermal dust and synchrotron around their mean values on the sky (see Sect. 2.6). As for the multifrequency power spectra, this results in an extension of the C_ℓ fiducial model by the four parameters $\{B_d,\, B_s,\, \gamma_d,\, \gamma_s\}$. The C_ℓ-moments model can be written as

  $$D_\ell^{\nu\nu'}(\text{moments}) = D_\ell^{\nu\nu'}(\text{fiducial}) + D_\ell|_{1\times1} + D_\ell|_{0\times2}\,, \tag{3.24}$$

  where the two last terms are defined in Eq. (3.16).

- **ℓ-wise CMB model.** This model, inspired by previous works (Dunkley et al., 2013; Planck Collaboration XI, 2016; Planck Collaboration V, 2020) lifts the scale dependence of the CMB in the C_ℓ-fiducial model by parameterizing the CMB amplitude in every ℓ-bin, thereby replacing r and $A_{\rm lens}$. The benefit of this is the retrieval of individual CMB multipole bins with error bars, which we used in the comparison of the cleaned CMB power spectra for the SO pipeline comparison project (see Sect. 4.4.1). We can write the ℓ-wise CMB model as in Eq. (3.23), with a new CMB model that reads

  $$D_\ell^c = A_\ell^c\,, \tag{3.25}$$

  where $\{A_\ell^c\}$ parameterize the CMB power at multipole ℓ. While this increases the total number of parameters to $N_{\rm bin} + 6$ (in our case, 33), which may potentially dilute the cosmological constraints found from the retrieved bandpowers, we find that this model leads to results on r and $A_{\rm lens}$ that are consistent with the C_ℓ-fiducial model, with little to no extra variance.

- **ℓ-wise sky model.** The power spectrum of polarized foreground emission may not scale like a simple power law. Moreover, the spectral emission laws may generally vary with angular scale. To include both effects, we implemented an ℓ-wise sky model in all three sky components as described in Appendix D of BICEP2/Keck Collaboration (2018), thus generalizing the C_ℓ fiducial model (3.23). To do so, we made the SED spectral indices and the dust-synchrotron correlation scale dependent,

  $$\beta_d \to \beta_d(\ell)\,, \quad \beta_s \to \beta_s(\ell)\,, \quad \epsilon_{ds} \to \epsilon_{ds}(\ell)\,, \tag{3.26}$$

  and re-parameterize the harmonic models by ℓ-wise component amplitudes,

  $$D_\ell^c = A_\ell^c\,, \quad D_\ell^d = A_\ell^d\,, \quad D_\ell^s = A_\ell^s\,, \tag{3.27}$$

  thus obtaining a total of $6\,N_{\rm bin}$ (in our case, 162) parameters. In practice, we use the fact that ℓ-modes decouple, and independently sample the six-dimensional posteriors for every ℓ. While this model considerably dilutes the parameter constraints with respect to simpler models, it may still be useful for comparison with other component separation methods.

To retrieve the marginal posteriors of CMB and foreground parameters, we use Bayesian sampling methods. In most models, where the number of jointly sampled parameters does not exceed 10, this is done with the MCMC sampler `emcee` (Foreman-Mackey et al., 2013). In the ℓ-wise CMB model, we use the nested sampler `PolyChord` (Handley et al., 2015a,b) that guarantees faster convergence for higher-dimensional models.

3.4.4 `BBPower` code implementation

The code of the C_ℓ pipeline is publicly available under the name `BBPower`[4]. Originally developed by David Alonso, the main work carried out in this Thesis consisted in implementing the parallelization on a cluster, implementing the ℓ-wise CMB and sky models described in the previous section, and continuously testing and validating all parts of the code using simulations. Here, we briefly discuss the code implementation, its modular structure, and its parallel performance tested on computing facilities at the National Energy Research Scientific Computing Center (NERSC)[5].

The code is mainly written in the `python` programming language, with some auxiliary scripts written in `bash`. In practice, the code operates in two production modes: covariance computation (C) and data analysis (D). Some of the `BBPower` core modules are used in both modes and can adopt their functionality to perform either C or D. The `BBPower` core modules are:

- **`BBPowerSpecter`.** This is the module that estimates pseudo-C_ℓ BB power spectra from polarization maps in multiple channels using the `NaMaster` code (Alonso et al., 2019). Its inputs are the Q and U maps including data splits, the sky mask, the frequency bandpass windows, and the beam transfer function in multipole space. Its outputs are the estimated cross-split C_ℓs. `BBPowerSpecter` is used in both modes C and D and corresponds to Stage 1 as displayed in Fig. 3.1.

- **`BBPowerSummarizer`.** This module can perform two tasks: the computation of noise-unbiased C_ℓs from cross-split spectra and, optionally, the estimation of the covariance from simulations. Its inputs are the cross-split C_ℓs from data and, optionally, the cross-split C_ℓs from a set of simulations. Its output are noise-unbiased C_ℓs or, alternatively, the covariance matrix. `BBPowerSummarizer` can be tasked to run either in mode D, estimating data C_ℓs (Stage 2 in Fig. 3.1), or mode C, estimating the C_ℓ covariance (Stage 3 in Fig. 3.1).

- **`BBCompSep`.** This module performs the C_ℓ-based component separation and parameter inference. Its inputs are the noise-unbiased data C_ℓs, the covariance matrix, and the parameter priors. Its output is the MCMC chain of parameter posterior samples. `BBCompSep` is only needed in mode D and corresponds to Stage 4 in Fig. 3.1.

Parts of the code are designed to run in parallel at the *Perlmutter* supercomputing cluster at NERSC. Parallelization is currently managed through several auxiliary scripts that distribute single instances of the pipeline (or parts thereof) over all available computing nodes. This effective parallel “for loop” applies to

---

[4] URL: github.com/simonsobs/BBPower

[5] URL: nersc.gov

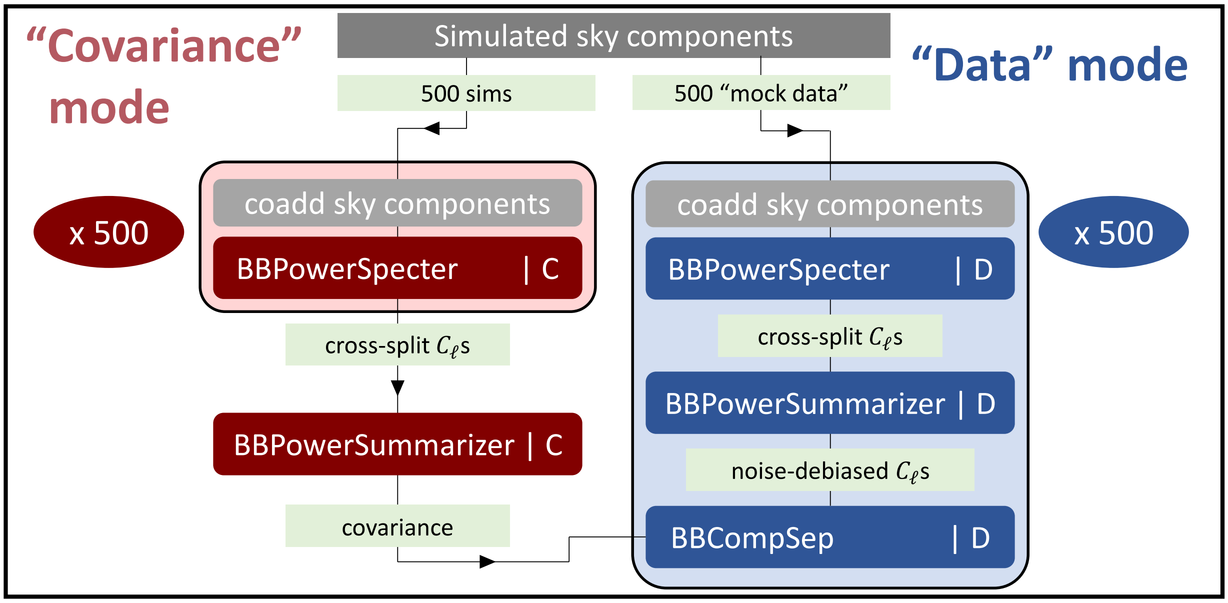

Figure 3.3: Schematic overview of the parallelization implemented in the `BBPower` code, highlighting the two operation modes to estimate covariances from simulations and perform inference on simulated "mock" data.

- **Covariance mode** (running time: about 20 minutes for 500 simulations at 15 computing nodes[6]): the computation of the cross-split C_ℓs from simulated maps (`BBPowerSpecter` in mode C)
- **Data mode** (running time: about 30 minutes for 500 simulations at 20 computing nodes): the computation of the cross-split and noise-unbiased C_ℓs from map data and the C_ℓ-based component separation (`BBPowerSpecter`, `BBPowerSummarizer`, and `BBCompSep`, all in mode D)

The computation of the covariance (`BBPowerSummarizer` in mode D) runs efficiently as a single process at *Perlmutter* within 30 minutes . Another auxiliary script included in both parallel loops performs the in-place coaddition of simulated sky components (see Sect. 4.2), so they do not need to be stored in all possible combinations. Figure 3.3 summarizes the parallelization of `BBPower`.

---

[6] The "running time" is the wall clock time that is needed at NERSC for the parallel processing of 500 simulations including four splits maps at six frequency bands and `HEALPix` resolution $N_{\rm side} = 512$, totaling to 432 GB of memory space.

3.5 Summary

Detecting a primordial polarization B-mode signal is considered a “smoking gun” for cosmic inflation and one of the main science goals of next-generation experiments such as SO. Primordial B-modes are known to be extremely faint and highly contaminated by gravitational lensing, Galactic foreground emission, and instrumental systematics. Inferring cosmological parameters from power spectra of the CMB anisotropies has been highly successful in the CMB field, motivating the SO Collaboration to base one of their main B-mode analysis pipelines on this technique.

The main focus of this chapter is the design of a CMB data analysis pipeline that robustly measures B-mode power spectra at intermediate to large scales on the partially masked sky as observed by the SO SATs. Measuring the power spectrum of polarization B-modes on the partial sky is challenging, since we need to overcome

- the coupling between different angular modes, potentially biasing power spectra depending on the input signal, and
- the power leakage from the (much larger) E-modes into B-modes, which may significantly increase the statistical uncertainty of the inferred value of the tensor-to-scalar ratio.

As an additional challenge, it is imperative for any B-mode inference pipeline to accurately characterize and subtract Galactic foregrounds, which are one of the main contaminants to the much fainter CMB signal.

As the main result of this chapter, we give an overall description of the SO B-mode pipeline and report on the specific design choices that were made during its development. We start by reviewing the PCL estimator that is implemented in the pipeline. We explain how the PCL estimator enables us to efficiently retrieve unbiased and minimum-variance power spectrum estimates in a broad range of realistic scenarios, and outline how our code uses analytic B-mode purification to minimize the spurious variance from mask-induced E-to-B leakage. We introduce the power spectrum likelihood, explain why we chose a simple Gaussian likelihood approximation, and briefly discuss its generalization. We present a set of data models implemented in the power spectrum likelihood, highlighting how they differ in their specific parameterization of the CMB and foregrounds. Two of them, the C_ℓ-fiducial and the C_ℓ-moments model, adapt to different levels of foreground complexity that we may encounter in real data, while the ℓ-wise CMB and ℓ-wise sky models are designed to retrieve power spectrum estimates of individual sky components. We describe the full SO pipeline and its various stages, highlighting tests and design choices that tightly couple with the pipeline development work as a major contribution to this Thesis. In this context, we describe the parallel code architecture that was designed as part of this Thesis.

As possible limitations of this pipeline, we identify several stages that can be generalized to account for more complex real-world data than we consider in this chapter. First, we do not model instrumental systematic effects, such as a frequency-dependent polarization angle, which are expected to affect real data at the power spectrum level (see Sect. 2.7). Second, this pipeline does not account for possible mode coupling and E-to-B leakage beyond the mask-induced effect accounted for by the `NaMaster` code. Instead, those can have multiple origins, such as atmospheric contamination, ground pickup, or detector gain drifts. Third, we do not include template-based delensing, which should significantly reduce the lensing B-mode power at intermediate and small scales, resulting in a lower statistical uncertainty. Finally, we ignore non-Gaussian contributions to the power spectrum covariance matrix, as expected from Galactic foregrounds (see Sect. 3.4.2).

Possible future improvements of the SO pipeline that address these limitations are being actively studied. To account for the most important systematic biases to power spectra beyond the ones from Galactic foregrounds, timestream simulations must be used to estimate the transfer function to quantify the loss of power at each angular scale. Ongoing work (Hertig et al., 2023) has achieved the implementation of map-based delensing in the SO pipeline. Another recent work (Abril-Cabezas et al., 2023) proposes a model to account for non-Gaussian contributions to the power spectrum variance due to Galactic foregrounds. We elaborate on these future perspectives in the Conclusions.

CHAPTER 4

The Simons Observatory degree-scale B-mode cleaning pipelines

Detecting a signal of gravitational waves from the very early Universe requires the subtraction of Galactic foreground emission from the CMB by exploiting spectral differences. As we explored in the previous chapter, a possible avenue to do so is by comparing B-mode power spectra at multiple frequencies with a parametric sky model, the result of which is the SO B-mode power spectrum pipeline. In this chapter, we broaden our perspective on B-mode analysis, in that we include other component separation techniques beyond the power spectrum pipeline. Given the large set of possible systematic contaminants, any single measurement of the tensor-to-scalar ratio must be carefully checked for robustness. One way of achieving this is by comparing several independent data analysis methods. SO's efforts to constrain primordial B-modes extend to three inherently different foreground cleaning algorithms. The C_ℓ pipeline described in Chpt. 3 is one of them.

This chapter is based on a recent paper project by the SO Collaboration, Wolz et al. (2023a) (hereinafter BB2023). This work aims at validating the three component separation pipelines (with two extensions) as to their ability to constrain the tensor-to-scalar ratio r with the SO SATs in the presence of complex Galactic foregrounds, ensuring that post-cleaning residuals are small enough to not bias r. This is a continuation of the work described in the SO Science Goals and Forecasts paper (SO Collaboration, 2019, hereinafter SO2019) that sets a baseline constraint on the statistical uncertainty of $\sigma(r) = 0.003$ at 68% central CL. Compared to SO2019, BB2023 assumes more complex foreground simulations and introduces more advanced cleaning pipelines that should meet the requirement of unbiased r for all foreground scenarios and at a precision of $\sigma(r) = 0.003$. Robustness requires that the results agree between the inherently different algorithms chosen for BB2023, which include C_ℓ-based parametric, map-based parametric, and needlet-based blind component separation. Compared to the real data analysis case, this project does not attempt the map-level removal of gravitational lensing B-modes, or the inclusion of instrumental systematic effects. Those additional challenges are left for future work, which we briefly address in the Conclusions. With the SO Collaboration expecting the first data in late 2023, this work is a timely assessment of the anticipated constraints on primordial B-modes.

This chapter is structured as follows. We start with a brief overview of the pipelines in Sect. 4.1 and discuss the sky simulations in Sect. 4.2. We then present the main r inference results in Sect. 4.3 and close with the pipeline comparison methods in Sect. 4.4.

4.1 Component separation pipelines

After setting the context of BB2023, let us describe the three SO component separation pipelines and the two extensions in more detail. All operate on partial-sky maps at the six SO frequency channels between 27 and 280 GHz that contain simulated CMB, Galactic foregrounds, and instrumental noise and output marginal posteriors of the inferred tensor-to-scalar ratio r and the lensing amplitude $A_{\rm lens}$. The pipelines differ either in their cleaning algorithm or in the modeling of foregrounds. We can categorize algorithms into parametric ones that use specific spectral foreground models, or blind algorithms that do not make any assumptions on foreground SEDs. Furthermore, we can distinguish between the data spaces where cleaning is performed: this can be pixel space, angular power spectra, or the space of needlets, which are basis functions on the sphere that are localized in real and harmonic space. Let us now briefly describe each one of them:

- **Pipeline A:** The main part of the work presented in this Thesis focuses on Pipeline A. This is the C_ℓ pipeline with the C_ℓ-fiducial model described in Sect. 3.4. It cleans foregrounds from multifrequency power spectra assuming constant spectral indices for Galactic foreground emission.

- **Pipeline A + moments:** This is the C_ℓ pipeline with the C_ℓ-moments model described in Sect. 3.4. It extends the parametric foreground model of Pipeline A by the moment parameters that account for spatially varying spectral indices in dust and synchrotron emission.

- **Pipeline B:** This is a blind cleaning pipeline based on the `NILC` algorithm (Basak & Delabrouille, 2012, 2013). It does not assume any spectral model for non-blackbody components. After transforming the maps to needlet space, the algorithm linearly combines multifrequency data to retrieve the cleaned CMB. The parameters r and $A_{\rm lens}$ are then inferred through a power spectrum likelihood.

- **Pipeline C:** This pipeline, introduced in Stompor et al. (2008), compares the multifrequency input maps with a parametric SED model that assumes spatially homogeneous emission. From the best-fit spectral parameter values, the pipeline constructs a cleaned CMB map and estimates the parameters r and $A_{\rm lens}$ through a power spectrum likelihood.

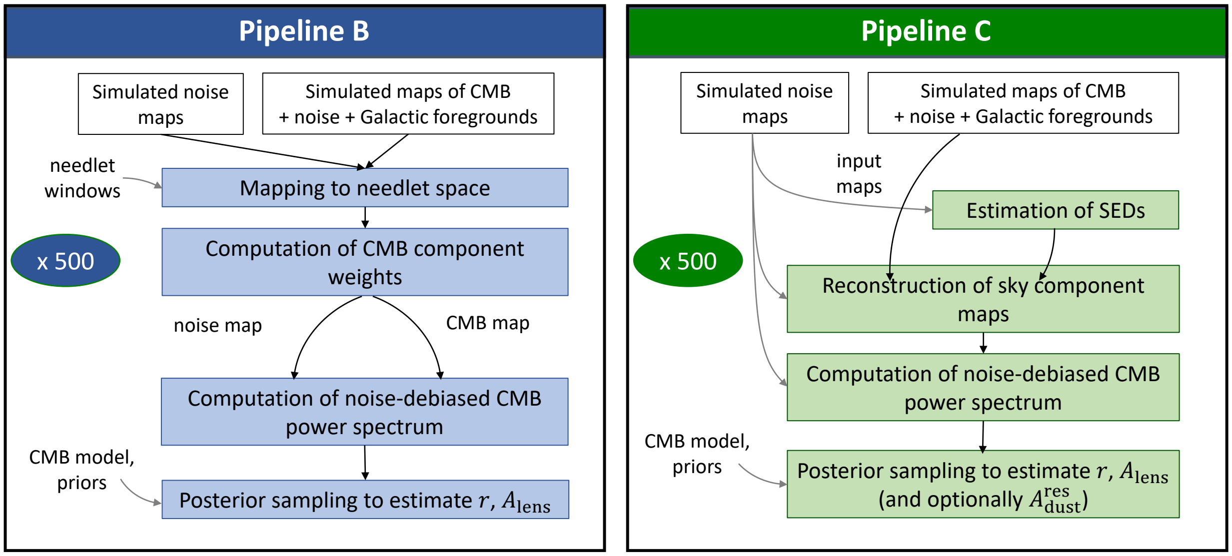

Figure 4.1: Schematics of Pipelines B and C (`NILC` and map-based parametric B-modes cleaning) used in BB2023.

- **Pipeline C + dust marginalization:** This extension of Pipeline C differs from the former at the post-cleaning level, when parameters are inferred from clean CMB maps. It marginalizes not only over r and A_{lens} but also over residual dust power estimated from the cleaned dust map that was constructed from the best-fit spectral model.

To better understand the results of the pipeline comparison, it is worth taking a closer look at the algorithms used in Pipelines B and C.

4.1.1 `NILC` cleaning

Let us first outline the principal architecture of Pipeline B as described in detail in Sect. 2.2 of BB2023. The `NILC` cleaning algorithm (Delabrouille et al., 2009; Basak & Delabrouille, 2012, 2013) reconstructs clean CMB maps from multifrequency maps by exploiting the spectral difference between the blackbody emission of the CMB and the emission laws of other sky components. Pipeline B operates in needlet space, allowing to separate different scales in harmonic space and simultaneously isolate features in pixel space (Narcowich et al., 2006). It uses the Internal Linear Combination (`ILC`) algorithm, a so-called "blind" method that does not assume models for the SEDs of any non-CMB components (see, e.g., Tegmark & Efstathiou, 1996).

The left panel of Fig. 4.1 shows the main analysis steps for Pipeline B. We first decompose the input data into five needlet windows, each one being sensitive to a different multipole range and a different region on the sky. Inside each window, we compute the `ILC` weights that allow a retrieval of the minimum-variance CMB signal from multiple frequency channels. These weights are normalized so that the sum over all needlet windows equals one, guaranteeing the conservation of the CMB signal amplitude. The final `NILC` CMB map is the sum of the five pixel maps that result from the needlet-weighted multifrequency maps. The same set of weights is used to obtain "noisy CMB" and "noise-only" maps from simulations. In the last two steps, we compute power spectra using the `NaMaster` code with B-mode purification to alleviate variance leakage from ambiguous modes (see Sect. 3.1.2 and Alonso et al., 2019), and infer r and A_{lens} with the Gaussian likelihood (Sect. 3.2.1).

4.1.2 Map-based parametric cleaning

Pipeline C performs map-based parametric cleaning as described in Stompor et al. (2008), using the numerical `ForeGroundBuster` code[1] (`FGBuster`, Poletti & Errard, 2023). The observed multifrequency data are modeled as a sum of noise and signal components that linearly map onto the frequency channels,

$$\boldsymbol{d_\nu} = \sum_c A_{c\nu}(\beta_c)\, \boldsymbol{s_c} + \boldsymbol{n_\nu}\,, \tag{4.1}$$

where $\boldsymbol{s_c}$ are the sky signals of components $c \in \{\text{CMB}, \text{dust}, \text{synchrotron}\}$ and $A_{c\nu}(\beta_c)$ is the mixing matrix. The elements of the mixing matrix are the spectral emission laws of the three sky components evaluated at the six SO frequencies, parameterized by the spectral indices $\{\beta_c\}$: a modified blackbody with spectral index β_d and fixed temperature $T_d = 20$ K for Galactic thermal dust, a power law with spectral index β_s for Galactic synchrotron, and a blackbody law in the case of the CMB. Input data are compared to this spectral emission model through a likelihood that assumes isotropic emission and includes statistical uncertainty through the pixel noise variance estimated from 500 noise-only simulations. The maximum-likelihood spectral indices are then inserted into the mixing matrix to retrieve the component maps.

The right panel of Fig. 4.1 shows the main analysis steps of Pipeline C. We start by estimating the spectral indices from multifrequency maps and use them to reconstruct maximum-likelihood component maps. We obtain component-wise noise maps by applying the previously found mixing matrix to the 500 noise-only simulations. We estimate B-purified, noisy component power spectra using the `NaMaster` code and obtain the final noise-debiased result by subtracting the average noise-only power spectra. Finally, we infer r and A_{lens} from the CMB-only power spectrum using a Gaussian power spectrum likelihood (see Sect. 3.2.1) with the CMB-only model from Eq. (3.10). For more details, we refer to Sect. 2.3 in BB2023.

---

[1] URL: github.com/fgbuster/fgbuster

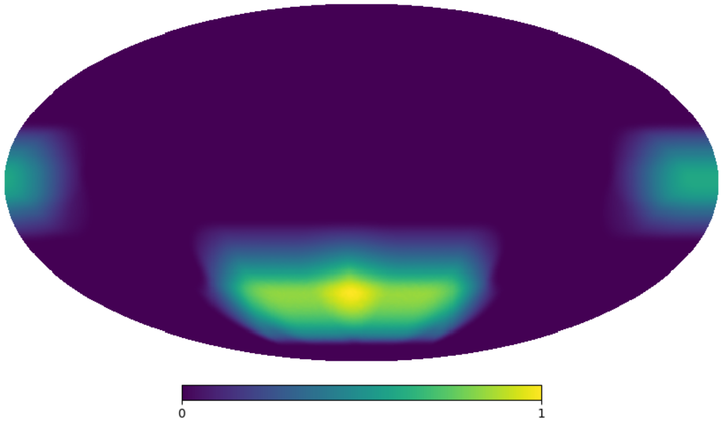

Figure 4.2: SO SAT sky mask used for the B-mode pipelines. This mask is displayed in equatorial coordinates and was apodized using a C^1 kernel of 10 degrees radius to optimize the performance of analytical B-mode purification.

4.2 Sky simulations

The simulations used in this work are coadded polarized sky maps in Q and U, containing the CMB signal, instrumental noise, and Galactic foregrounds. These are the major contributions expected to influence primordial B-mode inference, in addition to instrumental systematic effects (see Sect. 2.7). While including the latter would exceed the scope of this project, targeting them in future work is indispensable for robust data analysis and we will address this in the Conclusions.

All simulated maps use the "Hierarchical Equal Area isoLatitude Pixelisation of a 2-sphere" scheme (`HEALPix`, Górski et al., 2005) with $N_{\mathrm{side}} = 512$ (or a pixel size of 6'.87)[2]. The SAT sky mask, displayed in Fig. 4.2, effectively covers 10% of the sky in two distinct patches in the northern and southern Galactic hemisphere. The six frequency channels are simulated at the SO band centers at 27, 39, 93, 145, 225, and 280 GHz. We assume delta-like frequency bandpasses in order to accelerate the production of these simulations. However, as all pipelines are able to handle finite bandpasses, this approximation should not impact their performance. The sky signals are convolved with Gaussian beam window functions with full widths at half maximum (FWHM) as displayed in Tab. 4.1. In the following, we describe each simulated component in more detail.

- **CMB:** The polarized CMB simulations are Gaussian realizations of theory power spectra evaluated at the fiducial cosmology of *Planck* 2018 (Planck Collaboration VI, 2020) including gravitational lensing of amplitude $A_{\mathrm{lens}} \in \{0.5,\ 1\}$ and a tensor-to-scalar ratio $r \in \{0,\ 0.01\}$. We simulate 500 CMB Q and U maps for each of the four cosmologies.

---

[2] A `HEALPix` map has $N_{\mathrm{pix}} = 12N_{\mathrm{side}}^2$ pixels of the same area $\Omega_{\mathrm{pix}} = \pi/(3N_{\mathrm{side}}^2)$.

- **Noise:** We consider Gaussian noise realizations of the parametric power spectrum model presented in SO2019, given by

$$N_\ell = N_{\rm white}\left[1+\left(\frac{\ell}{\ell_{\rm knee}}\right)^{\alpha_{\rm knee}}\right]. \tag{4.2}$$

  The white-noise amplitude $N_{\rm white}$ and the $1/f$ noise parameters $\{\ell_{\rm knee},\ \alpha_{\rm knee}\}$ depend on the SO noise scenario and assume different values for each frequency channel, listed in Tab. 4.1. The "baseline" and "goal" noise cases span a range of predicted white noise levels for the SO SATs, while the "optimistic" and "pessimistic" $1/f$ noise cases mimic the effects of filtering, polarization modulation and atmospheric emission based on measurements from previous ground-based experiments (see Sect. 2.2 of SO2019). We account for an inhomogeneous scanning strategy by dividing the noise maps by the inverse square root of the hit counts shown in Fig. 4.2. For each of the four scenarios, we simulate 2000 noise Q and U maps (500 × 4 data splits[3]) at six frequencies.

- **Galactic foregrounds.** We simulated maps of polarized synchrotron and thermal dust emission based on models produced by `PySM` (Thorne et al., 2017; Zonca et al., 2021). We consider five diffuse foreground emission models of increasing complexity: Gaussian, `d0s0`, `d1s1`, `dmsm`, and `d10s5`[4], described hereinafter.

  Gaussian maps are random realizations of the C_ℓ-fiducial model (see Sect. 3.4.3) at spectral parameter values that were estimated from the `PySM` "nominal index" models of synchrotron and dust (see Sects. 2.1.1 and 2.2.1 of Thorne et al., 2017), evaluated at the SO sky patch. In thermodynamic temperature units, those fiducial values are (d: thermal dust at 353 GHz; s: synchrotron at 23 GHz): $A_d^{EE} = 56\,\mu K^2_{\rm CMB}$, $A_d^{BB} = 28\,\mu K^2_{\rm CMB}$, $\alpha_d^{EE} = -0.32$, $\alpha^d_{BB} = -0.16$; $A^s_{EE} = 9\,\mu K^2_{\rm CMB}$, $A^s_{BB} = 1.6\,\mu K^2_{\rm CMB}$, $\alpha^s_{EE} = -0.7$, $\alpha^s_{BB} = -0.93$. We simulated 500 Gaussian foreground Q and U maps in six channels assuming a fixed frequency scaling with $\beta_d = 1.54$, $T_d = 20$ K, and $\beta_s = -3$, consistent with Planck Collaboration IV (2020).

  The non-Gaussian foreground models contain anisotropic dust and synchrotron templates from `PySM3` (Zonca et al., 2021) based on real temperature and polarization data. The most notable differences among them are: a) `d0s0` assumes spatially constant SEDs; b) `d1s1` assumes spatially varying SEDs based on *Planck* `Commander` temperature data; c) `dmsm` has a larger spatial variation of β_s than `d1s1`, additional small-scale Gaussian fluctuations as realizations of a power-law power spectrum with slope $\alpha_s^{BB} = -0.6$ based on S-PASS data (Krachmalnicoff et al., 2018), and uses β_d and T_d templates smoothed to 2 degrees

---

[3] We only used data splits for Pipeline A and A+moments.

[4] `d10s5` is a preliminary template that had been released shortly before the publication of BB2023 and was therefore tested on 100 simulations instead of 500 simulations.

Table 4.1: Instrument and noise specifications used for the simulations in the SO B-mode pipeline comparison project. These levels correspond to homogeneous noise, while our default analysis assumes inhomogeneous noise maps weighted according to the SAT hit counts.

Frequency [GHz]	FWHM [arcmin]	Baseline Noise [μK-arcmin]	Goal Noise [μK-arcmin]	Pessimistic $\ell_{\rm knee}$	Optimistic $\ell_{\rm knee}$	$\alpha_{\rm knee}$
27	91	46	33	30	15	-2.4
39	63	28	22	30	15	-2.4
93	30	3.5	2.5	50	25	-2.5
145	17	4.4	2.8	50	25	-3.0
225	11	8.4	5.5	70	35	-3.0
280	9	21	14	100	40	-3.0

angular resolution; d) the `d10s5` model assumes dust emission and spectral templates based on the *Planck* Generalized Needlet Internal Linear Combination (`GNILC`) data set (Planck Collaboration IV, 2020), which feature larger spatial variation than `d1s1` and `dmsm`. We refer to pysm3.readthedocs.io/en/latest for more details on the non-Gaussian models.

4.3 Results on the tensor-to-scalar ratio

The five component separation pipelines were applied to the set of simulations described in Sect. 4.2 with the goal of inferring r and its statistical uncertainty. In the following, we discuss the results of several test settings, starting with the fiducial cosmology analysis that was conducted on the four least complex foreground scenarios. We then briefly address alternative cosmologies, noise inhomogeneity, and the `d10s5` scenario.

In Fig. 4.3, we show the results on the standard cosmology ($r = 0$, $A_{\rm lens} = 1$) for the two $1/f$ noise cases and in four different foreground scenarios, including Gaussian, `d0s0`, `d1s1`, and `dmsm`. The five pipelines are statistically consistent with $r = 0$ and among each other in the two simplest foreground scenarios, with biases of 1σ (2-3σ) emerging for Pipelines A, B, and C in the `d1s1` (`dmsm`) scenario and unbiased results for the extended pipelines. The statistical uncertainties vary between 2.1×10^{-3} and 3.6×10^{-3} for Pipelines A, B, and C and increase by about 25% (80-100%) for A+moments (C+dust marginalization), depending on the noise and foreground case. The effect of changing between the two white noise scenarios and between the two $1/f$ noise scenarios results in a respective increase of $\sigma(r)$ by about 30%. These results are in reasonable agreement with the forecasts from SO2019.

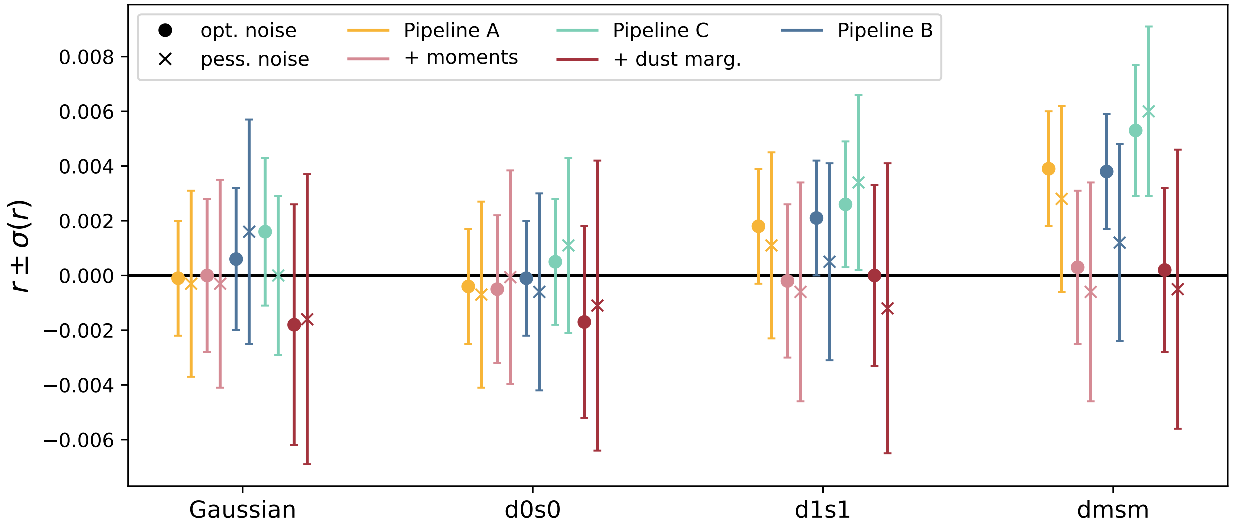

Figure 4.3: Mean posterior values and standard deviations of the tensor-to-scalar ratio r obtained by the SO degree-scale B-mode cleaning pipelines, as a function of increasing complexity in Galactic foregrounds. Results are averaged over 500 simulations, which include CMB with the ($r = 0$, $A_{\text{lens}} = 1$) fiducial cosmology, foreground emission and inhomogeneous noise at the baseline level with optimistic (*dot markers*) or pessimistic $1/f$ component (*cross markers*). This figure was directly adopted from BB2023.

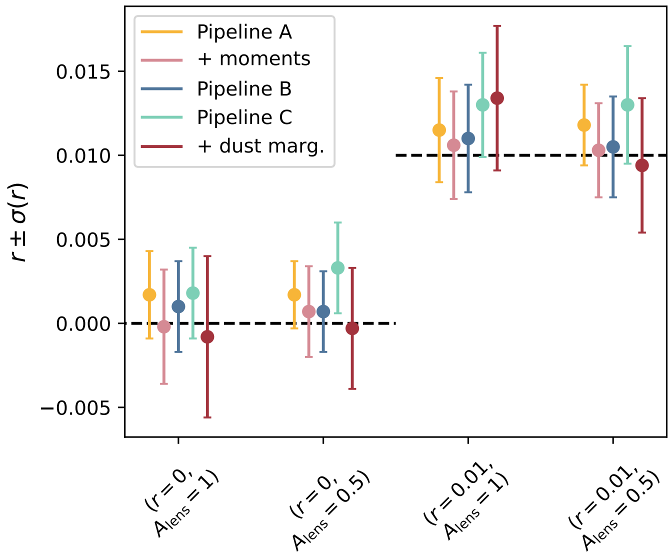

Figure 4.4: Mean posterior values and standard deviations of the tensor-to-scalar ratio r obtained by the SO degree-scale B-mode cleaning pipelines, as a function of input cosmology. Results are averaged over 500 simulations, which include CMB, foreground emission in the `d1s1` scenario and inhomogeneous noise in the baseline-optimistic case. This figure was directly adopted from BB2023.

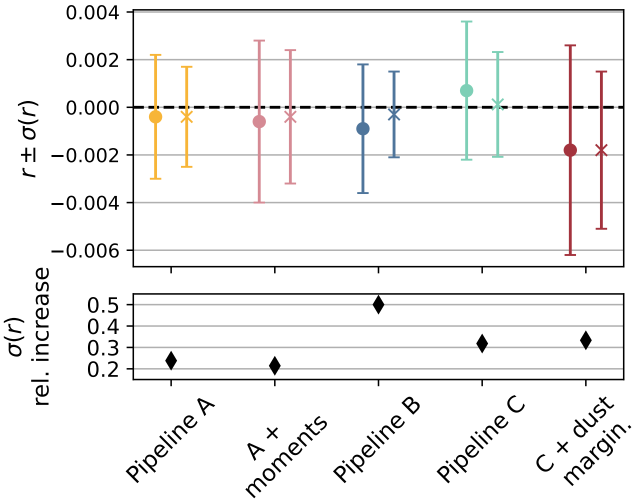

Figure 4.5: Effect of noise inhomogeneity on the inferred tensor-to-scalar ratio r. *Upper panel:* Mean posterior values and standard deviations of r obtained by the SO degree-scale B-mode cleaning pipelines. Results are averaged over 500 simulations, which include CMB with the standard cosmology, foreground emission in the `d0s0` scenario and noise in the baseline-optimistic case. *Lower panel:* Relative increase in $\sigma(r)$ when exchanging homogeneous by inhomogeneous noise. This figure was directly adopted from BB2023.

Aside from tightening constraints in the case of $r = 0$, the SO baseline requirements include the bias-free measurement of a primordial signal with $r = 0.01$ at the 2-3σ level considering a 50% removal of the gravitational lensing signal. To test this, we investigated three further scenarios ($r = 0.01$, $A_{\rm lens} = 1$), ($r = 0$, $A_{\rm lens} = 0.5$), and ($r = 0.01$, $A_{\rm lens} = 0.05$) in the baseline-optimistic noise case, assuming `d1s1` foregrounds. Figure 4.4 shows the results. We find that in all four cosmologies, the pipelines can measure r at or below a 1σ bias, confirming their robustness. We also find that including 50% delensing lowers $\sigma(r)$ by about 20 to 30% and assuming primordial gravitational waves at a level of $r = 0.01$ increases the error by about 40%, consistent with the expected increase in cosmic variance.

Another distinguishing feature between this work and SO2019 is the inhomogeneous noise described in Sect. 4.2. To check the impact of inhomogeneous noise on the r inference, we ran all pipelines on simulations that contained homogeneous noise in the baseline-optimistic case, CMB with the standard cosmology, and `d1s1` foregrounds and compared those with the corresponding inhomogeneous noise simulations. The results, shown in Fig. 4.5, reveal that noise inhomogeneity has different impacts on different pipelines, from a moderate rise of $\sigma(r)$ by about 20% for Pipeline A to a more pronounced 50% increase for Pipeline B.

We also investigated the `d10s5` foreground template, which, among other changes with respect to `dmsm`, has a more pronounced spatial variation in the dust amplitude and spectral index. We ran all pipelines on 100 simulations with the standard cosmology and inhomogeneous goal-optimistic noise, and obtain

$$\begin{aligned}
r &= (19.4 \pm 2.1) \times 10^{-3} && \text{(Pipeline A)} \\
r &= (2.5 \pm 3.1) \times 10^{-3} && \text{(A + moments)} \\
r &= (14.4 \pm 2.3) \times 10^{-3} && \text{(Pipeline B)} \\
r &= (22.0 \pm 2.6) \times 10^{-3} && \text{(Pipeline C)} \\
r &= (-1.5 \pm 5.1) \times 10^{-3} && \text{(C + dust marg.)}
\end{aligned} \tag{4.3}$$

We obtain a substantial, up to 9σ, bias on r for Pipelines A, B, and C and no significant bias for the two extended pipelines (below 1σ). For A+moments and C+dust marginalization, an unbiased measurement comes at the cost of 40% and 95% increase in the statistical uncertainty compared to the nominal pipelines A and C, respectively. The moderate increase in $\sigma(r)$ makes the moments pipeline the unbiased method with the lowest statistical error, which is at the level of the SO2019 baseline requirement. These results substantiate that measuring unbiased r in the presence of complex Galactic foregrounds, in particular with spatially varying spectral indices, relies on algorithms that specifically target foreground residuals.

To validate these results, we applied a set of pipeline comparison tests, which we discuss in the following section.

4.4 Pipeline comparison

Robust r constraints require selecting one or several foreground cleaning pipelines that are unbiased while achieving the smallest possible statistical error. To exclude any unexpected algorithm-specific bias, it is useful to compare the pipelines not just with regard to the inferred r value, but also to intermediate data products, such as angular power spectra.

In the following, we summarize three sets of pipeline consistency checks targeting these secondary data products. This includes comparisons of CMB power spectra, foreground models used in Pipelines A and A+moments, and channel weights derived from the three cleaning algorithms.

4.4.1 C_ℓ-based comparison

The first consistency test compares cleaned angular power spectra of the CMB as inferred by Pipelines A, B, and C. While clean CMB power spectra can be readily obtained from Pipelines B and C, they cannot in the case of Pipeline A, which directly infers r from multifrequency power spectra. We therefore replace the C_ℓ-fiducial model by the ℓ-wise CMB model (see Sect. 3.4.3),

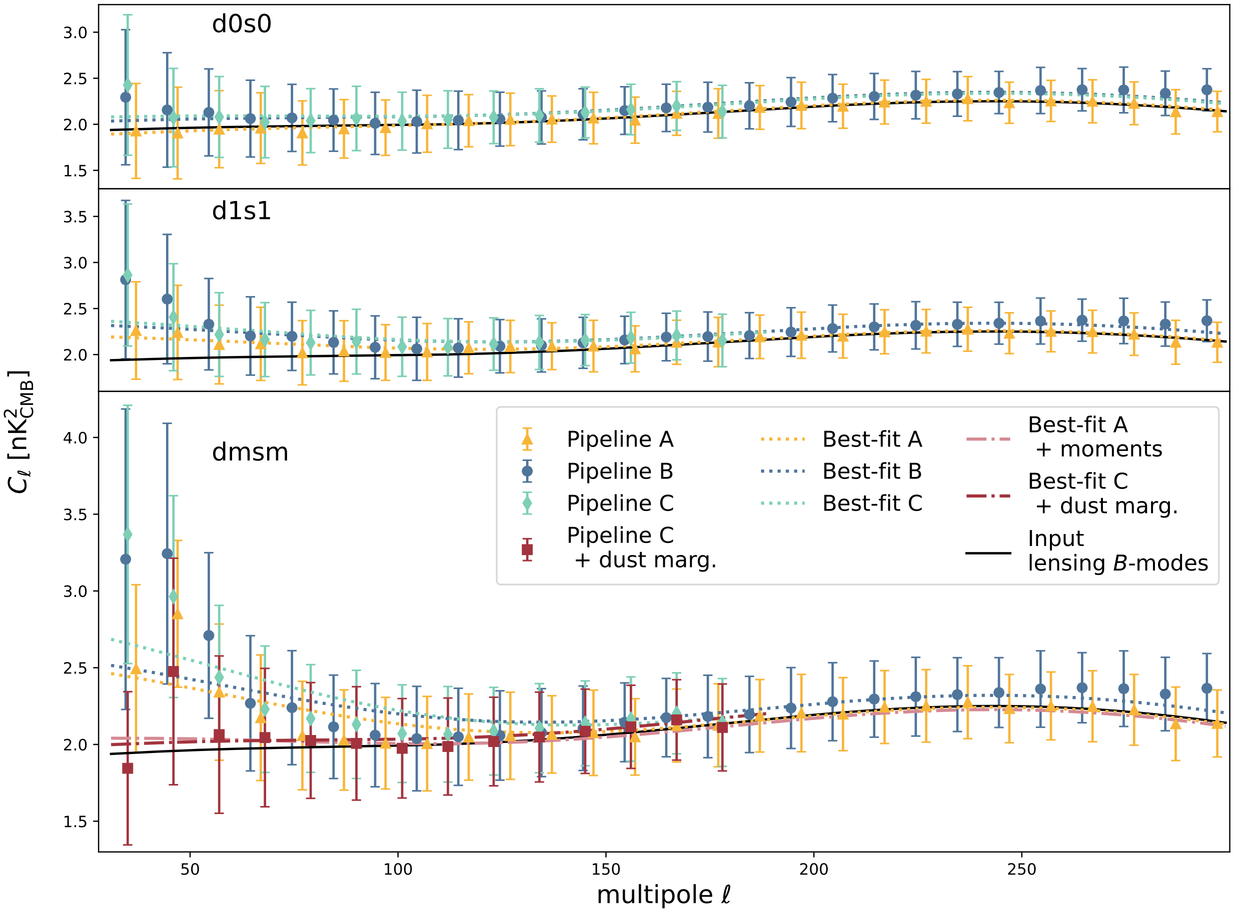

Figure 4.6: Clean CMB power spectra as retrieved from the SO degree-scale B-mode cleaning pipelines. Results are averaged over 500 simulations of CMB with ($r = 0$, $A_{\rm lens} = 1$), inhomogeneous goal-optimistic noise, and Galactic foregrounds of different levels of complexity. The error bars quantify the empirical standard deviation calculated from simulations. This figure was directly adopted from BB2023.

which instead infers the binned power spectrum amplitudes of the cleaned CMB while marginalizing over foreground parameters.

Figure 4.6 shows the binned CMB BB power spectra and their statistical uncertainties as retrieved by Pipelines A, B, and C from 500 simulations containing CMB with ($r = 0$, $A_{\rm lens} = 1$) and inhomogeneous noise in the goal-optimistic case. The three panels show the `d0s0`, `d1s1`, and `dmsm` foreground scenarios, respectively. We also compare the fiducal CMB power spectrum with the best-fit CMB models as found by Pipelines A, B, and C and, in the most complex `dmsm` scenario, show the best-fit models for both pipeline extensions. We also show the data obtained by the C+dust marginalization pipeline in the `dmsm` scenario. We find that in the simpler `d0s0` and `d1s1` scenarios, the data from the three nominal pipelines agree among each other and with the input model, with the largest deviations of up to 1σ found at $\ell \lesssim 100$. For `dmsm` at multipoles below 50, the bias grows to about 1.5σ. At $\ell \gtrsim 150$, Pipelines B and C mildly deviate from the fiducial model, which could originate from the fact that maps at different channels have different beam resolutions, and therefore require a convolution to a common resolution prior to component separation. The bias with pipelines B and C seen at large ℓ could indicate power leakage from performing this transformation. The extended pipelines A+moments and C+dust marginalization achieve an unbiased CMB reconstruction at the power spectrum level. In summary, we find that, as expected, degree-scale foreground residuals bias the r inference in complex foreground scenarios, and can be mitigated by using the C_ℓ-moments model or by marginalizing over the residual dust amplitude after map-based component separation. We discuss future perspectives for the B-mode cleaning pipelines in the Conclusions.

4.4.2 Model comparison

Besides looking at the marginalized r results, we can compare different models in terms of their overall fit to all model parameters, taking into account "Occam's razor" principle that simple models should be preferred. To do so, we apply the Akaike Information Criterion (AIC, Akaike, 1974) to the two models of Pipeline A. The C_ℓ-fiducial and C_ℓ-moments model are well suited for this comparison since they are nested and include foregrounds. The relative AIC can be interpreted as the logarithm of the relative model odds (Wagenmakers & Farrell, 2004), and is defined as

$$\Delta\mathrm{AIC} = 2\Delta\mathrm{k} + \chi^2_{\rm min}(\mathrm{moments}) - \chi^2_{\rm min}(\mathrm{fiducial})\,, \tag{4.4}$$

where $\Delta k = 4$ is the number of excess parameters of C_ℓ-moments compared to C_ℓ-fiducial. We calculate ΔAIC for a set of 100 simulated maps of coadded Galactic foregrounds in all five scenarios, noise in the goal-optimistic case, and CMB with ($r = 0$, $A_{\rm lens} = 1$).

Figure 4.7 shows the results. We find that the distribution of ΔAIC concentrates at values $\gtrsim 1$

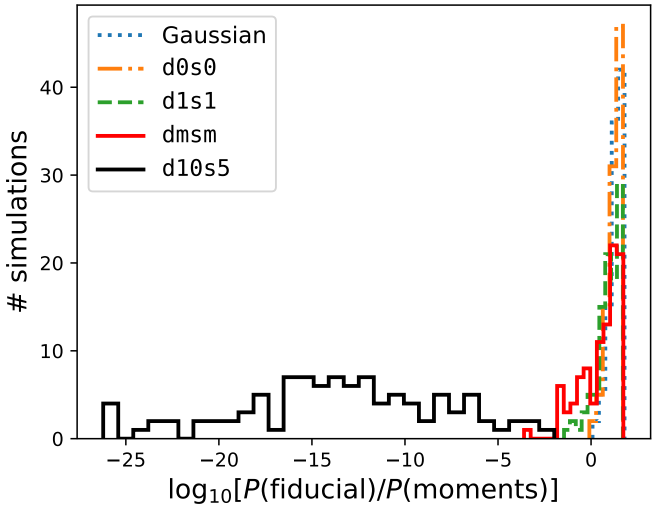

Figure 4.7: Empirical distribution of ΔAIC values obtained from comparing the C_ℓ-fiducial with the C_ℓ-moments model on 100 sky simulations of coadded CMB with ($r = 0$, $A_{\text{lens}} = 1$), inhomogeneous goal-optimistic noise, and Galactic foregrounds with different levels of complexity. Positive values on the abscissa denote preference for the fiducial model, negative values indicate preference for the moments expansion. This figure was directly adopted from BB2023.

for the four simple scenarios (Gaussian, `d0s0`, `d1s1`, `dmsm`), thus slightly preferring the C_ℓ-fiducial model, while `d10s5` foregrounds lead to model odds of $\Delta\text{AIC} < 10^{-5}$ for 78 out of 100 simulations, clearly preferring the C_ℓ-moments model. This result contradicts the intuitive expectation that the C_ℓ-moments model should be preferred already in the `dmsm` scenario, given the 2-3σ bias on r reported in Sect. 4.3. We conclude that a 2-3σ detection by the C_ℓ-fiducial model cannot be validated by model comparison metrics such as the AIC, at least for models with nine parameters or more. For `d10s5` instead, the AIC clearly identifies the 9σ result as a bias. We shall comment further on this result in Sect. 4.5.

4.4.3 Channel weights comparison

Although Pipelines A, B, and C use completely different component separation algorithms, we can define for each of them a set of channel weights that quantify how much each frequency channel contributes to the cleaned CMB signal. We first note that for Pipelines B and C, such weights are already part of the algorithms, in form of the `NILC` weights (see Eq. (10) in Delabrouille et al., 2009) calculated in every needlet window and the least-squares weights (see Eq. (8) in Stompor et al., 2008) computed at every map pixel, respectively. In case of Pipeline A, we do not directly retrieve maps, therefore no CMB map weights emerge from the algorithm. We can, however, take inspiration from *Planck*'s power-spectrum-based `SMICA` algorithm (Cardoso et al., 2008, see also

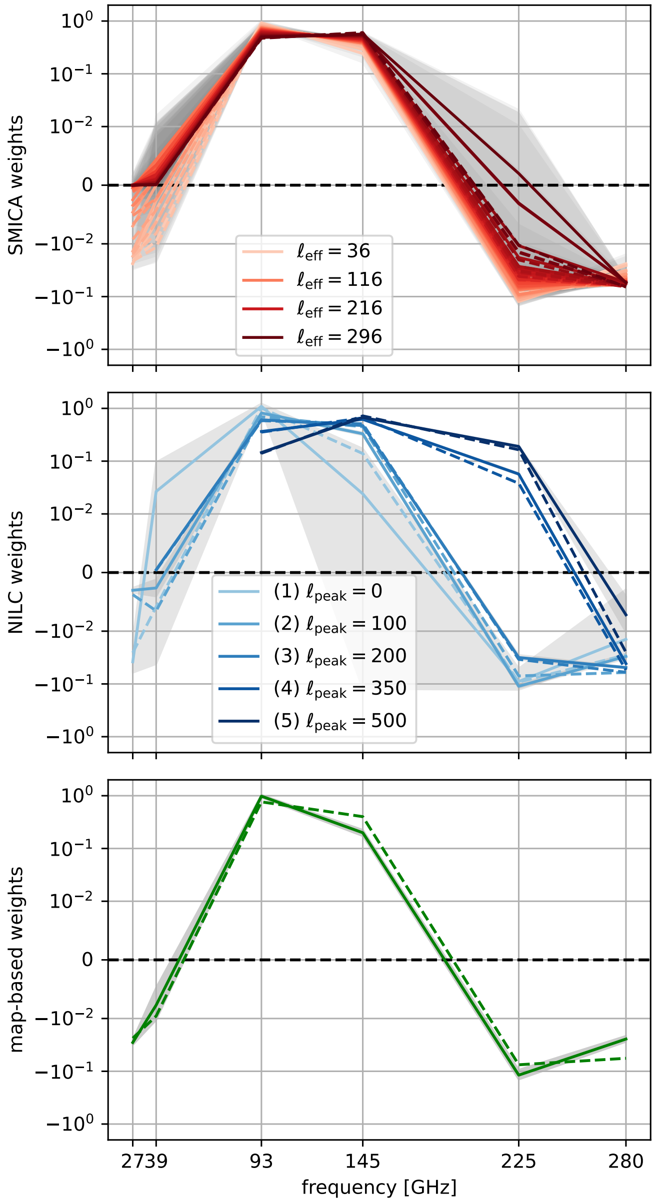

Figure 4.8: Channel weights for Pipelines A, B, and C. We show the `SMICA` weights for 27 different ℓ-bins calculated from noisy multifrequency C_ℓs (Pipeline A, *upper panel*), pixel-averaged `NILC` weights for five needlet windows (Pipeline B, *middle panel*) and pixel-averaged weights from parametric map-based component separation (Pipeline C, *lower panel*). Results are averaged over 100 simulations of coadded CMB with ($r = 0$, $A_{\mathrm{lens}} = 1$), and inhomogeneous noise in the goal-optimistic (*dashed lines*) and baseline-pessimistic case (*solid lines*). The semitransparent gray areas represent the channel weights' 1σ empirical standard deviation over simulations for baseline-pessimistic noise. This figure was directly adopted from BB2023.

Sect. 2.4.1), which synthesizes clean CMB maps from binned, noisy multifrequency power spectra, assuming that the signal and noise maps are Gaussian and uncorrelated between pixels. Those maps can be constructed for each ℓ-bin individually by multiplying the original maps with the `SMICA` weights

$$w_l = \frac{\boldsymbol{a}^T \cdot \hat{\boldsymbol{C}}_l^{-1}}{\boldsymbol{a}^T \cdot \hat{\boldsymbol{C}}_l^{-1} \cdot \boldsymbol{a}} \,, \tag{4.5}$$

where $\boldsymbol{a}$ is vector of length six filled with ones, corresponding to the CMB frequency scaling in thermodynamic temperature units, and $\hat{\boldsymbol{C}}_l^{-1}$ is the 6×6-dimensional inverse cross-frequency power spectrum matrix, which we calculate from noisy multifrequency power spectra.[5]

Figure 4.8 shows the channel weights for the three pipelines. We display Pipeline A's `SMICA` weights for all 27 ℓ-bins in the upper panel, Pipeline B's pixel-averaged `NILC` weights for all five needlet windows in the middle panel, and Pipeline C's pixel-averaged least-squares weights in the lower panel. We find that low and high frequencies are generally subtracted due to the large foreground contribution, while the middle frequencies carry a larger CMB contribution and are therefore added to the CMB reconstruction. Smaller scales ($\ell \gtrsim 300$) considered by the `NILC` weights tend to be larger at higher frequencies, possibly because small-scale lensing B-modes become more important than thermal dust emission. In general we find that all pipelines are in agreement, confirming the robustness of the three algorithms.

This series of consistency tests confirms that the bias seen in the r results (Sect. 4.3) is not related to algorithmic errors but the treatment of foreground residuals. We provide an extended interpretation of this in the following section.

4.5 Summary

One of today's frontiers in modern cosmology, the origin of cosmological fluctuations in the very early Universe, may be in reach thanks to near-future B-mode experiments such as SO. Given the challenge of detecting this tiny signal of unknown amplitude in the presence of much larger Galactic and instrumental contaminants, obtaining consistent results from different cleaning algorithms is indispensable for a robust analysis.

This chapter describes the SO component separation pipeline comparison project, which embeds the C_ℓ pipeline described in Chpt. 3 as one of the algorithms to constrain the amplitude of primordial gravitational waves from degree-scale B-modes. We consider the C_ℓ pipeline in the C_ℓ-fiducial model, assuming no spatial variation in the foreground spectral indices, and the C_ℓ-moments model accounting for small variations. The second pipeline uses the `NILC` algorithm, which is "blind",

---

[5] Noisy multifrequency power spectra are obtained by Pipeline A by computing the weighted mean of all cross- and auto-split power spectra (and not just the cross-split spectra, as described in Sect. 3.4).

that is, agnostic with respect to the spectral properties of possible contaminants of the CMB, and applies to needlets that are localized both in harmonic and pixel space (Basak & Delabrouille, 2013). The third pipeline is a parametric pixel-based cleaning pipeline based on the `FGBuster` code (Poletti & Errard, 2023; Stompor et al., 2008), which assumes spatially constant SEDs, and allows for a-posteriori marginalization over dust residuals in the cleaned CMB power spectrum.

We find that the three algorithms and two extensions agree in the simplest tested foreground scenarios and confirm the statistical sensitivity on r, $\sigma(r) = 0.003$ at 68% central CL, anticipated by the SO Goals and Forecasts paper SO Collaboration (2019). In the presence of complex foregrounds, a full reduction of the bias requires adequate extended model designs, such as the moment expansion to the C_ℓ pipeline and the option to marginalize over residual dust in the case of the map-based parametric pipeline. Thanks to these pipeline extensions, bias on r found with the simpler pipeline variants could be successfully identified as originating from simulated foregrounds, and discerned from a simulated primordial and lensing B-mode signal in all cases considered. We stress that choosing the more complex pipeline variants comes at the cost of larger statistical uncertainty and possible bias associated with volume effects, which are spurious shift in the parameter posterior distribution caused by a suboptimal choice of parameters (see Chpt. 6 for more details). Whether the more complex, extended models are to be chosen for the data analysis must therefore be judged on the basis of the performance of simpler models. Model comparison metrics are not always able to validate 2-3σ detections in r, at least when considering models with nine or more parameters, such as the ones used in the C_ℓ pipeline.

Although in these simulated Galactic foreground scenarios, the extensions to Pipelines A and C retrieve unbiased r values, this represents a significant challenge for real data analysis: it is not guaranteed that we can robustly distinguish between a primordial signal and residuals from Galactic foregrounds, or residual instrumental systematic effects, which are beyond the scope of the work presented in this chapter. Therefore, with the imminent arrival of the first data from SO, an extended set of robustness tests will be needed to exclude any potential source of systematic bias in r. We shall discuss future robustness tests in the Conclusions. The five pipeline designs are not definitive and efforts are ongoing to make the pipelines more optimal and robust. Examples are a hybrid approach between map-based and C_ℓ cleaning (Azzoni et al., 2023), `NILC`-based algorithms introducing more advanced treatment of spatially varying foreground emission, such as Multi-Clustering `NILC` (`MCNILC`, Carones et al., 2023).

CHAPTER 5

NN-based inference of the reionization optical depth from Planck maps

In the previous chapters, we focused on the estimation of the tensor-to-scalar ratio r for SO. As we pointed out there, inferring r is challenging due to the presence of non-Gaussian foreground residuals, and possibly also instrumental systematics, although the latter were not taken into account in said works. These residuals can bias the estimation of r, as demonstrated in the case of the `d10s5` model. We showed that marginalizing over residuals at the likelihood level can mitigate this bias, as we saw with Pipeline C in the extended "dust marginalization" option. This requires one to find a parametric model of the residuals, which generally may not be a trivial task. In the case of SO, we successfully used the power spectrum of the component-separated dust map, but this procedure is not guaranteed to correct for residuals in general. For future experiments such as CMB-S4 and LiteBIRD that aim at constraints of $\sigma(r) \sim 10^{-3}$, this becomes a critical issue: while a correct analytic model of foregrounds is mandatory, it is difficult to obtain. One possible way to overcome this issue is to use likelihood-free inference, which relies only on simulations and not on the existence of an analytical likelihood. This is a much simpler task, since designing and running simulations is usually less challenging than writing down the corresponding analytical model. A potential solution is employing neural networks (NNs), trained on simulations, to perform the inference task.

In this chapter, based on Wolz et al. (2023b), we investigate this approach by estimating the optical depth τ from *Planck* maps. The parameter τ measures the number density of free electrons that a CMB photon encounters in the reionized Universe after decoupling. It is known from high-redshift quasar data that the reionization of the Universe, following the birth of the first stars and galaxies, was completed by a redshift of $z \sim 5.3$ (Qin et al., 2021). CMB photons and free electrons undergo elastic Compton scattering, leaving the photon energy constant but changing the polarization fraction as the photons free-stream toward our instruments. As described in Sect. 2.3, a nonzero optical depth τ causes a bump in the CMB polarization power spectra at large scales

$\ell \gtrsim 20$. Thus, analogous to r, the τ parameter is constrained by the large-scale CMB polarization signal. A crucial difference between those two parameters, though, is their typical amplitude in the CMB power spectra: the peak variance of large-scale E-modes at $\ell \sim 4$ that is sensitive to the reionization optical depth is at least two orders of magnitude higher than that of primordial B-modes. This makes estimating τ a similar but easier inference problem than estimating r. A critical issue in the estimation of τ from *Planck* and WMAP data are large systematic effects that contaminate the polarization signal at low multipoles. For this reason, also *Planck* analyses (Pagano et al., 2020) use a simulation-based likelihood, given the difficulty to model residual systematic effects analytically.

In recent years, NN-based approaches to likelihood-free inference underwent a rapid development in cosmology, showing potential as an alternative tool for parameter estimation. Promising tools are being developed for many applications: from LSS simulations (Villaescusa-Navarro et al., 2022), to CMB lensing reconstruction (Caldeira et al., 2019), kSZ detection (Tanimura et al., 2022), or modeling and cleaning of Galactic foregrounds (Jeffrey et al., 2022; Wang et al., 2022; Casas et al., 2022; Krachmalnicoff & Puglisi, 2021). NN-based inference of cosmological parameters has seen significant progress in the context of observations of the LSS, where the complexity of the cosmological and astrophysical signals, together with the difficulty in the definition of optimal summary statistics, challenge analytical methods. Up to now, this approach has been tested on simulations (see e.g., Villaescusa-Navarro et al., 2022), with applications on real data still limited in number, although leading to promising results (e.g., Fluri et al., 2019). In this context, CMB data analysis could also benefit from the application of NN-based inference, helping overcome the limitations of traditional methods.

Full-sky space missions such as WMAP and *Planck* have been able to measure the reionization bump (see Sect. 2.3.2) through pixel-based and power-spectrum-based analysis methods. The WMAP nine-year data release cites $\tau = 0.089 \pm 0.014$ (Hinshaw et al., 2013), a value that later turned out to be biased high due to Galactic dust emission (Planck Collaboration XI, 2016; Natale et al., 2020). The *Planck* 2018 legacy polarization data products at large scales are known to be affected by residual contamination from instrumental systematic effects. *Planck*'s Low Frequency Instrument (LFI) polarization data at 70 GHz contain less large-scale systematics than the High Frequency Instrument (HFI) data at 100 GHz and 143 GHz, motivating the *Planck* Collaboration to perform map-based analysis on LFI data and cross-spectrum analysis on HFI data. The *Planck* 2018 legacy release cites $\tau = 0.063 \pm 0.020$ as inferred from LFI data and 0.051 ± 0.009 from HFI data (Planck Collaboration V, 2020). The cross-spectrum analysis method of *Planck* HFI data at 143 GHz and 100 GHz yields the tightest constraint to date (Pagano et al., 2020; de Belsunce et al., 2021), while avoiding bias from uncorrelated noise in the individual frequency channels.

As investigated in Planck Collaboration VI (2014) and Delouis et al. (2019), the most important systematic effects at *Planck* 143 GHz and 100 GHz are temperature-to-polarization (T-to-P) leakage due the analog-to-digital converter nonlinearity (ADCNL), uncertainties on the detector orientation and polarization efficiencies, T-to-P leakage due to bandpass mismatch and inaccurate Galactic foreground modeling, and a varying time constant associated with the heat transfer to the bolometers. In general, these systematic effects follow non-Gaussian statistical distributions and are expected to correlate among different channels, mainly because they are partially sourced by the temperature signal. Improved mapmaking algorithms, such as `SRoll2` (Delouis et al., 2019), and `NPIPE` (Planck Collaboration Int. LVII, 2020), have been designed to mitigate those systematics. However, while the `NPIPE` maps are largely filtered at low ℓ, the `SRoll2` map still have residuals that need to be taken into account in the estimation of τ. For this reason, Pagano et al. (2020) use an empirical likelihood built from realistic simulations (Planck Collaboration V, 2020; Gerbino et al., 2020). The difficulty of accounting for or modeling the systematic residuals in the `SRoll2` maps therefore calls for an alternative, likelihood-free methodology, for example NN-based inference.

This work represents the first map-level cosmological inference on CMB data that is entirely based on neural networks. We use NNs to infer the optical depth to reionization τ and its statistical uncertainty from polarized *Planck* multifrequency maps at 100 and 143 GHz at scales $\gtrsim$ 4 degrees, employing realistic simulations to build our model (training) and an independent subset to validate our findings. By adopting a suitable strategy for training we were able to constrain τ to a value in agreement with previous work, although with larger error bars. The remainder of this chapter is structured as follows. We present the simulations and data used in this work in Sect. 5.1, followed by the neural network inference method in Sect. 5.2. In order to validate this method, we apply it to a series of simulations and present the results in Sect. 5.3. We discuss the final results on the *Planck* `SRoll2` maps in Sect. 5.4.

5.1 Simulations and data

A neural network is a high-dimensional parametric function that maps a set of numbers called "input features" (such as the pixels of one or several CMB maps) to a set of output features (such as a τ value and its statistical uncertainty). In order to achieve the necessary arithmetic sophistication to perform complex tasks, NNs consist of a large number of independent elementary building blocks, called neurons, each of which performs a simple nonlinear operation that can be tuned by a set of free parameters called the NN weights. Neural networks can easily contain tens of thousands of independent weights that are optimized ("learned") during a training process as described below.

Several neurons in parallel can act as a set of independent functions called a layer. A NN can have many serially connected layers, with each layers' output acting as an input to the subsequent one. The first NN layer is connected to the input features, while the last layer gives us the desired output features.

Training a NN involves repeatedly exposing it to a set of training data, of which we know the truth value of interest (in our case, the input τ value). The degree of closeness between the NN output and the truth value is assessed by a "loss function", which we can choose according to the task at hand. The NN training consists of a series of "epochs", during which we compare the NN output to the truth and adapt the NN weights, iteratively walking down the gradient of the loss function until, in the ideal case, we reach a global minimum. During each epoch we use validation data that the NN has not been exposed to yet, and evaluate the loss function without changing the NN weights as an independent cross-check. The training process is complete once the loss function, evaluated at the validation data, consistently falls below a certain threshold.

After training, we may still wish to assess the performance of the finished NN, for which we then use yet another set, called test data. This is important because of a known risk related to NNs, known as "overfitting". This means that the network weights adapt excellently to the input features of the training data, while the NN still fails to make correct predictions on similar but independent test data sets.

The goal of our analysis is the estimation of τ from real `SRoll2` *Planck* maps. Therefore, we need a training set of simulated maps that are representative of the `SRoll2` data. As we discuss in more detail in Sect. 5.3, we explored several different strategies for training, starting with a single frequency channel, and then going to two channels. Also, as we shall see later, we use different types of simulations, initially training only on Gaussian simulations with *Planck* noise levels, then moving to simulated `SRoll2` maps with non-Gaussian systematics, and finally combining both in a hybrid "retraining" approach. To achieve all this, we needed a large number of simulations to perform NN training, validation, and testing. We generated simulated maps that include CMB emission, noise, and instrumental systematic effects, as well as possible spurious signals coming from our Galaxy. In this section, we describe the simulations, the data, and the sky masks needed to avoid the highly contaminated Galactic plane region.

5.1.1 Simulated CMB maps

As we saw in Sect. 2.3, CMB temperature and polarization anisotropies can be described by the six parameters of the ΛCDM model. Analyses of small-scale temperature data from the *Planck* 2018 legacy release place a 0.5% constraint on the parameter combination $10^9\, A_s\, e^{-2\tau} = 1.88 \pm 0.01$

(Planck Collaboration VI, 2020). Varying the two parameters $(A_s,\ \tau)$ simultaneously conditioned on $10^9\,A_s\,e^{-2\tau} = 1.884$, coherent with previous studies (Planck Collaboration Int. XLVI, 2016; Planck Collaboration V, 2020; Pagano et al., 2020; Planck Collaboration Int. LVII, 2020), we used the Boltzmann solver `CAMB` (Lewis et al., 2000) to generate a lookup table of EE power spectra computed with the ΛCDM model. In order to robustly train and validate a NN, we need a set of simulations that span the entire distribution of plausible “real-world” data, which should at least include all cosmologies that we deem realistic and wish our model to be sensitive to. Therefore, to build the simulated CMB maps, we discretized $\tau \in [0.01,\ 0.13]$ with step size $\Delta\tau = 5\times 10^{-4}$. Since the other ΛCDM parameters have no substantial impact on polarized CMB spectra at low multipoles, we fixed them to the *Planck* 2018 legacy best-fit values $H_0 = 67.32$ km/s/Mpc, $\Omega_b h^2 = 0.02237$, $\Omega_c h^2 = 0.1201$, $n_s = 0.9651$, $m_\nu = 0.06$. From the tabulated power spectra, we uniformly drew 200,000 samples based on which we generated 200,000 pairs of full-sky Stokes Q and U maps using the `HEALPix` package (Górski et al., 2005). We fixed the Q and U maps’ angular pixel resolution by choosing $N_{\rm side} = 16$ (or a pixel size of about 4 degrees) and smooth each map with a cosine beam window function (Benabed et al., 2009), in analogy with the procedure used to generate the *Planck* `SRoll2` maps (see Sect. 5.1.4). These large scales retained in our maps correspond to multipoles $\ell \lesssim 50$, where the reionization bump leaves an observable imprint in the CMB EE spectrum.

5.1.2 Simulated Gaussian noise

Planck maps contain Gaussian instrumental noise which, in pixel space, is well described by the covariance matrix published alongside *Planck*’s 8th Full Focal Plane simulations (`FFP8`, Planck Collaboration XII, 2016). We drew samples from them for the *Planck* 100 and 143 GHz polarization channels (Planck Collaboration VI, 2014; Planck Collaboration XIII, 2016), obtaining 200,000 Gaussian noise maps at pixel resolution $N_{\rm side} = 16$ for both channels, respectively. We coadded the training maps of CMB and noise to obtain 200,000 *Planck*-like simulations on the full sky, out of which we selected 190,000 for training and 10,000 for validation. For the testing phase, we drew new noise samples in the same fashion as before, but coadded CMB simulations with fixed input values $\tau = 0.05$, 0.06, and 0.07 and different seeds than the ones used for training and validation. In this way, we obtained three sets of 10,000 independent Gaussian test simulations with the fixed input cosmologies.

5.1.3 `SRoll2` simulations

Several updated mapmaking codes exist that include improved systematics cleaning algorithms of the low-multipole *Planck* maps, such as `SRoll2` (Delouis et al., 2019), and `NPIPE` (Planck Collaboration Int. LVII, 2020). The `SRoll2` algorithm, an upgraded version of the *Planck* Collaboration's `SRoll` algorithm (Planck Collaboration Int. XLVI, 2016), iteratively cleans systematics from *Planck*'s time-ordered data products. Major improvements in `SRoll2` encompass a new gain calibration model that accounts for second-order ADCNL, updated foreground templates, and an internal marginalization over the polarization angles and efficiencies for each bolometer. The `SRoll2` data products contain a significantly lower level of spurious systematic effects and a dipole residual power reduced by 50% with respect to the *Planck* 2018 legacy data, falling below the noise level. The `SRoll2` EE cross-frequency power spectrum between 100 GHz and 143 GHz is dominated by the cosmological signal at all scales that were considered in the analysis ($2 < \ell < 30$).

The `SRoll2` simulations are the result of applying the `SRoll2` cleaning algorithm to a set of 500 *Planck*-like realistic sky simulations containing modeled noise, foregrounds, and instrument systematics. We chose them as our reference for systematic effects present in the `SRoll2` *Planck* data. All simulated maps are cleaned from Galactic foregrounds through a template fitting procedure, as described in Pagano et al. (2020). In order to produce our training set, we started with 400 of the 500 original `SRoll2` simulations containing pairs of Q and U full-sky maps at pixel resolution $N_{\mathrm{side}} = 16$ and two channels at 100 and 143 GHz. To augment our original `SRoll2` simulation set, we randomly drew `SRoll2` maps from the original 400 maps (with repetition), keeping corresponding Q and U maps together. This allowed us to assemble a total of 200,000 `SRoll2` simulations. After coadding them with CMB simulations, we obtained a set of 200,000 polarized full-sky simulations, used for training and validation. For the testing phase, we made 3×100 copies of the 100 remaining, previously unused `SRoll2` maps and coadded them with $10,000$ CMB maps with fixed input $\tau = 0.05$, 0.06, and 0.07, respectively. In this way we obtained a set of $3 \times 10,000$ full-sky `SRoll2` test simulations.

5.1.4 *Planck* maps

The goal of this work is the analysis of the `SRoll2` *Planck* polarization data products (Delouis et al., 2019). They consist of Stokes Q and U maps at the 100 and 143 GHz HFI frequency channels, stored at pixel resolution $N_{\mathrm{side}} = 16$. The *Planck* maps are first smoothed by a pair of cosine beam window functions, and then cleaned from foreground contamination through a template fitting procedure (Pagano et al., 2020). Figure 5.1 shows the map products in Galactic coordinates. We note that close to the Galactic plane, Q and U on both channels are visibly contaminated by residual systematic effects, which we masked prior to the analysis in order to avoid bias. The arc-shaped features in the northern and southern Galactic hemisphere likely indicate residual gain variations caused by the ADCNL systematic effect. As shown by Delouis et al. (2019), these features show lower residual power than the CMB in the 100×143 GHz EE cross-spectrum but may still amount to a nonnegligible bias in cosmological analyses.

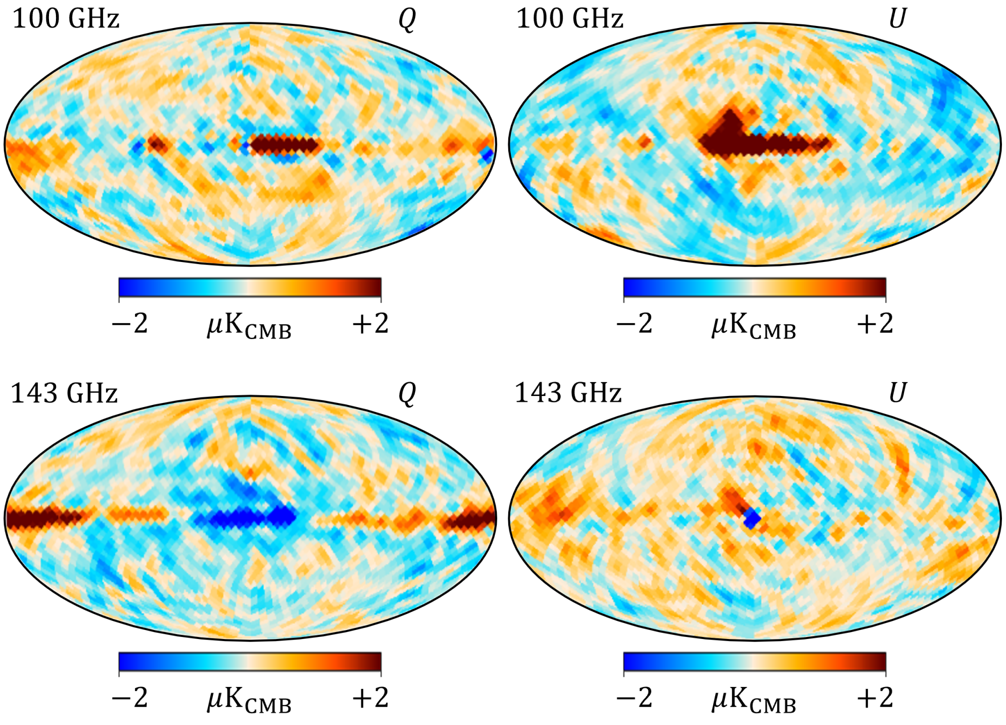

Figure 5.1: `SRoll2` data products of the *Planck* Q and U maps at frequencies 100 and 143 GHz, after component separation, used in this chapter, displayed in Galactic coordinates.

5.1.5 Masks

At low Galactic latitudes, the Milky Way emits polarized foreground radiation which dominates the CMB signal in intensity and polarization. Even after component separation, residuals of this emission need to be excluded from the analysis to avoid biasing cosmological analyses. We therefore applied masks to all maps described in the previous sections. We considered two of the binary polarization masks published in Pagano et al. (2020), retaining sky fractions of $f_{\text{sky}} = \{0.5, 0.6\}$. We smoothed them with Gaussian beams of corresponding full width at half maximum (FWHM) of 15 and 16 degrees, respectively, and apply a binary threshold, setting all pixels with a value larger than 0.5 to one and all others to zero. This procedure allows us to avoid fuzzy borders and mitigate groups of isolated masked pixels. The smoothed masks are shown in Fig. 5.2. Our baseline mask in this work is the $f_{\text{sky}} = 0.5$ smoothed mask, as it retains enough large-scale information to constrain τ but avoids excessive levels of foregrounds in the Galactic plane.

5.2 NN inference

NNs can have different architectures that are commonly categorized by how different layers of neurons are connected and arranged in relation to each other. Choosing a certain architecture

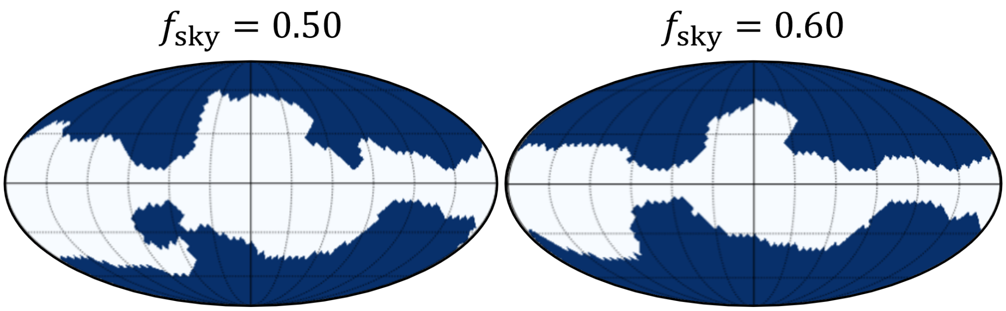

Figure 5.2: Smoothed version of the `SRoll2` sky masks at sky fractions 0.5 and 0.6 used in this chapter, displayed in Galactic coordinates.

can crucially affect the ability of a NN to perform certain types of tasks. For example, in a fully connected architecture, every neuron in a given layer feeds its output value to every neuron in the subsequent layer, thus maximizing the number of connections. Depending on the number of layers, such an architecture may excel at emulating complex arithmetic operations, but less so at recognizing sparse patterns in a large, noisy data set, such as a melody in a digital piece of music, or an object in a digital image. Another architecture is the convolutional neural network (CNN), which uses a much lower number of connections between subsequent layers, in a way so that local, sparse patterns can be much more efficiently identified. If we think of the input data as a digital image, then every layer is a convolutional filter that singles out specific, recurring visual patterns. For this reason, CNNs have become exceedingly popular in the context of digital image recognition.

5.2.1 CNN architecture for τ estimation

CNNs are the industry standard of pattern recognition in two-dimensional images, performing both classification (e.g., identifying families of objects) and regression tasks (e.g., estimating continuous parameters on maps). The success of CNNs in extracting low-dimensional information from visual input is due to a multilayer image filtering algorithm. This typically involves searching for distinct sets of local features in the original image (through convolution) and compressing the data (via a type of operation called "pooling"), going to lower and lower resolution, before inferring the desired summary statistic. In this work, we use CNNs to perform cosmological inference using simulation-based regression models that estimate τ from spherical images of the polarized microwave sky in the form of `HEALPix` maps. The NN architecture must therefore be able to filter images, but also massively compress the data in order to converge to a final two-number output. As we shall see in this section, this is achieved by combining several convolutional layers with a set of fully connected layers.

We want to retrieve information from data projected on the sphere, requiring convolutions on

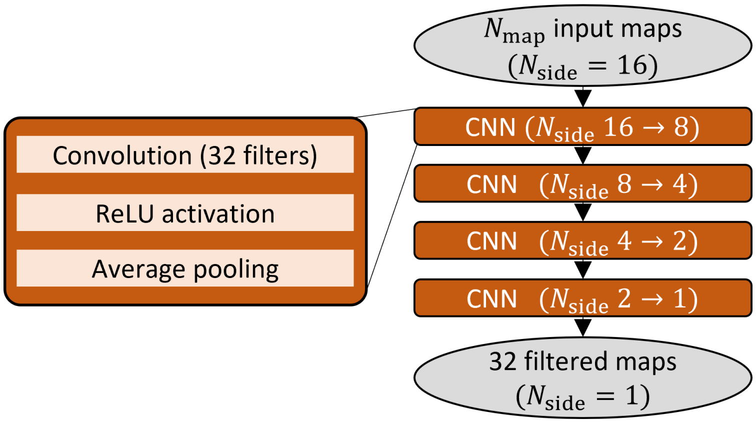

Figure 5.3: Schematic of the convolutional layers of the neural network used in this chapter. This represents the first part of the architecture, performing image filtering.

the spherical domain. To this end, we made use of the `NNhealpix`[1] algorithm which allows to build spherical CNNs taking advantage of the `HEALPix` tessellation. In particular, `NNhealpix` convolves pixels on the 2-sphere by only considering the first neighbors for each given pixel, and subsequently downgrades the map resolution by assigning the average over quartets of pixels to a single pixel at the next lower resolution, thus reducing the $N_{\rm side}$ parameter by one. This is known as "average pooling". We refer to Krachmalnicoff & Tomasi (2019) for additional details on how the algorithm works, as well as its advantages and disadvantages. In the work described in this chapter, we used `NNhealpix` in combination with the public `keras`[2] python package to build our CNN architecture, and to perform training, validation, and testing.

The first part of our CNN, performing image filtering, consists of four CNN building blocks, as illustrated in Fig. 5.3. We accept $N_{\rm map}$ input maps, which in our case represent one or two frequency channels and Stokes Q and U maps, hence $N_{\rm map} = 2$ or 4. Each convolutional layer introduces 32 filters with nine trainable pixel weights, respectively, and is followed by a Rectified Linear Unit (`ReLU`) activation layer. Mathematically, this means each image pixel p_i undergoes a linear transformation f followed by a nonlinear transformation g

$$p_i \mapsto p_i' = (f \circ g)(p_i)\,, \tag{5.1}$$

$$f(p_i) = p_i w_0 + \sum_{j=1}^{N_{\rm neigh}(i)} p_{k_j(i)} w_j\,, \tag{5.2}$$

$$g(x) \equiv \max(0, x)\,, \tag{5.3}$$

---

[1] `https://github.com/ai4cmb/NNhealpix`

[2] `https://keras.io`

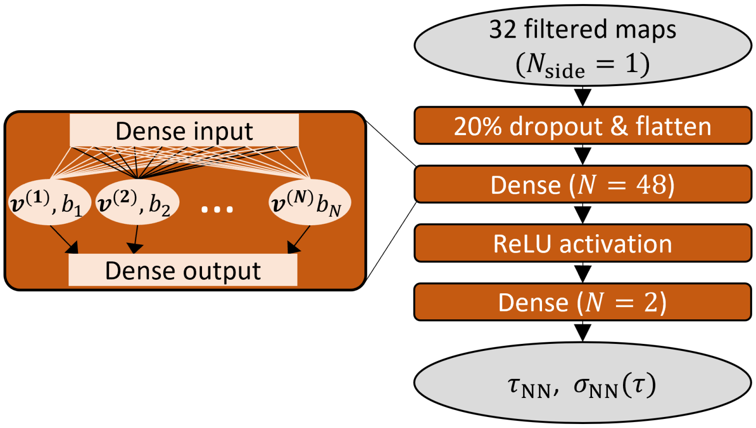

Figure 5.4: Schematic of the fully connected layers of the neural network used in this chapter. This represents the second part of the architecture, performing data compression.

where $k_j(i)$, $j = 1, \ldots, N_{\text{neigh}}(i)$ runs over the indices of all neighboring pixels in the `HEALPix` map (which can be either seven or eight, depending on the pixel location). Then, an "average pooling" degradation layer reduces the map resolution from N_{side} to $N_{\text{side}}/2$, assigning to every low-resolution pixel the average of its four children at the next higher resolution. Up to this point, the application of the four CNN building blocks transform the array of input maps at $N_{\text{side}} = 16$ (or $N_{\text{pix}} = 3072$ pixels) into an array of 32 filtered maps at $N_{\text{side}} = 1$ (or $N_{\text{pix}} = 12$ pixels). This represents the image filtering part, meaning the transformation of the original inputs into 32 maximally compressed feature maps that, ideally, retain all the desired (cosmological) information. We still need to "learn" the mapping from theses feature maps to the output numbers τ_{NN} and $\sigma_{\text{NN}}(\tau)$. Compression is achieved by two fully connected (or dense) layers. We address both of these operations in the following section.

A fully connected layer is a linear map from M-dimensional input feature space to N-dimensional output feature space and is commonly used for data compression (in which case $N < M$). A fully connected layer of output dimension N is said to contain N neurons associated to a vector of trainable weights that parameterize the layer. In each of its N neurons, a fully connected layer linearly contracts the input vector x of length M to a number by means of a weights vector $v^{(i)}$,

$$x_i \mapsto x'_{i'} = \sum_{j=1}^{M} v_j^{(i')} x_j \ . \tag{5.4}$$

The second part of our CNN, the data compression block, is shown in Fig. 5.4 and contains a dropout and flattening layer, a fully connected layer with 48 neurons, a `ReLU` nonlinear activation layer, concluded by a final fully connected layer with two neurons that outputs τ_{NN} and $\sigma_{\text{NN}}(\tau)$

as described in the following section. The dropout layer acts as a selective off switch for parts of the following fully connected layer, deactivating at random 20% of its 48 neurons at a time, thus mitigating the overfitting problem common for neural networks (Srivastava et al., 2014). With the described architecture the total number of weights that need to be optimized during training is $N_\mathrm{w} \approx 4.7 \times 10^4$.

5.2.2 Training

When we train a neural network, we effectively tune its many free parameters until the task at hand, such as estimating parameters from maps, would be optimally performed on the training data. In the following, we describe this procedure in detail.

At each training step we passed one batch of $N_\mathrm{batch} = 32$ training simulations through the network, meaning we simultaneously considered the results from all simulations that belong to a single batch. Input maps need to be masked with the same mask that is used in the analysis. The output values of the two neurons of the final layer, representing the estimated parameters $\tau_{\mathrm{NN}j}$, $\sigma_\mathrm{NN}(\tau)_j$ ($j = 1, \ldots, N_\mathrm{batch}$), as well as the truth values τ_j, are then inserted into the loss function (Jeffrey & Wandelt, 2020)

$$\mathcal{L}\left[\tau,\ \left(\tau_{\mathrm{NN}j},\, \sigma_\mathrm{NN}(\tau)_j\right)\right] = \\ \sum_{j=1}^{N_\mathrm{batch}} \left[(\tau_j - \tau_{\mathrm{NN}j})^2 + \left((\tau_j - \tau_{\mathrm{NN}j})^2 - \sigma_\mathrm{NN}(\tau)_j^2 \right)^2 \right] . \tag{5.5}$$

We then updated all N_w network parameters subject to minimizing this loss function. For doing so, we employed the `Adam` optimizer, a widely used stochastic gradient descent algorithm implemented in `keras`, for which we found an initial training rate of $LR = 10^{-3}$ and first- and second-moment exponential decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$ to be appropriate (see Kingma & Ba, 2014, for more details). Repeating the described procedure for the entire training set of size $N_\mathrm{train} = 190,000$ made up one training epoch[3]. We trained on a maximum of 45 epochs, using the `keras` callback function `ReduceLROnPlateau` to allow for learning rates to decrease by a factor of 0.1 if the loss of the validation set did not improve over the course of five epochs. Moreover, the callback function `EarlyStopping` allows for training to stop after a minimum number of epochs (in our case 20) without improvement in the validation loss. Using both of these callback functions allowed for a faster convergence and suppressed unwanted oscillations in the loss function during the training phase. Training on a 32-core Intel Xeon CPU node took about three hours, while training on eight NVIDIA Tesla V100 GPU cores took about 30 minutes.

---

[3] Among the total 200,000 simulations generated as described in Sect. 5.1, we actually used 190,000 to optimize the NN's parameters, while we used the remaining 10,000 as a validation set.

5.2.3 Testing

After training, we fix the neural network parameters, which in principle completes the model building. However, trained NNs may not perform well for two main reasons: firstly, the loss function may have not converged to its global minimum, causing model predictions to be biased. Secondly, the model may overfit the input, meaning that the network learns the training set's features with an excellent accuracy, but fails to make correct predictions on similar, independent test sets. One usually addresses both problems by testing the model's predictions on simulations that have not been fed into the network before. We used 2×3 test sets of 10,000 sky simulations with fixed input $\tau = \{0.05, 0.06, 0.07\}$, described in detail in Sects. 5.1.2 and 5.1.3.

We note that, by inferring only $\tau_{\rm NN}$ and $\sigma_{\rm NN}(\tau)$, we implicitly assumed Gaussian posteriors, which we exhaustively validated on simulations by checking for biases in the Gaussian mean and variance (see Sect. 5.3). If, instead, our algorithm had provided an entire, potentially non-Gaussian probability distribution function or higher statistical moments, we would have needed to perform more extensive sanity checks and indicate credible intervals instead of Gaussian standard deviations.

5.3 Results on simulations

Before arriving at the estimation of τ from the *Planck* `SRoll2` data, we considered several setups to train our CNN model, increasing the complexity of the training simulations. This allowed us to get valuable insight into the learning process. In particular, we started by training the CNN on a set of simulations including CMB with Gaussian noise (see Sect. 5.1.2), either on a single frequency channel, or on two channels. We then moved to simulations including non-Gaussian systematic effects (i.e., `SRoll2` simulations), trying different possible strategies to obtain unbiased τ estimates in the presence of complex residuals. Only once we achieved this, we applied our trained model to real *Planck* data. In all the cases presented in this section, we trained and tested the CNNs considering the $f_{\rm sky} = 0.5$ mask as our reference (see Fig. 5.2). A summary of all analysis cases, along with their corresponding results tables and figures, can be found in Table 5.1.

5.3.1 Gaussian training

As aforementioned, we first tested the ability of our CNN to estimate the value of τ considering only Gaussian noise. These simulations have noise amplitudes and pixel-pixel correlations directly estimated from *Planck* maps, and therefore serve as a good description of the Gaussian noise present in real data. At the same time, they lack for realism, since they do not include non-Gaussian residual systematic effects, contamination due to Galactic foregrounds, both known to be present on the *Planck* `SRoll2` maps. We therefore expected these models (which we refer to as "Gaussian

Table 5.1: References to results tables and figures in this chapter.

	Gaussian simulations	`SRoll2` simulations	*Planck* data
Gaussian NN (one channel)	Table 5.2; Fig. 5.5	Table 5.3; Fig. 5.5	
Gaussian NN (two channels)	Table 5.2; Figs. 5.5, 5.8	Table 5.3; Figs. 5.5, 5.8	Table 5.5; Figs. 5.9, 5.10
HL likelihood	Table 5.2	Table 5.3	
`SRoll2` training		Table 5.4	
`SRoll2` retraining		Table 5.4; Fig. 5.8	Table 5.5; Figs. 5.9, 5.10
Empirical likelihood			Table 5.5; Fig. 5.10

models") to induce a bias on τ when applied to the more realistic `SRoll2` simulations, or to real *Planck* data.

Single channel. We began by training our CNN on Stokes Q and U maps with Gaussian *Planck*-like noise and CMB at 143 GHz only, thus feeding $N_{\rm map} = 2$ maps to the network. In the left-hand side of Table 5.2, we show the results of testing $N_{\rm sims} = 10,000$ Gaussian simulations of CMB and noise generated with fiducial $\tau = 0.05$, 0.06, and 0.07, respectively. The average learned mean posterior values $\overline{\tau_{\rm NN}}$ are close to unbiased and deviate at the 0.2σ level. The average learned posterior standard deviations $\overline{\sigma_{\rm NN}(\tau)}$ are within 5% agreement with the sample scatter across simulations, $\sigma(\tau_{\rm NN})$.

To assess the performance of the Gaussian model also on non-Gaussian *Planck*-like maps, we tested this model on 10,000 `SRoll2` simulations generated with fiducial $\tau = 0.06$ (see Sect. 5.1.3). As illustrated in the upper right panel of Fig. 5.5, this leads to a bias of more than 1σ on $\tau_{\rm NN}$. These tests on a single frequency channel leave us with two conclusions: on the one hand, CNNs are able to correctly retrieve τ and its statistical uncertainty from a single *Planck*-like simulation of the 143 GHz channel containing correlated Gaussian noise. On the other hand, systematic effects present in the *Planck* `SRoll2` simulations bias the single-channel CNN inference, as expected. To improve our results, we added another frequency channel to the inference pipeline, seeking to mitigate this bias. We expected that combining two channels should lead to a lower error bar and a lower bias on the `SRoll2` simulations, in a similar way as cross-spectra achieve lower noise bias than auto-spectra.

Two channels. As a second test, we added the HFI 100 GHz channel in the training and testing procedures, simulated as CMB plus the corresponding Gaussian correlated noise, so that $N_{\rm map} = 4$ maps were fed into the neural network. The results from testing on Gaussian noise are

Table 5.2: τ predictions from $10,000$ Gaussian CMB + noise simulations generated with three different, fixed fiducial τ values. The results correspond to the Gaussian NN training on one and two channels, and the Bayesian inference with a power spectrum likelihood. We show the posterior mean $\tau_{\mathrm{NN/HL}}$ and standard deviation $\sigma_{\mathrm{NN/HL}}(\tau)$ averaged over all simulations, as well as the scatter of $\tau_{\mathrm{NN/HL}}$ over all simulations.

Test on Gaussian simulations									
	143 GHz			143+100 GHz			143×100 GHz		
	Gaussian training			Gaussian training			HL likelihood		
fiducial τ	$\overline{\tau_{\mathrm{NN}}}$	$\overline{\sigma_{\mathrm{NN}}(\tau)}$	$\sigma(\tau_{\mathrm{NN}})$	$\overline{\tau_{\mathrm{NN}}}$	$\overline{\sigma_{\mathrm{NN}}(\tau)}$	$\sigma(\tau_{\mathrm{NN}})$	$\overline{\tau_{\mathrm{HL}}}$	$\overline{\sigma_{\mathrm{HL}}(\tau)}$	$\sigma(\tau_{\mathrm{HL}})$
0.05	0.0508	0.0059	0.0066	0.0503	0.0054	0.0057	0.0496	0.0046	0.0047
0.06	0.0608	0.0065	0.0067	0.0600	0.0056	0.0059	0.0596	0.0048	0.0048
0.07	0.0712	0.0067	0.0070	0.0702	0.0057	0.0063	0.0697	0.0048	0.0049

shown in Table 5.2. We note two positive effects: firstly, the small bias observed for Gaussian noise on a single channel reduces to below 1% of a standard deviation. Secondly, the learned $\sigma_{\mathrm{NN}}(\tau)$ decreases by more than 10% when training on two frequency channels. Meanwhile, the prediction of the posterior standard deviation stays within 5% of the sample standard deviation of the inferred τ_{NN}. The same results are presented in Fig. 5.5 for fiducial $\tau = 0.06$, showing significant improvement of the two-channel CNN inference in the lower panels with respect to the one-channel results (upper panels). We proceeded to test this two-channel Gaussian model on the `SRoll2` simulations. As shown in the right panel of Fig. 5.5, for fiducial $\tau = 0.06$, the addition of a second channel allows for a significant reduction of the systematic-related bias in τ_{NN} and to a better statistical constraint. This leads us to conclude that CNNs are able to recognize common features across channels, combining the information to reduce the statistical uncertainty and the bias due to uncorrelated systematic effects.

The corresponding quantitative results, for all the three τ values used during testing, are listed in Table 5.3: adding a second channel in the Gaussian training model leads to improved results on the `SRoll2` test simulations for all considered values of τ. However, a residual bias is still present, especially for $\tau = 0.05$, when the CMB signal is smallest.

Moreover, we noticed that, when applied to the `SRoll2` test maps, the models trained on Gaussian simulations returned values of $\sigma_{\mathrm{NN}}(\tau)$ that disagree with the actual spread of estimates $\sigma(\tau_{\mathrm{NN}})$, with the latter being up to about 25% larger. This implies that the learned error is not accurate in this case, hence could not be used to describe the uncertainties of our inferred τ values on `SRoll2` maps. We address this issue in Sect. 5.3.4.

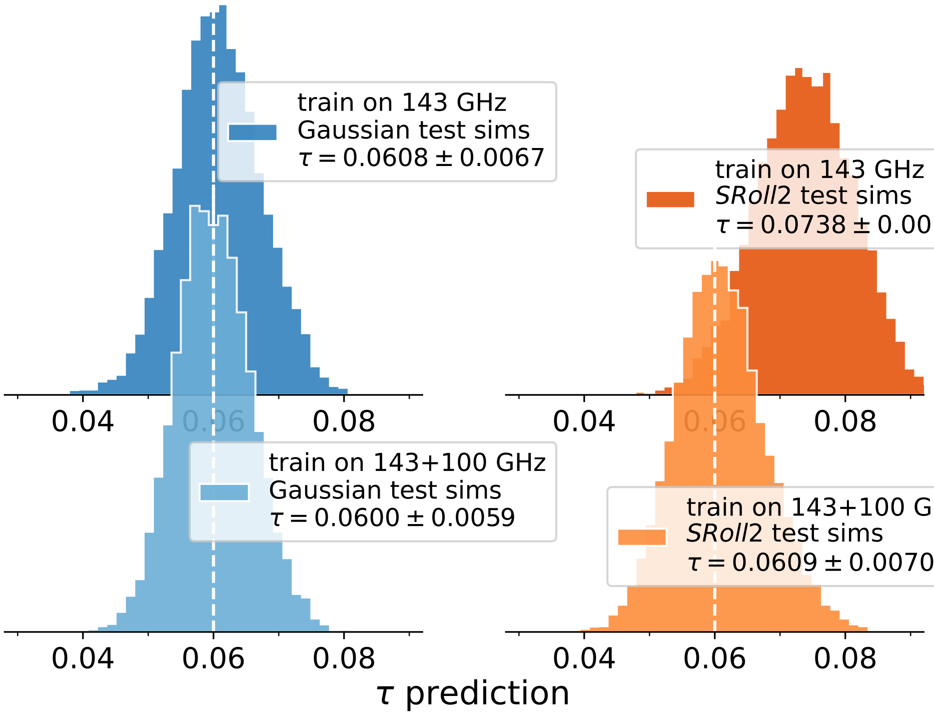

Figure 5.5: Predictions of $\tau_{\rm NN}$ from 10,000 simulations with input $\tau = 0.06$, containing either CMB with Gaussian noise (*left panels*) or CMB with `SRoll2` noise + systematics (*right panels*). The two rows denote different CNN models trained on CMB with Gaussian noise on a single frequency channel (*top*), on two frequency channels (*bottom*).

Table 5.3: Same as Table 5.2 but testing on CMB and `SRoll2` simulations instead of CMB and Gaussian noise simulations.

Test on `SRoll2` simulations									
	143 GHz Gaussian training			143+100 GHz Gaussian training			143×100 GHz HL likelihood		
fiducial τ	$\overline{\tau_{\rm NN}}$	$\overline{\sigma_{\rm NN}(\tau)}$	$\sigma(\tau_{\rm NN})$	$\overline{\tau_{\rm NN}}$	$\overline{\sigma_{\rm NN}(\tau)}$	$\sigma(\tau_{\rm NN})$	$\overline{\tau_{\rm HL}}$	$\overline{\sigma_{\rm HL}(\tau)}$	$\sigma(\tau_{\rm HL})$
0.05	0.0669	0.0065	0.0074	0.0536	0.0055	0.0067	0.0478	0.0050	0.0079
0.06	0.0738	0.0067	0.0076	0.0609	0.0056	0.0070	0.0585	0.0050	0.0073
0.07	0.0813	0.0069	0.0074	0.0690	0.0057	0.0071	0.0688	0.0049	0.0069

5.3.2 Comparison with Bayesian inference from cross-QML C_ℓ estimates

In this section we compare NN inference results with results coming from a standard Bayesian approach applied to E-mode power spectra. In particular, we considered QML estimates (see Sect. 3.1.3) of the 100×143 GHz EE cross-spectrum and drew posterior samples using the HL likelihood (Hamimeche & Lewis, 2008, see also Sect. 3.2.2). The HL likelihood provides a good approximation to the non-Gaussian distribution of the exact power spectrum likelihood, which markedly differs from Gaussianity at the low multipoles $2 \leq \ell \lesssim 30$ that are most relevant for constraining τ. Evaluating the HL likelihood requires a power spectrum covariance matrix, which we obtained directly from simulations of Gaussian noise and CMB realized with the same τ values used for generating the test simulations (Sect. 5.1). For the HL likelihood we assumed a theoretical model of the CMB E-modes, computed with `CAMB`, considering the multipole range $2 \leq \ell \leq 30$, and sampling only for the τ parameter, keeping $10^9 A_s e^{-2\tau} = 1.884$ fixed. Our final results are the best-fit value τ_{HL}, the standard deviation $\sigma_{\text{HL}}(\tau)$ of the posterior, and the scatter $\sigma(\tau_{\text{HL}})$ computed from the set of test simulations.

We ran the HL likelihood on $3 \times 10,000$ Gaussian sky simulations with input $\tau = 0.05$, 0.06, and 0.07. As shown in the last three columns of Table 5.2, we find unbiased best-fit results with average posterior standard deviation $\overline{\sigma_{\text{HL}}(\tau)}$ and best-fit parameter scatter $\sigma(\tau_{\text{HL}})$ of about 0.0048. We note that the uncertainties derived from sampling the HL likelihood are about 20% lower than the ones from NN estimates. Part of the scatter of τ_{NN} comes from the intrinsic stochastic nature of the training process. We could reduce this scatter by taking the average over multiple NN models (as discussed in Sect. 5.3.4). Nevertheless, these results reveal that although we were able to retrieve unbiased τ values with NNs from Gaussian simulations, our estimator does not achieve minimum variance. Further development of the method, including an optimization of the convolution algorithm on the sphere, the NN architecture, and the training procedure, are required and will be explored in future work in the light of improving the estimator's variance.

In addition to Gaussian simulations, we applied the cross-spectrum inference pipeline on $3 \times 10,000$ `SRoll2` simulations and show the corresponding results in the last three columns of Table 5.3. We stress that the HL likelihood contains the same covariance matrix as before, calculated from Gaussian simulations. This is done in analogy with the case of Gaussian NN training applied to `SRoll2` simulations, therefore neglecting the presence of systematic effects. We retrieve biased estimates on τ, confirming our expectation that the power spectrum model implemented in the likelihood is an inaccurate representation of the `SRoll2` simulations, which include spurious non-Gaussian signals. Interestingly, this affects the NN and HL estimates in different ways, leading to biases in opposite directions for $\tau = 0.05$ and 0.06. To study the relative behavior of the two estimators, it is instructive to look at a one-by-one comparison of the NN and HL results on the same 10,000 test simulations, as presented in Fig. 5.6 for $\tau = 0.06$. The scatter plot of the estimated

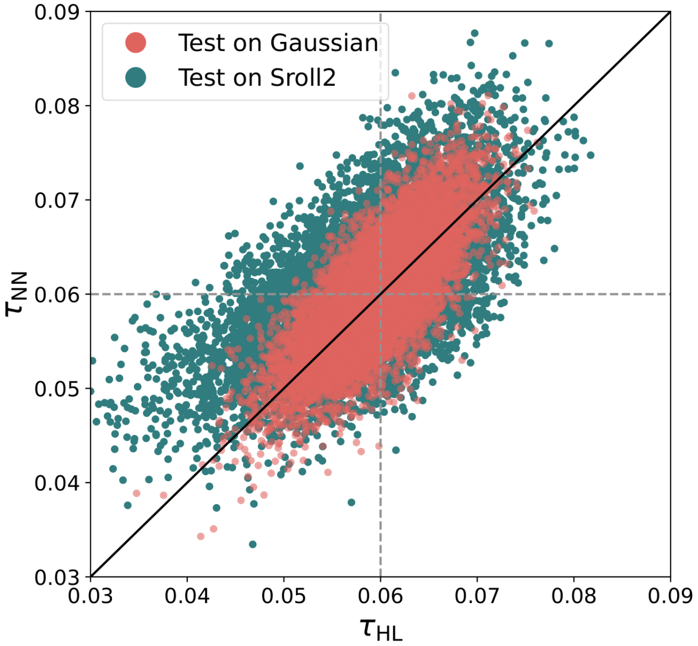

Figure 5.6: Per-simulation comparison between the HL likelihood estimate τ and the NN estimate τ_{NN} for a test set of 10,000 simulations realized with $\tau = 0.06$. Gaussian simulations are shown in bright red, `SRoll2` simulations in dark green. The correlation coefficients between both estimators are 76% (Gaussian) and 63% (`SRoll2`).

.

τ_{NN} and τ_{HL} on Gaussian simulations and on `SRoll2` simulations are shown in bright red and dark green, respectively. In the Gaussian case the correlation of the estimated τ values is at a level of about 76%, while for `SRoll2` it is at about 63%. We conclude that map-level systematic effects, which are partially unaccounted for in the estimates, decrease the correlation and increase the differences between τ_{HL} and τ_{NN} when changing from Gaussian to `SRoll2` test simulations. This indicates that spurious non-Gaussian signals impact the two estimators in different ways.

5.3.3 Training including systematic effects

As previously seen, the two-channel Gaussian training allowed to improve our τ estimates on `SRoll2` simulations. However, the continued occurrence of bias, even though small, motivated us to evolve the training setup by including systematic effects in the training set. Our goal was to achieve fully unbiased results before applying our NN models to real *Planck* maps. In this section we explore two possible ways of including systematics: training on `SRoll2` simulations from the very beginning and performing a `SRoll2` minimal retraining update on previously trained Gaussian

Table 5.4: τ predictions from $10,000$ CMB + `SRoll2` test simulations generated with three different fiducial τ values. These results correspond to two frequency channels, either training on `SRoll2` from the start or retraining on `SRoll2` maps. Displayed are the average posterior mean, average predicted standard deviation $\overline{\sigma_{\rm NN}(\tau)}$, and the scatter $\sigma(\tau_{\rm NN})$ calculated across the test simulations.

Test on `SRoll2` simulations						
	143+100 GHz `SRoll2` training			143+100 GHz `SRoll2` retraining		
fiducial τ	$\overline{\tau_{\rm NN}}$	$\overline{\sigma_{\rm NN}(\tau)}$	$\sigma(\tau_{\rm NN})$	$\overline{\tau_{\rm NN}}$	$\overline{\sigma_{\rm NN}(\tau)}$	$\sigma(\tau_{\rm NN})$
0.05	0.0526	0.0059	0.0066	0.0508	0.0077	0.0091
0.06	0.0622	0.0062	0.0070	0.0606	0.0079	0.0088
0.07	0.0722	0.0064	0.0070	0.0707	0.0081	0.0087

networks.

Training on `SRoll2` simulations. The `SRoll2` simulations (Delouis et al., 2019) are designed to accurately describe *Planck*'s Gaussian noise component and non-Gaussian polarization systematics. Motivated by this, we trained a CNN from the start on the 200,000 `SRoll2` training simulations described in Sect. 5.1.3. As usual, we used 190,000 simulations to perform weight optimization, and 10,000 for validation. We trained on *Planck*'s 143 GHz and 100 GHz channels simultaneously and used the same hyperparameter values as for the Gaussian training, described in Sect. 5.2.2. We stress that even though artificially augmented by forming new channel pair combinations, the `SRoll2` training set was essentially built from 400 sampled skies only. We tested on $3 \times 10,000$ `SRoll2` simulations with fixed $\tau = 0.05$, 0.06, and 0.07, generated from the remaining 100 independent realizations that the CNN did not "see" during training.

Table 5.4 shows the results obtained with this approach. For the three input τ values we find a positive bias of about 0.4σ. For $\tau = 0.06$, the average learned error $\overline{\sigma_{\rm NN}(\tau)} = 0.0062$ is slightly larger than for the two-channel Gaussian training but smaller than the scatter $\sigma(\tau_{\rm NN}) = 0.0070$. We see similar results both for the Gaussian CNN and the HL likelihood inference (see Table 5.3). As in the case of Gaussian NN training, the learned error does not agree with the `SRoll2` simulation scatter, therefore it cannot be used to infer the statistical uncertainty on $\tau_{\rm NN}$.

We ascribe the main reason for the bias on τ to overfitting. Figure 5.7 illustrates this problem. We compared the τ predictions on a set of 10,000 test simulations with the ones coming from 10,000 training simulations. The results show a bias and standard deviation of $\Delta\tau = 0.0023 \pm 0.0069$ for the test set, while the training set is unbiased, with $\Delta\tau = 0.0001 \pm 0.0068$. This is clear evidence for overfitting: while the model performs well on the 400 `SRoll2` simulations that the training set is built from, these are not enough to generalize to the remaining 100 `SRoll2` simulations used to

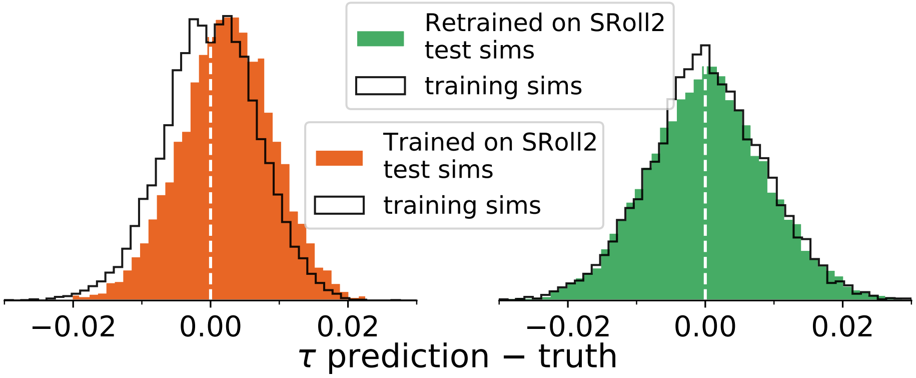

Figure 5.7: Neural network accuracy in predicting the true τ input from $10,000$ simulations. Step-filled histograms show the results on unseen test simulations, black outlines show the results on a subset of the actual training simulations. We compare a network exclusively trained on `SRoll2` simulations (*left panel*) with a Gaussian network retrained on `SRoll2` simulations (*right panel*).

build the test set, leading to the observed bias on τ in the latter case.

Minimal retraining with `SRoll2` simulations. We recognized the bias described above as a critical problem that needed to be addressed. The obvious option, training on a considerably larger simulation set, was unavailable to us due to the limited number of `SRoll2` realizations. Therefore, we applied a transfer learning technique to inform our previously trained Gaussian networks on the `SRoll2` systematics. As shown in the previous sections, our Gaussian NN model is not affected by overfitting issues and, if trained on two channels, performs reasonably well even on `SRoll2` simulations. This motivated us to leverage the existing results on Gaussian networks to solve the overfitting issue with as little changes as possible. To this end, we chose the approach of retraining the two-channel Gaussian model on the full set of `SRoll2` training simulations, while targeting two specific goals:

(i) The retrained CNN should learn to extract information on the systematic effects present in the `SRoll2` simulations and update its CNN weights just enough to achieve fully unbiased results on the `SRoll2` training set.

(ii) At the same time, we wanted to ensure that the information already learned was not destroyed during the new training phase (an issue sometimes referred to as “catastrophic forgetting”, see e.g., Kirkpatrick et al., 2017; Ramasesh et al., 2021), avoiding going back to the overfitting situation described in the previous section.

We achieved this by performing what we call **minimal retraining**: we chose the hyperparameters of the NN such that we obtained unbiased results on the `SRoll2` test simulations while making

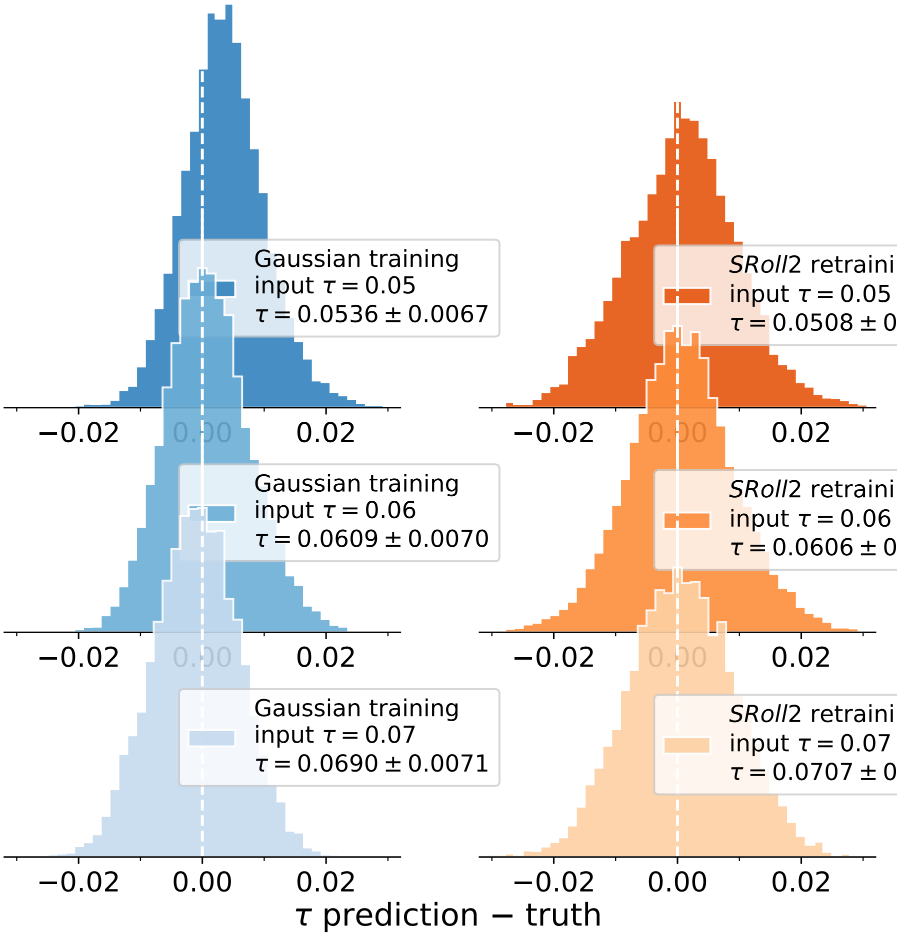

Figure 5.8: Predictions of τ_{NN} on 10,000 `SRoll2` simulations with input $\tau = 0.05$, 0.06, and 0.07 (*first, second, and third row, respectively*). The two columns display two different NN models trained on two channels of Gaussian simulations (*left panels*) and retrained on two channels of `SRoll2` simulations (*right panels*). All results correspond to $f_{\mathrm{sky}} = 0.5$.

minimal changes to the original network. We found an optimal setup with five retraining epochs, a learning rate of LR = 10^{-7}, and no additional changes to the original network architecture.

The right panel of Fig. 5.7, in analogy to the left panel, compares the distribution of $\Delta\tau$ from the `SRoll2`-retrained model on training simulations (black contours), or test simulations (green filled histogram). We find both histograms to be in good agreement, indicating that unlike the `SRoll2`-trained model, the retrained model does not suffer from overfitting, thus achieving our goal (ii) defined above. Table 5.4 on the right-hand side lists the results of the `SRoll2`-retrained model on `SRoll2` test simulations. We find $\overline{\tau_{\mathrm{NN}}} = 0.0508$, 0.0606, and 0.0707 for the respective input values of $\tau = 0.05$, 0.06, and 0.07. This amounts to a bias below $\Delta\tau = 8 \times 10^{-4}$, or $\lesssim 0.1\sigma$. In Fig. 5.8, we show a comparison of the results on `SRoll2` test sets obtained by Gaussian and

`SRoll2`-retrained CNNs. The reduction of the bias is evident, in particular for $\tau = 0.05$. Therefore, we chose the retrained approach as our baseline model to estimate τ on real *Planck* data. At the same time, this approach brings an increase in $\sigma(\tau_{\mathrm{NN}})$, an effect not seen with the `SRoll2` training procedure described in Sect. 5.3.3[4]. This could be the consequence of the typical variance-bias trade-off observed between statistical estimators: with minimal retraining we are able to achieve unbiased estimates (goal (i) above) at the cost of a larger $\sigma(\tau_{\mathrm{NN}})$. In addition to that, we are still unable to retrieve values of the learned $\sigma_{\mathrm{NN}}(\tau)$ that agree with $\sigma(\tau_{\mathrm{NN}})$ for `SRoll2` simulations (and therefore also for *Planck* data). We conclude that, except for the case in which we test the Gaussian model on Gaussian simulations, we cannot use the learned error as an estimate of the uncertainty of the inferred τ_{NN}.

5.3.4 NN errors

The loss function in Eq. (5.5) provides an estimate for the posterior standard deviation $\sigma_{\mathrm{NN}}(\tau)$. However, as seen in the previous sections, the learned $\sigma_{\mathrm{NN}}(\tau)$ tends to underestimate the actual spread of the inferred values of τ_{NN} on test set maps, especially in the case of `SRoll2` maps. We therefore proceeded to empirically estimate our errors from simulations.

In doing so, we needed to make an additional consideration: training a NN is an intrinsically stochastic procedure that relies upon the use of a stochastic optimizer, randomly initialized NN weights and random realizations of the maps in the training set. This results in the fact that each NN prediction can be described as the sum of two random variables: $\tau_{\mathrm{NN}} = \tau + \Delta_{\mathrm{NN}}$, and therefore

$$\sigma^2(\tau_{\mathrm{NN}}) = \sigma^2(\tau) + \sigma^2(\Delta_{\mathrm{NN}}) + 2\,\mathrm{Cov}(\tau, \Delta_{\mathrm{NN}})\,, \tag{5.6}$$

where the first source of uncertainty, $\sigma(\tau)$, is due to noise and cosmic variance of test simulations or observed data, while the second, $\sigma(\Delta_{\mathrm{NN}})$, represents the stochasticity of the NN estimator. These two terms are sometime referred to as "aleatory" and "epistemic" error, respectively (Hüllermeier & Waegeman, 2021).

We can measure the uncertainty related to the NN stochasticity by training an ensemble of models, all based on the same architecture and hyperparameters, but with different initial weights and training set realizations. Our estimate of $\sigma(\Delta_{\mathrm{NN}})$ is given by the standard deviation of the models' τ predictions when tested on a single test map. In practice, we define the "model ID" of a trained NN as the fixed random seed controlling the initialization of network weights. We generated a new training set with map realizations (of CMB, noise, and potentially systematics) that were fully determined by the model ID. Following this recipe, we created 100 independent Gaussian training sets and used them to train 100 Gaussian networks. Repeating this procedure with 100

---

[4] Compare the fourth column in Table 5.4 with the seventh column in Table 5.3

Table 5.5: Results from *Planck* data on two different sky masks, using Gaussian NNs, `SRoll2`-retrained NN models, and the empirical C_ℓ-based likelihood presented in Pagano et al. (2020). The NN results are averaged over 100 models, and $\sigma(\tau_{\rm NN})$ is computed from 10,000 simulations with input $\tau = 0.058$.

Predictions on *Planck* `SRoll2` data						
	143+100 GHz Gaussian training		143+100 GHz `SRoll2` retraining		143x100 GHz C_ℓ likelihood	
$f_{\rm sky}$	$\tau_{\rm NN}$	$\sigma(\tau_{\rm NN})$	$\tau_{\rm NN}$	$\sigma(\tau_{\rm NN})$	τ	$\sigma(\tau)$
50%	0.0588	0.0063	0.0579	0.0082	0.0566	0.0062
60%	0.0593	0.0059	0.0583	0.0078	0.0577	0.0054

`SRoll2` training sets, we retrained the set of 100 Gaussian networks to obtain 100 `SRoll2`-retrained networks. Using a single test map with input $\tau = 0.06$, we find $\sigma(\Delta_{\rm NN}) \simeq 0.0024$ for Gaussian NN models tested on Gaussian maps, and $\sigma(\Delta_{\rm NN}) \simeq 0.0034$ for minimally retrained NN models tested on `SRoll2`. In both cases this represents about 40% of the corresponding value of $\sigma(\tau_{\rm NN})$ reported in Tables 5.2 and 5.4, respectively.

We can reduce the impact of the NN stochasticity by taking, for each test map, the ensemble average of the τ estimates over the 100 trained NNs. By doing so, for the case with $f_{\rm sky} = 0.5$ and input $\tau = 0.06$, we find $\sigma(\tau_{\rm NN}) \simeq 0.0054$ for Gaussian models applied to Gaussian maps and $\sigma(\tau_{\rm NN}) \simeq 0.0083$ for retrained models applied to `SRoll2` simulations.

We also evaluated the correlation coefficient between the predictions of pairs of models (j, k), tested on the same 10,000 simulations, for both Gaussian and `SRoll2` training and testing, respectively. In both cases, we find $\rho_{jk} \simeq 0.84$, in agreement with what is expected if Eq. (5.6) holds and the models' epistemic errors are uncorrelated,

$$\mathrm{Cov}(\Delta^j_{\rm NN}, \Delta^k_{\rm NN}) = \delta^{\mathcal{K}}_{jk}\sigma^2(\Delta_{\rm NN})\,, \tag{5.7}$$

where $\delta^{\mathcal{K}}$ is the Kronecker symbol. In the following section we describe how we applied our CNN models to *Planck* maps to infer the value of τ from real data, estimating its uncertainty from simulations and using the ensemble average over 100 trained models to reduce the impact of the NN stochasticity.

5.4 Results on *Planck* data

As shown in Sects. 5.3.3 and 5.3.4, by retraining on the `SRoll2` simulations, we are able to obtain a CNN-based model that yields unbiased results on unseen `SRoll2` test simulations generated with

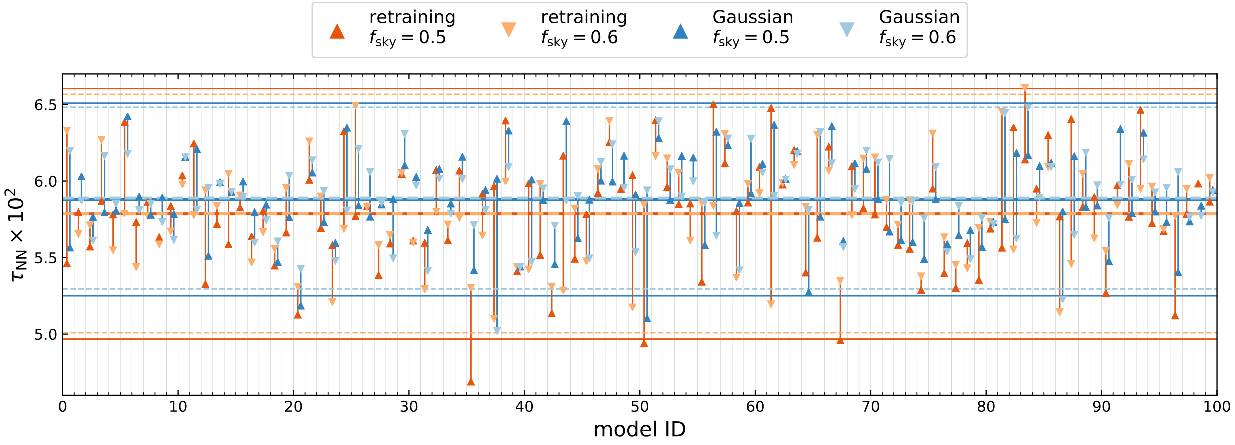

Figure 5.9: NN predictions of τ from *Planck* 100+143 GHz data, resulting from training 100 equivalent models with different random initial weights and random seeds for training data, considering Gaussian two-channel training (*blue tones*) versus `SRoll2` retraining (*orange tones*), and $f_{\rm sky} = 0.5$ (*downward triangles*) versus $f_{\rm sky} = 0.6$ (*upward triangles*). Colored triangle markers show the best-fit values for the single models and horizontal lines in the corresponding colors show the ensemble average of τ (*bold lines in the middle*) ± the 68% central CL (*thin lines on the bottom and the top*).

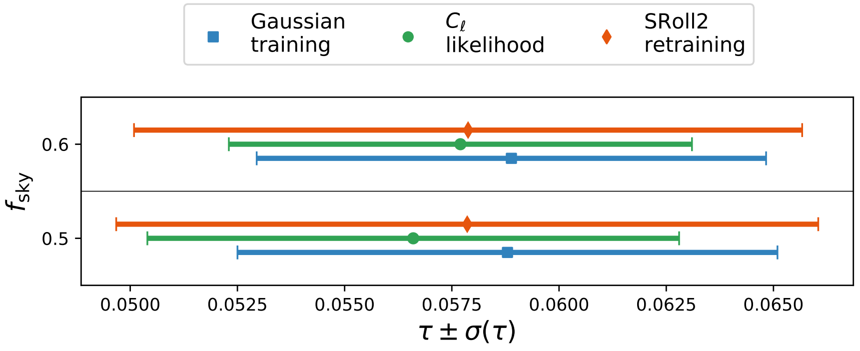

Figure 5.10: Results on τ obtained from *Planck* `SRoll2` data. The values correspond to Table 5.5.

fixed $\tau \in \{0.05, 0.06, 0.07\}$. Having thus confirmed the robustness of our method, we moved to real *Planck* data and proceeded to predict τ from the 100 and 143 GHz `SRoll2` HFI maps.

We obtained our baseline τ estimate by taking the average of the inferred values from the 100 minimally retrained NNs applied to *Planck* data for a sky mask with $f_{\rm sky} = 0.5$, resulting in a mean estimate of $\tau_{\rm NN} \simeq 0.0058$. Figure 5.9 shows the obtained τ values for each of these NN models. Following the conclusions of the previous sections, since the learned $\sigma_{\rm NN}(\tau)$ is inadequate as an error prediction, we estimated the uncertainty from simulations. In practice, we generated a set of 10,000 `SRoll2` simulations realized with $\tau = 0.058$ and average the $\tau_{\rm NN}$ estimates over 100 networks. Afterwards, we computed the standard deviation over 10,000 simulations. Our final inference on *Planck* maps in this baseline case results in:

$$\tau_{\rm NN} = 0.0579 \pm 0.0082 \quad (\textit{Planck}\, 100 + 143\,\text{GHz})\,. \tag{5.8}$$

This value is in very good agreement with the τ estimates obtained with an empirical likelihood based on cross-QML power spectra, presented in Pagano et al. (2020) (hereafter P2020), applied to the same *Planck* maps and constructed from the same `SRoll2` simulations that we use in this chapter. In particular, P2020 obtained $\tau = 0.0566^{+0.0053}_{-0.0062}$ on the $f_{\rm sky} = 0.5$ sky mask. We note that the uncertainty from our NN method is about 30% larger. As previously described, this is due to the fact that our NN estimator does not reach minimum variance and that we relied on the retraining strategy leading to larger errors. However, the fact that we obtain a τ value in agreement with the literature while using an inherently different inference approach that is, for the first time, fully based on NNs, represents a remarkable result of this work.

We also applied the Gaussian NN model to *Planck* data, deriving the best-fit parameter value and error bars analogously. We note that, although the Gaussian model leads to results that are mildly biased by up to about 0.5σ when applied to `SRoll2` maps with low CMB input signal ($\tau = 0.05$), the bias is below 0.15σ when $\tau = 0.06$, as displayed in the fifth column of Table 5.3. In this case, using the same $f_{\rm sky} = 0.5$ mask, we obtained $\tau_{\rm NN} = 0.0588 \pm 0.0063$. The statistical uncertainty is lower for this second method, as we omitted retraining on systematics, and similar to the one obtained from the empirical likelihood presented in P2020.

Lastly, as a robustness test, we applied these same methods to a second sky mask, with a larger sky coverage of $f_{\rm sky} = 0.6$. The parameter estimates remain stable for both the retrained and the Gaussian model, with lower uncertainties. A model-by-model comparison of the NN predictions for every single network on $f_{\rm sky} = 0.6$ and $f_{\rm sky} = 0.5$ is displayed in Fig. 5.9. A summary of our results on *Planck* maps is shown in Fig. 5.10 and Table 5.5.

5.5 Summary

In this chapter, we present the first NN-based cosmological parameter inference that operates directly on CMB polarization maps. We estimated the optical depth τ from the largest angular scales of the *Planck* `SRoll2` polarization maps, which contain significant non-Gaussian residual contamination from instrumental and foreground systematics, which are difficult to write down in a model. Pagano et al. (2020) approach this issue by means of an empirical power spectrum likelihood built from `SRoll2` simulations. In our novel approach, we try to solve this by using a NN-based inference algorithm on the map domain that maps Stokes Q and U maps to a τ value, merely requiring realistic simulations for NN training. Our convolutional network algorithm is fully compatible with spherical maps, using the `NNhealpix` code (Krachmalnicoff & Tomasi, 2019). We considered training on a single channel and two channels, and explored training on Gaussian maps with *Planck* noise levels, `SRoll2` simulations, and a hybrid minimal retraining approach. We validated the NN pipeline against unseen simulations and, in the case of Gaussian data, a Bayesian C_ℓ-based inference pipeline.

By extensively testing our NN algorithms on simulations, we obtained several key results that we summarize in the following.

- When trained and applied to Gaussian simulations, the NN models retrieve unbiased τ values directly from maps. Training them on two frequency channels that share the same cosmological signal leads to substantial improvement in accuracy and precision compared to one channel alone. This allows the straightforward combination of data sets to help reduce the impact of noise and systematic effects without the need for a joint model, and is a key advantage of the NN approach. However, on Gaussian maps, the NN estimator does not reach the optimality of the QML power spectrum estimator.

- Testing on the non-Gaussian `SRoll2` simulations, we find that residual systematic effects mildly bias both the Gaussian two-channel NN model and the HL likelihood. To address this, we incorporated instrumental systematic effects by minimally retraining the Gaussian NN models on a limited set of 400 `SRoll2` simulations, leading to unbiased τ estimates with an error increased by 30%. Lastly, we found that the empirical τ standard deviation retrieved from the test simulations overestimated the real uncertainty due to an intrinsic NN stochasticity. To overcome this issue, we evaluate the final error on τ through simulations, by taking the ensemble average of 100 NN models.

- After validating the performance of the NNs on simulations, we applied our trained models to *Planck* `SRoll2` data at 100 and 143 GHz. For the minimally retrained model, which is the one that leads to fully unbiased results on the `SRoll2` simulations, we obtain $\tau_{\rm NN} =$

0.0579 ± 0.0082. The value is in excellent agreement with P2020, which uses an `SRoll2` simulation-based likelihood but relies on an inherently different analysis method. Our final simulation-based uncertainty estimate on τ_{NN} is about 30% larger. This is because our NN estimator is intrinsically suboptimal and minimal retraining further enlarges its variance.

- To test the robustness against systematic-induced "unknown unknows", we repeated the *Planck* analysis with the Gaussian network, which achieved good results on `SRoll2` simulations for $\tau = 0.06$ while being agnostic to non-Gaussian map features. We obtain $\tau_{\mathrm{NN}} = 0.0588 \pm 0.0063$, in agreement with the estimate reported in the literature, and with a similar level of uncertainty. As a second robustness check, we used a mask with a larger $f_{\mathrm{sky}} = 0.6$ and find consistent results, confirming the stability of our τ_{NN} estimates.

The main limitations of our NN estimator are its intrinsic suboptimality (meaning it does not reach minimum variance even if trained and tested on Gaussian simulations), the slightly biased NN estimate of statistical uncertainty, and the insufficient `SRoll2` training data for the model to learn the features of the `SRoll2` data without overfitting. We expect that future improvements regarding the intrinsic suboptimality can be achieved through algorithmic developments. Further refinements in the algorithm are also needed to achieve unbiased NN-based uncertainty estimates, which we did not use for our final results on real data. While we circumvented the `SRoll2` overfitting problem by means of the minimal retraining method, we anticipate that larger non-Gaussian training sets are indispensable to achieve a significantly higher precision on τ with *Planck* data. It is important to stress that obtaining reliable results on real data required a significant effort to validate and test our models on different setups and to develop training strategies that can effectively cope with systematic effects. This highlights the fact that NN models developed to perform well on simplified simulations cannot always be straightforwardly applied to real data and need careful consideration of the training and validation procedures.

Nonetheless, the consistent and robust results we obtain demonstrate that NNs represent a promising tool that could complement standard statistical data analysis techniques for CMB observations, especially in cases where the Gaussian CMB signal is contaminated by spurious effects that cannot be analytically described in a likelihood model. This is particularly relevant for the ongoing search for primordial gravitational waves, constrained by large- to intermediate-scale B-modes which are targeted by a number of near-future experiments such as SO, LiteBIRD, or CMB-S4. For a further outlook on improvements and applications, we refer to the Conclusions.

CHAPTER 6

Efficient parameter marginalization for LSS analysis

In the matter-dominated Universe, linear perturbations of the matter density, widely believed to originate from primordial quantum fluctuations, kept growing gravitationally and eventually formed the cosmic large-scale structure, which consists of massive clusters and superclusters hosting dark matter halos and galaxies that we observe today. Cosmologists describe the formation of the LSS by models of hierarchical growth which, once implemented in numerical simulations, are able to predict the LSS morphology to remarkable accuracy. The growth of cosmic structure is highly sensitive to the parameters of the ΛCDM model, such as the fractional densities of dark energy, dark matter, baryons, and neutrinos, each of which influence the growth dynamics in their own characteristic way. Real data of the LSS can therefore be used for cosmological inference that complements CMB data analysis in that it constrains late-time gravitational dynamics (Press & Schechter, 1974; Bond et al., 1996; Springel et al., 2005). A particular challenge for future LSS analyses is the tension affecting the measurement of the amplitude of matter fluctuations, $S_8 \equiv (\sigma_8 \Omega_m / 0.3)^{0.5}$, where σ_8 is the root mean square of matter fluctuations on an 8 Mpc$/h$ scale, and Ω_m is the fractional energy density in non-relativistic matter. In particular, the latest prediction from *Planck* CMB data finds its value (Planck Collaboration VI, 2020) to be about 2σ to 3σ higher than the analogous measurement from surveys such as the Kilo-Degree Survey (KiDS), the Dark Energy Survey (DES), and Hyper Suprime-Cam (HSC) (Heymans et al., 2021; DES Collaboration, 2022, 2018; MacCrann et al., 2015; Hamana et al., 2020).

As we go to smaller scales, our models become more and more sensitive to small-scale astrophysical processes, such as star formation, baryonic feedback from massive stars, or mass accretion by black holes at the centers of active galactic nuclei. Since including such processes in detail would be too expensive for present-day numerical simulations, they are often included as effective parameters. Another example of an effective parameterization is galaxy bias, a phenomenological model that quantifies the poorly understood relation between the two-point statistics of the galaxy distribution and the total matter distribution (including dark matter). While these “astrophysical systematics” may not be of immediate interest to the cosmological analysis, they must be jointly constrained with the cosmological parameters, which often leads to degeneracies between large-scale cosmology and small-scale astrophysics.

As one of the most common LSS data analysis techniques, tomographic "3x2pt" analysis jointly investigates three different observables that we can effectively use to constrain the morphology of the LSS. These are:

- **Galaxy clustering autospectra:** angular power spectra between galaxy number densities observed in different tomographic bins at various redshifts ("redshift bins"), realized by means of observations with different photometric filters,
- **Weak lensing shear autospectra:** angular power spectra of small distortions in the observed shapes of galaxies ("galaxy shear") between different redshift bins, caused by the weak gravitational lensing effect ("weak lensing") from intervening massive structures along the line of sight,
- **cluster-shear cross-spectra:** angular cross-power spectra between the galaxy number density in one redshift bin and the weak lensing shear in another one.

Future tomographic 3x2pt analyses have the potential to break degeneracies between cosmological and astrophysical parameters and shed light on known cosmological tensions, such as the one regarding the S_8 parameter. To be able to push current limits on the cosmological parameters, current and future LSS analyses require accurate knowledge of the astrophysical systematics, in particular galaxy bias. Current experiments, such as DES (DES Collaboration, 2022), model such systematics by $O(20)$ redshift-dependent parameters that are included in the data model together with the cosmological parameters and then marginalized over. Ignoring astrophysical systematics would introduce critical inaccuracies, especially in the S_8 parameter.

In Bayesian inference, effective parameters that are not of primary interest to the analysis are sometimes called "nuisance parameters". Marginalizing over many nuisance parameters, however, leads to long convergence times if we use traditional parameter sampling methods, such as MCMC, which rely on sampling the full high-dimensional posterior distribution. In the case of 3x2pt analysis of the LSS, this may result in convergence times of up to a few weeks for state-of-the-art data sets, critically delaying the overall data analysis. This calls for more efficient methods that take into account large sets of nuisance parameters without this numerical overhead. Addressing this issue is the main focus of this chapter.

In this final chapter, we present a recent work, Hadzhiyska, Wolz, et al. (2023), that introduces an analytical method, based on the Laplace approximation, to efficiently marginalize over nuisance parameters in Bayesian inference. We apply this to real and simulated future data of the LSS of the Universe, where alleviating this issue is of immediate relevance to current and near-future data analysis. Specifically, the contributed work carried out in the context of this Thesis has been the writing of a code that implements an analytical approximate likelihood for 3x2pt analysis of the LSS. We stress that our method is not strictly limited to this application but can be readily applied to other Bayesian inference problems where the marginalization over nuisance parameters considerably slows analyses down.

This is also relevant for CMB polarization analysis, specifically the search for primordial B-modes. As we show in Chpt. 3, for near-future experiments such as SO, standard power spectrum analysis methods rely on marginalizing over foreground systematics to obtain robust results. A similar exercise is likely needed to optimally account for instrumental systematic effects described in Sect. 2.7. More sensitive experiments such as LiteBIRD and CMB-S4 are expected to force up the number of nuisance parameters even more. Therefore, analytical marginalization may become a time-critical requirement for future CMB experiments.

This chapter is structured as follows. We start by presenting our new method in Sect. 6.1, summarizing its advantages and disadvantages, and deriving the mathematics. We then discuss volume effects, a problem that can arise when marginalizing over many nuisance parameters, and identify them in the context of our approximation. In Sect. 6.2, we apply our method to LSS data analysis and present results and conclusions.

6.1 New efficient bias parameter marginalization

In order to solve the problem of computational efficiency that comes with marginalizing over many nuisance parameters, we propose a new analytical technique based on the Laplace approximation. This method is a general semi-analytical formalism that assumes that upon fixing the key parameters (in our case, the cosmological ones), we can locally describe the joint probability density of the remaining nuisance parameters by a multivariate Gaussian. The Laplace approximation is, under certain conditions that we discuss below, a good approximation to the exact marginal posterior. It comes with the clear advantage of faster convergence, since it reduces a high-dimensional sampling problem to a low-dimensional sampling plus a (fast to evaluate) maximization problem. Like standard numerical marginalization, this analytical marginalization may lead to so-called "volume effects" that bias the analysis, requiring us to carefully choose our set of parameters and priors. Moreover, as this method is an approximation, we need to know when we can safely apply it.

Our goal is to explore the posterior distribution of a set of parameters that can be divided into parameters $\boldsymbol{\Omega}$ of cosmological interest, and nuisance parameters $\boldsymbol{n}$ that quantify systematic biases we wish to marginalize over. The full (joint) posterior is proportional to the likelihood $P(\boldsymbol{d}|\boldsymbol{\Omega}, \boldsymbol{n})$ times the prior $P(\boldsymbol{\Omega}, \boldsymbol{n})$ via Bayes theorem, and is given by

$$P(\boldsymbol{\Omega}, \boldsymbol{n}|\boldsymbol{d}) \propto P(\boldsymbol{d}|\boldsymbol{\Omega}, \boldsymbol{n})\, P(\boldsymbol{\Omega}, \boldsymbol{n})\,. \tag{6.1}$$

We find the marginal cosmological posterior by integration over the nuisance parameters,

$$P(\boldsymbol{\Omega}|\boldsymbol{d}) = \int \mathrm{d}\boldsymbol{n}\, P(\boldsymbol{\Omega}, \boldsymbol{n}|\boldsymbol{d})\,. \tag{6.2}$$

The marginal posterior is the main quantity of interest that we wish to evaluate as fast and accurately as possible. To do so, we use the Laplace approximation, as described the following.

6.1.1 Laplace approximation

The Laplace approximation makes use of the fact that most unimodal distributions can be sufficiently well described by a Gaussian centered at their maximum. Defining the χ^2 function as the negative log-posterior modulo an irrelevant normalization constant K_1,

$$\chi^2(\boldsymbol{\Omega}, \boldsymbol{n}) \equiv -2 \ln P(\boldsymbol{\Omega}, \boldsymbol{n}|\boldsymbol{d}) + K_1\,, \tag{6.3}$$

the Laplace approximation can be written as

$$\chi^2(\boldsymbol{\Omega}, \boldsymbol{n}) \simeq \chi^2(\boldsymbol{\Omega}, \boldsymbol{n}_*) + \boldsymbol{\Delta n}^T \mathcal{F}(\boldsymbol{\Omega}, \boldsymbol{n}_*) \boldsymbol{\Delta n}\,, \tag{6.4}$$

where $\boldsymbol{n}_* = \boldsymbol{n}_*(\boldsymbol{\Omega})$ maximizes the full posterior locally, meaning at a fixed set of cosmological parameter values, $\boldsymbol{\Delta n} \equiv \boldsymbol{n} - \boldsymbol{n}_*$. Moreover,

$$\mathcal{F}_{ij}(\boldsymbol{\Omega}, \boldsymbol{n}) \equiv \frac{1}{2} \frac{\partial^2 \chi^2}{\partial n_i \partial n_j}(\boldsymbol{\Omega}, \boldsymbol{n}) \tag{6.5}$$

is the (negative) Hessian matrix of the log-posterior in the subspace of the nuisance parameters, locally evaluated at $\boldsymbol{n}_*(\boldsymbol{\Omega})$ given a fixed cosmology $\boldsymbol{\Omega}$. The Laplace approximation is thus a second-order Taylor expansion in the nuisance parameters at a fixed cosmology around the local posterior maximum $\boldsymbol{n}_*$. We can now take the Laplace approximation (6.4) and marginalize over the nuisance parameters to find

$$-2 \ln P(\boldsymbol{\Omega}|\boldsymbol{d}) \simeq \chi^2(\boldsymbol{\Omega}, \boldsymbol{n}_*) + \ln \det \mathcal{F}(\boldsymbol{\Omega}, \boldsymbol{n}_*) + K_2\,, \tag{6.6}$$

where K_2 is another irrelevant constant.

Equation (6.6) describes the marginal posterior in the local Laplace approximation at fixed $\boldsymbol{\Omega}$. Its accuracy depends on how well the local posterior follows a Gaussian distribution. Let us have a closer look at the right hand side of (6.6). The first term is identical to the frequentist "profile likelihood", obtained by locally maximizing the nuisance parameters at a given set of cosmological parameters (Cole et al., 2013). If the local distribution of data is sufficiently close to Gaussian, we can obtain confidence levels on $\boldsymbol{\Omega}$ by means of the χ^2 distribution (Feldman & Cousins, 1998). By construction, the profile likelihood is centered at the best-fit parameters and is therefore independent of the choice of the nuisance parameters (Hamann et al., 2007; Herold et al., 2022; Campeti & Komatsu, 2022), as we discuss in the following section. The second term, dubbed "Laplace term" (also known as "Occam's razor term") measures the volume in nuisance parameter space enclosed by the Gaussian approximation. The Laplace term is associated with volume effects that introduce a dependence between the cosmological parameters and the choice of the nuisance parameters with their associated priors. Calculating the Laplace approximation is especially efficient using common minimization algorithms, as finding the optimal step size in these algorithms often requires evaluating the Hessian of the function being minimized. Therefore, the matrix $\mathcal{F}$ entering the Laplace term in Eq. (6.2) is already a product of minimizing the function $\chi^2(\boldsymbol{\Omega}, \cdot)$ to find the point $\boldsymbol{n}_*$.

6.1.2 Volume effects

It is instructive to study the Laplace approximation when the data follows a Gaussian distribution and the nuisance parameters enter the data model in a linear way. In this case, the Laplace approximation is exact and can be evaluated analytically, meaning that the local posterior (at fixed cosmological parameters $\boldsymbol{\Omega}$) is exactly Gaussian in $\boldsymbol{n}$. The Laplace term neither depends on the nuisance parameters $\boldsymbol{n}$ nor on the data $\boldsymbol{d}$, but it generally varies with the cosmological parameters, acting as an $\boldsymbol{\Omega}$-dependent correction to the profile likelihood. This induces a so-called "volume effect". To gain some intuition on what a volume effect entails, let us now consider noise-dominated data and assume the priors are wide and therefore negligible. Naïvely, one would expect the data to favor no region of parameter space whatsoever. In the noise-dominated regime, the profile term is given by the sum of least squares between the data and the model, which averages to the number of data points and does not depend on the model, as expected. However, the Laplace term, as we argued above, depends on the cosmological parameters, even in the absence of data. Therefore, volume effects tend to be important if the model parameters are poorly constrained by both the data and the priors.

In general, volume effects are shifts in the peak of the marginal posterior $P(\boldsymbol{\Omega}|\boldsymbol{d})$ with respect to the profile likelihood $P(\boldsymbol{\Omega}, \boldsymbol{n}_*|\boldsymbol{d})$. It arises in multivariate Bayesian inference when marginalization favors regions of parameter space that cover a larger volume of the probability density. The size and direction of volume effects depend on the definition of the nuisance parameters and, if informative, the associated priors. Another simple example is fitting noisy data to a power-law model of the form nx^{Ω}, where x is an independent variable that takes values larger than one. Let us assume the noise makes the data scatter around zero. If we try to fit the power-law model, Ω prefers large negative values, since the amplitude parameter n can then take a larger range of possible values that provide a reasonable fit to the noisy data. Re-defining the model to $n(x/x_0)^{\Omega}$, where x_0 is larger than all values assumed by x, the best-fit Ω will instead favor large positive values.

The finding that volume effects reside in the Laplace term is closely related to the concept of Jeffreys priors. The Jeffreys prior (Jeffreys, 1946) is a common choice to eliminate the dependence on model parameterization, and thus to partially mitigate the impact of volume effects. It is defined as

$$2 \ln P_J(\boldsymbol{\theta}) \equiv \ln \det F(\boldsymbol{\theta}) \,, \tag{6.7}$$

where $F(\boldsymbol{\theta})$ is the Fisher matrix of the parameters $\boldsymbol{\theta}$. While the Fisher matrix generally differs from the Hessian matrix $\mathcal{F}$ defined above (the former is the data average of the latter), it is straightforward to show that both are identical for Gaussian data that have a parameter-independent covariance and a model that linearly depends on parameters. If we take the Jeffreys prior at fixed $\boldsymbol{\Omega}$, we find that:

- the full Laplace approximation (6.6) and the marginal posterior (6.2) are equal,
- the marginal posterior including the Jeffreys prior, and the profile term (first term on the right hand side of Eq. (6.6)) are equal.

Thus, in this limit, the volume effects residing in the Laplace term (second term on the right hand side of Eq. (6.6)) would be exactly canceled if we chose to use a Jeffreys prior. In this scenario, we can also confirm the general tendency that we found above: volume effects become negligible if the constraints on nuisance parameters are either driven by the data (in which case the profile term dominates over the Laplace term), or by the priors (in which case the Laplace term is constant and does not vary with $\boldsymbol{\Omega}$).

6.2 Application to LSS analysis

In this section, we apply our new method to the problem of marginalizing over astrophysical systematic effects in two-point analyses of the large-scale structure of the Universe. A significant contribution to this work, carried out in the context of this Thesis, was the implementation of the analytical approximate likelihood in a `python` code framework.

Current and future LSS data analysis is based on observables that trace the matter and galaxy distributions, in particular galaxy number counts and galaxy weak lensing shear. Galaxies, being subject to many complex astrophysical processes during their evolution, do not accurately trace the underlying cosmological quantity that is the matter distribution. This uncertainty is quantified by a large set of parameters $\boldsymbol{b}$ that include galaxy bias, intrinsic noise in the shape of galaxies, and intrinsic alignment. Next-generation experiments rely on tightly constraining those bias parameters to obtain accurate cosmological results, but the large number of bias parameters needed (currently, more than 20 bias parameters versus six cosmological parameters for DES, (DES Collaboration, 2022)) brings a high numerical cost that is expected to grow with the sensitivity of the instrument.

One of the most widely used LSS analysis techniques relies on radially subdividing the three-dimensional volume into two-dimensional slices of different redshifts ("tomographic redshift bins") and measuring angular power spectra between those redshift bins. Crucially, in tomographic power spectrum analysis, the theory model $C_\ell^{\boldsymbol{XY}}$ between two tracers $\boldsymbol{X}$ and $\boldsymbol{Y}$ generally depends on the bias parameters $\boldsymbol{b}$ through a second-order polynomial, with coefficients that only depend on the cosmological parameters, $\boldsymbol{\Omega}$:

$$C_\ell^{XY}(\boldsymbol{\Omega}, \boldsymbol{b}) = \epsilon^X \epsilon^Y C_\ell^{X_0 Y_0}(\boldsymbol{\Omega}) + \sum_i b_i^X \epsilon^Y C_\ell^{X_i Y_0}(\boldsymbol{\Omega}) + \{X \leftrightarrow Y\} + \sum_{i,j} b_i^X b_j^Y C_\ell^{X_i Y_j}(\boldsymbol{\Omega}) \,. \tag{6.8}$$

Here, the sums go over different bias terms related to a single biased tracer, and ϵ^X is either one or zero, depending on whether or not that tracer has an "unbiased" component that directly traces the matter distribution. $\{X \leftrightarrow Y\}$ on the right-hand-side of Eq. (6.8) implies a repetition of the preceding term with X and Y exchanged. Assuming a Gaussian likelihood with a parameter-independent covariance matrix (a common assumption valid under sufficiently general conditions, as shown in Kodwani et al. (2019)), this leads to a quartic dependence on the bias parameters in the log-likelihood. Hence, we do not expect the Laplace approximation, which is quadratic by construction, to exactly reproduce the bias-marginalized posterior.

We make use of the simple analytical form of the likelihood given by Eq. (6.8), and implement a two-level parameter sampler that distinguishes between bias and cosmological parameters. At the outer level, we sample over the five-dimensional space of cosmological parameters and, at every step, evaluate the angular matter power spectrum over all unique pairs of redshift bins using the `Halofit` code (Smith et al., 2003). At the inner level, we use a Newton-Raphson minimizer to compute the profile likelihood and the Laplace term in the subspace of bias parameters for a fixed cosmology. Combining both levels, we obtain the five-dimensional profile likelihood and our analytical approximation to the marginalized posterior. To compare these methods with the established "brute-force" sampling, we also sample the full joint posterior using the `cobaya` MCMC sampler (Torrado & Lewis, 2019, 2021) and numerically marginalize over the bias parameters. The main work that we carried out in the context of this Thesis, was to implement this two-level likelihood starting from an existing one-level likelihood that had been designed to sample over the full, high-dimensional parameter distribution. In particular, our work entailed writing the code for the data model, defined by (6.8), and re-organizing the full parameter vector into a vector of nuisance parameters to analytically marginalize over, and a vector of cosmological parameters to include in the MCMC sampler.

To test our method in realistic current and near-future scenarios, we consider simulated data of a future high-sensitivity experiment similar to LSST (LSST Dark Energy Science Collaboration, 2012), and the first-year data release of DES (DES Collaboration, 2022). Let us briefly summarize both data sets:

- **LSST-like mock data:** This is a simulated idealized 2x2pt data set, consisting only of the clustering-clustering and clustering-shear power spectra. It is important to test our method in the low-noise regime, where the inferred posterior strongly depends on degeneracies between cosmological and bias parameters, and where the final error budget is more dominated by these effects. By excluding the shear-shear autocorrelation, the model is more sensitive to the complexity of the galaxy bias parameterization. To define the clustering and shear samples we follow the same procedure outlined in Nicola et al. (2023), using six redshift bins for galaxy clustering and five redshift bins for galaxy shear, with associated number densities and redshift distributions in line with expectations for LSST (see The LSST Dark Energy Science Collaboration et al., 2018; Alonso et al., 2015, for more details). For simplicity, we use a Gaussian covariance to describe the uncertainties of the resulting data vector, calculated assuming a sky fraction $f_{\rm sky} = 0.4$. The LSST data vector was generated assuming a true cosmology with parameters $(\Omega_m, \Omega_b, h, n_s, \sigma_8) = (0.3, 0.05, 0.7, 0.96, 0.8)$.

- **DES-Y1 data:** We use the clustering-clustering, clustering-shear, and shear-shear power spectra and covariance matrix provided in García-García et al. (2021), constructed from the DES-Y1 data. DES is a five-year photometric survey which has observed $5000\,\mathrm{deg}^2$ of the sky using five different filter bands from the 4m-aperture Blanco Telescope at the Cerro Tololo Inter-American Observatory (CTIO) in Chile. The galaxy clustering sample is divided into five redshift bins, defined in Elvin-Poole et al. (2018), and we employ the fiducial redshift distributions released by DES to model the angular power spectra. For the galaxy shear analysis, we use the official sample used in the DES-Y1 analysis (Zuntz et al., 2018), including all cuts and definitions of the four tomographic bins (Hoyle et al., 2018). Galaxy shapes were determined using the METACALIBRATION algorithm (Huff & Mandelbaum, 2017; Sheldon & Huff, 2017). We refer to Nicola et al. (2021) for further details regarding the estimation of the shear power spectra and covariance.

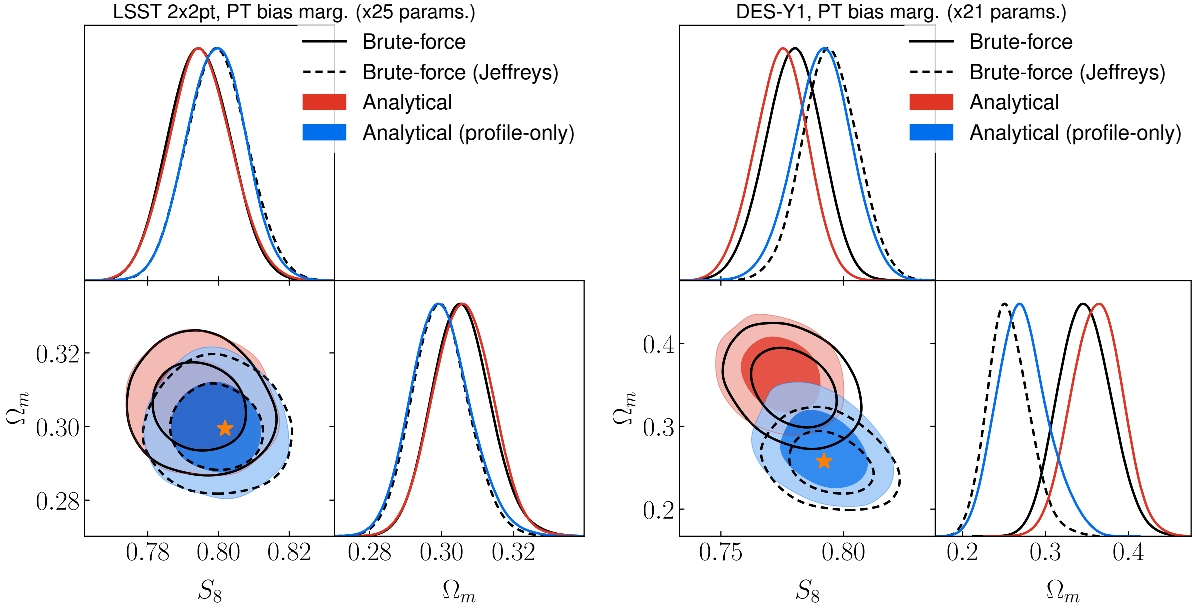

Figure 6.1: Demonstration of volume effects for simulated and real LSS power spectrum data, modeled with six cosmological parameters and 25 and 21 bias parameters, respectively. This plot shows the 2D marginal posterior (with contours indicating the 68% and 95% credible levels) of the rescaled amplitude of matter fluctuations, S_8, and the fractional matter density Ω_m, which are the two cosmological parameters that are most sensitive to astrophysical systematics. To recover the best-fit parameter values (denoted with a star), when applying brute-force marginalization, one needs to employ a Jeffreys prior (compare solid with dashed black line). Similarly, when analytically marginalizing over the bias parameters, one needs to remove the Laplace term to avoid biasing the constraints (compare blue and red contours). *Left panel:* constraints for a 2x2pt analysis with a simulated LSST-like data set. *Right panel:* DES-Y1 data set. In the latter case, the volume effects become more pronounced, as the data has less constraining power.

Figure 6.1 shows the cosmological results after applying our new method to both LSS data sets described above. Both panels show the posterior of the matter density parameter Ω_m and the rescaled matter fluctuation parameter S_8 after marginalizing over the bias parameters using four different methods. We compare the results of the standard, brute-force marginalization with and without a Jeffreys prior (solid black line and dashed black line, respectively), with our new analytical method including both profile and Laplace terms (red area), and including only the profile term (blue shades). Orange stars denote the best-fit parameter values. In general, these results confirm our conclusions from the previous section, stating that in specific cases, volume effects are quantified by the Laplace term and can be mitigated by marginalizing over a Jeffreys prior.

In the LSST test case shown on the left panel of Fig. 6.1, we find that these conclusions hold almost exactly, and the exact brute-force method coincides perfectly with the analytical method. We find a clear shift of about 0.8σ in the posterior, caused by the volume effects associated with marginalizing over the 25 bias parameters.

The right panel of Fig. 6.1 shows the DES-Y1 case. The lower sensitivity of this data set compared to LSST should reduce its ability to constrain all 21 bias parameters, and increase the impact of volume effects. We find that the Laplace approximation starts to fail, as the corresponding contours manifest slight deviations of about 0.3σ to 0.5σ with respect to the exact marginalized constraints. At the same time, as expected, the lower sensitivity of DES causes larger volume effects, leading to shifts of up to 3σ with respect to the best-fit parameters, indicated by the orange star.

The observed difference between our analytical results and the brute-force marginalized posterior in the DES-Y1 scenario allows a preliminary validity assessment of our method. In general, as the Laplace term becomes larger (and with it the volume effects), higher-order terms tend to become relevant, leading to a growing discrepancy between the exact marginal posterior and the Laplace approximation. As a “rule of thumb”, we can therefore consider the following: if a significant ($> 1\sigma$) shift is found between the marginalized contours obtained using the profile likelihood and those obtained accounting also for the Laplace term, the approximation may start to fail, and a brute-force marginalization over the nuisance parameters using a Jeffreys prior is needed to obtain accurate results. Nonetheless, even in those cases, the Laplace approximation can be useful as an efficient, preliminary characterization of the credible regions that can then be refined by more accurate sampling methods. Another good criterion for testing the accuracy of the Laplace approximation in the case of linear parameters is to explore the difference between the Laplace approximation treatment, computed with the Hessian matrix $\mathcal{F}$ (defined in Eq. (6.5)), and the Jeffreys prior treatment, computed with the Fisher matrix F, since this difference should be non-zero only when the parameters are non-linear and would otherwise demarcate the breakdown of our approximation.

Importantly, our analytical method brings substantial improvements in the convergence time by a factor 2.7 in the LSS case (17.2 hrs for brute-force marginalization including a Jeffreys prior vs. 6.4 hrs using the Laplace approximation) and by a factor 5.6 in the DES-Y1 case (45.5 hrs vs 8.1 hrs) with none or only mild method-related bias, respectively.

6.3 Summary

The cosmic LSS holds rich information on the late-time growth of cosmic structure, complementing the cosmological information ingrained in the CMB. Future 3x2pt data analyses constrain $O(5)$ cosmological parameters but rely on marginalizing over $O(20)$ nuisance parameters that quantify "astrophysical systematics", for example galaxy bias parameters. Numerical methods, such as MCMC, take a long time to converge risk becoming a bottleneck for future LSS data analysis, which calls for more efficient methods.

In this chapter, we proposed an analytical method to efficiently marginalize over high-dimensional sets of nuisance parameters in cosmological inference. We investigated the issue of volume effects biasing cosmological parameters in the case of Gaussian data and found approximate relations between the marginal posterior in the Laplace approximation, the profile likelihood, and the exact marginal posterior using a Jeffreys prior. We applied our new method to tomographic power spectrum data in LSS analysis, where the need of marginalizing over astrophysical systematics prevents the fast inference of cosmological parameters. Our results bring a speedup by a factor of about 3 to 5 and a good agreement with exact methods. We offer a simple "rule of thumb" to help judging the accuracy of our method on a case-by-case basis.

The main limitations of our analytical marginal likelihood approximation are volume effects, its breaking down in highly nonlinear cases, and its lack of generality regarding the shape of the likelihood. The first issue, volume effects, influence the exact marginal likelihood and reside mostly in the Laplace term of the analytical approximation, but can be mitigated approximately by using the profile likelihood. We propose to use the Jeffreys prior for Gaussian data where the nuisance parameters depend on the model in a general nonlinear way. Secondly, our method may break down if the parameter dependence of the data model is highly nonlinear, which can be roughly estimated from the size of the Laplace term. Thirdly, this method only applies to Gaussian data described by model parameters that reside in the data model, and not in the covariance matrix. In the case of non-Gaussian distributions, it is often possible to apply a transformation to the nuisance parameters that Gaussianizes the posterior (e.g., via normalizing flows) without introducing any pathologies, such as singularities. This may extend the range of applications of this method and is worth investigating in future studies.

Conclusions

In this Thesis, we present state-of-the-art methods to analyze cosmological data sets, focusing on CMB polarization B-modes from the primordial Universe, CMB E-modes from reionization, and galaxy clustering statistics from the cosmic LSS. Thanks to past CMB experiments, cosmologists have, for the first time, a single model that consistently describes the time evolution of the Universe at supergalactic scales: from a hot, dense plasma filled with the fundamental particles of the standard model, all the way until the formation of the first stars, galaxies and the cosmic LSS that we observe today. Near-future CMB experiments target the first detection of polarized B-modes of primordial origin, which would serve as a powerful probe for models of the very early Universe including inflation, less than 10^{-32} seconds after the big bang.

In the current era of precision cosmology, systematic effects dominate the error budget in most cosmological data sets. The need for controlling those systematics leads to a natural tendency of growing complexity in real data sets. Examples are Galactic foregrounds that dominate polarization B-mode measurements, galaxy bias that crucially affects LSS analysis, as well as instrumental systematics. The increase in data complexity requires more sophisticated data reduction and analysis methods. The power spectrum is still a highly useful and versatile statistic and finds its application in *Planck* data analysis, the SO BB pipelines described in Chpts. 3 and 4, and LSS data analysis as described in Chpt. 6. However, in many cases, we cannot predict the broad variety of poorly known non-Gaussian systematic effects from first principles, but must infer them empirically from real data. This calls for alternative analysis strategies that tend to be less bottom-up and more adapted to a specific problem, such as phenomenological models in the style of the C_ℓ-moments model (see Chpt. 3), hybrid cleaning methods in the style of the SO map-based pipeline with dust marginalization (see Chpt. 4), or NN-based likelihood-free inference methods (see Chpt. 5).

In the following, we summarize the main results of this Thesis, and outline future prospectives.

In Chpt. 3, we introduce the BB power spectrum pipeline designed to infer primordial gravitational waves with SO. The pipeline features a parametric model of coadded BB power spectra with lensed CMB, Galactic dust, and synchrotron emission, in order to infer cosmological and foreground parameters by means of a Bayesian sampling algorithm. In the context of this Thesis, our main work consisted of the parallel implementation of this pipeline, the co-development and testing of the `BBPower` code, and the validation of the simulation-based covariance matrix.

Although the C_ℓ pipeline performs well on the simulated sky, algorithmic developments are ongoing to ensure that the complexity of real data can be controlled in a robust way and does not bias cosmological parameters. We anticipate several future refinements that account for additional systematic effects, such as the filtering of atmosphere- and ground-related contamination, or foreground-related non-Gaussianity in the power spectrum covariance matrix. Atmospheric and ground contamination requires the timestream-level filtering of the sky signal and is expected to lead to a significant loss in large-scale power, coupling of modes, as well as potential E-to-B leakage. Upcoming studies aim at capturing these effects in an empirical transfer function approach (see e.g., Sect. 7.3 of BICEP2/Keck Collaboration, 2016a). As shown in a recent study (Abril-Cabezas et al., 2023), modeling non-Gaussianity from Galactic foregrounds in the BB power spectrum covariance is expected to not affect the statistical uncertainty on r, while achieving a better accuracy of the likelihood. Other future works aim at boosting the pipeline performance, such as the inclusion of map-based delensing, the inclusion of external foreground channels to improve the cleaning performance, or the inclusion of a Jeffreys prior to mitigate volume effects during the inference. Including map-based delensing in the SO pipelines will considerably reduce the cosmic variance by partially subtracting the lensing B-modes from the conjectured primordial signal, and is expected to improve $\sigma(r)$ by at least 30% (see Sect. 4.3). This extension to the SO C_ℓ pipeline is currently under development (Hertig et al., 2023). Including external frequency channels (e.g., from *Planck* or WMAP) is crucial for SO to achieve the necessary foreground sensitivity during the first-year analysis, and requires further work, specifically on the construction of the joint data likelihood. Finally, the C_ℓ-moments parameters may lead to volume effects if the real foregrounds can only weakly constrain them. In future analyses, a Jeffreys prior, as explained in Chpt. 6, may be used to prevent these volume effects.

Chapter 4 describes the SO B-mode pipeline comparison project, assessing the performance of three component separation algorithms that target a measurement of primordial B-modes, and published in Wolz et al. (2023a). These are the SO C_ℓ pipeline in the C_ℓ-fiducial and the C_ℓ-moments model variants, a "blind" `NILC` cleaning pipeline (Basak & Delabrouille, 2013), and a parametric map-based cleaning pipeline (Stompor et al., 2008) with an optional "dust marginalization" option. We find that the three algorithms and two extensions agree in the simplest tested foreground scenarios and confirm the statistical sensitivity on r anticipated by SO Collaboration (2019). In the presence of simulated complex foregrounds, fully mitigating the bias requires extended pipeline designs, such as the C_ℓ-moments model or the dust marginalization option.

With the imminent arrival of the first data from SO, scrutinizing the robustness of a possible r

detection is crucial. Besides ensuring mutually consistent results from the three B-mode pipelines (and their extensions), more robustness tests are needed to exclude any potential source of bias on r, such as Galactic foreground residuals and unchecked instrumental systematic effects. Examples of possible robustness tests include a) comparing r values inferred on different sky regions within the SO patch as a check for signal isotropy, b) removing single channels from the SO data analysis as a check for frequency-dependent systematics, c) adding external data that overlap with the SO sky patch (e.g., WMAP, *Planck*, S-PASS (Krachmalnicoff et al., 2018), FYST/CCAT-Prime (CCAT-Prime Collaboration et al., 2023)) as a check for instrument-specific systematics, d) cross-correlating SO CMB maps with external foreground tracers (e.g., HI maps, the *Planck* 353 GHz channel, or the S-PASS 2.3 GHz channel) as checks for residual foreground contamination, or, for the same purpose, e) comparing measures of non-Gaussianity (e.g., Minkowski functionals) between SO pure-B-mode CMB maps and simulated foreground maps (e.g., Hervías-Caimapo & Huffenberger, 2022; Vansyngel et al., 2017; Zonca et al., 2021) .

Further algorithmic developments are expected to make the inference of r by the SO B-mode pipelines more robust and precise. One example is the hybrid component separation method proposed in Azzoni et al. (2023), which combines a map-based parametric component separation step with a power-spectrum-based cleaning step applying the C_ℓ-moments model on the cleaned CMB component map. Step one removes the spatially constant part of foreground emission, while the residuals expected to originate mainly from the spatial variability of foreground SEDs can be accurately cleaned in step two. The SO `NILC` pipeline may adopt a more advanced treatment of spatially varying foreground emission in order to achieve fully unbiased results on the most complex foreground scenarios. An example of such an advanced method is Multi-Clustering `NILC` (`MCNILC`, Carones et al., 2023), which performs the `NILC` variance minimization within separate sky regions that are chosen to have similar foreground B-mode emission properties.

In Chpt. 5, based on Wolz et al. (2023b), we present the first cosmological inference on CMB polarization maps that is performed entirely by neural networks, estimating the optical depth to reionization, τ, from *Planck* data, which contain spurious non-Gaussian instrumental systematics that are hard to model analytically. We approached this by a NN-based parameter estimator, which does not require an analytical data model but relies solely on simulations to train a regression model. In order to find the best training method, we started from simulated maps at a single frequency that contained coadded CMB with Gaussian correlated noise. Step by step, we then moved to more complex setups that involved two frequency channels of coadded CMB, noise, and systematic effects. We obtain unbiased results when training and testing on Gaussian simulations. When including a second channel, we observe a significant improvement on both the estimator's accuracy and its precision, which shows that NN models can effectively combine multiple data sets. Applying our method on real *Planck* data, we find $\tau_{\mathrm{NN}} = 0.0579 \pm 0.0082$ with the unbiased retrained NN model, in full agreement with literature results, albeit with a 30% larger error bar. The main limitations of our NN estimator are its intrinsic suboptimality, a slight bias on the "learnt" statistical uncertainty $\sigma_{NN}(\tau)$, and the limited size of training data that prevents the NN model from learning *Planck*'s systematics without overfitting.

Our consistent and robust results demonstrate that NNs represent a promising tool that could complement standard statistical data analysis techniques for future, systematics-dominated CMB data, such as large-scale B-modes. We anticipate that several improvements will be necessary before the NN estimator can be used to robustly estimate the tensor-to-scalar ratio r from polarized maps. First, the NN estimator may be optimized by improving the `NNhealpix` implementation of pixel-based convolution on the 2-sphere (Krachmalnicoff & Tomasi, 2019), and similarly, by optimizing the NN architecture. While we tested our τ estimator on one and two frequency channels, typical near-future experiments include $O(10)$ channels, which will require considerable efforts in training and testing new models. Second, in order to inform accurate NN estimators about potential large-scale B-mode systematics, real data sets must be studied extensively so that accurate simulations can be built. Well-tested NN pipelines will potentially be able to provide efficient and robust alternatives to established methods whenever analytical models are difficult to construct.

Next-generation CMB data analysis offers ample opportunities for prospective applications of NN-based estimators. One great potential of NN estimators, the straightforward combination of multiple data sets to perform joint analysis, may be of use for the delensing of future B-mode data with SO, which will depend on combining low-resolution SAT maps and high-resolution LAT maps. Moreover, cross-correlations between LSS and CMB data promise to set new constraints on late-time cosmology, including dark energy and the total neutrino mass, but currently rely on expensive simulations to estimate cross-probe covariances. NNs may potentially provide faster and cheaper alternatives. Analogously, the robust measurement of primordial B-modes relies upon combining data at different frequencies and from multiple experiments, which may be significantly simplified and accelerated with the aid of NN estimators. Finally, NNs could be used to assess the level of spurious systematics within clean CMB B-mode maps, for example by complementing traditional estimators when characterizing foreground-related residual non-Gaussianity.

Finally, in Chpt. 6, we describe a novel analytical method, based on the Laplace approximation and published in Hadzhiyska, Wolz, et al. (2023), to efficiently marginalize over large numbers of nuisance parameters in Bayesian inference. Accurately estimating cosmological parameters often depends on marginalizing over systematic effects that we cannot calibrate a priori, for example galaxy bias parameters in tomographic power spectrum analysis of the LSS. We apply our method to simulated and real galaxy clustering and galaxy shear data and find that our method agrees well with current standard methods, while improving on the running time by a factor of 3 to 5. As a second result of this work, we find that the Jeffreys prior in the subspace of nuisance parameters is closely related to the second-order term in the Laplace approximation. Volume effects, shifts in the marginal posterior with respect to the global posterior maximum, can be mitigated by constructing the full posterior with a Jeffreys prior. In the context of our method, this approximately corresponds to considering only the lowest-order term, also known as the "profile likelihood". We propose a "rule of thumb" to judge the validity of the Laplace approximation.

This work offers a promising perspective for the efficient marginalization of nuisance parameters. This is vital for precision cosmology, which relies upon our ability to keep systematic effects under ever-tighter control, requiring growing numbers of nuisance parameters. In particular, combining multiple cosmological probes that each might introduce their own set of respective nuisance parameters requires fast inference algorithms. Our method could, for example, be useful in cross-correlation studies of the CMB and the LSS to assess late-time observables, such as the total mass of neutrinos. As a current limitation, this method only applies to Gaussian data described by model parameters that reside in the data model, and not in the covariance matrix. In the case of non-Gaussian distributions, it is often possible to apply a transformation to the nuisance parameters that Gaussianizes the posterior (e.g., via normalizing flows) without introducing unwanted behavior, such as singularities. This may extend the range of applications of this method and is worth investigating in future studies.

Many challenges and chances lie ahead of us in today's era of precision cosmology. Future data will be of unparalleled precision and their accuracy will be subject to systematic effects. Controlling these effects is challenging, and requires ever-more flexible analysis frameworks and problem-tailored solutions. Examples are CMB foreground cleaning pipelines targeting primordial B-modes, NN-based parameter estimators for the optical depth τ, or phenomenological models that quantify galaxy bias in analyses of the cosmic LSS. The anticipated diversity of estimators, models, and analysis pipelines calls for a large variety of validation strategies. Going forward, it will be crucial to devise tests that assess the robustness of those new methods, make them easier to interpret, and help with selecting the right model for a given task.

The coming decade will see cosmological experiments that may allow us to tackle some of the great open questions in cosmology. Examples are the nature of dark energy and dark matter, which may be constrained by upcoming experiments that measure the cosmic LSS and galaxy clustering properties, such as DESI (DESI Collaboration, 2016), the Vera C. Rubin Observatory (Ivezić et al., 2019), and Euclid (Amendola et al., 2018), in combination with CMB experiments, such as SO and CMB-S4 (Abazajian et al., 2019). Those experiments might also be able to shed light on the apparent inconsistency of some standard-model cosmological parameters (the Hubble constant,

H_0, or the rescaled matter fluctuation parameter, S_8) between measurements of early- and late-time probes. Finally, high-sensitivity CMB polarization experiments such as SO, LiteBIRD (LiteBIRD Collaboration, 2023), and CMB-S4, may be able to find evidence for the inflationary paradigm and further constrain the space of viable models, possibly exploring the hitherto unknown particle content of the Universe within the first 10^{-32} seconds after the big bang.

APPENDIX A

Chi-squared goodness of fit

As a particularly versatile, simple, and fast to evaluate robustness check, the χ^2 goodness-of-fit test is suited to detect biases either at the sky model level, the estimator stage, or the likelihood level. For general Gaussian data with known covariance, the χ^2 test assesses the statistical agreement among the data model $\boldsymbol{m}(\boldsymbol{\theta})$, the data $\boldsymbol{d}$ and the inverse covariance Σ^{-1}. It is defined as

$$\chi^2(\boldsymbol{d}|\boldsymbol{\theta}) \equiv [\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{\theta})]^T \Sigma^{-1} [\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{\theta})] \,, \tag{A.1}$$

which, for Gaussian data, equals twice the negative log-likelihood, $-2\ \ln(L(\boldsymbol{d}|\boldsymbol{\theta}))$ (minus a constant term $-\ln\det(\Sigma)$). Pearson (1900) and Cramér (1946) showed that the minimum χ^2 statistic follows a χ^2 distribution with $N - P$ degrees of freedom,

$$T(\boldsymbol{d}) \equiv \min_{\boldsymbol{\theta}} \chi^2(\boldsymbol{d}|\boldsymbol{\theta}) \sim f_{N-P} \,, \tag{A.2}$$

where N and P are the dimensions of data and parameters, respectively. Note that this statement is independent of the functional form of the parameter model, and t is quickly evaluated by numerical methods, such as gradient descent. When testing set of simulations $\{\boldsymbol{d}^{(i)}\}$, we can compute the empirical probabilities to exceed, PTE $= \int_{t(\{\boldsymbol{d}^{(i)}\})} f_{N-P}(t')\,\mathrm{d}t'$, which converge to a uniform distribution $U(0,1)$ under the null hypothesis. We can use the minimum χ^2 method to test a model's goodness of fit, or the accuracy of a given inverse covariance matrix as compared to a set of simulated data. On the other hand, the χ^2 goodness of fit is not very restrictive for models with many parameters, in which case it generally allows for large variations in the data while still fitting reasonably well. This is an undesirable feature when models with different degrees of freedom shall be compared, in which case measures such as the AIC (Akaike, 1974), the Bayesian Information Criterion (BIC, Schwarz, 1978), or Bayesian model comparison methods (Jeffreys, 1939; Trotta, 2008) might be better alternatives.

BIBLIOGRAPHY

Abazajian, K. et al. 2019, arXiv e-prints, arXiv:1907.04473

Abitbol, M. H. et al. 2021, J. Cosmol. Astropart. Phys., 2021, 032

Abril-Cabezas, I. et al. 2023, in prep.

Akaike, H. 1974, IEEE Trans. Automat. Contr., 19, 716

Ali, A. M. et al. 2020, J. Low Temp., 200, 461

Alonso, D., Bull, P., Ferreira, P. G., Maartens, R., & Santos, M. G. 2015, Astrophys. J., 814, 145

Alonso, D., Sanchez, J., Slosar, A., & LSST Dark Energy Science Collaboration. 2019, Mon. Not. R. Astron. Soc., 484, 4127

Amendola, L., Appleby, S., Avgoustidis, A., et al. 2018, Living Rev. Relativ., 21, 2

Ashton, P. C. et al. 2018, Astrophys. J., 857, 10

Azzoni, S., Abitbol, M., Alonso, D., et al. 2021, J. Cosmol. Astropart. Phys., 2021, 047

Azzoni, S., Alonso, D., Abitbol, M. H., Errard, J., & Krachmalnicoff, N. 2023, J. Cosmol. Astropart. Phys., 2023, 035

Basak, S. & Delabrouille, J. 2012, Mon. Not. R. Astron. Soc., 419, 1163

Basak, S. & Delabrouille, J. 2013, Mon. Not. R. Astron. Soc., 435, 18

Battaglia, N., Ferraro, S., Schaan, E., & Spergel, D. N. 2017, J. Cosmol. Astropart. Phys., 2017, 040

Benabed, K., Cardoso, J. F., Prunet, S., & Hivon, E. 2009, Mon. Not. R. Astron. Soc, 400, 219

Bennett, C. L., Hill, R. S., Hinshaw, G., et al. 2003, Astrophys. J., Suppl. Ser., 148, 97

BICEP2 Collaboration. 2014, Phys. Rev. Lett., 112, 241101

BICEP2/Keck Collaboration. 2016a, Astrophys. J, 825, 66

BICEP2/Keck Collaboration. 2016b, Phys. Rev. Lett., 116, 031302

BICEP2/Keck Collaboration. 2018, Phys. Rev. Lett., 121, 221301

BICEP2/Keck Collaboration. 2021, Phys. Rev. Lett., 127, 151301

BICEP2/Keck Collaboration. 2022, arXiv e-prints, arXiv:2203.16556

BICEP2/Keck Collaboration & Planck Collaboration. 2015, Phys. Rev. Lett., 114, 101301

Bond, J. R., Kofman, L., & Pogosyan, D. 1996, Nature, 380, 603

Bunn, E. F. 2011, Phys. Rev. D, 83, 083003

Calabrese, E., Hložek, R., Battaglia, N., et al. 2014, J. Cosmol. Astropart. Phys., 2014, 010

Caldeira, J., Wu, W. L. K., Nord, B., et al. 2019, Astron. Comput., 28, 100307

Campeti, P. & Komatsu, E. 2022, Astrophys. J., 941, 110

Cardoso, J.-F., Martin, M., Delabrouille, J., Betoule, M., & Patanchon, G. 2008, arXiv e-prints, arXiv:0803.1814

Carlstrom, J. et al. 2019, in Bull. Am. Astron. Soc., Vol. 51, 209

Carones, A., Migliaccio, M., Puglisi, G., et al. 2023, Mon. Not. R. Astron. Soc., 525, 3117

Casas, J. M., Bonavera, L., González-Nuevo, J., et al. 2022, Astron. Astrophys., 666, A89

CCAT-Prime Collaboration, Aravena, M., Austermann, J. E., et al. 2023, Astrophys. J., Suppl. Ser., 264, 7

Chluba, J., Hill, J. C., & Abitbol, M. H. 2017, Mon. Not. R. Astron. Soc., 472, 1195

Choi, S. K. & Page, L. A. 2015, J. Cosmol. Astropart. Phys., 2015, 020

Choi, S. K. et al. 2020, J. Cosmol. Astropart. Phys., 2020, 045

Cole, S. R., Chu, H., & Greenland, S. 2013, Am. J. Epidemiol., 179, 252

Coulton, W., Ota, A., & van Engelen, A. 2020, Phys. Rev. Lett., 125, 111301

Coulton, W. R. & Spergel, D. N. 2019, J. Cosmol. Astropart. Phys., 2019, 056

Cramér, H. 1946, Mathematical Methods of Statistics (PMS-9), Volume 9 (Princeton: Princeton University Press)

Crowley, K. T. et al. 2018, Proc. SPIE Int. Soc. Opt. Eng., 10708, 107083Z

de Belsunce, R., Gratton, S., Coulton, W., & Efstathiou, G. 2021, Mon. Not. Roy. Astron. Soc., 507, 1072

de la Hoz, E. et al. 2023, Mon. Not. Roy. Astron. Soc., 519, 3504

Delabrouille, J. & Cardoso, J. F. 2007, arXiv e-prints, astro

Delabrouille, J., Cardoso, J. F., Le Jeune, M., et al. 2009, Astron. Astrophys., 493, 835

Delouis, J.-M., Allys, E., Gauvrit, E., & Boulanger, F. 2022, Astron. Astrophys., 668, A122

Delouis, J. M., Pagano, L., Mottet, S., Puget, J. L., & Vibert, L. 2019, Astron. Astrophys., 629, A38

DES Collaboration. 2018, Phys. Rev. D, 98, 043526

DES Collaboration. 2022, Phys. Rev. D, 105, 023520

DESI Collaboration. 2016, arXiv e-prints, arXiv:1611.00036

Draine, B. T. & Hensley, B. S. 2021, Astrophys. J., 919, 65

Dunkley, J., Calabrese, E., Sievers, J., et al. 2013, J. Cosmol. Astropart. Phys., 2013, 025

Durrer, R. 2008, The Cosmic Microwave Background (Cambridge University Press)

Efstathiou, G. 2004, Mon. Not. R. Astron. Soc., 349, 603

Efstathiou, G. 2006, Mon. Not. R. Astron. Soc., 370, 343

Eicker, F. 1966, Ann. Math. Stat., 37, 1825

Einstein, A. 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften, 142

Elvin-Poole, J., Crocce, M., Ross, A. J., et al. 2018, Phys. Rev. D, 98, 042006

Eriksen, H. K., Banday, A. J., Gorski, K. M., & Lilje, P. B. 2004, Astrophys. J., 612, 633

Eriksen, H. K., Jewell, J. B., Dickinson, C., et al. 2008, Astrophys. J., 676, 10

Feldman, G. J. & Cousins, R. D. 1998, Phys. Rev. D, 57, 3873

Fernández-Cobos, R., Marcos-Caballero, A., Vielva, P., Martínez-González, E., & Barreiro, R. B. 2016, Mon. Not. R. Astron. Soc., 459, 441

Fluri, J., Kacprzak, T., Lucchi, A., et al. 2019, Phys. Rev. D, 100, 063514

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, Publ. Astron. Soc. Pac., 125, 306

Freese, K., Frieman, J. A., & Olinto, A. V. 1990, Phys. Rev. Lett., 65, 3233

Friedmann, A. 1922, Zeitschrift für Physik, 10, 377

Gallardo, P. A. et al. 2018, Proc. SPIE Int. Soc. Opt. Eng., 10708, 107083Y

García-García, C., Ruiz-Zapatero, J., Alonso, D., et al. 2021, J. Cosmol. Astropart. Phys., 2021, 030

Gerbino, M., Lattanzi, M., Migliaccio, M., et al. 2020, Front Phys, 8, 15

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, Astrophys. J., 622, 759

Grain, J., Tristram, M., & Stompor, R. 2009, Phys. Rev. D, 79, 123515

Guillet, V., Fanciullo, L., Verstraete, L., et al. 2018, Astron. Astrophys., 610, A16

Guth, A. H. 1981, Phys. Rev. D, 23, 347

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, Astrophys. J., 622, 759

Hadzhiyska, B., Wolz, K., Azzoni, S., et al. 2023, Open J. of Astrophysics, 6, 23

Hamana, T., Shirasaki, M., Miyazaki, S., et al. 2020, Publications of the Astronomical Society of Japan, 72, 16

Hamann, J., Hannestad, S., Raffelt, G. G., & Wong, Y. Y. Y. 2007, J. Cosmol. Astropart. Phys., 2007, 021

Hamimeche, S. & Lewis, A. 2008, Physical Review D, 77

Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015a, Mon. Not. R. Astron. Soc., 450, L61

Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015b, Mon. Not. R. Astron. Soc., 453, 4384

Hansen, F. K. & Gorski, K. M. 2003, Mon. Not. Roy. Astron. Soc., 343, 559

Harper, S. E., Dickinson, C., Barr, A., et al. 2022, Mon. Not. R. Astron. Soc., 513, 5900

Hazumi, M. et al. 2020, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 11443, Space Telescopes and Instrumentation 2020: Optical, Infrared, and Millimeter Wave, ed. M. Lystrup & M. D. Perrin, 114432F

Hensley, B. S. et al. 2022, Astrophys. J., 929, 166

Herold, L., Ferreira, E. G. M., & Komatsu, E. 2022, Astrophys. J.l, 929, L16

Hertig, E. et al. 2023, in prep.

Hervías-Caimapo, C. & Huffenberger, K. M. 2022, Astrophys. J., 928, 65

Heymans, C., Tröster, T., Asgari, M., et al. 2021, Astron. Astrophys., 646, A140

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, Astrophys. J., Suppl. Ser., 208, 19

Hivon, E., Górski, K. M., Netterfield, C. B., et al. 2002, Astrophys. J., 567, 2

Howlett, C., Lewis, A., Hall, A., & Challinor, A. 2012, J. Cosmol. Astropart. Phys., 1204, 027

Hoyle, B., Gruen, D., Bernstein, G. M., et al. 2018, Mon. Not. R. Astron. Soc., 478, 592

Hu, W. 2003, ICTP Lect. Notes Ser., 14, 145

Hu, W. & White, M. 1997, New Astron., 2, 323

Huff, E. & Mandelbaum, R. 2017, arXiv e-prints, arXiv:1702.02600

Hüllermeier, E. & Waegeman, W. 2021, Machine Learning, 110, 457

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, Astrophys. J., 873, 111

Jeffrey, N., Boulanger, F., Wandelt, B. D., et al. 2022, Mon. Not. Roy. Astron. Soc., 510, L1

Jeffrey, N. & Wandelt, B. D. 2020, in 34th Conference on Neural Information Processing Systems

Jeffreys, H. 1939, Theory of Probability (Oxford, England: Clarendon Press)

Jeffreys, H. 1946, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 186, 453

Kamionkowski, M., Kosowsky, A., & Stebbins, A. 1997, Phys. Rev. D, 55, 7368

Kaplinghat, M., Knox, L., & Skordis, C. 2002, Astrophys. J., 578, 665

Kesden, M., Cooray, A., & Kamionkowski, M. 2002, Phys. Rev. Lett., 89, 011304

Kingma, D. P. & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., et al. 2017, Proceedings of the National Academy of Science, 114, 3521

Knox, L. 1997, Astrophys. J., 480, 72

Kodwani, D., Alonso, D., & Ferreira, P. 2019, Open J. Astrophys., 2, 3

Kogut, A., Spergel, D. N., Barnes, C., et al. 2003, Astrophys. J., Suppl. Ser., 148, 161

Krachmalnicoff, N., Baccigalupi, C., Aumont, J., Bersanelli, M., & Mennella, A. 2016, Astron. Astrophys., 588, A65

Krachmalnicoff, N., Carretti, E., Baccigalupi, C., et al. 2018, Astron. Astrophys., 618, A166

Krachmalnicoff, N. & Puglisi, G. 2021, Astrophys. J., 911, 42

Krachmalnicoff, N. & Tomasi, M. 2019, Astron. Astrophys., 628, A129

Leach, S. M., Cardoso, J. F., Baccigalupi, C., et al. 2008, Astron. Astrophys., 491, 597

Lemaître, A. G. 1931, Mon. Not. R. Astron. Soc., 91, 483

Lesgourgues, J. 2011, arXiv e-prints, arXiv:1104.2932

Lewis, A. & Challinor, A. 2006, Phys. Rep., 429, 1

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

Lewis, A., Challinor, A., & Turok, N. 2002, Phys. Rev. D, 65, 023505

Linde, A. D. 1983, Phys. Lett. B, 129, 177

Linde, A. D. 2008, Lect. Notes Phys., 738, 1

LiteBIRD Collaboration. 2023, Progress of Theoretical and Experimental Physics, 2023, 042F01

LSST Dark Energy Science Collaboration. 2012, arXiv e-prints, arXiv:1211.0310

MacCrann, N., Zuntz, J., Bridle, S., Jain, B., & Becker, M. R. 2015, Mon. Not. R. Astron. Soc., 451, 2877

Martin, J., Ringeval, C., & Vennin, V. 2014, Physics of the Dark Universe, 5, 75

Mather, J. C., Fixsen, D. J., Shafer, R. A., Mosier, C., & Wilkinson, D. T. 1999, ApJ, 512, 511

McAllister, L., Silverstein, E., & Westphal, A. 2010, Phys. Rev. D, 82, 046003

Mortlock, D. J., Challinor, A. D., & Hobson, M. P. 2002, Mon. Not. Roy. Astron. Soc., 330, 405

Namikawa, T. et al. 2022, Phys. Rev. D, 105, 023511

Narcowich, F. J., Petrushev, P., & Ward, J. D. 2006, SIAM J. Math. Anal., 38, 574

Natale, U., Pagano, L., Lattanzi, M., et al. 2020, Astron. Astrophys., 644, A32

Nicola, A., García-García, C., Alonso, D., et al. 2021, J. Cosmol. Astropart. Phys., 2021, 067

Nicola, A., Hadzhiyska, B., Findlay, N., et al. 2023, arXiv e-prints, arXiv:2307.03226

Okamoto, T. & Hu, W. 2003, Phys. Rev. D, 67, 083002

Pagano, L., Delouis, J. M., Mottet, S., Puget, J. L., & Vibert, L. 2020, Astron. Astrophys., 635, A99

Page, L., Hinshaw, G., Komatsu, E., et al. 2007, Astrophys. J., Suppl. Ser., 170, 335

Pearson, K. 1900, Lond. Edinb. Dublin philos. mag. j. sci., 50, 157

Peebles, P. J. E. 1993, Principles of Physical Cosmology (Princeton: Princeton University Press)

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

Planck Collaboration I. 2016, Astron. Astrophys., 594, A1

Planck Collaboration I. 2020, Astron. Astrophys., 641, A1

Planck Collaboration Int. LVII. 2020, Astron. Astrophys., 643, A42

Planck Collaboration Int. XLVI. 2016, Astron. Astrophys., 596, A107

Planck Collaboration Int. XXX. 2016, Astron. Astrophys., 586, A133

Planck Collaboration Int. XXXII. 2016, Astron. Astrophys., 586, A135

Planck Collaboration IV. 2020, Astron. Astrophys., 641, A4

Planck Collaboration IX. 2020, Astron. Astrophys., 641, A9

Planck Collaboration V. 2020, Astron. Astrophys., 641, A5

Planck Collaboration VI. 2014, Astron. Astrophys., 571, A6

Planck Collaboration VI. 2020, Astron. Astrophys., 641, A6, [Erratum: Astron. Astrophys. 652, C4 (2021)]

Planck Collaboration VII. 2020, Astron. Astrophys., 641, A7

Planck Collaboration VIII. 2020, Astron. Astrophys., 641, A8

Planck Collaboration X. 2016, Astron. Astrophys., 594, A10

Planck Collaboration X. 2020, Astron. Astrophys., 641, A10

Planck Collaboration XI. 2016, Astron. Astrophys., 594, A11

Planck Collaboration XI. 2020, Astron. Astrophys., 641, A11

Planck Collaboration XII. 2016, Astron. Astrophys., 594, A12

Planck Collaboration XIII. 2016, Astron. Astrophys., 594, A13

Planck Collaboration XXV. 2016, Astron. Astrophys., 594, A25

Planck Collaboration XXX. 2016, Astron. Astrophys., 586, A133

POLARBEAR Collaboration. 2014, Phys. Rev. Lett., 113, 021301

POLARBEAR Collaboration. 2017, Astrophys. J., 848, 121

Poletti, D. & Errard, J. 2023, FGBuster: Parametric component separation for Cosmic Microwave Background observations, Astrophysics Source Code Library, record ascl:2307.021

Press, W. H. & Schechter, P. 1974, Astrophys. J., 187, 425

Qin, Y., Mesinger, A., Bosman, S. E. I., & Viel, M. 2021, Mon. Not. Roy. Astron. Soc., 506, 2390

Ramasesh, V. V., Lewkowycz, A., & Dyer, E. 2021, in International Conference on Learning Representations

Rees, M. J. & Sciama, D. W. 1968, Nature, 217, 511

Reichardt, C. L. 2016, in Astrophysics and Space Science Library, Vol. 423, Understanding the Epoch of Cosmic Reionization: Challenges and Progress, ed. A. Mesinger, 227

Riotto, A. 2003, ICTP Lect. Notes Ser., 14, 317

Robertson, H. P. 1935, Astrophys. J., 82, 284

Rubiño-Martín, J. A., Rebolo, R., Tucci, M., et al. 2010, in Astrophys. Space Sci. Proc., Vol. 14, Highlights of Spanish Astrophysics V, 127

Rybicki, G. B. & Lightman, A. P. 1985, Radiative Processes in Astrophysics (New York, NY: Wiley)

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

Salatino, M. et al. 2018, in SPIE Astronomical Telescopes + Instrumentation 2018

Sayre, J. T., Reichardt, C. L., Henning, J. W., & SPTpol Collaboration. 2020, Phys. Rev. D, 101, 122003

Schwarz, G. 1978, Ann Stat, 6, 461

Seljak, U. & Zaldarriaga, M. 1997, Phys. Rev. Lett., 78, 2054

Sheldon, E. S. & Huff, E. M. 2017, Astrophys. J., 841, 24

Smith, K. M. 2006, Phys. Rev. D, 74, 083002

Smith, K. M., Hanson, D., LoVerde, M., Hirata, C. M., & Zahn, O. 2012, J. Cosmol. Astropart. Phys., 2012, 014

Smith, K. M. & Zaldarriaga, M. 2007, Phys. Rev. D, 76, 043001

Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, Mon. Not. R. Astron. Soc., 341, 1311

SO Collaboration. 2019, J. Cosmol. Astropart. Phys., 2019, 056

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, J Mach Learn Res, 15, 1929

Starobinskij, A. A. 1992, JETP Lett., 55, 489

Starobinsky, A. 1980, Phys. Lett. B, 91, 99

Stompor, R., Leach, S., Stivoli, F., & Baccigalupi, C. 2008, Mon. Not. R. Astron. Soc., 392, 216

Sunyaev, R. A. & Zeldovich, I. B. 1980, Annu. Rev. Astron., 18, 537

Tanimura, H., Aghanim, N., Bonjean, V., & Zaroubi, S. 2022, Astron. Astrophys., 662, A48

Tegmark, M. 1998, Astrophys. J., 502, 1

Tegmark, M. & de Oliveira-Costa, A. 2001, Phys. Rev. D, 64, 063001

Tegmark, M. & Efstathiou, G. 1996, Mon. Not. R. Astron. Soc., 281, 1297

The LSST Dark Energy Science Collaboration et al. 2018, arXiv e-prints, arXiv:1809.01669

Thorne, B., Dunkley, J., Alonso, D., & Næss, S. 2017, Mon. Not. R. Astron. Soc., 469, 2821

Torrado, J. & Lewis, A. 2019, Cobaya: Bayesian analysis in cosmology, Astrophysics Source Code Library, record ascl:1910.019

Torrado, J. & Lewis, A. 2021, J. Cosmol. Astropart. Phys., 2021, 057

Tristram, M., Banday, A. J., Górski, K. M., et al. 2022, Phys. Rev. D, 105, 083524

Trotta, R. 2008, Contemp. Phys., 49, 71

Tsujikawa, S. 2003, in 2nd Tah Poe School on Cosmology: Modern Cosmology

Vansyngel, F., Boulanger, F., Ghosh, T., et al. 2017, Astron. Astrophys., 603, A62, [Erratum: Astron. Astrophys. 618, C4 (2018)]

Villaescusa-Navarro, F., Genel, S., Anglés-Alcázar, D., et al. 2022, Astrophys. J., Suppl. Ser., 259, 61

Wagenmakers, E. & Farrell, S. 2004, Psychon Bull Rev., 11, 192

Wald, R. M. 1984, General Relativity (Chicago, USA: Chicago Univ. Pr.)

Walker, A. G. 1937, Proc. London Math. Soc., 42, 90

Wang, G.-J., Shi, H.-L., Yan, Y.-P., et al. 2022, Astrophys. J., Suppl. Ser., 260, 13

White, M. J. & Hu, W. 1997, Astron. Astrophys., 321, 8

Wishart, J. 1928, Biometrika, 20A, 32

Wolz, K., Azzoni, S., Hervias-Caimapo, C., et al. 2023a, arXiv e-prints, arXiv:2302.04276

Wolz, K., Krachmalnicoff, N., & Pagano, L. 2023b, Astron. Astrophys., 676, A30

Zaldarriaga, M. & Seljak, U. 1997, Phys. Rev. D, 55, 1830

Zonca, A., Thorne, B., Krachmalnicoff, N., & Borrill, J. 2021, J. Open Source Softw., 6, 3783

Zuntz, J., Sheldon, E., Samuroff, S., et al. 2018, Mon. Not. R. Astron. Soc., 481, 1149