SISSA
Scuola
Internazionale
Superiore di
Studi Avanzati

Physics Area - PhD course in
Astrophysics and Cosmology

Foreground Challenge to CMB Polarization: Present Methodologies and New Concepts

Candidate:
Farida Farsian

Advisors:
Nicoletta Krachmalnicoff
Carlo Baccigalupi

Academic Year 2019-20

Abstract

In this thesis, I focus on the issue of contamination to the polarization of the Cosmic Microwave Background (CMB) anisotropies from diffuse Galactic foregrounds, which is known to be one of the greatest challenges to the detection of the curl (B) modes of the signal, which might be sourced by cosmological gravitational waves. I take parallel approaches along these lines. I apply the most recent techniques capable of parametrizing, fitting, and removing the main known Galactic foregrounds in a multi-frequency CMB dataset to one of the forthcoming powerful CMB polarization experiments, the Large Scale Polarization Explorer (LSPE). I presented the result of the complete simulation done for the parametric component separation pipeline of this experiment. On the other hand, I explored the latest Machine Learning and Artificial Intelligence algorithms and their application in CMB data analysis, specifically component separation and foreground cleaning. I start the investigation of the relevance of Neural Networks (NNs) in the understanding of the physical properties of foregrounds, as it is necessary before the foreground removal layer, by implementing a novel algorithm, which I test on simulated data from future B-mode probes. The results of the implemented NN's prediction in discerning the correct foreground model address the high accuracy and suitability of this model as a preceding stage for the component separation procedure. Finally, I also investigate how different NNs, as a generative model, could be used for reconstructing CMB anisotropies where the removal is impossible, and data have to be abandoned in the analysis. Lots remain to be done along each of these three investigations, which have been published in scientific journals, and constitute the basis of my future research.

Acknowledgements

First of all, I would like to thank my supervisors Prof. Carlo Baccigalupi and Dr. Nicoletta Krachmalnicoff for their advice, support and the technical and fruitful discussions. They taught me very valuable skills such as the scientific research method, professional and concise writing, and problem solving methodologies. I would like to give my special thanks to Dr. Davide Poletti for the useful discussions.

I would also like to express my gratitude to my parents, who always encouraged me to follow the path of science and knowledge. To my dad who thought me what is critical thinking and how to be curious about all the phenomena that take place around us. He showed me how being able to ask questions is one of the most important tools in life. To my mom who has always been able to support me and give me self-confidence with her deep emotions and whose words always relieved me from all the concerns.

My serene thanks go to my dear azizam, Francesco Franchina, with whom I had many technical discussions about my thesis. Not only he gave his best to support me emotionally but he was a good company from which learnt his precise way of looking at the problems.

I would also like to thank my friends, colleagues, and Cofficemates, Dr. Tommaso Ronconi (Tommino), Dr. Paolo Campeti (Campeti power), Hasti Khoraminezhad (Hasti Banoo), Dr. Gabriele Parimbelli (Grandpa Gabri), Matteo Nurisso (Nurissimo), Dr. Riccardo Murgia (Rikixeddu), with whom I shared a lot of joyful moments and that were next to me in hard times, on whom I could always rely on and feel home even kilometres far from home. And special thanks to all the APC group members who helped in creating a very friendly and active environment. I sincerely thank my dear friends of the "Cosmology circle", with which we started to study cosmology seriously all together in Iran, to their support and advocacy I owe part of my passion and progresses, Dr.Mona Jalilvand, Golshan Ejlali, Dr.Aghile Ebrahimi, Dr.Farbod Hasani, Dr.Alireza Vafai Sadr, Abbas Khanbeigi, Mohammadreza Ayramlou, Amirnezam Amiri and Niloofar Khorshid.

Contents

Chapter 1

Introduction

The Cosmic microwave background (CMB) is made of photons that decoupled from the hot and dense phase of the early Universe, about ∼380,000 years after the Big Bang. On the basis of the knowledge that we have at this moment, the CMB represents one of the main pillars of the standard cosmological model. Since its discovery, in the middle 60s of the past centuries, the CMB and its anisotropies have been observed by satellites, balloons, and ground-based telescopes [see 1–5, and references therein]. These observations shed light on the different aspects of our understanding of the evolution and composition of the Universe, from small to large cosmological scales. Several mysteries remain, such as the quasi-exponential expansion due to transient vacuum energy, an era known as Inflation, in the very early Universe, at energy not accessible by ordinary or foreseen laboratories. The CMB radiation is predicted [6] and found to be linearly polarized, [4, 5, 7] and its pattern can be decomposed into curl-free, and divergence-free quantities called E- and B-modes, respectively [8]. Primordial Gravitational Waves (GWs) produced by the Inflationary era in the early Universe are sources of the CMB B-mode anisotropies, and represent the main observational target of ongoing and future CMB probes [see 9, and references therein]. The GW contribution to B-modes, parametrized by its amplitude relative to primordial scalar perturbations, the tensor-to-scalar ratio, r, induces anisotropies at the degree and super-degree scale. To date, there is no detection for r, while only upper limit exist, $r < 0.044$ (at 95% confidence level) [10].

The search for primordial GWs is complicated by the fact that the polarized emissions from the Galaxy are comparable or dominant with respect to the CMB B-modes signal in all the frequencies and all over the sky, representing one of the main obstacles in observing CMB polarization anisotropies [11–13]. This occurrence has stimulated the creation of a new layer of CMB data analysis algorithms, the component separation, which has the goal of subtracting the Galactic foregrounds, with confidence and accurately estimated errors, despite of the complexities of the Galactic signal which is removed from the multi-frequency datasets.

In this Thesis, we push forward several fronts along this line of investigation. First, we consider the most advanced component separation procedures in order to assess their predicted performance in one of the forthcoming B-mode probe, the Large Scale Polarization Explorer (LSPE). It follows from our analysis that the LSPE will be able to set an upper limit for tensor-to-scalar ratio r at the level of 10^{-2} and detect a $r = 0.03$ with 95% confidence limits. In parallel, and looking at the challenges described earlier, we are motivated to investigate innovative algorithms based on Neural Networks (NNs) as a subset of Machine Learning (ML) and Artificial Intelligence (AI). We start the construction of a new layer of analysis of multi-frequency CMB datasets, to be exploited prior to component separation, designed to discern the right physical parametrization of foregrounds when the polarized Galactic emissions vary across the sky: we introduce the implement a NN model for the foreground recognition in the context of CMB B-mode data analysis for the first time. The established model is able to classify the foreground models correctly, in the absence or presence of the noise, with higher than 90% accuracy. Moreover, in Comparison with χ^2 information shows advantages in terms of accuracy. Finally, we consider a different algorithm based on NNs, the so-called Generative Adversarial Networks (GANs), in order to check the possible potential for generating the CMB signal when it is missing due to harsh removal due to a high level of contamination. We quantify how much these techniques are able to fill-in the CMB signal in the masked regions due to point sources.

The Thesis is organized as follows: we review the fundamental physics of the CMB radiation in Chapter 2. In Chapter 3, we describe the contamination of the CMB B-mode polarization signal, discuss the different Galactic emissions that play the dominant role in this occurrence, and set the current status of the CMB B-mode observations, as well as the expectations from future probes. Chapter 4 is dedicated to the component separation forecasts for the LSPE experiment,

using the most advanced component separation techniques. In Chapter 5, we illustrate the NN algorithm, which is able to discern different CMB foreground models in different regions of the sky, as a pre-processing phase with respect to the component separation pipeline in terms of accuracy and efficiency as a foreground classifier. In Chapter 6, we study the GAN in the context of reconstructing the CMB signal where is missing due to point sources removal. The latter three Chapters have been taken from an equal number of original papers. Each of them represents the starting point for future investigations, as we stress in the relevant Chapter, as well as in Chapter 7, where we summarize the results obtained in the Thesis and indicate the desired forthcoming steps along each of these lines.

Chapter 2

Physics of the Cosmic Microwave Background

The purpose of this Chapter is to introduce the scientific context in which the work for this Thesis has been carried out. First, we highlight the standard model of Cosmology in Section 2.1. We explain how the CMB is generated, and also how it has been observed across decades, in Section 2.1.1. After that. in Section 2.1.2 we describe the inflationary paradigm for introducing the science case which motivates the data analysis efforts presented in the following Chapters. Finally, Section 2.2 reviews the sources of CMB anisotropies and their scientific relevance, besides the direct imprint of inflationary perturbations.

2.1 The standard cosmological model

The current Standard Model of Cosmology, also called the "Concordance Cosmological Model", is based on the "Big Bang" theory, which assumes that the Universe started from a very hot and dense state, 13.8 billions years ago. This model predicts that the Universe has been expanding over time with an expansion rate depending on the types of matter and energy in it. The first evidence of cosmological expansion has been obtained by Hubble [14, 15], observing the recession velocity of nearby galaxies. He discovered that galaxies are moving away from each other with speed proportional to their distance; therefore, the spectrum of light emitted by them is redshifted. In other words, he observed that the light emitted by galaxies has progressively more reddening because of the Doppler

effect, increasing with their distance. Based in this observation, he formulated what it is known today as the Hubble low, $v = H_0 D$, where v and D are the velocity and proper distance from galaxy to the observer respectively, measured typically in Mega Parsec (Mpc); H_0 is the Hubble's constant where the subscript 0 shows the value of H (Hubble parameter) at the time of observation. Two principles state that the Universe is homogeneous and isotropic at cosmological scales, corresponding to about hundreds of Mpc or more, which means that there is no preferred location or direction in any observable we may access. Both expansion and Large Scale Structure (LSS) formation follow the Einstein's General Relativity (GR). The cosmological metric in GR consists of 4 dimensions, 3 spatial dimensions plus time explaining expansion on large scales, as well as LSS formation following local curvature due to gravitational collapse. By considering the cosmological principles and GR, the four-dimensional space-time can be written as Friedmann-Lemaitre-Robertson-Walker (FLRW) metric [16]:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + sin^2\theta d\phi^2)\right] . \tag{2.1}$$

Where (r, θ, ϕ) are comoving (i.e. at rest with respect to the cosmological expansion) coordinates and t is the cosmological proper time, $a(t)$ is known as scale factor which is a dimensionless quantity describing the cosmological expansion, usually assumed to have the value $a_0 = a(t_0) = 1$ at the present time t_0. K is the curvature parameter, assuming positive, negative, or a null value, meaning respectively that the metric is close, open, or flat.

The standard model of cosmology is also known as the "ΛCDM Model", where Λ refers to the Cosmological constant and CDM stand for Cold Dark Matter [17]. The cosmological constant is a particular case of the broad range of models corresponding to the Dark Energy (DE), responsible for the accelerated expansion of the Universe, which was discovered by observation of supernovae type Ia by Riess et al. [18], Perlmutter et al. [19]. Instead, the CDM is referred to non-relativistic and non-baryonic particles that move slowly respect to the speed of light (*cold*) and interact with ordinary matter and electromagnetic radiation very weakly. The existence of CDM was proposed to describe the shape of the galaxy rotation curves [see e.g. 20, and references therein], the galaxy cluster dynamics [21], and several other processes associated to LSS.

The expansion of the Universe as we described before is described by the Friedmann equations, determining the behavior of the scale factor $a(t)$ in relation to the matter and energy content of the Universe, as well as its curvature:

$$H^2 = \frac{8\pi G}{3}\rho_{tot} - \frac{K}{a^2}\,. \tag{2.2}$$

G is the Netwon gravitational constant, the Hubble parameter $H \equiv \frac{1}{a}\frac{da}{dt}$ explains the expansion of universe and $\rho_{tot} \equiv \Sigma_i \rho_i$ is the total energy density, summed over each ρ_i indicating the energy density of each cosmological component. The current value for the Hubble parameter is $H_0 = 100.hkms^{-1}Mpc^{-1}$ with $h = 0.67$ and known with sub-percent precision [22, 23]. Typically, the energy densities of the different components of the universe ρ_i are expressed in units of so-called critical density ρ_c:

$$\Omega_i \equiv \frac{\rho_i}{\rho_c} = \frac{8\pi G\rho_i}{3H^2}\,. \tag{2.3}$$

Each cosmological component is assumed to behave as a perfect relativistic fluid, characterized by a specific Equation of State (EoS) between pressure and energy density:

$$\omega_i = \frac{p_i}{\rho_i}\,. \tag{2.4}$$

In the ΛCDM framework, the Universe is composed by three components: DE which is described as a cosmological constant within current constraints (characterize by its own abundance parameter Ω_Λ), non-relativistic matter consisting of DM and baryionic matter ($\Omega_m = \Omega_{DM} + \Omega_b$), and relativistic photon and neutrino components ($\Omega_r = \Omega_\gamma + \Omega_\nu$). The EoS of the cosmological constant is $\omega_\Lambda = -1$ while for non-relativistic matter $\omega_b = \omega_{DM} = 0$. The photon contribution to the radiation component has an EoS $\omega_\gamma = 1/3$, following the assumption of thermal equilibrium in the energy density distribution. This EoS is valid also for the neutrinos, as they are characterized by a mass making them effectively in a relativistic regime in all the relevant cosmological epochs, $m_\nu \ll 1$ eV. As we will see, observations show that the Universe has a global zero curvature ($K \simeq 0$), meaning that the total density ρ_{tot} should be equal to the critical density ρ_c which means $\Omega_{tot} = \rho_{tot}/\rho_c \simeq 1$. Based on the latest measurements, the current values for the relative densities of DE, DM, baryonic matter and radiation are: $\Omega_\Lambda \simeq 68.3\%$, $\Omega_{DM} \simeq 26.8\%$, $\Omega_b \simeq 4.9\%$, and $\Omega_r \simeq 10^{-5}$ [22].

We are interested in probing Cosmology at very high energy scales, right after the Big bang, in order to investigate physical mechanisms at the edge of our

knowledge. This regime is known as the Planck scale. The Planck energy, time, and length scales can be obtained by combining the Newton, Planck and light speed constants, obtaining extreme values, such as the Planck length $\sim 10^{-35}$ m and, in terms of time, the Planck time $\sim 10^{-44}$ seconds. As we discuss with more detail in Section 2.1.2, according to the current picture of the cosmological model at those extreme epochs, a process named Inflation took place and lasted just for about $\sim 10^{-33}$ seconds, expanding the size of Universe in an accelerated, quasi-exponential manner, by a factor of at least $\sim 10^{26}$ [24, 25]. Inflation is known to solve three main problems of the pre-inflationary cosmology: flatness, horizon, and magnetic-monopole problem [26, 27], associated to the remarkable vanishing value of the observed curvature, the apparent homogeneity in the distribution of structures, as well as the absence of relics from phase transitions. Moreover, by describing the Inflation as triggered by a scalar field known as the Inflation, it has been realized how its quantum fluctuations appear to be the seeds of the observed LSS. It can be seen how the emergent spectrum of inflationary fluctuations in the energy density is characterized as a Gaussian random process, where the only quantity which matters is the variance at each cosmological wavenumber, $k = 2\pi/lambda$, which is described as a power law with exponent n_s [28] which is know today to be close by lower with respect to 1 by about 4%.

One of the most important evidence of the Big bang theory and the standard model of cosmology is the Cosmic Microwave Background (CMB) radiation, which we will describe extensively in the next Section.

2.1.1 The Cosmic Microwave Background

The standard model of cosmology, as we described in the previous Section, predicts the existence of thermal radiation as a remnant of the hot and dense stage of the very early Universe, known as CMB. Following Inflation, the hot and dense gas made of relativistic and non-relativistic components, started to cool down and get rarefied due to the expansion. Within the first seconds, baryons formed and de-coupled out of thermal equilibriuum with the cosmic plasma when the temperature of the Universe dropped below ~ 100 keV, while relativistic particles like electrons, positrons and photons remained in equilibrium. After that, with the cooling continuing, the nuclei of elements started to form, but due to the low speed of nuclear reactions with respect to the Universe expansion, only helium and

hydrogen nuclei were produced at an appreciable level. This era is the Big Bang Nucleosynthesis, which lasts for less than one thousand seconds.

Later, the Universe gets too cold to provide the needed energy for the nuclear reaction of the elements heavier than helium. As the expansion goes on and the temperature drops to $\sim$ 1 eV, the fraction of neutral hydrogen gets very small. Photons are tightly coupled to electrons via Compton scattering and electrons to photons via Coulomb scattering. Although the binding energy of neutral hydrogen atom is $\epsilon_0 = 13.6$ eV, and we expect the ionized hydrogen converts to neutral atoms at that energy, the high ratio of photon/electron prohibits any electron from staying bound to nuclei. Hydrogen atom formation happens only when the temperature comes down to $\sim$ 0.1 eV, and photons are no longer in thermal equilibrium with the matter. Recombination is the era in which electron and atomic nuclei construct the neutral atoms, and the Universe for the first time becomes transparent. Before the recombination, the mean free path of photons was of the order of length corresponding to one atom, meaning opacity to radiation. After atom formation, the photons can travel freely and their mean free path increasis up to cosmological scales. This era is the origin of CMB photons and is associated with the Last Scattering Region (LSR), the most distant portion of the Universe which we can access through electromagnetic radiation. Recombination happens at redshift $z \approx 1100$, in other words, when the Universe was around 380,000 years old. The collision with electrons before the last scattering guarantees the fact that photons were in equilibrium and they keep a black body spectrum.

The CMB was predicted for the first time by Gamow [29] and Alpher and Herman [30], but its discovery was postponed to about 17 years later [1]. They measured isotropic microwave radiation in one single frequency turns out to be the most important evidence of the Big bang theory. The black-body shape of the CMB temperature spectrum was reported weeks after the launch of Cosmic Background Explorer satellite from the National Aeronautic Space Administration, [COBE, NASA, see 2, and references therein], showing an almost perfect black-body shape. Figure 2.1 indicates the observational data of Far Infrared Absolute Spectrophotometer (FIRAS) instrument of COBE on top of the theoretical black-body curve. The COBE mission ended in 1994 [31] and represented a milestone along the way for the next generation of balloon and ground based experiments during the 90s. The Wilkinson Microwave Anisotropy Probe (WMAP) [3], supported again by NASA, was launched in 2001. WMAP had five frequencies from 22.8 GHz to 93.5

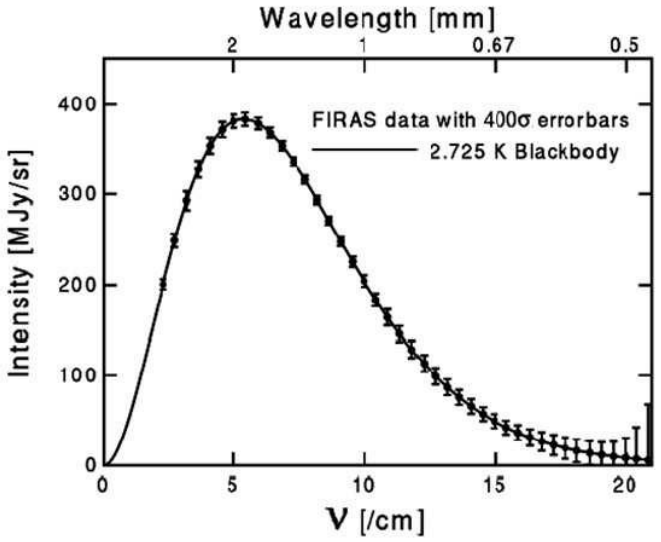

FIGURE 2.1: The FIRAS measurements of the CMB radiation on top of a theoretical black-body curve. Figure from [33]. The errorbars are artificially augmented by a factor of 100, in order to make them graphically visible.

GHz and took data for nine years. In 2009, the Planck satellite from the European Space Agency (ESA), equipped with nine frequencies from 30 GHz to 857 GHz, was launched in 2009 and in operation till 2013 [32]. Planck consisted of two different instruments the High Frequency Instrument (HFI) and Low Frequency Instrument (LFI). As we anticipated, satellites represent only one component of all efforts concerning CMB observations; a number of ground based and balloon borne observatories operated in parallel to satellites. It is difficult to quote properly all efforts along these lines, but the reader can look at the repository available at `lambda.gsfc.nasa.gov` for a complete list of past and operating observations. Planck observed the CMB temperature power spectrum with very high precision, in terms of sensitivity and angular resolution, reaching a few arcminutes. It also probed polarization down to an angular scale corresponding to about 10 arcminutes, while leaving the high precision polarization observation to the more sensitive instruments corresponding to the next generation of CMB experiments, which we will describe extensively in Chapter 3.

2.1.2 The Inflationary Paradigm

In this Section we describe the main features of the Inflation, which we already anticipated. Our goal is to review how the early Universe is thought to produce cosmological perturbations with the properties of tensors, i.e. gravitational waves,

and their abundance with respect to scalar (density) perturbations. Indeed, the search for cosmological gravitational waves is the main target of CMB experiments where the methodologies developed in this Thesis found their motivation and application.

The pre-inflationary cosmology was afflicted by the horizon and flatness problems. As it is well known, see e.g. Dodelson [17], the horizon is the maximum comoving distance traveled by light since the beginning. At the time of recombination, the size of the horizon corresponded to $\sim 2^\circ$ in the sky today, which means that the objects in the distance larger than the horizon didn't exchange information and they were not causally connected. This was in conflict with several evidences, such as the fact that the CMB temperature exhibits changes of one part per a hundred thousands all over the sky. Therefore, the horizon problem corresponds to the question: how did the CMB photons coming from different regions with super-horizon distances thermalize and share such similar temperatures?

On the other hand, the flatness problem refers to the curvature density, Ω_K (explained in Equation 2.3), which is probed by CMB measurements as we will see, and found to be smaller than $\sim 10^{-2}$ [22]. Since the curvature density has a direct dependency with the scale factor, proportional to a^{-2}, we can compute this parameter at the Planck time, and find $\Omega_K \leqslant 10^{-63}$. So the question would be: why did the Universe favor a flat geometry, i. e. $K = 0$, in its initial state?

In 1980 Alan Guth introduced an answer to these questions: prior to the Big Bang, there was a phase of very rapid accelerated expansion driven by a scalar field, with a negative-pressure equation of state $\omega \approx -1$. This scalar field, named Inflaton, would be also responsible for quantum fluctuations generating cosmological density and metric perturbations. One year earlier, the accelerated expansion and the generation of gravitational waves from quantum fluctuations was studied by [34]. Starobinsky proposed a simple model of inflation where the scalar field ϕ was obeying a dynamics dictated by the potential $V(\phi)$.

We describe here the fluctuations produced by the Inflation, which become later the LSSs. The background metric depends only on time, whereas the metric fluctuations depend on both space and time, see Durrer [35] for reviews. Since the metric is a symmetric tensor, there are ten fluctuating degrees of freedom. There are four degrees of freedom that correspond to scalar metric fluctuations. The other four correspond to vectors, and the last two to tensors. A gauge freedom,

due to reference frames different from the comoving one at the linear level, reduce the physically relevant scalar and vector perturbations to 2, and tensors to 1, In longitudinal gauge, the perturbed metric can be written as

$$ds^2 = -(1+2\Psi)dt^2 + a^2(t)[(1-2\Phi)\gamma_{ij} + 2h_{ij}]dx^i dx^j \,, \tag{2.5}$$

where Ψ and Φ are the gauge invariant scalar Bardeen potentials [36] and h_{ij} is transverse and traceless, $\nabla_i h_{ij} = 0$ which consists of the tensorial part of the perturbations and describes gravitational waves. For a perfect fluid or quasi-Newtonian matter made of non-relativistic particles, the two Bardeen potentials are equal (with the exception of the sign) and analogue to the Newtonian gravitational potential. The metric perturbations are related to perturbations in the energy momentum tensor of matter by Einstein's equations. We also define the curvature perturbation $\mathcal{R}$ which is an important quantity for inflationary models in terms of the Bardeen potentials:

$$-\mathcal{R} = \frac{2}{3(1+\omega)}\left[\Psi + H^{-1}\Phi\right] \,. \tag{2.6}$$

The Equation above is similar to 2.4, where ω is the EoS of the entire content of the stress energy tensor. $\mathcal{R}$ is proportional to the perturbation of the spatial Riemann curvature in the comoving gauge. Following the calculation in [35], the power spectrum of the curvature fluctuation, $\mathcal{R}$, on super Hubble scales can be written as

$$\mathcal{P}(k) = \Delta^2_{\mathcal{R}} \left(\frac{k}{k_*}\right)^{(n_s-1)} \,, \tag{2.7}$$

where $n_s - 1$ is the spectral tilt, k_* is an arbitrary pivot scale and Δ_R is the amplitude of the power spectrum at this scale. The spectral tilt and the amplitude depend on the details of the inflationary model. In the FRLW metric, the equation of motion for a scalar field is the Klein Gordon equation which is written as

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0 \,. \tag{2.8}$$

The Friedmann equation mentioned in Equation 2.2, without curvature, takes the form

$$H^2 = \frac{8\pi G}{3}\rho = \frac{8\pi G}{3}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right) \simeq \frac{8\pi G}{3}V(\phi) \,, \tag{2.9}$$

where $\dot{\phi}^2/2$ is the kinetic term and V is the potential of the scalar field.

A key property of inflationary models is the capability of performing a slow roll of the potential mean value down to the minimum of the potential, in order to produce the amount of expansion which is required to solve the horizon and flatness problems. One can introduce two slow roll parameters

$$\epsilon \equiv -\frac{\dot{H}}{H^2} = \frac{m_p^2}{16\pi}\left(\frac{V_{,\phi}}{V}\right)^2 \simeq \frac{3}{2}\frac{\dot{\phi}^2}{V} \ll 1\,, \tag{2.10}$$

and

$$\eta \equiv \frac{m_p^2}{8\pi}\left(\frac{V_{,\phi\phi}}{V}\right) = \frac{V_{,\phi\phi}}{3H^2}\,, \tag{2.11}$$

where $V_{,\phi}$ and $V_{,\phi\phi}$ are the first and second derivatives of the potential and m_p is the Planck mass. Inflationary models need to satisfy both conditions $\epsilon \ll 1$ and $\eta \ll 1$, in order to keep the slow roll regime for inflation [37]. It is possible to show that in this regime, the curvature fluctuations obey the relation

$$\Delta^2_{\mathcal{R}} = \left.\frac{H^2}{\pi\epsilon m_p^2}\right|_{H=k_*/a}\,, \qquad n_s - 1 = -6\epsilon + 2\eta\,, \tag{2.12}$$

where $|_{H=k_*/a}$ means that the Hubble parameter which is varying very slowly should be evaluated when the pivot scale k_* corresponds to the Hubble expansion rate H, a moment which is also known as horizon exit. During Inflation, the quantum fluctuations which oscillate at constant amplitude as long as $k/a \gg H$, and those in the tensor perturbations h_{ij} correspond to cosmological gravitational waves. Once the scale factor has grown sufficiently, reaching $k/a \simeq H$, the oscillations freeze and when $k/a \ll H$, h_{ij} (in Equation 2.5) becomes constant. It is possible to show that the constancy of cosmological gravitational waves on scales larger than the horizon keeps being valid also after Inflation ends. Therefore, their detection would be not only a unique evidence of inflationary field but also it would be a most important carrier of information about the features of Inflation itself. It is possible to show that the power spectrum of gravitational waves produced in this way during single field slow roll inflation is given by

$$\mathcal{P}_h(k) = \Delta^2_h\left(\frac{k}{k_*}\right)^{n_t}\,, \tag{2.13}$$

with

$$\Delta^2_h = \left.\frac{16H^2}{\pi m_p^2}\right|_{H=k_*/a}\,, \qquad n_t = -2\epsilon\,. \tag{2.14}$$

Therefore, by having the amplitude of both tensor and scalar modes, we can define the observable parameter, tensor to scalar ratio, r:

$$r = \frac{\Delta_h^2}{\Delta_{\mathcal{R}}^2} = 16\epsilon = -8n_t\,. \tag{2.15}$$

For a more complete review of a variety of inflationary models, we refer to [38].

2.2 CMB Anisotropies

In this Section, we will introduce the main properties of the CMB anisotropies, consisting in their link to cosmological perturbations, and focusing on the decomposition of their polarization. At the end, we review some of the main secondary sources for them, mainly due to LSS formation.

2.2.1 Temperature anisotropies

By indicating with T the temperature of the CMB, the signature of quantum fluctuations generated during inflation can be seen on the CMB temperature field at the level of $\Delta T/T \sim 10^{-5}$. Figure 2.2 shows the temperature anisotropies of the CMB, observed by the Planck satellite [4]. The map is characterized by fluctuations on all visible scales, from the entire sphere, down to the arcminute, and we will outline the main phenomenology leading to those imprints in this Section. At any given time t, the CMB temperature field T is a function of position $\mathbf{x}$ and direction $\mathbf{n}$, $T(\mathbf{x}, t, \mathbf{n})$. The direction dependence can be expanded on the sphere as:

$$T(\mathbf{x}, t, \mathbf{n}) = \bar{T}(t) \sum_{\ell} \sum_{m=-\ell}^{\ell} a_{\ell m}(\mathbf{x}, t) Y_{\ell m}(\mathbf{n})\,, \tag{2.16}$$

where $Y_{\ell m}$ are the scalar spherical harmonics, ℓ and m determine the wavelength and the shape of the mode. The expansion coefficients are given by $a_{\ell m}$. $\bar{T}(t)$ is the average of T over all the directions in the spherical coordinate system at the given time. The CMB temperature at present is $T_0 \sim 2.7255$ Kelvin. The statistical isotropy means that for different ℓ's and m's, the random variables $a_{\ell m}(\mathbf{x}, t)$ are

not correlated. Therefore we can write:

$$\langle a_{\ell m}(\mathbf{x},t)a^*_{\ell' m'}(\mathbf{x},t)\rangle = \delta_{\ell\ell'}\delta_{mm'}C_\ell(t), \tag{2.17}$$

where the latter term represents the variance at each given scale ℓ, and the set of values for all scales is known as the CMB angular power spectrum for temperature. Since the CMB, as a result of inflationary perturbation, is nearly Gaussian, as it is confirmed by Planck observations, [39], most of its characterizations are coded in angular power spectra. By taking the average over every ℓ and m, we obtain

$$C_\ell = \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2\,. \tag{2.18}$$

Given the Gaussian nature of the field the sample variance which affects the unique observation we have of the CMB is given by

$$\Delta C_\ell = \sqrt{\frac{2}{2\ell+1}}C_\ell\,. \tag{2.19}$$

The actual values of the coefficients of the CMB temperature power spectrum are obtained by integrating the photon geodesic along the Line Of Sight (LOS) [see e.g. 40, and references therein]. By solving Einstein's equations for the evolution of the geometry and the Boltzmann equation for the evolution of the photon distribution function to first order, we can derive the temperature fluctuation spectrum for a given initial spectrum of curvature fluctuations, which has the expression

$$\frac{\ell(\ell+1)C_\ell}{2\pi} = \int \frac{dk}{k}\Theta^2_T(k,\ell)\mathcal{P}(k)\,, \tag{2.20}$$

where $\Theta^2_T(k,\ell)$ is the transfer function which encodes all the changes of the fluctuations from the initial time, when the primordial spectrum of density perturbations $\mathcal{P}(k)$, defined in 2.7, were produced. The above expression, of course, transfers the power of the primordial density perturbations only, while, as we will see later, the other kinds of perturbations, namely tensors, also affect CMB anisotropies. The transfer function is a result of the integration of the Boltzmann equation. We report here just the more significant highlights of these calculations, referring to Durrer [35] for a complete treatment. The LOS integration gives:

$$\frac{\Delta T}{T}(\mathbf{n}) = [\frac{1}{4}\,\delta^{(\gamma)}_g + \mathbf{V}^{(b)}.\,\mathbf{n} + (\Psi+\Phi)](t_{dec}+\mathbf{x}_{dec}) + \int_{t_{dec}}^{t_0}\partial_t(\Psi+\Phi)(t,\mathbf{x}(t))dt\,. \tag{2.21}$$

Here $\delta_g^{(\gamma)}$ is the radiation density fluctuation, $\mathbf{V}^{(b)}$ is the baryon peculiar velocity and Ψ and Φ (the same as Equation 2.5) are the Bardeen potentials. All these parameter are evaluated at the time of photon decoupling from photon-baryon fluid ($t_{dec} + \mathbf{x}_{dec}$). The first and the third term in the square bracket together combine to the ordinary Sachs-Wolfe effect on large angular scales [41], meaning the effects due to gravitational potensials on the Last Scattering Region, as well as their evolution along the line of sight (Integrated Sachs-Wolfe, ISW). Meaning that, the photon coming to the observer at $(\mathbf{x}_0, t_0)$ from a direction $\mathbf{n}$, starting on the last scattering surface from the Bardeen potential at the position $(\mathbf{x}_{dec})$. One can show that, on large angular scales, i.e. larger than the scale subtended by the horizon at decoupling, i.e. about 2 degrees, the temperature fluctuations from adiabatic inflationary initial fluctuations are given by

$$\frac{\Delta T}{T_0}(\mathbf{x}_0, t_0, \mathbf{n}) = \frac{1}{3}\Psi(t_{dec} + \mathbf{x}_{dec})\,, \tag{2.22}$$

meaning that on those scales, temperature CMB anisotropies are a faithful tracer of the inflationary initial conditions, via the Sachs-Wolfe effect, shown in Figure 2.3 with pink color. The power imprinted by Inflation on all scales represents the source of the oscillatory part of the spectrum. Indeed, a characteristic process of the CMB consists in the Baryon Acoustic Oscillations (BAO). At the time of recombination, in the photon-baryon fluid, radiation pressure resists the gravitational compression of the fluid into potential wells, and sets up acoustic oscillations in the fluid. Due to gravity which is powered by the initial conditions imprinted by Inflationary perturbations, the baryon density grows and due to photon pressure response, they start a bouncing process. The regions with compression and rarefaction represent the hot (overdensity) and cold (underdensity) zones of CMB temperature. The baryon density controls the relative heights of overdensity and rarefaction peaks [42]. The integral part in Equation 2.21 corresponds to the ISW, which we'll comment in more detail later in the Chapter, and corresponds to the green line in Figure 2.3). The second term in the square bracket of Equation 2.21 is the Doppler effect, which is caused by photons emitted from electrons with non-zero peculiar velocity in the direction of the emission (blue line in Figure 2.3).

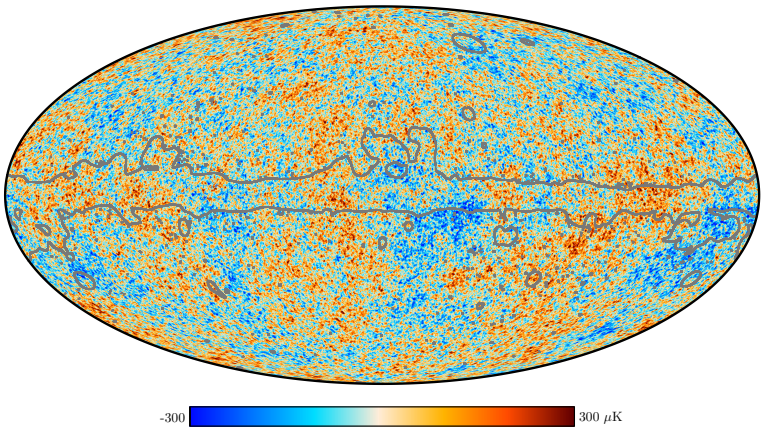

FIGURE 2.2: The 2018 Planck map of the temperature anisotropies of the CMB. The gray outline shows the extent of the galactic mask. Plot taken from the Planck legacy archive. From [11].

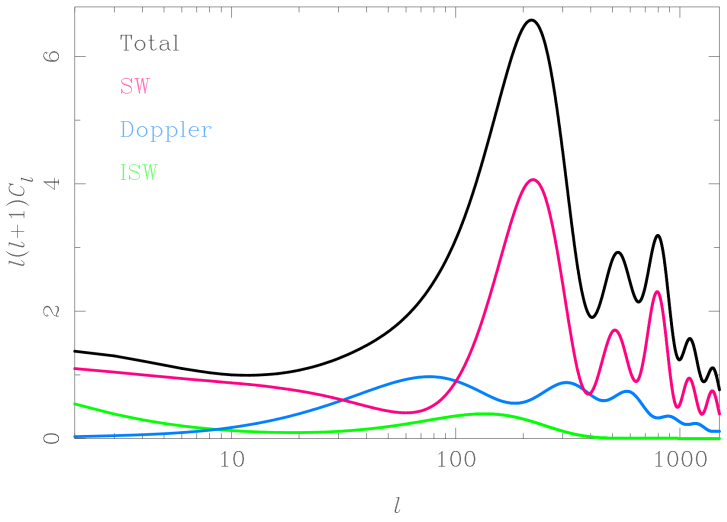

FIGURE 2.3: The variance (C_ℓ coefficients, defined in the text) as a function of the angular multipole, related to the angle at which the anisotropy is probed by the approximate relation $\ell \simeq 220/\theta$ [degrees]. The main contributions of the various terms in Equation 2.21, discussed in the text, are shown. The green, blue and pink represent ISW, Doppler and SW effects respectively, while black line describes their sum. The units of the spectrum are arbitrary. Plot from [43].

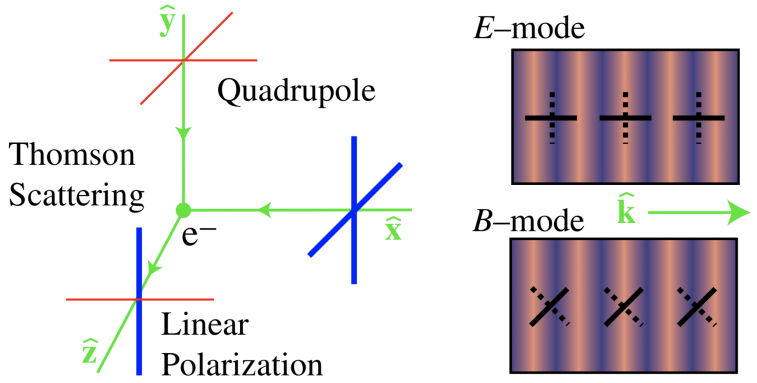

FIGURE 2.4: Left: Thomson scattering of radiation with a quadrupole anisotropy generating linear polarization. Blue colors (thick lines) represent hot and red colors (thin lines) cold radiation. Right: Polarization in the $\hat{x} - \hat{y}$ plane along the outgoing $\hat{z}$ axis. The component of the polarization that is parallel or perpendicular to the wavevector $\mathbf{k}$ is called the E-mode and the one at 45° angles is called the B-mode. Plot from [42]

2.2.2 Polarization Anisotropies

Just like the scattering producing temperature anisotropies, the process responsible for polarizing the CMB radiation is the Thomson scattering between photons and electrons at the epoch of recombination [6]. A quadrupole temperature anisotropy in the radiation generates a net linear polarization from Thomson scattering and a quadrupole can be generated causally by the motion of photons when the Universe is optically thin to Thomson scattering. Therefore, the order of magnitude and orientation of the polarization anisotropies are proportional to the magnitude and orientation of the radiation's quadrupole. The left panel of Figure 2.4 shows how Thomson scattering of radiation with quadrupole anisotropy, coming from cold and hot spots of the CMB, generates linear polarization. Here we make use of well-known Stokes parameters (I, Q, U, and V) in order to describe the polarization field of monochromatic electromagnetic wave. The parameter I quantifies the relative intensity of the wave, corresponding to $\Delta T/T$ as described above, Q measures the linear polarization in the direction of x axis, where U gives the same information along the axis rotated by 45°. Finally, V is the parameter that encodes circular polarization, which is usually ignored in cosmology, since the Thomson scattering is not expected to produce any circular polarization. CMB polarization can be described by the Q and U parameters, which are coordinate dependent [44], and transform under the rotation angle along the axis perpendicular to the wave. Q and U, in terms of spherical harmonics, can be written

as

$$(Q \pm iU)(\hat{n}) = \sum_{l,m} a^{\pm 2}_{l,m}\, Y^{\pm 2}_{l,m}\,(\hat{n}), \tag{2.23}$$

where at this time we exploit the tensor spherical harmonics, $Y^{\pm 2}_{l,m}$. We can also define the polarized intensity $P = \sqrt{Q^2 + U^2}$. Figure 2.5 shows the Polarized intensity of the CMB, observed by Planck [4].

2.2.2.1 The E and B modes

The CMB polarization pattern can also be decomposed into an alternative base, the E and B-modes, gradient and curl components, with odd and even behavior with respect to parity transformation, respectively [8, 45]. In the real space, they are defined as:

$$E(\hat{n}) = \sum_{l,m} a^{E}_{l,m}\, Y_{l,m}\,(\hat{n}), \qquad B(\hat{n}) = \sum_{l,m} a^{B}_{l,m}\, Y_{l,m}\,(\hat{n})\,. \tag{2.24}$$

In the harmonic space, $a^{E}_{l,m}$ and $a^{B}_{l,m}$ are defined as

$$a^{E}_{l,m} = -(-a^{2}_{l,m} + -a^{-2}_{l,m})\,/2\,, \qquad a^{B}_{l,m} = i(-a^{2}_{l,m} + -a^{-2}_{l,m})\,/2\,. \tag{2.25}$$

Unlike Q and U, B and E-modes are coordinate-independent on the sphere. The right panel of Figure 2.4 demonstrates the difference between E and B-modes patterns. In principle, we can calculate all the combination of temperature T, and polarization E, B power spectra along with their correlations: C^{TT}_{ℓ}, C^{EE}_{ℓ}, C^{BB}_{ℓ}, C^{TE}_{ℓ}, C^{TB}_{ℓ}, C^{EB}_{ℓ}. Under parity transformation T and E transform differently respect to B, therefore, due to parity conservation C^{TB}_{ℓ}, C^{EB}_{ℓ} will be zero. In Figure 2.6, we show C^{TT}_{ℓ}, C^{EE}_{ℓ}, C^{BB}_{ℓ}, C^{TE}_{ℓ}, power spectra in blue, orange, green and red respectively. As we have seen in Section 2.2.1, the reason for the oscillations in the TT and EE power spectra is related to the acoustic oscillations of photon-baryon fluid inside the gravitational wells. The polarization signal arises from the gradient of the peculiar velocity of the photon fluid (ν_γ). However, there is a difference in the oscillation phase of TT and EE in such a way that the peak of TT is the opposite of EE. This phase shift is due to the fact that the monopole and the quadrupole are the main contributors to TT and EE power spectra, respectively. The difference in power between temperature and polarization is due to the efficiency of Thomson scattering in converting into polarization a portion of the incoming

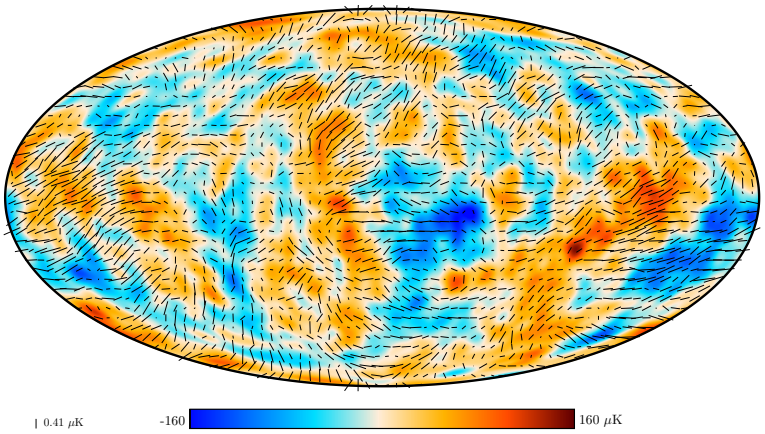

FIGURE 2.5: The 2018 Planck map of the polarized CMB anisotropies, shown as rods whose direction and length represent the direction and amplitude of polarized CMB. The colored background is the map of intensity anisotropies, smoothed to 5 degrees. Plot taken from the Planck legacy archive [see 32, and references therein].

intensity quadrupole. In the next Section, we will discuss different features of the BB power spectrum and its sources. Before concluding, we highlight the region at $l < 10$, which is affected by what is known as the reionization bump. Reionization happens around redshift $z \simeq 7$ [22], when free electrons produced through reionization of the intergalactic medium are hit by CMB photons along their path to us. The effect boosts anisotropies at large angular scales, corresponding to the angle subtended by the horizon at reionization, as a second scattering region, closer to the observer with respect to the recombination. The effect powers up polarization anisotropies, as the new era of Thomson scattering reprojects anisotropies from the quadrupole coming from the last scattering at recombination. The reionization is characterized by one very relevant parameter, corresponding to the optical depth from here to the beginning of the process, $\tau = \int n_e \sigma_T dt/a(t)$, where n_e is the free electron density, and σ_T the Thomson scattering cross section. Current measurements [22] set this parameter to be $\tau \simeq 0.0544$.

2.2.3 Sources of B-mode Polarization

The CMB B-mode polarization can be generated by two main mechanisms: $(i)$the gravitational lensing of CMB photons by LSS, (ii) the primordial gravitational waves which are generated by Inflation in the very early universe.

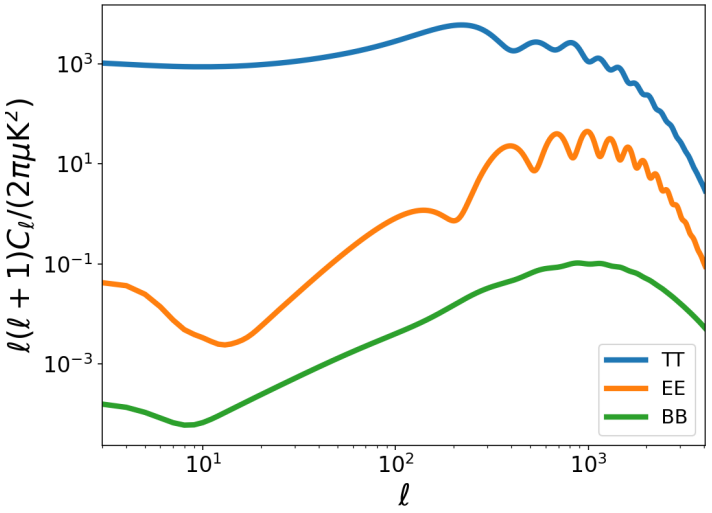

FIGURE 2.6: CMB TT, EE and BB power spectra, for a cosmological model with $r = 0.01$, and including CMB lensing.

2.2.3.1 Gravitational Lensing

The gravitational potentials produced by the LSS are able to bend the CMB photons and as a result, the hot and cold spots of the CMB temperature field distort around the foreground sources, generating an effect which is commonly known as CMB lensing. Not only CMB Lensing changes the temperature field but also the polarization. Lensing conserves surface brightness, therefore it only affects anisotropies and since the fluctuations are already first order, the lensing effect is second order in the perturbation theory. [46]. Looking in direction $\mathbf{n}$, we actually see the temperature fluctuation not as it was at position $\mathbf{n}r_*$ but at position $(\mathbf{n} + \boldsymbol{\alpha})r_*$, where $\boldsymbol{\alpha}$ denotes the deflection angle. To first order in perturbation theory the deflection angle [47] is given by:

$$\boldsymbol{\alpha} = -2 \int_0^{r_*} dr \frac{\chi(r_* - r)}{\chi(r_*)\chi(r)} \nabla_\perp \Psi(t(r), r, \vartheta, \varphi) \equiv \nabla_\perp \phi(\vartheta, \varphi)\,, \tag{2.26}$$

where Ψ is the Bardeen potential, $\mathbf{n}$ is given by (ϑ, φ) and $\nabla_\perp$ is the gradient on the sphere of photon directions. The lensing potential is given by

$$\phi(\vartheta, \varphi) = -2 \int_0^{r_*} dr \frac{\chi(r_* - r)}{\chi(r_*)\chi(r)} \Psi(t(r), r, \vartheta, \varphi)\,. \tag{2.27}$$

The distribution of the deflection angle is approximately Gaussian, with a standard deviation of about 2 arcmintes, for the standard cosmological model [48]. The

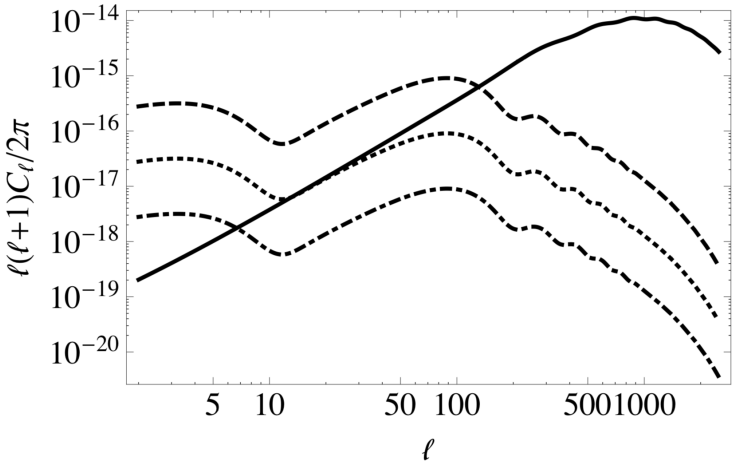

FIGURE 2.7: Theoretical B-mode polarization power spectra from lensing of E-modes, CMB lensing (solid) and from the tensor modes, primordial gravitational waves for $r = 0.1$ (dashed), $r = 10^{-2}$ (dotted) and $r = 10^{-3}$ (dot-dashed). The y-axis of this Figure should be multiplied by $(2.725 \times 10^6 \mu\text{K})^2$ to be normalized in physical units and compared with Figure 2.6. Plot from [35].

lensing potential can be expanded into spherical harmonics following the same procedure of Equation 2.16 to 2.18, which leads to the definition of the lensing power spectrum $C_\ell^{\phi\phi}$. The effect is detected today by a multitude of experiments, [see 49, and references therin]. The CMB lensing distorts the polarization tensor, causing a leakage of E to B-modes [47]. In Figure 2.7, the solid line shows the B-mode polarization power spectra from CMB lensing which completely dominates the power at arcminute scales. In terms of scientific relevance, the CMB lensing signal is fundamental for investigating the dark cosmological components through LSS formation, and it is a powerful tool to understand and constrain dark energy and modified gravity models [50]. In order to detect the CMB B mode generated by primordial gravitational waves, CMB lensing removal in B modes is necessary, through the so called de-lensing class of algorithms and techniques [51].

2.2.3.2 Primordial Gravitational waves

As we mentioned in Section 2.1.2, different inflationary models suggest the existence of primordial gravitational waves due to the exponential expansion of the Universe in the early phases [6]. Since primordial gravitational waves are generated by the tensor perturbations, they can source B-mode polarization in the CMB power spectrum. The observable parameter which relate the amplitude of primordial gravitational waves to energy scale of inflation model is tensor-to-scalar

ratio r as we described in Section 2.1.2, Equation 2.15: it is defined on the basis of the slow roll parameters of the Inflation, as it is described by Equation

$$r = 16\epsilon = \left[0.1 \times \frac{V}{(2 \times 10^{16}\ \mathrm{GeV})^4}\right] . \tag{2.28}$$

Therefore, cosmological gravitational waves might have tremendous implications for cosmology and high energy physics, as they are directly related to the production during inflation, and the details of the latter. In Figure 2.7, the CMB B-mode polarization power spectrum generated by gravitational waves for different values of r is plotted. However, B-modes polarization due to gravitational waves has not yet been discovered and only upper limits exist so far for its amplitude, corresponding to $r < 0.06$ (at 95% confidence level) [52]. The next generation of CMB experiments either ground-based or satellite, as we will discuss more in Section 3.4, aim to push this limit to the $r \leq 0.001$ limit.

2.2.4 Other Secondary Anisotropies

We close this Chapter by quoting the most important effects sourcing CMB anisotropies after recombination, i.e. along the LOS to us, known as secondary anisotropies. In the last Sections, we mentioned the CMB lensing, which is secondary anisotropy caused by gravitational lensing. Here we listed the other main mechanisms.

2.2.4.1 ISW

As we anticipated in the Section 2.2.1, the dynamics in the gravitational potential may lead to contribution to CMB temperature anisotropies through the ISW. Such a dynamics happens for example when the equation of state changes in time. Therefore, the ISW effect has contributions both from the early and late time Universe. During recombination, the Universe is almost completely matter dominated; therefore, Ψ in Equation 2.21 evolves due to the non-negligible presence of radiation, which gets rapidly diluted (early ISW). Also, at late times, the Universe becomes dark energy dominated, making the gravitational potentials evolving again (late ISW). The two aspects of the ISW appear on the size of the horizon at the relevant time, i.e. a few degrees (early ISW) or the entire sky (late ISW). The effect is now detected with great confidence [53].

2.2.4.2 Rees-Sciama Effect

The non-linear regime of ISW is known as Rees Sciama (RS) effect [54]. If we consider a single isolated structure, its potential changes due to its own evolution through gravitational collapse. causing photons to experience a different potential perturbation in and out the same structure, gathering a net effect on their temperature. These fluctuations have a very small effect on the temperature of CMB photons and they are very hard to detect.

2.2.4.3 Sunyaev-Zel'dovich (SZ) Effect

The SZ effect happens if CMB photons pass through a cluster of galaxies and scatter off by the hot electron gas present in the galaxy clusters and intercluster medium via inverse Compton [55]. This process creates spectral distortion in the CMB (y-type) and also generates additional temperature fluctuations [56]. There are two main types of SZ effect, classified by the different physical process involved: Thermal Sunyaev-Zel'dovich (tSZ) and Kinetic Sunyaev-Zel'dovich (kSZ) [57].

The tSZ occurs when CMB photons interact with a gas of hot electrons at a temperate $T_{CMB} \ll T_{gas}$ which modifies the Planck spectrum. By passing through a hot plasma, the low energy Rayleigh-Jeans regime of the photon spectrum is depleted and the high energy, Wien part is enhanced. An important property of the tSZ effect is that it is independent of the redshift of the galaxy cluster [58]. The kSZ effect arises when the scattering of CMB photons with electrons takes place in a bulk motion of a non-linear perturbation with respect to the CMB. Essentially it is a Doppler term related to electron velocity projected along the line of sight [59].

Chapter 3

Diffuse Foreground Contamination for B-mode polarization

As we already mentioned, in recent years the observation of the CMB B-mode polarization became one of the greatest challenges for detecting the imprint of primordial gravitational waves generated during the inflationary era, as we mentioned in Chapter 2. One of the most important obstacles on the way of CMB B-mode polarization observation is represented by the foreground electromagnetic radiation of astrophysical objects along the LOS, and more specifically, the diffuse polarization from our own Galaxy. This Chapter has three main purposes. First, in Section 3.1, and 3.2 we will explain the level of foreground contamination for CMB B-mode signal and then introduce the foreground emissions responsible for this contamination. Second, in Section 3.3, we will mention the common methods for controlling and removing the foreground emission from the CMB signal. At the end of the Chapter, in Section 3.4, we describe state of CMB observations and future prospects for B-mode observations.

3.1 Contamination to B-modes

The evidence for the significance of the contamination from the diffuse polarized Galactic foreground to CMB B-mode polarization is now widely accepted, Krachmalnicoff et al. [see 13, and references therein]. Figure 3.1 shows the CMB power

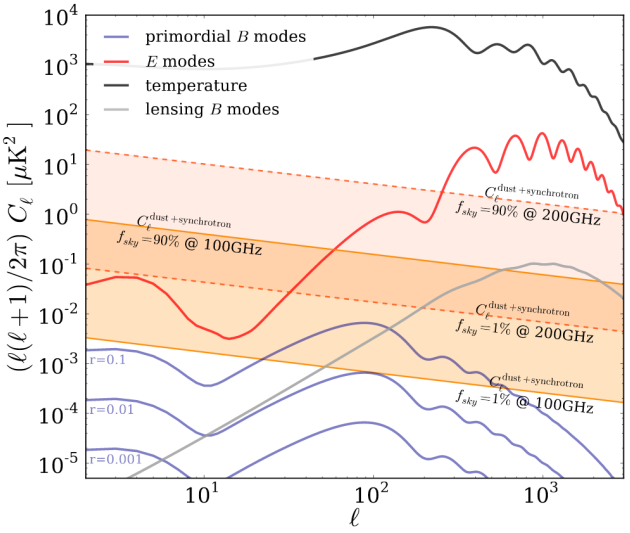

FIGURE 3.1: Contamination from diffuse CMB foregrounds onto the CMB B mode power spectrum. The figure shows the angular power spectra of total intensity (black line), E-modes (red line), lensing B modes (gray line) and the primordial B modes (blue lines) for different values of r. The level of the power spectra of the two dominant Galactic foregrounds, to be described later in the Chapter, known as synchrotron and thermal dust, are shown as an orange shaded area in comparison with primordial B-modes. From [60].

spectra in temperature (black line), E-modes (red line), lensing B-modes (gray line), primordial B-modes (blue lines) for different tensor to scalar ratio r, and the total contribution of polarized B-mode foregrounds (dust plus synchrotron, which we'll describe in the following), expected on the cleanest 1-90 % of the sky, at 100 and 200 GHz (orange shaded areas). As we mentioned, the best constrain that we have for r is less than 0.06; therefore we can immediately conclude that polarized B-mode foregrounds are comparable or dominant with respect to the primordial B-modes signal. This is consistent with the latest evidences from Planck 2018 [11], where the level of CMB polarization signal in all the frequency channels is sub-dominant respect to diffuse polarized Galactic emissions. In Figure 3.2, the polarized intensity rms (root mean square) amplitude of synchrotron and thermal dust emissions as a function of Planck frequencies are compared with the CMB rms for a ΛCDM model with reionization optical depth corresponding to $\tau = 0.05$. In this plot, the sum of foregrounds is evaluated over three different masks with $f_{sky} = 0.83, 0.52, 0.27$. Thus, the foreground challenge for the B-mode observation is very serious and needs an accurate treatment at the level of ancillary studies, and data analysis, for which we are going to present our contribution in Chapter 5.

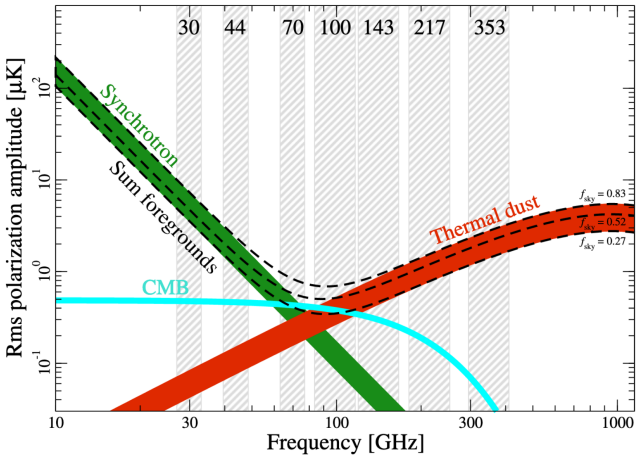

FIGURE 3.2: Polarized intensity *rms* amplitude of synchrotron and thermal dust emissions as a function of frequency, for areas of the sky excluding the Galactic plane, with highlighted areas corresponding to Planck frequencies. The green band indicates polarized synchrotron emission, and the red band indicates polarized thermal dust emission. The cyan curve shows the CMB *rms* for a ΛCDM model with $\tau = 0.05$, and is strongly dominated by E-mode polarization. The dashed black lines indicate the sum of foregrounds evaluated over three different masks with $f_{sky} = 0.83$, 0.52, and 0.27. From the Planck 2018 results [11].

3.2 Polarized CMB foregrounds

In this Thesis, we focus on the diffuse B mode foregrounds from our own Galaxy, representing the most important challenge to the detection of primordial B-modes from cosmological gravitational waves. The Galaxy is filled with a large scale Galactic magnetic field, which bends the trajectory of cosmic ray particles, mostly electrons, emitting polarized synchrotron. Also, dust grains heated back by starlight constitutes a quasi-thermal emission, known as thermal dust. They are known to be the ones with largest polarized emission, for obvious reasons for synchrotron, and because the dust grains have a magnetic dipole, which also gets aligned with the local direction of the Galactic magnetic field. In addition, we will also consider the Anomalous Microwave Emission (AME), possibly associated to the spinning dust grains.

3.2.1 Synchrotron emission

The synchrotron radiation is generated by cosmic ray electrons accelerating in the Galactic magnetic field. This emission depends on the number and energy spectral

index of their population ($N(E) \propto E^{-p}$) where p is the electron-energy distribution index. It also depends on the strength of the magnetic field, B, which can vary across the Galaxy, resulting different spectral behaviour for the synchrotron emission. This emission dominates over the CMB at frequencies $\lesssim 70$ GHz and possesses a steep Spectral Energy Distribution (SED) due to the corresponding energy distribution of electrons. The propagation of the electrons in the uniform magnetic field results the linear polarization of the synchrotron radiation characterized by the fraction [61]:

$$f_s = \frac{p+1}{p+7/3} = \frac{3\beta_s + 3}{3\beta_s + 1}\,. \tag{3.1}$$

By considering the synchrotron spectral index $\beta_s \approx -3$, synchrotron can be polarized up to $\approx 70\%$. The observed value of synchrotron polarization ($\leq 20\%$) is lower due to non-uniform magnetic field directions along line of sight. The mean polarization fraction can have different values and morphology across the sky, and this value drops to about 4% near the Galactic plane and can rise up to 20 % at high Galactic latitude ($|b| > 50°$) [11]. In Figure 3.3, the polarized intensity of synchrotron emission at 30 GHz is shown. This map describes the strong intensity of synchrotron in the Galactic plane as well as the North and South Galactic spurs as observed by Planck [11]. At the present level of knowledge, Krachmalnicoff et al. [see 13, and references therein], the synchrotron SED can be parameterized as a simple power-law in brightness temperature. Nonetheless, the energy distribution of electrons may be responsible for a curvature in the SED, which departs from a pure power-law. Thus, a general model for synchrotron emission can be written as

$$T(\hat{n}, \nu) = A_s(\hat{n}) \left(\frac{\nu}{\nu_0}\right)^{\beta_s(\hat{n}) + C(\hat{n}) \log(\nu/\nu_0)}\,, \tag{3.2}$$

where A_s is synchrotron amplitude at the pivot frequency ν_0, β_s is the synchrotron spectral index, and C parameterizes SED curvature. In general, all quantities are functions of the sky direction $\hat{n}$. The synchrotron spectral index has a typical value $\beta_s \approx -3$, with a variation between -2.98 and -3.12 in the sky, on the degree scale [62]. In another recent work, the synchrotron spectral index variation has been found to be in the range between -2.5 and -4.4, with a mean value of $\beta_s \simeq -3.2$; this has been obtained by considering low frequency channels from 2.3 to 33 GHz, combining radio observations by the S-band Polarization All Sky Survey (S-PASS, see [63]), WMAP and Planck data [13]. Non-zero curvature is suggested by cosmic ray energy spectrum at frequencies above 23 GHz in total intensity, resulting in

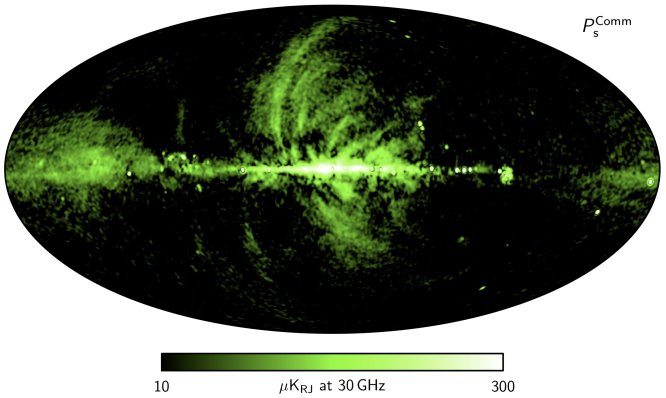

FIGURE 3.3: Planck observed polarized synchrotron amplitude map at 40' FWHM resolution. Plot from [11].

$C = -0.052 \pm 0.005$ [64]. Krachmalnicoff et al. [13] have derived an upper limit to the curvature value in polarization: the reported value is between 0.07 and 0.14 depending on the considered sky region and angular scales.

3.2.2 Thermal dust emission

Polarized thermal dust emission [see 65, and references therein] comes from interstellar dust grains which are mostly made of graphites, silicates, and Polycyclic Aromatic Hydrocarbons (PAHs), and they tend to align perpendicularly to the Galactic magnetic field, therefore emitting partially linearly polarized radiation. Dust grains are heated back by starlight and possess a modified black body SED, known as the grey body, with a temperature T_d with values around 20 K and varying across the sky. The SED is also described by a multiplicative emissivity correction ν^{β_d}, which determines the deviation from a pure black body, with β_d assuming values around 1.6 and a variation between 1.53 and 1.67 across the sky.

Dust grains have different shapes, and polarization is expected only from the ones with non-spherical shapes and a preferential axis of alignment. The magnetic moments of the dust grains will preferentially align with the ambient magnetic field, therefore as these aligned grains rotate, they emit polarized radiation [66]. The thermal dust polarization radiation has a dependency on the frequency, since the degree of alignment varies according to the size of dust grains. Similar to the synchrotron radiation, thermal dust polarization fraction on the Galactic plane is few percent while at the intermediate latitude can reach up to $\sim 20\%$ [65].

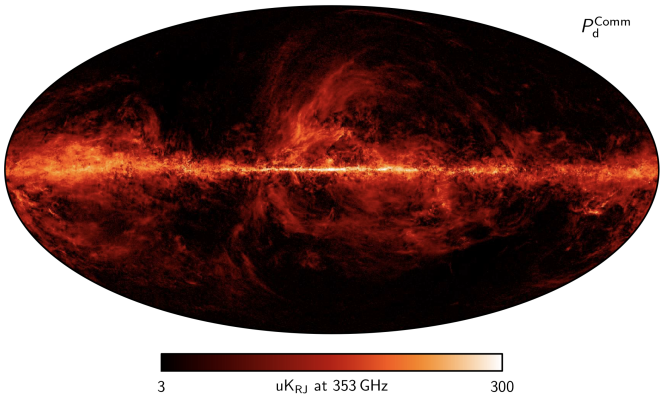

FIGURE 3.4: The polarized intensity of the thermal dust as observed by Planck, at 5 arcminutes FWHM resolution. Plot from [11].

Thermal dust emission dominates the polarized sky radiation at frequencies $\gtrsim$ 70 GHz [see 11, and references therein]. Figure 3.4 shows the distribution of polarized intensity of thermal dust emission which can be seen across the sky, as reconstructed by Planck. The analytic form of the brightness emission of the SED can be written as:

$$T(\hat{n},\nu) = A_d(\hat{n}) \left(\frac{\nu}{\nu_0}\right)^{\beta_d(\hat{n})} B(\nu, T_d(\hat{n}))\,, \tag{3.3}$$

where A_d defines the dust amplitude varying across the sky at the pivot frequency ν_0, and B represents the standard black body spectrum at the temperature T_d and frequency ν [65]. The aforementioned values for the dust spectral index β_d and T are based on the simplistic assumption that thermal dust emission is generated by one population of dust grains. There are different modellings of dust grain and populations which can fit the spectra. One of the models which is mostly adopted in modern literature consist of two populations of dust grains, shining differently at different frequencies; therefore the model has two spectral indices $\beta_{d\,1,2} = (1.67, 2.70)$ and two dust temperatures $T_{1,2} = (9.4, 16)$ [67].

It is also known that thermal dust and synchrotron emission are spatially correlated as shown in [68, 69]. They are anti-correlated in such a way that there is a rise in polarization SED towards low frequencies by a synchrotron component while in the higher frequencies thermal dust lift the polarization SED up [70].

3.2.3 Anomalous Microwave Emission

In total intensity, the AME has been observed by the Q, U and I Joint Observatory in TEnerife (QUIJOTE) and Planck in the frequency range $\approx$ 10-60 GHz [71]. A possible explanation of this emission is represented by the spinning of the dust grains, which rotate at GHz frequencies and emit electric dipole radiation if they have an electric dipole moment [72], or magnetized dust grains and free-floating ferromagnetic material [73]. The AME SED is expected to exhibit a bell shape form, characterized by a peak at around 30 GHz, associated to the frequency of the grain rotation. If the AME is polarized, its polarization fraction must be very small, at the level of a per cent [71]. QUIJOTE [74] has constrained the AME polarization to be less than 2.8% with a 95% confidence level in the Perseus molecular complex. In another paper [75], the upper limit correspond to 0.39%, and by combining the data with WMAP, it tightened the constrain to 0.22% for the W43 molecular complexes. Note that the aforementioned limits are measured for specific regions and cannot be applied to the whole sky. Remazeilles et al. [76] have shown that neglecting 1% polarized AME can bias the extracted r value, particularly for satellite missions.

The AME parameterization is based on the paper by Ali-Haimoud et al. [77]. The spinning dust grains with angular velocity ω of electric dipole moment μ can radiate as follows

$$P = \frac{2}{3}\frac{\mu_{\perp}^2 \omega^4}{c^3}\,, \tag{3.4}$$

where P is the radiation power and $\mu_{\perp}$ is the perpendicular component of μ with respect to the angular velocity. The emissivity of electric dipole radiation per Hydrogen (H) atom can be calculated through

$$\frac{I_\nu}{n_H} = \frac{1}{d\pi}\int_{a_{min}}^{a_{max}} da \frac{1}{n_H}\frac{dn_{gr}}{da} 4\pi\omega^2 f_a(\omega) 2\pi \frac{2}{3}\frac{\mu_{a\perp}^2 \omega^4}{c^3}\,, \tag{3.5}$$

where $\omega = 2\pi\nu$. The term $\frac{1}{n_H}\frac{dn_{gr}}{da}$ determines the grain size distribution function, which gives the number of dust grains per unit size per H atom, $\mu(a)$ is the electric dipole moments as a function of grain size and $f_a(\omega)$ is the angular velocity distribution function which depends upon the grain radius and environmental condition. This function is calculated for a cold neutral medium in the simulations we adopt. In this work, we adopt the standard model of the AME, constituting of simulated polarized maps with thermal dust polarization angles and nominal

AME intensity. The assumption of a complete mixture of small and big grains leads to consider the same angles as thermal dust. The AME polarization can be written as:

$$Q_{ame} = fI_{\nu}cos(2\gamma_{353}), \qquad U_{ame} = fI_{\nu}sin(2\gamma_{353})\,, \tag{3.6}$$

where f is polarization fraction. In this work, we have considered a global 2% polarization fraction, within the limits observed by WMAP, Planck, and QUIJOTE in Perseus.

3.3 Component separation methods

The component separation problem usually addresses the following question: given a number of observations of the sky at different frequencies, how can we isolate the CMB signal from all the different astrophysical processes contributing to the total observed emission? There are several ways to answer to this question but usually that is treated on the basis of statistical approaches, which assume that the observed emission just follows the linear superposition of a number of independent components.

Many different approaches have been studied in order to tackle component separation. Some methods are based on the internal template subtraction [78, 79], some exploit statistical independence of the sky components [80, 81], others invoke the maximum-entropy principle [82], or perform a parametric fit to the data [83–85]. The reconstruction of the different foreground emissions in a given CMB experiment are largely based on these approaches, Akrami et al. [see 11, for reviews]. In general, different component separation methods can be divided into three main classes: Parametric, Blind, and Template-fitting, and we give a basic description of those in the following Sections.

3.3.1 Parametric fitting

In the parametric approach to component separation it is assumed that the functional form of the frequency scaling for all relevant components is known, and all the prior knowledge and physical modelling of different foregrounds are exploited. The relative simplicity of its implementation makes this approach widely used, and as we mentioned already, the Galactic foregrounds from Planck have been

obtained using this technique. Moreover, it is characterized by a flexibility, and elegant mathematical casting, through a maximum likelihood form [86]. On the other hand, in the case of low signal to noise ratio on the pixel scale, the numerical efficiency drops and parameter estimation can face challenges. Also, given a large number of pixels in the CMB experiments with high sensitivity, the parametric technique should perform the non-linear and high-dimensional parameter fitting, which is computationally expensive. We describe here the approach developed by Stompor et al. [87] which is currently used for quantifying the science outcome of future B-mode probes [see e.g. 9, and references therein]. We refer to these papers for further details on this approach, limiting ourselves to the definition of quantities of relevance for the present work. The data model is usually written as

$$d_p(\nu) = \sum_c a_p^c(\nu) s_p^c + n_p(\nu) \equiv \mathbf{A}_p s_p + n_p \,, \tag{3.7}$$

where d_p contains measured signal at each frequency ν and sky direction p, summed over all components whose amplitude is written as s_p^c. $\mathbf{A}_p$ is the mixing matrix which contains the parametric SED model to fit, depending in principle on the sky direction, and n_p represents the noise. The component separation process consists in obtaining an estimate $\tilde{s}_p = \mathbf{W}_p d_p$ of the components, by means of a kernel operator $\mathbf{W}_p$, given by:

$$\mathbf{W}_p \equiv (\mathbf{A}_p^T \mathbf{N}_p^{-1} \mathbf{A}_p)^{-1} \mathbf{A}_p^T \mathbf{N}_p^{-1}, \qquad \mathbf{N}_p \equiv {n_p}^T n_p \,, \tag{3.8}$$

where $\mathbf{N}_p$ represents the noise correlation matrix. The kernel operator is the result of the maximization of the likelihood:

$$-2 \log \mathcal{L} = -\sum_p (d_p - \mathbf{A}_p s_p)^T \mathbf{N}_p^{-1} (d_p - \mathbf{A}_p s_p) \,, \tag{3.9}$$

which is valid in the case in which the noise is block diagonal, i.e. correlations are allowed between Stokes parameters in a given pixel only. The χ^2 is defined as:

$$\chi^2(p) = \sum_{\nu=1}^{N_{band}} \left(\frac{d_\nu - s_\nu(p)}{\sigma_\nu(p)} \right)^2 \,, \tag{3.10}$$

where $\sigma_\nu(p)$ represents the uncertainty due to the presence of noise. The corresponding approach to component separation has been implemented into the

publicly available code called ForeGround Buster (`FGBuster`)[1].

3.3.2 Blind component separation and template fitting

As opposed to parametric fitting, blind methods apply minimal assumptions about the components and suppose that the sky components are statistically independent. For these methods, one of the emissions is the signal of interest (CMB), and the rest are the unwanted foregrounds. The Internal Linear Combination (ILC) [80] belongs to this category. This scheme uses only the CMB column of mixing matrix $\mathbf{A}$ in Eq 3.7 to minimize the variance of the cleaned map. The main advantage of this method is represented by the minimal assumptions concerning the unknown or poorly known astrophysical components and the instrumental noise. The only requirement is that the component of interest must have a known emission law, and it must un-correlated to the contaminating foregrounds. This approach is widely used in CMB experiments like WMAP [88], but it introduces a well-known bias that comes from an empirical correlation between the CMB and the foregrounds [89].

Also in the category of blind approaches, the Independent Component Analysis that (ICA) uses just the statistically independent property of the different components to recover the full mixing matrix $\mathbf{A}$. This methodology has been cased in the spectral domain, through the Spectral Matching ICA (SMICA) method [90] solving the problem of measuring parameters of modelled multi-component spectral covariances, using empirical covariances computed on multi-detector data sets and in this way, works in a smaller dimension space which contains foregrounds.

Finally, the template fitting is a pixel based algorithm which does not model the emission laws, and the analysis is done by maximizing the likelihood over CMB and the amplitudes of each foreground component. [91]. The Spectral Estimation Via Expectation-Maximization [92] builds foreground templates with a combination of a subset of the input frequency maps to add more constraints and to reduce the foreground subspace size consequently.

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

3.4 Status of B-mode CMB observations and future probes

An intense and global effort is currently ongoing towards the measurement of the B-mode polarization. Lensing B-modes have been detected for the first time by the South Pole Telescope [SPTpol, see 93, and references therein] through cross-correlation, and directly by POLARBEAR [94]. Moreover, they have been observed by the Planck satellite [95], the Background Imaging of Cosmic Extragalactic Polarization 2 (BICEP2) [5], the Atacama Cosmology Telescope (ACT) [96]. On the other hand, only upper limits exist so far for the amplitude of the cosmological GWs, corresponding to $r < 0.06$ (at 95% confidence level) by the combination of BICEP2/Keck and Planck data [52]. Only recently, a tighter constrain, $r < 0.044$, has been presented by adding Planck 2018 polarization data to BICEP2/Keck 2015 data release [10].

Figure 3.5 shows the status of CMB measurements and their accuracy limits. Planck 2018 (in black) determines the TT and EE power spectra very well while SPT and BICEP2 and Keck array has been used to have a better picture in the BB power spectrum.

In relation to foreground studies, a significant progress on the knowledge of the B-mode emission has been possible through low frequency surveys. SPASS [63] is surveying the entire southern sky at 2.3 GHz, aiming at characterizing the synchrotron polarized emission for progressing on the knowledge of the CMB foregrounds, as well as the Galactic magnetism. C-BASS measures both the brightness and the polarization of the sky at 5 GHz by two telescopes, one at the Owens Valley Observatory (OVRO) in California, and the other in South Africa [97]. QUIJOTE is an ongoing project, started in 2012 in the northern hemisphere at the Teide observatory, by having four different channels 11, 13, 17, and 19 GHz. The focus of this experiment is on measuring CMB polarization and Galactic emissions, from synchrotron and AME [98].

Future CMB experiments will mostly focus on the detection of primordial gravitational waves and the improvement of our knowledge from reionization history, which we can extract from CMB polarization. The list of all the upcoming CMB experiments is long. We will report here the ones which are more relevant in our work, to be considered in the following Chapters.

Large Scale Polarization Explorer (LSPE): LSPE is the first large scale experiment after Planck which is designed to observe the B-modes of CMB polarization. LSPE is made by two instruments: (i) a ground-based telescope, the Survey TeneRIfe Polarimeter (STRIP), observing at 44 GHz, plus a 95 GHz channel for atmospheric measurements from the Teide observational site in the North pole, (ii) a balloon-borne instrument, Short-Wavelength Instrument for the Polarization Explorer (SWIPE), which will observe the sky from the Arctic stratosphere at 145, 220, and 240. STRIP will observe the sky, starting from summer 2021 for two years, instead, for SWIPE two weeks flight is scheduled during winter 2020/21. In Chapter 4, we explain the scientific goals and specifications of the LSPE, base on a recent publication [99], which represent the context in which we tested our implementation of the parametric component separation.

Lite (Light) satellite for the studies of B-mode polarization and Inflation from cosmic background Radiation Detection: LiteBIRD is the first satellite mission after Planck with the goal of observing the CMB polarization over the full sky with unprecedented precision. LiteBIRD is selected by JAXA[2] with international collaboration of ESA[3] and NASA[4] and Canada. Three years of a full-sky survey from a Lagrangian point L2 are planned by launching LiteBIRD in 2027 at 15 frequency bands between 34 and 448 GHz. The most important scientific target of LiteBIRD is achieving $\delta r < 0.001$, following B-mode observations in at large angular scale which corresponds to $2 \leq \ell \leq 200$ in terms of multiple, when δr is the total error on tensor-to-scalar ratio r. The high number of frequency channels would enable LiteBIRD to study and model CMB galactic foregrounds precisely. [100, 101]

Simons Observatory (SO): the SO will be put in the Atacama Desert in Chile and will observe the south pole sky measuring the CMB temperature and polarization anisotropies in six different frequencies from 27 GHz to 280 GHz. SO will include three small-aperture 0.5 m telescopes (SATs) and one large-aperture 6 m telescope (LAT) in the initial phase. SATs will observe 10% of the sky at the largest angular scales observable from Chile, while the LAT will scan 40% of the sky at arcminute angular resolution. SO's target is to measure primordial tensor-to-scalar ratio, r, at the level of $\sigma(r) = 0.003$. For the moment, SO is at the last stages of construction and it is planned to be in operation from 2021-22 [102, 103].

[2]Japan Aerospace Exploration Agency
[3]European Space Agency
[4]National Aeronautics Space Administration

CMB-S4: The next generation of ground-based CMB experiments is called "stage VI", CMB-S4 that is consists of highly sensitive telescopes at the South Pole, the high Chilean Atacama plateau, and possibly northern hemisphere sites. CMB-S4 targets very high sensitivity by putting all of these telescopes together and having more than 500,000 detectors. Its final aim is to detect the primordial gravitational waves for $r \geq 10^{-3}$ with 5σ. It has been predicted that by finishing 2020, CMB experiments will start entering the "stage VI" and by 2024, CMB-S4 will be complete. Apart from detecting primordial gravitational waves, CMB-S4 has the following subjects in its scientific goals: determining the number and masses of the neutrinos, constraining possible new light relic particles, providing precise constraints on the nature of dark energy, and testing general relativity on large scales [60]. The gray shaded area in Figure 3.5 shows the predicted improvement of CMB-S4 in measuring the EE and BB power spectra.

European Low Frequency Survey (ELFS): The ELFS is a proposal to the European Research Council, which would focus observations and foreground control, especially at low frequencies, in the interval ranging from 5 to 120 GHz. It would be all-sky, constituted by two Large (6 meters) telescopes, with resolution increasing from 20 arcminutes at 5 GHz to about a few arcminutes at 120 GHz. The survey would complement the CMB-S4 and LiteBIRD, and would solve the problem of monitoring the Galactic synchrotron completely.

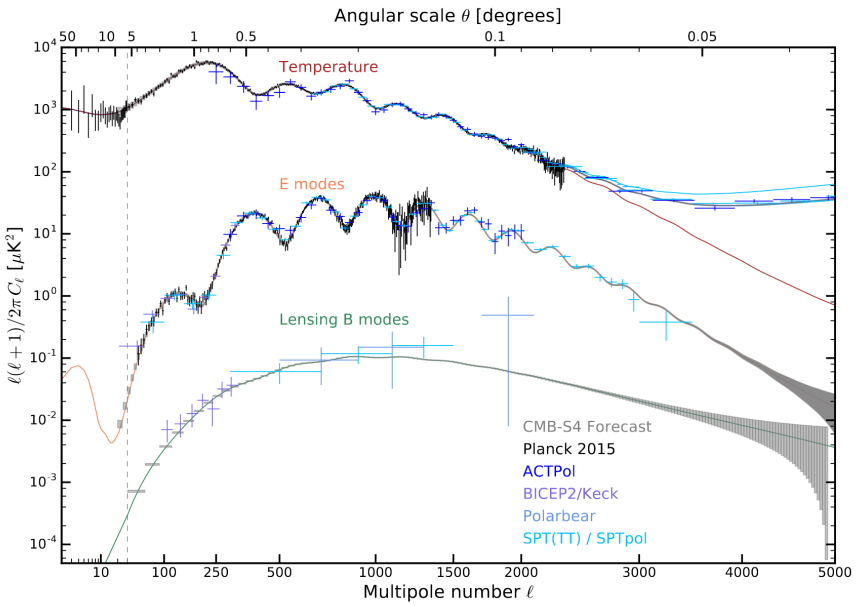

FIGURE 3.5: Current measurements of the angular power spectrum of the CMB temperature and polarization anisotropy from different experiments. Planck 2015, ACTPol, BICEP2/Keck, Polarbear, and SPT in black, dark blue, purple, light blue, and cyan respectively. The gray shaded area shows the predicted improvement of CMB-S4 for a ΛCDM with $r = 0$ cosmological model. From [60]

Chapter 4

Component separation for the LSPE experiment

In this Chapter we present the first amongst our set of original investigations. We focus on the most advanced parametric component separation technique presented in the previous Chapter and use it for evaluating the foreseen capabilities of one of the forthcoming B-mode probes. In Section 4.1, we give a broad overview on the scientific goals of the LSPE experiment, instrument specifications and project timeline. Section 4.2 specified to the explanation of the simulation done considering the foreseen instrument specifications, while in Section 4.3 the details of the component separation runs are shown. Finally, we describe the results and conclude in Sections 4.4, and 4.5. The content of this Chapter is based on the LSPE project's paper [99].

4.1 Overview of the experiment

As we described in Section 3.4, LSPE will focus on observation of the CMB B-mode polarization, with the following main objectives:

- detection of CMB B-modes polarization at a level corresponding to a tensor to scalar ratio r=0.03 with 99.7% confidence;
- setting an upper limit to tensor to scalar ratio r=0.01 at 68% confidence level;

- improving the measurement of the optical depth to the CMB τ, observed from the large scale E-mode CMB polarization;
- investigating the *low-ℓ anomaly*, a series of anomalies observed in the large angular scales of the CMB polarization, including lack of power, asymmetries and alignment of multiple moments [104, 105],
- preparing wide maps of polarized foreground produced in our Galaxy by synchrotron and thermal dust emission, which will be important to map the Galactic magnetic field and to study the properties of the ionized gas and the diffuse interstellar dust in the Milky Way;
- studying the quality of the atmosphere at Teide Observatory (Tenerife) for CMB polarization measurements.

LSPE has a wide range of frequency to control the CMB foregrounds. It consists of two instruments: Survey TeneRIfe Polarimeter (STRIP) and Short-Wavelength Instrument for the Polarization Explorer (SWIPE). STRIP is a ground-based telescope which will be observing at 43 GHz at the Teide Observatory (Tenerife) for two years, investigating polarized synchrotron. Moreover, STRIP will have another frequency at 95 GHz that will be specific to do atmospheric studies. Figure 4.2 shows the schematic view of STRIP instrument which is a coherent polarimeter array with three-axis allowing the rotation of the optical assembly around the boresight direction. STRIP will be in operation from April 2021 to April 2023.

On the other side, SWIPE is a balloon-borne mission that will be working at 145, 210 and 240 GHz in a night Arctic stratospheric flight for two weeks. SWIPE has been optimized to be very sensitive to CMB polarization with one broadband channel matching the peak of CMB brightness. Since SWIPE has higher frequencies (210 and 240 GHz), the understanding and cleaning of thermal dust emission will be within its main scientific goals. In Figure 4.1, the optical system of SWIPE is shown. The detectors arrays are cooled at 0.3 K by a large wet cryostat, which also cools the polarization modulator and the entire telescope. SWIPE measurements are currently scheduled for Winter 2021/22.

The scanning strategies of both SWIPE and STRIP is shown in Figure 4.3 and the combination of two scanning strategies will produce full-frequency coverage over 37% of the sky. The specifications of STRIP and SWIPE are written in Table 4.1.

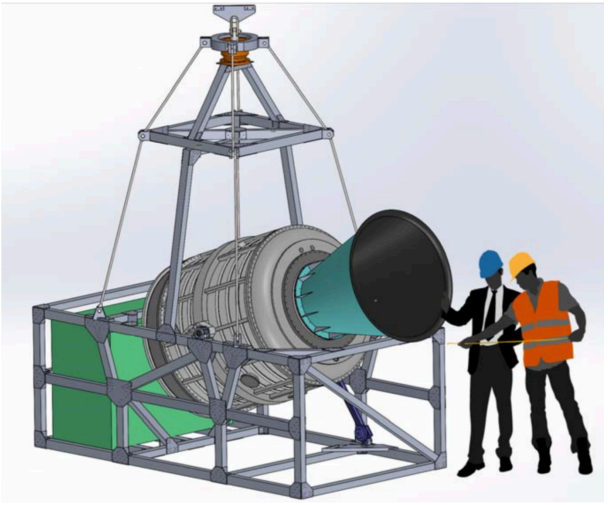

FIGURE 4.1: SWIPE optical system overview. The instrument that contains the optical elements is equipped to a large Helium cryostat.

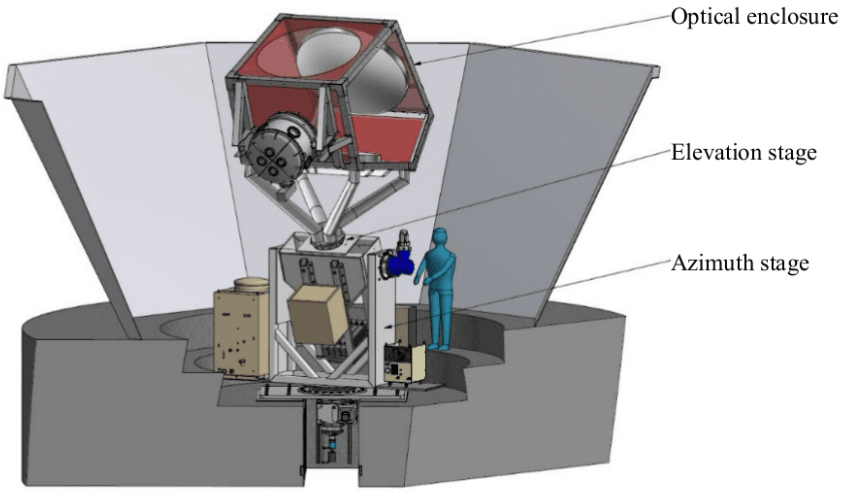

FIGURE 4.2: STRIP schematic optical system overview. This three-axis mount will be installed in the Teide observatory.

Instrument	STRIP		SWIPE		
Site	Tenerife		balloon		
Freq (GHz)	43	95	145	210	240
Bandwidth	17%	8%	30%	20%	10%
Angular resolution FWHM (arcmin) .	20	10		85	
Number of detectors $N_{\rm det}$	49	6	162	82	82
Observation time	2 years		8 - 15 days		
Sky coverage $f_{\rm sky}$	37%		38%		
Map sensitivity $\sigma_{Q,U}$ (μK$_{\rm CMB}$ · arcmin)	102	777	10	17	34

TABLE 4.1: LSPE baseline instrumental parameters.

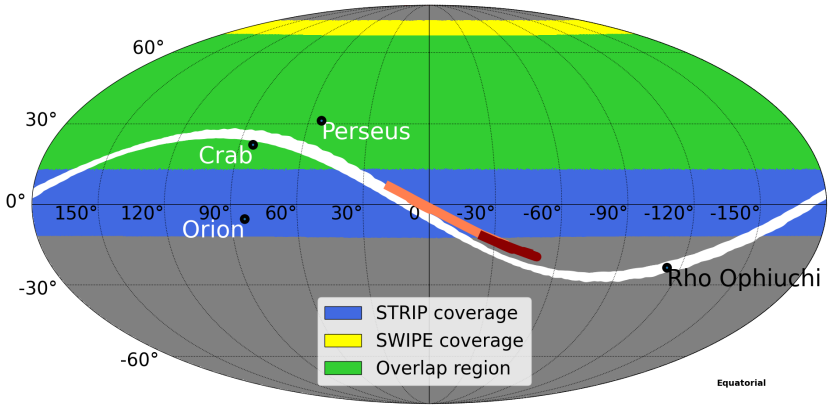

FIGURE 4.3: LSPE scanning strategy in Equatorial coordinates, The yellow and blue areas represent the SWIPE and STRIP sky coverage respectively, while the green area shows their overlap. The map also shows the position of the Crab and Orion nebulas, of the Perseus molecular cloud and the trajectories of Jupiter (orange), Saturn (dark red) and the Moon (white) from April 2021 to April 2023.

4.2 Simulations

In this Section, we explain the procedure concerning the simulations of the sky maps needed for the component separation pipeline. Our multi-frequency sky maps consist of 1000 CMB realizations and the basic model of thermal dust and synchrotron as polarized foregrounds. The CMB maps are simulated by CAMB code with the cosmological parameters based on Planck 2018 paper [22] and the optical depth $\tau = 0.06$ and tensor-to-scalar ratio $r = 0$. For the thermal dust and synchrotron emissions, we exploit the publicly available package `Python Sky Model (PySM)` [1] which generates the full-sky simulation in intensity and polarization [106]. The synchrotron brightness is modeled as a power law decaying in frequency with a constant spectral index $\beta_s = -3$ as we explained in equation 3.2. The dust component is modeled as a grey body, i.e. an almost thermal component at a temperature of $T_d = 20\,\mathrm{K}$, heated back by starlight, represented by a frequency dependent optical depth, with spectral index $\beta_d = 1.54$ in Equation 3.3. We also include noise realizations according to the SWIPE and STRIP sensitivity. The *rms* of the sensitivity maps are reported in Table 4.1. We did not consider any correlated noise from a pixel to another in our work.

Together with the LSPE bands, we also consider the observations of the Planck satellite between 30 and 353 GHz [4], and the ones of QUIJOTE at 11 GHz

[1] https://github.com/bthorne93/PySM_public

[98, 107]. In order to evaluate the power of the LSPE, we have tested different setups, by considering different combinations of LSPE bands with Planck LFI, HFI and QUIJOTE low-frequency band. In the minimal case, we have contemplated LSPE frequencies and only 30 GHz band from Planck which is reported in Table 4.3. All maps, including noise realizations, use the Hierarchical Equal Latitude Pixelization (`HEALPix` and its Python package (`healpy`)[2] [108] with resolution parameter $N_{\rm side} = 128$, corresponding to about half a degree in the sky. In the component separation runs, in order to deal with frequency channels at different resolutions, we smoothed all the component maps to a Gaussian beam with 85 arcminutes FWHM, which is the largest beam associated to LSPE channels.

4.3 Component separation pipeline

For the analysis presented in this work, we consider the component separation apparatus represented by the `ForeGroundBuster`[3], which is currently used to assess the foreground cleaning capabilities of a number of CMB B-mode probes [9, 85]. This pipeline makes use of parametric component separation method as we explained in Section 3.3.1. The method fits the multi-frequency maps, in each pixel observed by both STRIP and SWIPE, for CMB signal, amplitude, spectral indices and curvature of synchrotron, temperature, amplitude and emissivity of dust.

The component separation procedure is performed only on polarization maps and on the masked region which is the overlap of SWIPE and STRIP. This method recovers the value of the spectral indices β_s, β_d (a single value for the full map), as well as the amplitude of the thermal dust $A_d(\hat{n})$, and synchrotron emission signal $A_s(\hat{n})$. In this procedure, after component separation, we extract the CMB signal which has the noise instrumental effect. The EE and BB power spectra are calculated for these CMB maps and after that, by subtracting the extracted CMB map from the input one, the residual maps and noise power spectra are computed, respectively. In the upper panel of Figure 4.4 we show one sample of CMB output map after the component separation. In lower panel of this Figure, the average of 1000 residual maps has been shown. Very faint residual of the Galactic foregrounds can be seen in the residual map which are the consequence

[2] https://github.com/healpy/healpy
[3] https://github.com/fgbuster/fgbuster

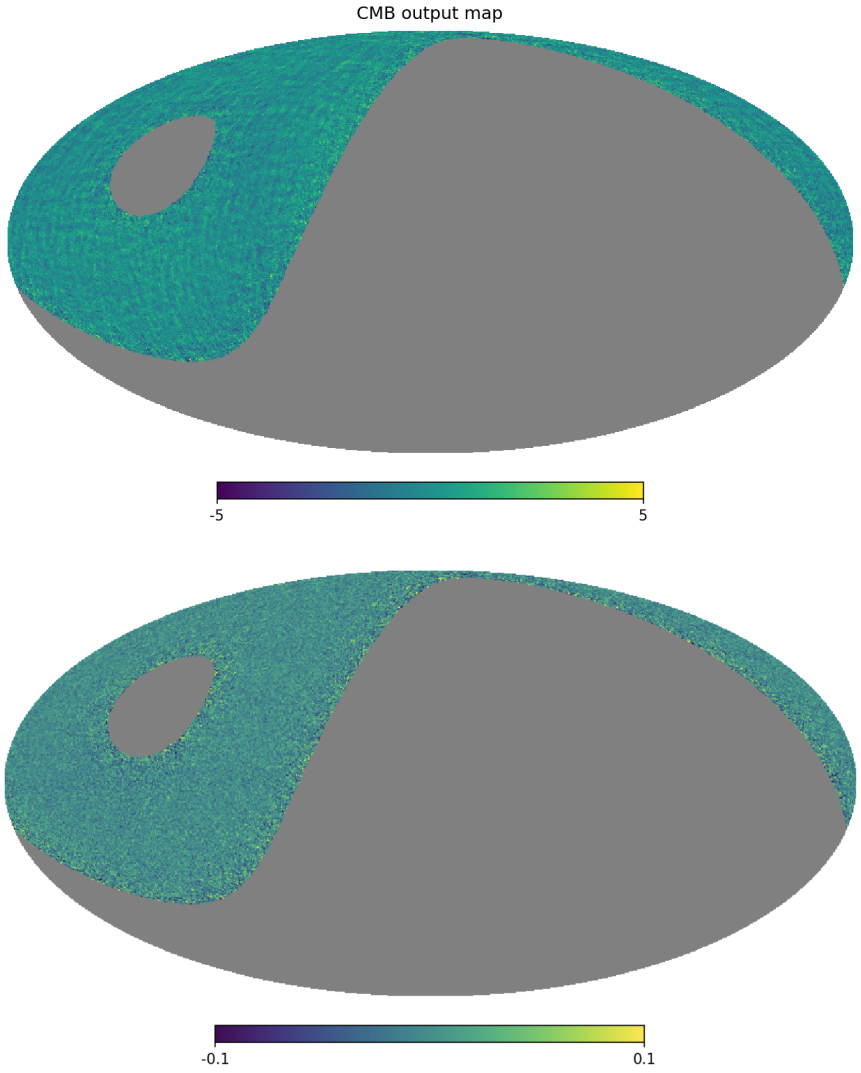

FIGURE 4.4: Upper panel: one realization of CMB map after components separation for the baseline case. Lower panel: the average of 1000 residual (input minus output) CMB maps.

of considering the simplest models for foreground emissions and the limitation of the fitting procedure adopted here.

By considering the sensitivity of STRIP at 43 GHz, it has been important to include low frequency channel of QUIJOTE 11 GHz and Planck 30 GHz, in order to constrain the spectral index of synchroton emission β_s. However, as we will explain in the next Section, for thermal dust spectral index β_d, SWIPE has the most sensitive channel of the LSPE and one could exclude Planck HFI channels without increasing the error bar on the results.

4.4 Results

In this Section, we present the result of component separation pipeline for the baseline and minimal case. A key element of component separation is the **W** matrix (in Equation 3.8), which is the linear operator that mixes the frequency maps in the component maps, taking into account the sensitivity and the contribution of each frequency to each astrophysical component. The elements $W_{i,j}$ of the **W** matrix (often referred to as *weights*) admit negative values for frequencies that must be subtracted in order to solve for the astrophysical component.

Frequency bands and weights for each component are shown in Table 4.2 for our baseline. From this table it is clear that the 145 GHz channel is the most important one for reconstructing the CMB, because of the relevance of w_{CMB}, clearly adding sensitivity to the currently available datasets. The Table also shows quantitatively the relevance for what concerns the wings of the frequency interval for fitting and subtracting foregrounds. As we mentioned before, the LSPE 210 and 240 GHz are the most important channels for constraining thermal dust foregrounds and having larger numbers of w_{Dust}; on the other hand, for synchrotron the QUIJOTE 11 GHz and Planck 30 GHz are necessary. In Table 4.3, we show the minimal setup of our baseline, where we have used just the 30 GHz channel of Planck and LSPE frequencies. This table confirms our results with the same scheme of weights distribution; by decreasing the number of frequencies, the weights change in the component separation solver. Moreover, in Table 4.4 we show the accuracy of the component separation in terms of dust and synchrotron spectral indices. As an illustration, Figure 4.6 indicates the distribution of β_s and β_d as a result of the component separation. Fig. 4.5 shows the polarization CMB power spectra obtained by averaging 1000 simulations where the error-bars come from the noise power spectrum after component separation. The cut-off at $\ell \approx 110$, where error-bars increase dramatically, reflects the angular resolution of the combined instruments.

4.5 Summary

We conclude here this Chapter, as an application of parametric fitting component separation, to a very relevant and forthcoming CMB experiment, characterized

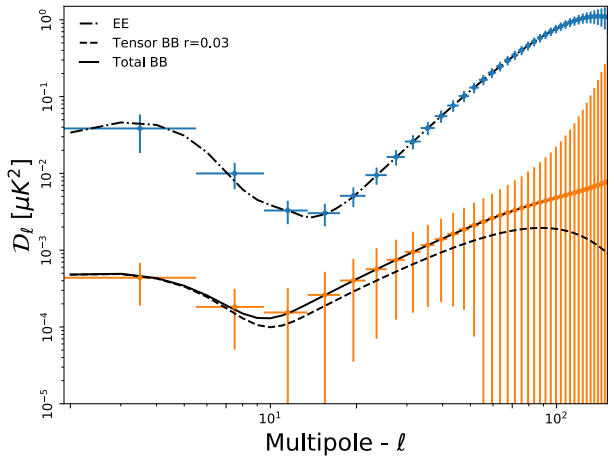

FIGURE 4.5: CMB E-modes (blue) and B-modes (orange) power spectra averaged over 1000 simulations after component separation simulated for the LSPE experiment.

Band (GHz)	Probes	w_{CMB} $\times 10^3$	w_{Dust} $\times 10^3$	w_{Synch} $\times 10^3$
11	Q	-1.1	0.24	56
30	P	2.52	-1.1	18
43	ST	4.43	-1.9	8.03
44	P	1.92	-0.82	3.2
70	P	2.85	-1.1	0.86
100	P	14	-5.3	0.41
143	P	26.4	-7.4	-2.2
145	SW	1226	-330	-107
210	SW	-130	204	5.6
217	P	-7.1	8.5	0.48
240	SW	-150	130	14
353	P	-9.9	6.5	1.2

TABLE 4.2: Component separation weights for each component in each channel, where P, Q, ST and SW stand for: Planck, QUIJOTE, STRIP and SWIPE respectively. The highest value for w_{CMB} shows that 145 GHz channel is the most important one in extraction of CMB signal.

by a marked multi-frequency, the LSPE. We have implemented an end-to-end simulation of the CMB and foreground components, along with running a full set of the components separation pipeline for the LSPE project. As a result, we foresee that the LSPE will be able to set an upper limit for tensor-to-scalar ratio r at the level of 10^{-2} and detect a $r = 0.03$ with 95% confidence level. This improvement in the CMB polarization and its B-mode component is extremely important. On the other hand, it is very important for more and more realistic studies to include more complicated Foreground models such as variation of dust and synchrotron spectral indices across the sky, synchrotron power-law with curvature, and considering two

Band (GHz)	Probes	w_{CMB} $\times 10^3$	w_{Dust} $\times 10^3$	w_{Synch} $\times 10^3$
30	P	-15	2.7	870
43	ST	-2.6	-0.45	390
145	SW	1400	-410	-1600
210	SW	-190	240	28
240	SW	-202	160	340

TABLE 4.3: Component separation weights for each component in each channel for the minimal case: Planck + LSPE (STRIP, SWIPE).

Parameter	Mean	σ
β_d	1.539	0.001
β_s	-2.999	0.002

TABLE 4.4: Dust and synchrotron spectral indices obtained by parametric component separation. The component separation algorithm fits for a single value in each map. The uncertainties are derived from the standard deviation of 1000 realizations of the noise in the maps.

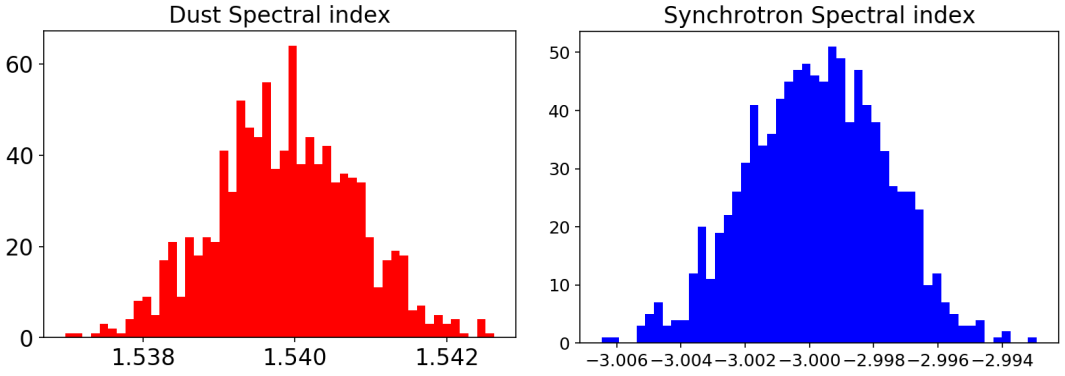

FIGURE 4.6: Histograms of dust β_d (left) and synchrotron spectral index β_s (right) distribution as the result of 1000 running of component separation pipeline.

populations of dust for thermal dust emission. Moreover, one could add AME component as a polarized component with a contribution less than a percent. These variations are present in the available sky simulations, and we will test these cases as well. On the other hand, in view of more complicated foreground models to consider, an interesting question arises, precisely consisting of how to be able to recognize, in a pre-component separation phase, the right modelization of the foregrounds to fit in the sky. Given this aspect's novelty, which will end up in one of the initial brick of a pre-processing component separation phase, dedicated to learning the physical processes in the sky before fitting them, we focus on this aspect in the next Chapter.

Chapter 5

Neural Networks for foreground model recognition

As we anticipated, in this Chapter, we are going to build up the case for the relevance of NNs in the context of the CMB data analysis and, in particular, component separation. In Section 5.1, we describe our motivation and goals from this project, and before going to the details of the work, we introduce the required concepts in the NNs fields in Section 5.2. After that, in Sections 5.3 and 5.4, we describe the implemented NN architecture designed for our specific problem and the detail of multi-frequency maps simulations. Later, we present our results by testing our NN in different cases, in the presence or absence of noise from Section 5.5 to 5.7; also, we compare our result with the state of the art of component separation method in terms of accuracy in Section 5.8. Finally, we bring up our conclusions and summarize the Chapter in Section 5.9. The results presented in this Chapter are based on the paper by Farsian et al. [109].

5.1 Motivation

As we explained in Chapter 3, in the last few years, it has become clear that one of the greatest challenges for the detection of primordial B-modes is represented by the control and removal of the diffuse emission from our own Galaxy. As a matter of fact, Galactic polarized radiation has an amplitude larger than the cosmological signal on the degree and super-degree scales, at all frequencies and in all the sky regions [see 11–13, and references therein]. In order to face this challenge and be

able to extract a clean cosmological signal, future CMB probes are characterized by a multi-frequency coverage, with very high sensitivity detectors in all the frequency channels. Along this line, and following the example of the LSPE, which we described in the previous Chapter, several observatories are currently being built. As we described in some more detail in Chapter 3, the SA [102] is observing along with others, and the SO [103] will start operations in 2022. On a longer term, the CMB-S4 [60] and LiteBIRD [110] are designed to reach an accuracy, after foreground subtraction, corresponding to the capability of detecting a B-mode signal with r as low as 10^{-3} with a high confidence level.

The computational procedure dedicated to the removal of diffuse foregrounds from the CMB signal is known as component separation, and consists of combining multi-frequency observations in order to reconstruct clean maps of the CMB as well as maps of each foreground emission. In particular, typical methods for component separation are based on parametric fitting of the multi-frequency maps, where the parameters are represented by the amplitude and frequency scaling of the different foreground components [87, 111]. A crucial aspect, and a necessary pre-processing of data with respect to component separation, in order to reach these extreme precision, is represented by the need of an accurate modeling of the foreground emissions and how the relevant parametrization might vary across the sky, as it is clearly shown in recent and comprehensive analyses concerning proposals of future satellite missions [112]. An incorrect or inaccurate modeling of Galactic emissions could indeed lead to high residuals in the final CMB maps, preventing the measurement of the faint B-mode cosmological signal [60].

This issue can be thought as a model recognition problem, which represents one of the most important applications of Artificial Intelligence (AI). NNs and Machine Learning (ML) in general, as a subset of AI, can be very useful in Cosmology and specifically in the CMB field. In particular, NNs are non-linear mathematical tools characterized by many parameters which are able to model different problems with high complexity. For this reason, they are widely used in scientific applications. In the recent years, several works include applications in this direction, ranging from estimating cosmological parameters from dark matter [113], to real-time multi-messenger astronomy for the detection of the GW signal from black hole merger [114] and weak lensing reconstruction via deep learning [115]. Recent applications, specific to CMB, include: foreground removal from CMB temperature maps using an MLP neural network [116], convolutional neural networks for

cosmic string detection in CMB temperature map [117], lensing reconstruction [118] and convolutional NNs on the sphere [119].

In this Chapter, we present a new NN application concerning the classification of the appropriate foreground model across the sky, identifying the physical parametrization which describes better a multi-frequency dataset in the different sky regions. This classification has to be seen as a pre-processing to the component separation phase, in order to instruct the latter with the proper functions to be exploited for the fitting. As a case study, in terms of frequency coverage, angular resolution, and sensitivity, we have considered the specifications of the complete frequency coverage of the LiteBIRD satellite [110] and the low frequencies channels of QUIJOTE [98]. For testing our NN model, we have focused on the analysis on the diffuse Galactic emissions which dominate the low frequency range, i.e. 70 GHz or less, in the CMB B-mode observations. Our goal is to study if a pre-processing model recognition phase is possible, and with which efficiency and accuracy.

5.2 Neural Network basic concepts

The concept behind NNs is quite old even though only in the last decade converts to a very hot topic and the related literature is flourishing [120]. The reason behind this popularity is the always increasing amount of available data along with the advancement of technology and computational power. The idea and development of the first NNs come from biological neurons back to 50's. The first NN, so-called *Perceptron* was designed by [121], by formalizing how biological neurons work.

5.2.1 Perceptron

The term *perceptron* was coined during the '50s and its way of working was already incredibly similar to how modern NNs work. Since the *perceptron* can be considered an ancestor of the modern NNs and since they have a lot in common, we'll describe how a simple *perceptron* works.

The whole *perceptron* can be represented by a function that takes a vector of inputs $\bar{x}$ and gives a binary output. For each input the perceptron has one internal parameter called *weight* denoted by $\bar{w}$ and a *bias* term b, that encodes the prior

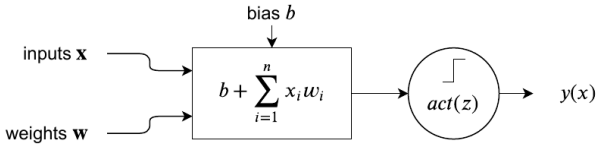

FIGURE 5.1: Schematic view of the perceptron.

knowledge. The *perceptron* does one very simple thing: it *weights* the external input $\bar{x}$ with the internal parameters $\bar{w}$, it sums them up along with b and it applies to the result an *activation function* that determines the final output of the *perceptron*. Nowadays *non-linear activation functions* are the most used although there is a vast literature about them [122].

The activation function of the perceptron is called Heaviside – a simple step function – and is defined as follows:

$$act(z) = \begin{cases} 1 & z \geq 0 \\ 0 & z < 0 \end{cases}$$

The output of the single perceptron will be given by

$$act(b + (\bar{x}^T \bar{w})) = act\left(b + \sum_{i=1}^{n} x_i w_i\right)$$

This definition is inspired by biological neurons and their electrophysiology: they receive some inputs, combine them and then, through the equivalent of the activation function, decide whether or not to propagate the signal [121].

5.2.2 Deep Neural Network

Like in the animal brains, we can achieve interesting flexibility of the model when we link multiple perceptrons together. In particular, during the years the community started following the approach of structuring the NNs in a layered fashion [123] although, in the recent years, researchers are exploring new ways of optimizing and structuring differently the NNs in order to gain in performance [124], [125].

What makes a neural network "deep" is actually the number of layers between the *input layer* and the *output layer* of the network. These layers are called *hidden*.

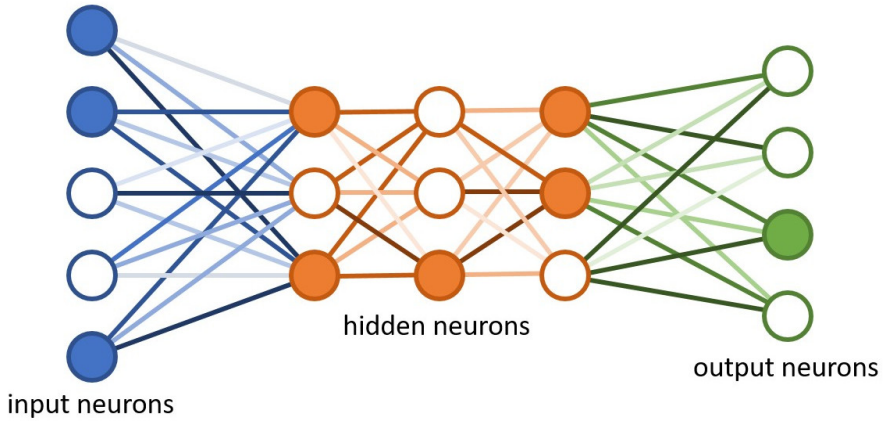

FIGURE 5.2: Schematic representation of a Deep Neural Network ans its layers.

The Deep Neural Network (DNN) are those kind of NNs with multiple (more than two) hidden layers. Schematic graph of a DNN is shown in Figure 5.2 with the input, hidden, and output layers.

The basic structure of NN is a *neuron*. Neurons are organized in layers; in each neuron a linear combination of all the elements of the previous layer is computed. These linear combinations are activated through a non-linear *activation function*, and the outputs of this operation become the inputs of the following layer. In the *input layer*, neurons take the value of the elements of the input x, while in the output layer the neurons take the values of the elements of $\tilde{y}$. For a general description of NN architectures can be found in [126].

In a layered NN, the outputs of the previous layer are the inputs to the next layer. With the exception of a few more sophisticated structures, NNs usually form an acyclic graph, known as *feed-forward network*.

It can be proven that a NN is a *universal approximator* of functions [127], this means that it is possible to approximate with arbitrary precision any measurable function depending on the number of neurons present in the NN. It is worth mentioning that choosing a *non-linear activation function* allows the NN to approximate even *non-linear* behaviours. This is usually a common practice, especially when using DNN.

5.2.3 Training of a Neural Network

Generally speaking, NNs are algorithms that recognize underlying relationships in a set of data [128]. Given a function f, that maps an input x into an output y, the goal of a NN is to find the best approximation f^* of f. In order to do that, the NN recursively applies non-linear functions to linear combinations of the input elements. In this way, the function f^* depends on several parameters θ (the coefficient of the linear combinations) which need to be optimized in order to get $f^*(\theta) \approx f$. This is done through a training set, i.e. a set of data for which the real output $y = f(x)$ is known: by computing the NN output $\tilde{y}$ for the elements of the training set, and by minimizing the distance between y and $\tilde{y}$, the best values for the NN parameters θ are found. The optimization is done numerically, usually with a *stochastic gradient descent* (SGD) method, which searches for the optimal parameters in the directions where the *gradient* is lower. The stochasticity is used to add noise to the trajectory and to avoid getting stuck into a local minima [129]. All these information are encoded in the function that determines the distance between y and $\tilde{y}$ which is called *loss function* [130].

The set of θ values which constitutes the best approximation of f is obtained through an iterative process, where the NN runs on the training set elements and the minimum of the loss function is found. Minimizing such *loss function* is not easy due to the high dimensionality of the problem and the underlying high non-linearity. The standard procedure that set a turning point in the world of NNs and that helps to solve efficiently this problem is called *backpropagation* [131]. The backpropagation algorithm computes, for each unit of the NN, the derivative of the error with respect to the weights in order to come up with the *gradient* of the error. Once the *gradient* has been computed, we use an *optimization algorithm* that minimizes the error and updates the weights of the NN accordingly. The values of the θ parameters are updated at each *epoch*. The number of epochs is one of the NN hyper-parameters and simply defines the number of iterations that are needed before the minimum of the loss function is reached. Given the very large number of parameters that a NN needs to optimize, over-fitting may occur; in this case, the NN approximates well the function f on the training set but it is unable to generalize to another set of data. In order to monitor the performance of the NN during the training, usually, a subset of the training set will be separated and used as the *validation set*. The validation set that has the known real output $y = f(x)$, can show how is the accomplishment of the NN on the unseen data and

if it is falling to the over-fitting problem or not. To avoid over-fitting, a typical approach is to introduce the so-called *dropout*, i.e. a mechanism for which, in each epoch, some of the neurons of the NN are randomly switched off. This prevents the NN to rely on any specific parameter and allow it to mitigate over-fitting.

After training the NN model with the training dataset, the saved model should be checked on an unseen set of data. This unseen dataset is so-called *test set*, which is independent of the training set but follows the same probability distributions, therefore if the NN model is able to fit the training data, it also fits the test set well.

5.3 Neural Network architecture

In this work, we have used NNs to recognize the actual parameterization of Galactic foregrounds in the sky. There exist several NN architectures. In this work we make use of the so-called *fully connected* ones. The architecture in this network is such that all the neurons, in one layer are connected to the neurons in the next layer. We have built the NNs in the *Keras*[1] environment, which is a *Python* library, with *Tensorflow*[2] backend. We have considered two NN architectures, which correspond to the problems we want to analyze, as described in the following.

As the purpose of this work is to solve a classification problem (assigning each pixel in the sky to a specific foreground model), the output of the NN is a vector where each element gives the probability that the input pixel belongs to any of the considered classes (models). The dimension of the output vector depends on how many possible sky models are considered, as explained in the following Sections.

5.3.1 Architecture for Binary classification

In our problem, the input of the NN are vectors of dimension 2×17. Each element of this vector represents the amplitude of the sky signal in a given pixel at the different considered frequencies (17 in total) for one of the polarization Stokes parameters. The two vectors of 17 elements each for Q and U are then stacked together to get the 34 elements long input vector.

[1] https://keras.io
[2] https://www.tensorflow.org

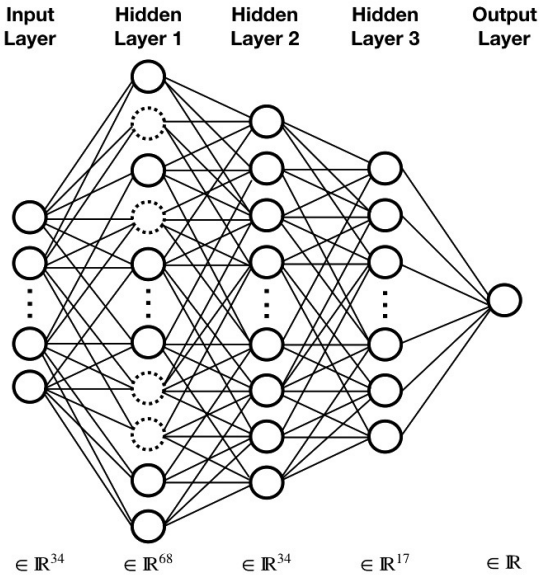

FIGURE 5.3: Schematic NN architecture used for binary classification: each circle represents a neuron, and the dashed circles indicate the application of dropout to a layer.

In the first considered case, we have trained the NN to perform a binary classification, meaning that its goal is to assign to each pixel in the sky one out of two possible foreground models. As we specified above, the NN input layer has dimension of 34, after that 3 hidden layers are present, including 68, 34, and 17 neurons each, with *tanh* as an activation function. In order to prevent overfitting, a dropout layer with a dropout rate = 0.5 is applied on the layer with the largest number neurons. Since we are in the case of binary classification, the output layer, activated with a *sigmoid* function, has, in this case, dimensionality 1, corresponding to the probability of the input to belong to the first class. Figure 5.3 shows the schematic architecture of our binary classifier. The loss function is defined as a binary-crossentropy function: $L = -(zlog(p) + (1 - z)log(1 - p))$, where p is the predicted probability for each input to belong to the specific class and z is the binary indicator associated to the two classes (0 or 1). We have used *Adadelta* optimizer with learning rate = 1.0 which is implemented in Keras. Adadelta is an adaptive learning rate method which dynamically changes the learning rate based on a moving window of gradient updates. Therefore this method needs no manual tuning of the learning rate during the learning and turns out to be robust to noisy gradient information [132].

5.3.2 Architecture for Multi classification

As we explain in the following Sections, we have also considered a case where the NN has to distinguish among four different sky models. Due to the enhanced complexity with respect to the binary classification, we increase the number of layers and neurons accordingly. In this case, the NN has 5 hidden layers with 272, 136, 68, 34, and 17 neurons, with *tanh* activation function. As before, a dropout layer with a dropout rate = 0.5 is applied to the first hidden layer with 272 neurons. The output layer is a multi-classification, with *softmax* as activation function. A Sparse-Categorical-Cross-entropy is chosen as loss function, corresponding to $L = -\sum_{c=1}^{M} z_{o,c} log(p_{o,c})$, where M is the number of categories for classification, p is the predicted probability for specific observation (o) of category c, and z represents the correct class indicator for that observation (o). The same optimizer as the binary classification is considered.

5.3.3 Hyper-parameters

The values of the hyper-parameters describing the architecture of a NN apparatus is usually determined empirically. That is the case of the number of layers and the number of neurons per layer. A large number of these quantities ensure performance, at the expense of computational efficiency and speed. Usually, large values of hyper-parameters are chosen and progressively reduced while keeping the performance stable, reaching minimum value which is then frozen in the NN apparatus. Moreover, the model is prone to overfitting problem by having large but not necessary number of neurons and layers. In our work, we have tried several NN configurations, and have selected, for both the cases of binary or multi-classification, the architecture which showed the best performance with the least number of parameters to be optimized during training. See [126] and references therein for a general description of the hyper-parameter definition and derivation for NNs.

5.4 Simulations

In this Section we describe the set up adopted to simulate the sky maps used to train and test our NN. As anticipated, we focus on low frequency foregrounds, and

we consider all the frequency channels covered by the future LiteBIRD satellite [110] plus the two lowest frequency bands with specifications of the QUIJOTE telescope [98]. Our results are conservative in this sense, because more powerful low frequency observations are being planned [133] and would results in more low frequency channels to be combined with LiteBIRD, and with more sensitivity. On the other hand, in this work we choose to see which results are achievable with the existing data. The corresponding frequencies, together with sensitivities and angular resolutions for all the considered channels are summarized in Table 5.1.

Therefore, the sky emissions included in our simulations are CMB, Galactic synchrotron, thermal dust and polarized AME. All the components are simulated using the `PySM`.

In particular, the CMB maps are generated as random Gaussian realizations of the Planck best fit Λ Cold Dark Matter (ΛCDM) power spectrum [134]. For dust, we have used the PySM template, rescaled as a modified blackbody, as in Equation (3.3), with constant spectral index $(\beta_d, T_d) = (1.54, 20K)$. For synchrotron, we have considered two different models. In the first one, the template is extrapolated in frequency with a simple power-law model. The spectral index is spatially varying, considering a Gaussian distribution with mean value β_s = -3 and standard deviation equal to 0.2. In the second case, a curvature is included in the synchrotron SED, with a constant value of $C = -0.052$, as indicated by Kogut [64] with 23 GHz as the pivot frequency; this setup is also compatible with the recent analysis by Krachmalnicoff et al. [13]. Finally, as explained in the following Sections, we have also included, in some specific cases, the AME polarized signal, assumed to have a 2% polarization fraction. The noise is simulated uniformly in the sky, through Gaussian realizations with standard deviations given by the parameters listed in Table 5.1. In Table 5.2 we have shown the summary of considered foreground models and their parameterizations. As an illustration of the relative relevance of the various components, in Figure 5.4 we plot the rms of their polarized intensity, in brightness temperature units and gridding the sky with 4 degree pixels, for all the sky emissions and frequencies considered in this work. We have applied Planck 2018 component separation common mask in polarization with $f_{sky} = 78\%$.

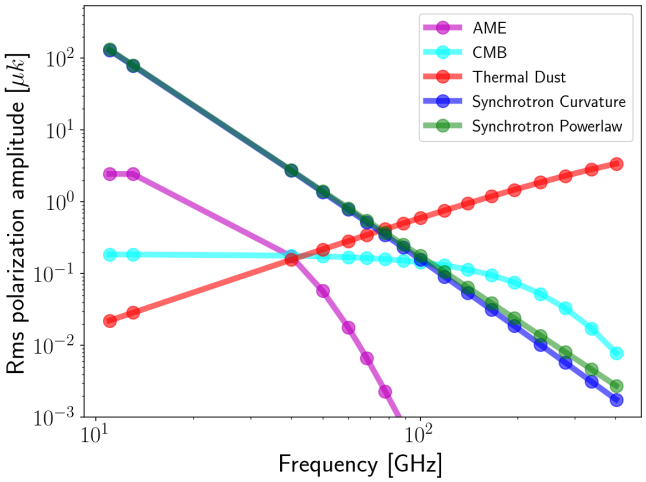

FIGURE 5.4: Polarized intensity rms amplitude as a function of frequency and different foreground component models which is used in our simulation. The plot refers to a sky fraction $f_{sky} = 78\%$ with 4 degrees pixel gridding, and units are in brightness temperature.

Experiment	Frequency [GHz]	Sensitivity [μK-arcmin]	FWHM [arcmin]
Quijote	11.0	840.0	55.2
	13.0	840.0	55.2
LiteBIRD	40.0	36.1	69.2
	50.0	19.6	56.9
	60.0	20.2	49.0
	68.0	11.3	40.8
	78.0	10.3	36.1
	89.0	8.4	32.3
	100.0	7.0	27.7
	119.0	5.8	23.7
	140.0	4.7	20.7
	166.0	7.0	24.2
	195.0	5.8	21.7
	235.0	8.0	19.6
	280.0	9.1	13.2
	337.0	11.4	11.2
	402.0	19.6	9.7

TABLE 5.1: Frequencies and instrumental specifications for QUIJOTE and LiteBIRD. The values are consistent with recent studies, [135], [9], respectively.

Foreground models	Parameterization
Synchrotron power-law	$\mu(\beta_s) = -3$, $\sigma(\beta_s) = 0.2$
Synchrotron curvature	$\mu(\beta_s) = -3$, $\sigma(\beta_s) = 0.2$, C = -0.052
Thermal dust	$\beta_d = 1.54$, $T_d = 20K$
AME	$f_p = 2\%$

TABLE 5.2: Summary of the foreground models considered in this work. The parameterization is based on Equation 3.2 for synchrotron, Equation 3.3 for thermal dust and Equation 3.6 for AME. $\mu(\beta_s)$ and $\sigma(\beta_s)$ are the synchrotron spectral index mean and standard deviation, respectively. f_p represents the polarization fraction.

5.5 Discerning two different foreground models (Binary classification)

We first use the binary classifier described in Section 5.3.1 to distinguish between two different foreground models. In particular, in the first case, we train the NN in order to understand whether low frequency data are fitted better by a synchrotron model, which does or does not include curvature of the spectral index (see Equation 3.2). Next, we focus on the case in which a pure power-law describes the synchrotron emission, and the NN is trained to recognize the presence of polarized AME.

5.5.1 Synchrotron with and without curvature

We have trained the NN with four sets of simulated multi-frequency maps. Each set consists in 34 maps, i.e. 17 frequencies for Stokes Q and U emissions. In each set we have included the emission coming from the CMB, polarized thermal dust and synchrotron simulated as described in Section 5.4. In two sets of maps, the synchrotron emission is scaled in frequency with a pure power-law, while in the remaining two a curvature is added to the spectral index. We have considered a different random realization of the CMB emission for each set of maps, as well as a different realization of the synchrotron spectral index spatial variation, which is taken from a Gaussian distribution with mean -3 and standard deviation 0.2. The synchrotron curvature in the two sets of maps is constant, with a value of $C = -0.052$, and 23 GHz as the pivot frequency. All the maps have been simulated at $N_{side} = 1024$, meaning that in total we have about 5×10^7 vectors, each of

which consists of 34 elements, which are used for training. Among these, we have randomly selected 20%, which are not used for optimizing the NN weights, but as a validation set, as it is typically done for validating the performance of a NN. The size of the training set has been chosen in order to find the optimal balance between NN performances and computational costs. Since we have considered all the pixels in the sky maps for training the NN, and given the high level of non-stationarity of the Galactic signals, the vectors in the training set cover a very large dynamic range, of about four orders of magnitude. As it is done in preparing the data for NN training, we normalized each input vectors in the range between -1 and 1 as follows: the minimum and maximum value for each input vector are computed; the minimum value is subtracted to the vector elements, and the result is divided by the difference between maximum and minimum. This procedure is shown in the following equation

$$std(x) = \frac{x - min(\mathbf{x})}{max(\mathbf{x}) - min(\mathbf{x})}, \quad scaled(x) = std(x)(1 - (-1)) + (-1) \tag{5.1}$$

Where $\mathbf{x}$ and x are a vector and a member of the vector from training and validation set respectively. And $scaled(x)$ is the scaled member of the vector inside the desirable range [-1, 1].

In order to further generalize the training set and make it substantially different from the test one, we have shifted the amplitude of each Galactic component. In particular, we have applied a multiplication to both the templates of synchrotron and thermal dust (at 23 and 353 GHz respectively): in each template, each pixel in Q and U is multiplied by a random value drawn from a Gaussian distribution with standard deviation equal to 30% of the amplitude of the pixel itself. The multi-frequency maps are then obtained by applying the correct frequency scaling to these modified templates.

In Figure 5.5 we show the training history with the accuracy reached by the NN as a function of epochs. Since we are working on a classification problem, in this case the accuracy represents the percentage of elements in the training (or validation) set which are classified correctly. We recall that the NN outputs the probability for each input pixel to belong to each considered class and that each pixel is assigned to the class that has the highest probability.

Once the NN is trained, we can apply it to the test set. In particular, we have built test maps, by making use of the PySM library, that include CMB, synchrotron,

and thermal dust. Maps of the test set have been generated at $N_{side} = 16$ and without the modulation of the foreground templates in order to make our test set considerably different from the training one. In some regions, the synchrotron emission has been scaled in frequency with a simple power-law, in others, we have modified the SED by including a running parameter of the spectral index. An example of a test set map is reported in Figure 5.6: in the pixels belonging to the red regions the synchrotron SED is a pure power-law, while in the blue region a curvature is added. The color scales in Figure 5.6 report the output of the NN, i.e. the probability that each pixel belongs to the correct class. In particular, pixels shown with darker colors are those where the NN assigned the correct class, while pixels with lighter colors are those where the NN has missed the right foreground model. For sake of clarity, in the right panel of Figure 5.6, we show, in white, the pixels where the NN has made an incorrect prediction. The achieved accuracy (i.e. the percentage of correctly classified pixels) is about 98%. We have tried different combinations of patterns for synchrotron power-law and curvature in the sky, assessing that the accuracy reached by the NN is stable and does not depend on the considered sky configuration.

We have also investigated the physical properties of those pixels where the NN assigned the wrong model. In particular, we have found that when the relative amplitude of the synchrotron emission over dust is small, the NN has the tendency to misclassify the model. This happens for example in the region near Galactic coordinate (230°, +40°) where the synchrotron amplitude is known to be extremely weak, or on the Galactic plane where dust emission is very bright. We have quantified this effect in Figure 5.7, where we show the fraction of misclassified pixels as a function of the relative amplitude of synchrotron over dust emission. In particular, we have considered a map at $N_{side} = 256$ (corresponding to about 7.8×10^5 pixels) where we have scaled the synchrotron emission with a pure power-law on the whole sky. For each pixel, we have computed the synchrotron over dust amplitude at the frequency of 11 GHz and for the total polarized intensity. We have applied a binning on this ratio such that in each bin we have the same number of pixels (about 1600). A threshold corresponds to each bin, and we have counted the ratio of misclassified pixels over the total number of pixels with $\log(A_{synch}/A_{dust})$ below the threshold. The results in Figure 5.7 show that when the synchrotron over dust amplitude is small, the fraction of misclassified pixels increases, up to about 38%, while for the pixels where synchrotron emission is high compare to dust, the faction of misclassified pixels decreases dramatically.

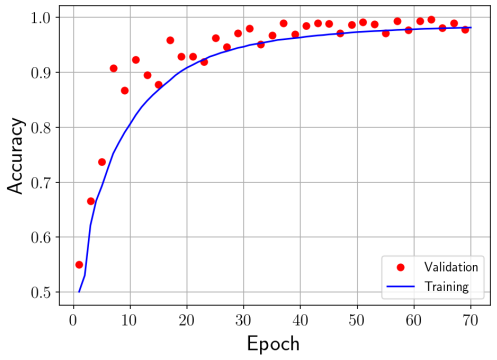

FIGURE 5.5: NN accuracy across training with respect to epochs for binary classification between synchrotron with and without curvature, in the noiseless case.

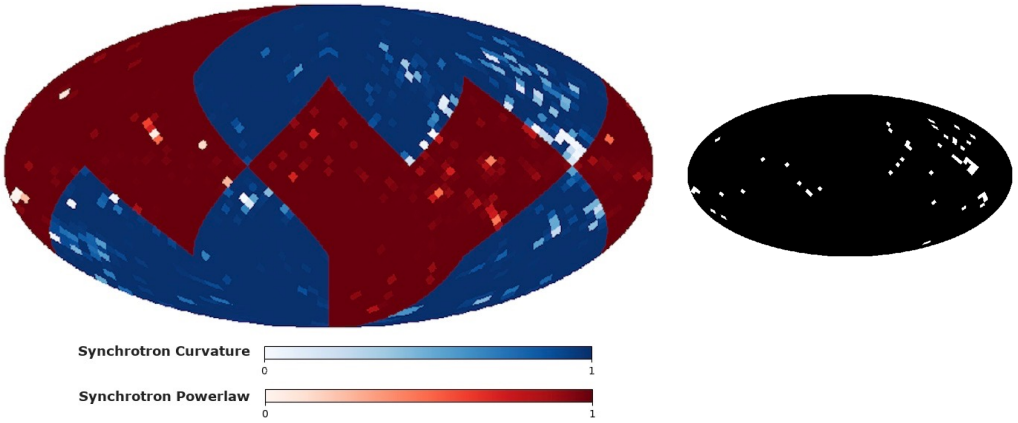

FIGURE 5.6: Left panel: NN prediction on test set for the binary classification for Galactic synchrotron with (blue regions) and without curvature (red regions) in the ideal case of noiseless maps. The color bar shows the NN probability assigned to the corresponding model across the sky. White pixels are those where the incorrect model is assigned. Right panel: For clarity, correct (black) and incorrect (white) pixels are also shown with a binary color scale.

5.5.2 Synchrotron and AME

We have used the same NN architecture developed for binary classification with the goal of identifying those pixels where AME polarized radiation is present in the sky. The two models considered in this case correspond therefore to Galactic synchrotron with a pure power-law SED, or synchrotron plus polarized AME component with the specifications described in Section 5.4.

For what concerns the training, we have followed a procedure analogue to the one presented in the previous Section. The training consists of four sets of maps; in two of them we have simulated the sky emission by considering the presence of CMB,

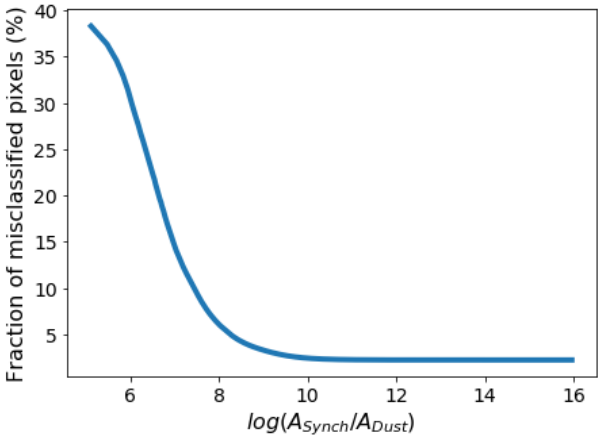

FIGURE 5.7: The faction of misclassified pixels respect to relative amplitude of synchrotron power-law to dust emission for binary classification of the Galactic synchrotron with and without curvature case at 11 GHz.

synchrotron and thermal dust radiation, while in the remaining two we have also included polarized AME. As before, the total number of vectors in the training set is about 5×10^7 and the templates of foreground emissions (dust, synchrotron and AME) have been modified by applying the multiplication factor as described in the previous Section. Results are presented in Figure 5.9, where AME is present are shown in green. In the ideal noiseless case, the NN is able to correctly classify the foreground model in about 97% of the cases. We highlight that pixels where the NN fails in classifying correctly the foreground models are those where the AME emission is faint with respect to the synchrotron one. In Figure 5.10 we report the fraction of misclassified pixels as a function of the relative amplitude of AME over synchrotron at 40 GHz (the frequency closest to the AME peak), similarly to what we have done for Figure 5.7. The results show that, as expected, the smaller AME amplitude is compared to synchrotron, the higher is the fraction of misclassified pixels, up to about 40%.

5.6 Discerning four different foreground models (Multi classification)

We now extend the study performed so far and consider a more complex case in which the NN is trained to classify four different foreground models in the simulated sky. In this case we have used the NN architecture described in Section 5.3.2. As before, we have built our simulated maps by including CMB and thermal dust,

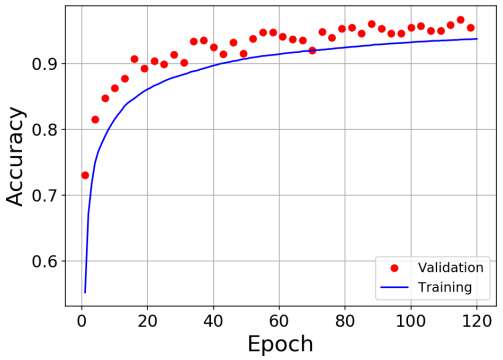

FIGURE 5.8: NN accuracy across the training with respect to epochs for binary classification between synhrotron power-law and AME, in the noiseless case.

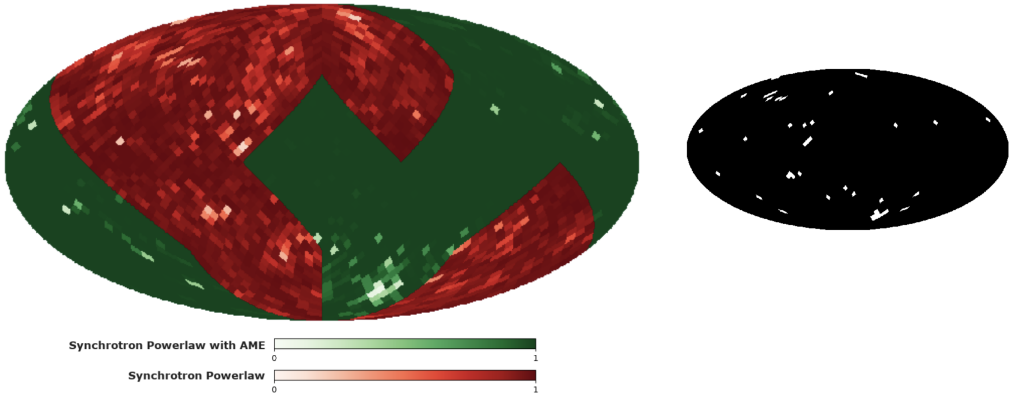

FIGURE 5.9: Left panel: NN prediction on test set for the binary classification for power-law synchrotron only (red regions) and without AME (green regions). The color bar shows the NN probability assigned to the corresponding and correct synchrotron model across the sky. Right panel: White pixels are those that the incorrect model is indicated by the NN with the highest accuracy. These results are for the noiseless case.

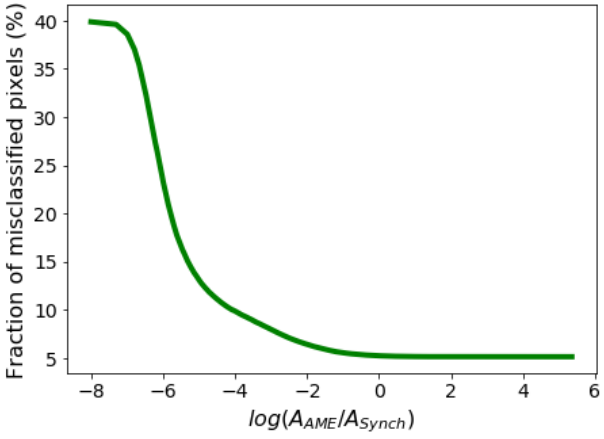

FIGURE 5.10: The fraction of misclassified pixels respect to relative amplitude of AME to synchrotron power-law emission for binary classification of the Galactic synchrotron in the presence of AME or not at 40 GHz.

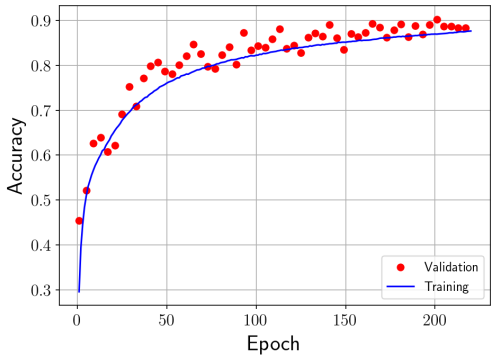

FIGURE 5.11: NN accuracy across training with respect to epochs for multi-class classification between synchrotron with and without curvature, with and without AME, in the noiseless case.

while the low frequency foregrounds include synchrotron with or without a curved SED and possibly AME.

The training set has been generated from four sets of maps as before, for a total of about 5×10^7 vectors used for optimizing the NN weights. The training history is shown in Figure 5.11: the NN reaches about 87% of accuracy on the training set after 220 epochs. It is worth noticing that as a result of the enhanced complexity in the simulations, the NN training takes more time to optimize weights.

Results on a test map are shown in Figure 5.12. In this case the sky is divided into four different regions, corresponding to the four models that the NN has to classify: synchrotron with a pure power-law SED (red), synchrotron with running of the spectral index (blue) and presence of polarized AME (green when AME is added to the synchrotron power-law model and purple when it is added to synchrotron with curvature). As before, color bars report the probability obtained by the NN that a given pixel belongs to the correct class, with lighter colors showing pixels where the NN has been assigned with the incorrect foreground model. The reached accuracy on the test set is at the level of about 93% and as before it does not depend on the specific pattern of models in the sky.

In Table 5.3 we report a summary of the performance of the NN in the different considered configurations. We notice that in some cases the accuracy reached on the test set is higher than the one on the training set, as it could happen as a consequence of having exploited dropout during training.

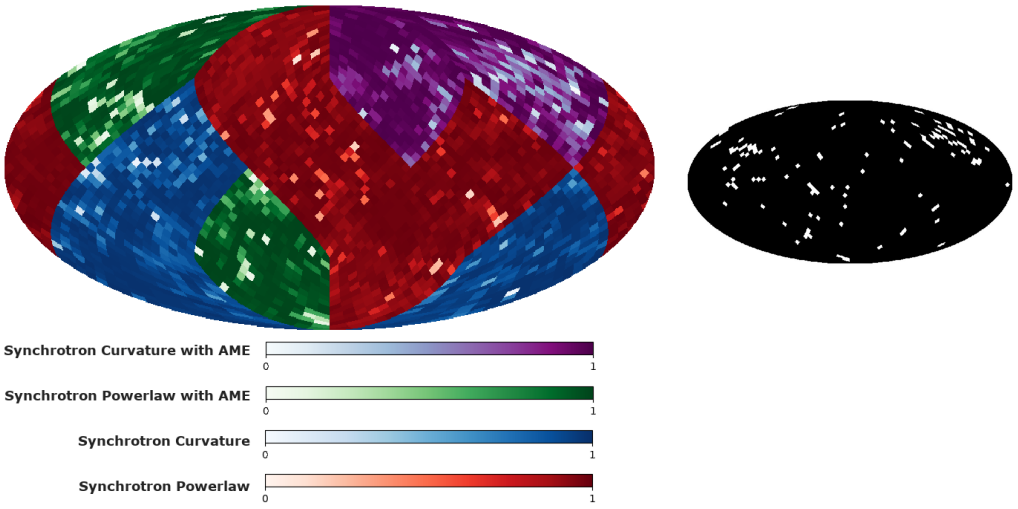

FIGURE 5.12: Left panel: NN prediction on test set for the multi-class classification for pure power-law synchrotron with (green regions) and without AME (red regions), curvature with (purple) and without (blue) AME. The color bars show the NN probability assigned to the corresponding and correct synchrotron model across the sky. Right panel: white pixels are those where an incorrect model is indicated by the NN with the highest accuracy. These results are for the noiseless case.

5.7 Classification in presence of noise

We have tested the performances of our NNs when instrumental noise is present on maps. In particular, we have considered the specification of the LiteBIRD and QUIJOTE experiments, with the sensitivities reported in Table 5.1 and uniform white noise distribution across the sky.

Our first approach has been to change only the test sets, by adding noise on the test maps, but keeping the weights of the NNs unchanged, therefore with the values optimized with the noiseless training. The first column of Table 5.4 reports the accuracy reached on the test sets for the three classification schemes we considered: binary classification for synchrotron models, presence of AME, and multi-classification. For the binary classification, we reached acceptable accuracy; While the accuracy drops significantly, reaching about 68% in the more complex multi-classification case.

In order to get better results, we have trained the NN with noise in the training set. We have considered two different approaches. In the first one, we have added one noise realization on the multi-frequency maps used previously as the training set. We have then taken the NN trained previously on noiseless data, and performed a

second phase of training with the noisy training set. In this way, the NN shows a remarkable improvement in accuracy, being able to reach $\sim 90\%$ on the test set for the multi-classification. In the second approach, we have built new training sets, consisting in 100 maps for each model at low resolution ($N_{side} = 16$), resulting in 400 sets of maps included in the training set, corresponding to more than 1 million pixels. Similarly to the previous case, the accuracy is pretty high, at the level of about 93%, proving that, during training, the NN is able to learn the noise properties and take those into account during the model classification.

In Figures 5.13 and 5.14 we show the results on the test map for the binary classification, for the case in which the training has been done with noiseless simulations (upper panels) and the one where the training set was obtained from low resolution maps (lower panels). Figure 5.13 is specified for the recognition of Galactic synchrotron emission with curvature or not in presence of noise while Figure 5.14 shows the classification of Galactic synchrotron emission in presence of AME or not when the test maps are noisy.

Instead, Figure 5.15 demonstrates the same setting for the multi-model classification. The upper panel shows the NN, trained with noiseless dataset, prediction on the noisy test map; while the lower panel shows the clear improvement of the NN prediction by training with noisy training set. A summary of all the results is reported in Table 5.4.

5.8 Comparison with chi-squared information

In this Section, we compare quantitatively the information retrieved via our NN apparatus with the ordinary goodness of fit represented by a χ^2 test following a parametric component separation analysis. We adopt in the rest of the work for calculating the χ^2 after component separation, using the same input maps used so far for the NN. We restrict this analysis to the classification in the simplest cases of pure power-law or curved SED for synchrotron, i.e. the first case analyzed in the previous Section, in the binary classification mode. We run `FGBuster` on the skies used to test the NN in the presence of noise, and calculate the χ^2 accordingly. For all the pixels we fit two different models: in one case, the parameters to fit with `FGBuster` are synchrotron, dust amplitudes and synchrotron spectral index, while in the other case, in addition to those, we also fit for synchrotron curvature.

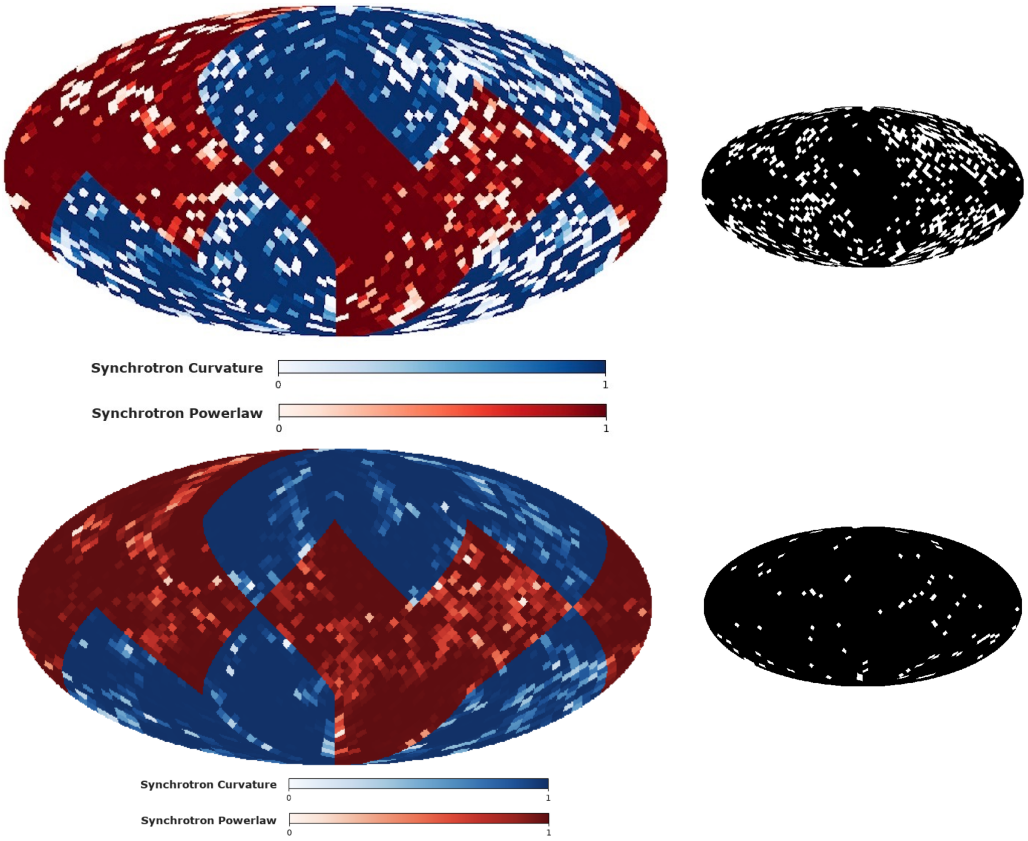

FIGURE 5.13: The effect of including the noise in the training set. The color scales for the considered models are the same as Figure 5.6. The upper panels indicate the NN accuracy on the noisy test set, and lower panels show the NN accuracy on the same noisy test set after re-training with 100 noise realizations at $N_{side} = 16$; As before, white pixels in the right panels are those where the NN indicates the incorrect model with the higher probability.

Since the parameterization of two synchrotron models is different, in order to have a fair comparison between the two χ^2 tests, we have computed the reduced χ^2 taking into account the degrees of freedom.

From the reduced χ^2 we compute the probability for each pixel to belong to the correct model that we show in the upper panel of Figure 5.16. As usual, darker colors indicate the pixels where thanks to the χ^2 computation we retrieve the correct model, while lighter colors are for those pixels where the classification is wrong. We compare the results obtained from the χ^2 with those of the NN (lower panel of 5.16, in the case where we have re-trained the NN with 100 noise realization at $N_{side} = 16$ (see Section 5.6). The reached accuracy calculated from reduced χ^2 is at the level about 73%, while the NN is able to distinguish two models with of 97% accuracy. This clearly shows the gain in using a NN for model recognition.

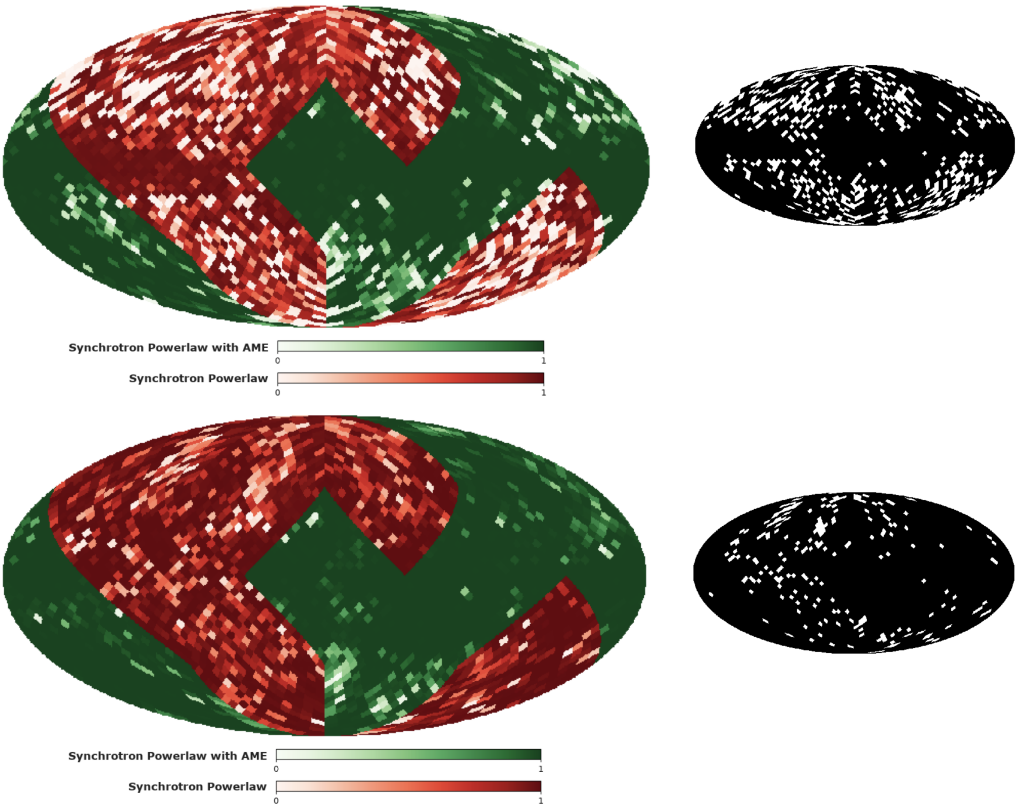

FIGURE 5.14: The effect of including the noise in the training set. The color scales for the considered models are the same as Figure 5.9. The upper panels indicate the NN accuracy on noisy test set and lower panels show the NN accuracy on the same noisy test set after re-training with 100 noise realizations at $N_{side} = 16$; As before, white pixels in the right panels are those where the incorrect model is indicated by the NN with the higher probability.

In Figure 5.17 we also show the difference between the χ^2 values computed in each pixel for the two different cases (with or without fitting for curvature) across the sky. As it is clear, the difference between the two reduced χ^2 is very close to zero in the region where the sky signal is low (greenish regions at intermediate and high Galactic latitudes). These are the regions where the χ^2 analysis leads to a higher probability of misclassification of the foreground model, due to the low signal-to-noise ratio. The same effect does not seem to affect the NN classification so strongly.

5.9 Summary and outlook

As outlined in this Chapter, throughout this PhD Thesis, we have started to investigate the relevance of NN in recognizing the physical properties of the diffuse

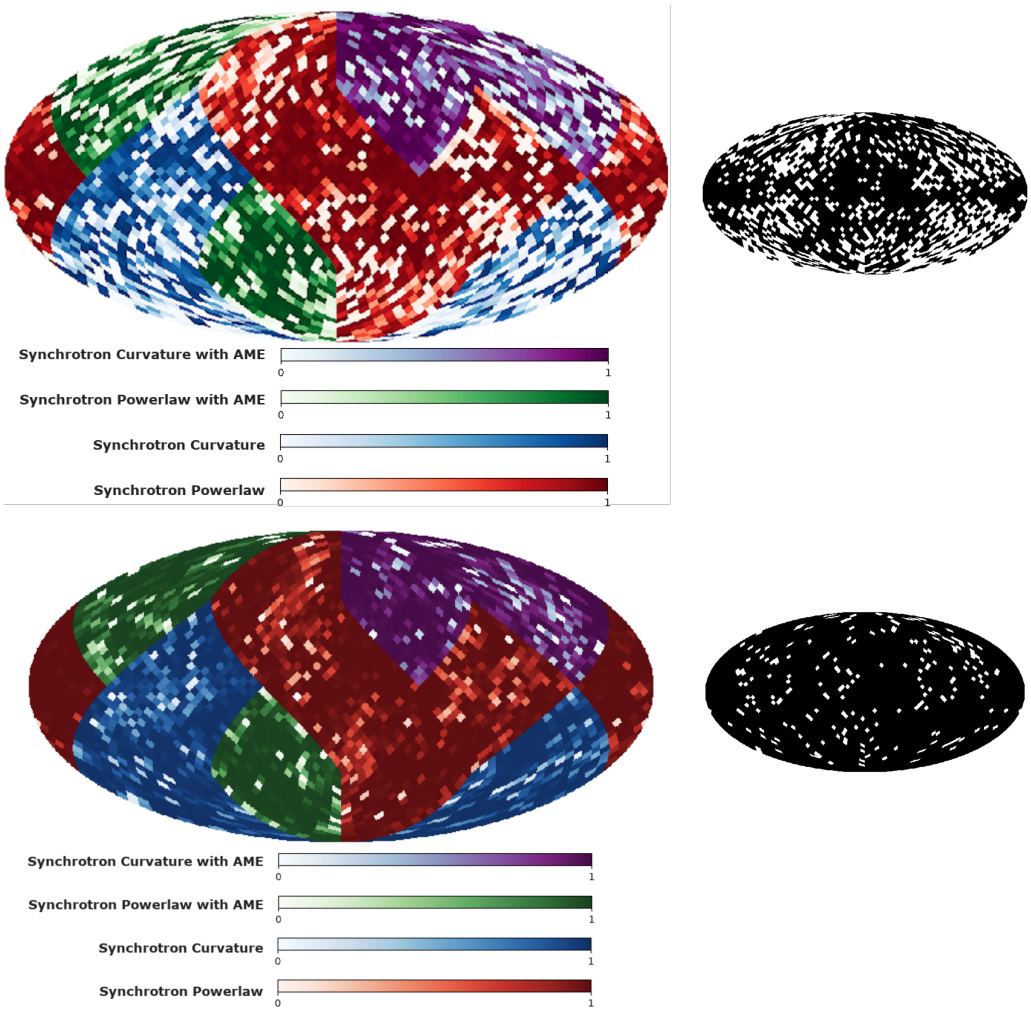

FIGURE 5.15: The effect of including the noise in the training set. The color scales for the considered models are the same as Figure 5.12. The upper panels indicate the NN accuracy on the noisy test set, and lower panels show the NN accuracy on the same noisy test set after re-training with 100 noise realizations at $N_{side} = 16$; As before, white pixels in the right panels are those where the NN indicates the incorrect model with the higher probability.

linearly polarized emission from our own Galaxy at microwave frequencies, which represents the main astrophysical contaminant to the measurement of the CMB B-mode polarization sourced by GWs in the early Universe. The problem is particularly challenging and urgent, due to the scientific relevance of the cosmological signal, and the difficulty in disentangle it from the much brighter foreground emission.

We want to summarize the main aspect of this work here and the future possibilities, as we think, this is the beginning of a new layer for foreground recognition and cleaning. The latter is usually performed via parametric fitting, which implies the necessity of identifying the physical parameters describing the foreground model

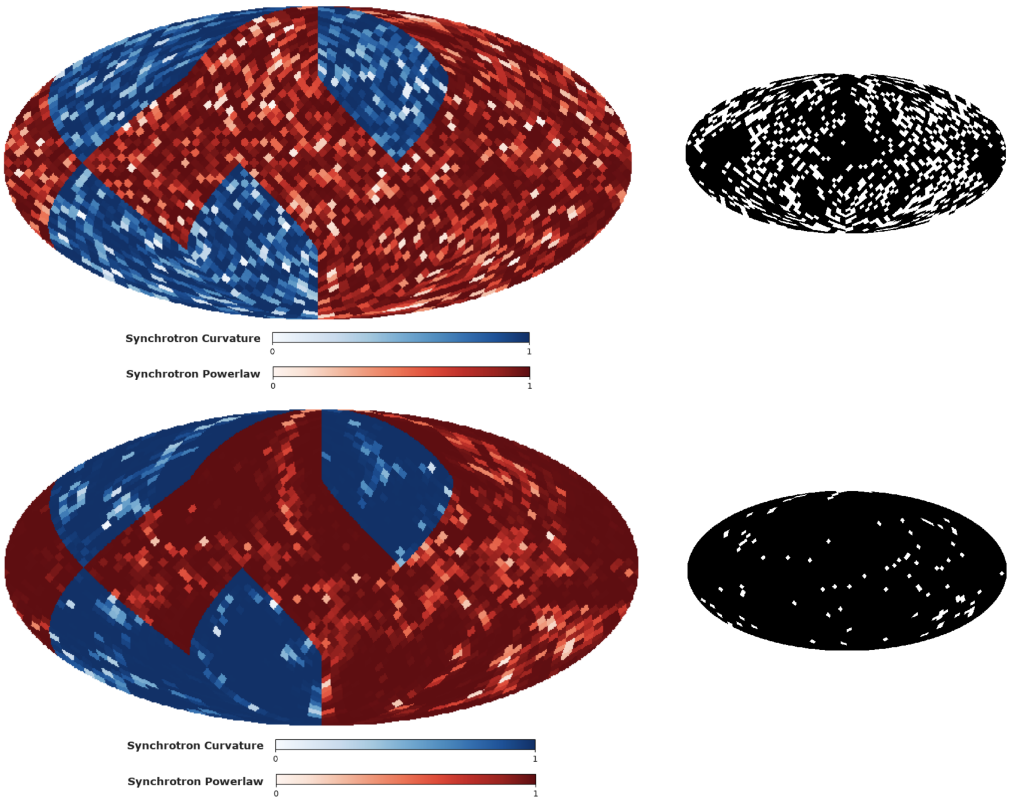

FIGURE 5.16: Comparison of χ^2 analysis with NN prediction in the presence of noise on test maps. Upper panels: on the left, we report the probability for each pixel to belong to the correct model as obtained via the χ^2 approach. In the red regions the correct model is represented by a synchrotron power law SED, while in the blue region a curvature is present. Lighter pixels are those where the χ^2 analysis leads to a wrong model classification (also shown in white in the right panel). Lower panels: same as the upper panels, but in this case the probability has been obtained via the NN approach. This comparison shows the advantage of using a NN approach, leading to a correct classification on about 97% of the pixels with respect to about 73% when the χ^2 information is used.

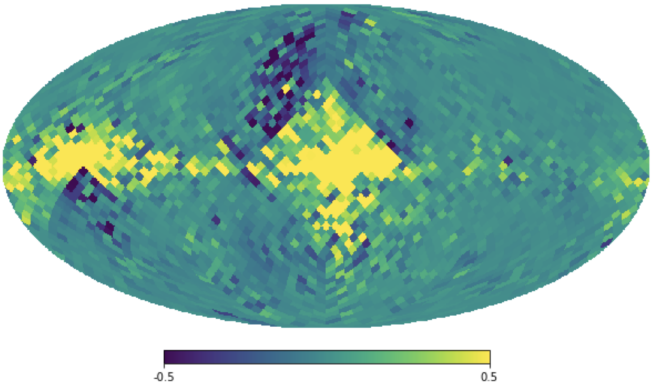

FIGURE 5.17: Reduced χ^2 difference for each pixel, obtained when the fit is done considering pure power-law SED for synchrotron and when the curvature is included. Pixels at intermediate and high Galactic latitudes (in green) are those where χ^2 is unable to distinguish between the two models.

in each portion of the sky, fitting and marginalizing them on the basis of a suitable multi-frequency coverage. On the other hand, foreground physical properties and model do vary in the sky, in a manner which is currently only partially revealed by observations, and yet crucial, because the right parametrization of them is necessary to perform a good fitting and to prevent the presence of large foreground residual in the CMB maps which could bias the scientific results.

Therefore, we study the possibility to identify the right physical parametrization of foregrounds, varying across the sky, in a pre-foreground cleaning phase. We do it with NNs, trained on simulations, and applied to test cases. We focus on the properties of Galactic synchrotron and AME, which have a rich phenomenology, resulting in possible different parametrization across the sky. We take care of making the simulations substantially different from observations, by explicitly and microscopically altering the training set with respect to the test one, at each resolution element. We find a good performance of the NN in recognizing the right parametrization of foregrounds, which achieve better results with respect to a standard χ^2 test on the goodness of fit, making our results interesting and suitable for future studies.

The combination of the simulations based on the specification of the QUIJOTE telescope and the LiteBIRD satellite, with a good coverage of the relevant frequencies, are analyzed in the binary and multi-class classification modes, i.e. when two and four models have to be recognized in the sky, respectively. In all cases, the rate of success in recognizing the right foreground model is equal or larger than 90%. This is true even in the case where four foreground models have to be recognized, namely pure power-law SED with or without curvature for synchrotron, with and without AME. We compare the NN information concerning model recognition with the χ^2 distribution following a parametric component separation assuming a given model, implemented and ran through the publicly available `FGBuster` code. We find that the NN performs better with respect to the χ^2, in particular at intermediate and high Galactic latitudes.

Moreover, we have repeated the same analysis done in this Chapter for a more complex case, and results are explained in Appendix A. In particular, we considered the thermal dust emission with a variable spectral index all over the sky, while in the simulation presented in this Chapter, this parameter was constant. It is interesting to see even by adding complexity to the foreground models, the NN

Sky models	Accuracy on training set	Accuracy on test set
Pure power-law & Curvature	99%	98%
AME & Pure power-law	93%	97%
AME & Pure power-law & Curvature	87%	93%

TABLE 5.3: Accuracy on training and test sets of the NN for different sky models in the basic configuration without noise.

Sky models	Acc on test set for training with noiseless data	Acc for Re-Training with $N_{side} = 1024$, 1 noise realization	Acc for Re-Training with $N_{side} = 16$, 100 noise realizations
Pure power-law & Curvature	82%	95%	97%
AME & Pure power-law	78%	92%	94%
AME & Pure power-law & Curvature	68%	90%	93%

TABLE 5.4: Accuracy of the NN for the binary and Multi-classification in presence of noise with different approaches for training.

accuracy for predicting sky models reaches higher than 90% for both noisy and noiseless cases.

We believe that these results are quite interesting, and a promising first step into the construction of a model recognition layer of data analysis in B-mode CMB measurements. Further lines of investigation for this work can be done by the extension of other foreground models, as well as the inclusion of possible realistic systematic effects. We will come back to these comments, and future directions, in the final Chapter of this Thesis.

Chapter 6

Generative Adversarial Networks for CMB maps

We will now turn to another application acting on multi-frequency CMB maps, and dealing with the control of the other foreground emission, due to point-like extra-Galactic sources, following their removal at the level of maps. We don't discuss here the actual point source removal of sources; for reviews of algorithms, testing, results and catalogues following point source extraction, Ade et al. [see 136, and references therein]. Instead, we focus on the post-processing concerning point source removal, i.e. the treatment of the regions where the sources have been removed in order to regain the sky coverage. As in the previous chapter, here we consider, implement, test an innovative approach, by considering GANs to fill in the masked regions due to point source removal. In Section 6.1, we describe the context and motivation. We explain the basic concepts concerning the generative models and the NN architecture in Section 6.2, and its implementation and testing in Section 6.3. We discuss our methodology and results in Section 6.4, and we conclude the chapter by summarizing the results and indicating the future prospects of the project in Section 6.5. The content of this Chapter is based on the paper: [137].

6.1 Motivation

Following point source removal, in order to avoid biases in the evaluation of the CMB angular power spectrum of the available sky fraction, a class of algorithms is

studied to replace the missing sky fraction with a statistical realization of the underlying CMB signal, known as "In-painting" or "Filling-in". Inpainting methodologies have also been used in CMB fields in order to estimate and recover specific parts of the sky. One of the most used methods in the CMB community is Gaussian Constrained Realizations (GCR) [138], which is based on the reconstruction of the Gaussian random field from its residual respect to the mean value of the field. Bucher and Louis [139] have made use of GCR to present a solution for filling in the CMB masked regions both on a spatially flat sky with periodic boundary conditions and on the pixelized sphere. Kim et al. [140] discuss this method in the context of pixel dataset of the order of Planck, with millions of pixels, commenting on its computational cost. They propose the same methodology, operating differently, in the harmonic space. Akrami et al. [11] have also exploited the Gaussian realization method with limited prior to restore the missing parts. Gruetjen et al. [141] applied inpainting to cut-sky CMB power spectrum and bispectrum estimators.

On the other hand, ML and specifically Deep Learning (DL) have been proposed as a solution to various problems concerning computationally expensive portions of data analysis in Cosmology [142–145]. For a comprehensive and inspiring reference about ML applications in cosmology, see Ntampaka et al. [146] and references therein. Also, ML and NNs have been used to improve different aspects of CMB analysis such as: cosmic string detection with tree-based machine learning in CMB data [147], predicting CMB dust foreground using galactic 21 cm data via NNs [148], Inpainting Galactic Foreground Intensity and Polarization maps using Convolutional Neural Network [149]. Moreover, and recently, the inpainting problem was addressed via DL by Yi et al. [150]; they have used another method as a subset of DL, known as the Variational AutoEncoders (VAE) in order to fill in the point source masked regions for the map-based CMB analysis.

GANs are a branch of deep NNs which are able to generate new realizations of a data set learning from a given training set. They are able to preserve the same statistics of input vectors. These networks are widely used in image inpainting applications and Image-processing. There is a vast literature on image inpainting by using GANs which address different capabilities and challenges [151]. They have been used in cosmology as well, especially in cases characterized by an high computational cost like LSS N-body simulations [152], detecting the 21cm emission from cosmic neutral hydrogen (HI) simulations [153] and generating weak lensing

convergence map [154]. Considering that GANs learns to replicate all statistical properties without any priory assumption, there is still a lot of room to investigate with GANs. One clear advantage of using GANs is they are capable of replicating either Gaussian or non-Gaussian distributions where other algorithms like GCR assume Gaussian distribution. Another advantage of GANs is related to their efficiency and flexibility. After training a GAN model on the real or simulated data, they can generate a large volume of data for high resolution maps. Moreover, their performance does not depend on the shape of the masked regions, and they are able to handle areas with different shapes and sizes. In the rest of this chapter, we are going to study the application of GANs in the context of filling in the CMB maps following point source masking. Relying on the GANs' advantages, we would like to check if GANs are able to fill-in the masked regions of the CMB intensity map. Later, we will compute the inpainted map's intensity power spectrum and check if GANs hold our desirable statistics.

6.2 GAN architecture

We have used a specific type of NNs, GAN, to inpaint masked regions of the sky in CMB maps following a point source removal process. In this Section, the basic concepts of GAN are defined, and after that, we explain our applied architecture and loss function.

6.2.1 Basic Concepts

Generally speaking, GAN is made of two models that play competitive roles: a generative model G and a discriminative model D. The role of the Discriminator is to distinguish between actual and generated (fake) data while the Generator has the responsibility of creating data in such a way that it can fool the Discriminator [155].

The generator is a NN $G(\mathbf{z}, \theta_G)$ where θ_G indicates the network's parameters and $\mathbf{z}$ is a noise vector drawn from a given prior distribution P_z. G is a *generative model* that tries to learn the mapping between P_z and the distribution P_G learnt from the data distribution P_{data}. The output of G, given $z \sim P_z$, is a sample of

the distribution P_G. This sample, coming from the estimated distribution of P_G, will resemble the samples drawn from the data distribution P_{data}.

The discriminator is a NN $D(\mathbf{x}, \theta_D)$ where θ_D indicates the network's parameters and $\mathbf{x}$ is a vector of data. D is a *classifier* that outputs the probability for the vector $\mathbf{x}$ to come from the training distribution P_{data} or from the learnt distribution P_G.

The general *loss function* of GAN architecture, considering the complete data set, by the combination of the generator and discriminator's loss function will be as follow:

$$\min_G \max_D V(D,G) = \min_G \max_D (\mathbb{E}_{x\sim P_{data(x)}}[log(D(x))] + \mathbb{E}_{z\sim P_{z(z)}}[log(1-D(G(z)))]). \tag{6.1}$$

where $\mathbb{E}$ is the expectation function. From this equation, we can observe how D tries to maximize the number of correct classifications of the original data – the part $\log(D(x))$ – and the generated data – the part $\log(1 - D(G(z)))$ – while G, on the other hand, by tuning its internal parameters θ_G, tries to minimize the number of correctly classified fake images producing more likely results. From the formula above, one could derive the individual loss functions that may be adapted to the specific problem's domain [156].

The training phase is finished when neither of the two models can get better results by adjusting their parameters; in other words, the Discriminator will not be able to distinguish between the real and fake data anymore. At this level, the Generator has learned to produce good enough data with characterization coming from the real data. This status is so-called *Nash equilibrium*. [157]

6.2.2 Applied Architectures

We have used a modified version of GAN architecture proposed by [158], which is called *Context Encoder*. This network aims to reconstruct the missing part(s) of an arbitrary image. By having this network as our baseline, we have adapted the architecture of Discriminator and Generator with respect to our target. Moreover, we have changed the type of applied Loss function, which we explain in the next Section. First of all, we describe the type of our input images for Generator and

Discriminator, which are CMB patches, and the applied mask. We can show the masked patch as follows:

$$\mathcal{P}_m = \mathcal{P}_i \odot \mathcal{M}. \tag{6.2}$$

where $\mathcal{P}_i$ is the complete (real) CMB patch as input and $\mathcal{M}$ is the mask, which includes value 0 for masked pixels and 1 for the rest and the sign, $\odot$, is the element-wise product operator, therefore $\mathcal{P}_m$ is the masked CMB patch.

Then to inpaint the masked patch, the Generator, which we indicate as $\mathcal{G}$ needs two inputs $\mathcal{P}_m$ and $1-\mathcal{M}$ that makes it aware of where it should inpaint. We can call the inpainted patch $\mathcal{P}_g$ where:

$$\mathcal{P}_g = \mathcal{G}(\mathcal{P}_m, (1-\mathcal{M})). \tag{6.3}$$

Finally, the Discriminator $\mathcal{D}$ should predict if either a patch is real or fake, so:

$$P = \mathcal{D}(\mathcal{P}_\diamond), \tag{6.4}$$

where $\diamond$ can be either i, meaning real CMB patches as input, or g, meaning fake CMB patch inpainted by the Generator, and P is prediction vector.

The optimized architecture for image and video recognition, image classification, and analysis is the *Convolutional Neural Networks* (CNN); So we have used fully convolutional architecture both for Discriminator and Generator as well. Not including *fully connected* layers, specifically in Generator, gives our network the flexibility advantage. By just having convolutional layers, we are able to give an arbitrary squared patch of CMB as an input to our network. Typically, the bottleneck layer of a Generator is formed by a fully connected layer. The fully connected layer is essential for the usual image recognition problem because by including only the convolutional layer, the information cannot propagate directly from a corner of the feature map to the other part. But this is not the case for our analysis since we are inpainting CMB patches locally, and only the neighborhood information is needed, since the long range correlation of the CMB field is not very significant. In Figure 6.1, we show the schematic architecture of our GAN. The details of *hidden layers* and different considered cases about Discriminator

and Generator architecture are explained separately in the following paragraphs. We have tried three diverse architectures with various depths. It came out that the deepest architecture was prone to overfitting, so here we address the two architecture with the best results.

6.2.2.1 Discriminator

For the first case, given a CMB patch 64×64 pixels, which can be the real CMB or inpainted by the Generator, we use three 2D convolutional layers with *Leaky ReLU* as the *activation function* and following three *batch normalization* layers. Rectified Linear Unit (ReLU) is one of the most popular activation functions which has the form of $f(x) = max(0, x)$. Besides having many advantages such as faster performance, ReLU has some caveats, which the most important one is prone to create dead neurons because if the units are not activated initially, they stay inactive with zero gradients. This problem can be solved by adding a small negative gradient in $x < 0$ part of the function; the result is the so-called Leaky ReLU [159]. Batch normalization is a technique usually applied for deep neural networks in order to improve the speed and stability. It is used to normalize the previous layer's activation, which causes decreasing the training epochs [160]. Our convolutional layers' kernel size is equal to 3, and we are using stride=2 for moving the filter. Here, the output layer is a simple dense layer with one value 0 or 1 to classify the fake or real image.
The second case has the same architecture, as mentioned in the first case, but with four 2D convolutional layers and the following batch normalization layers.

6.2.2.2 Generator

The Generator includes two important parts, the *encoder* and *decoder*. The encoder has the role of learning the features and structures of the given image by convolutional layers, and then the decoder takes the responsibility of reconstructing the missing area by using deconvolutional layers. For both cases, the input and output layers have the same shape of CMB patch, 64×64 pixels. For the first case, hidden layers consist of nine 2D convolutional layers. First, there are four 2D convolutional layers with Leaky ReLU as the activation function. Then, the batch normalization (Encoder part) of the Generator. Given an input image with a size of 64×64, we use these four layers to compute a feature representation with

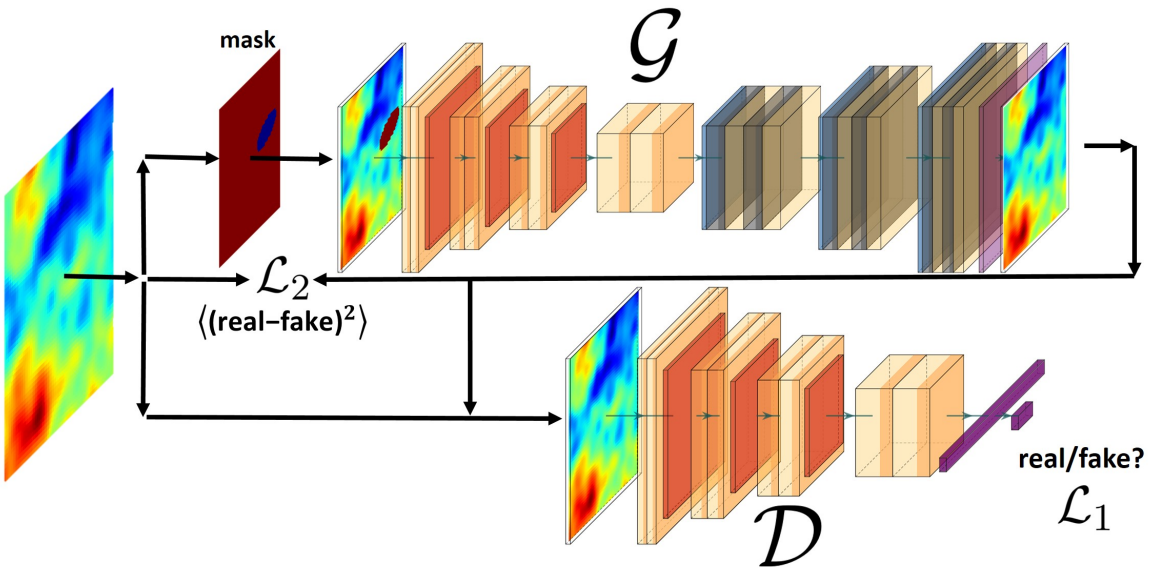

FIGURE 6.1: Schematic view of the training flow. $\mathcal{G}$ and $\mathcal{D}$ show the Generator and Discriminator architecture while $\mathcal{L}_1$ and $\mathcal{L}_2$ are traditional GAN loss and MSE loss function respectively. The modified loss we use for our network is $\mathcal{L} = \alpha\mathcal{L}_1 + (1-\alpha)\mathcal{L}_2$.

dimension $16 \times 16 \times 72$. Next, four 2D convolutional layers, in the decoder part of the Generator, are *up-convolution* which is simply up-sampling following by a convolutional layer [161] with ReLU activation function, The last convolutional layer with *tanh* activation function returns the inpainted CMB patch.
The second case has the same architecture as mentioned in the first one, but with eleven 2D convolutional layers (five convolutional for the encoder and five deconvolutional for the decoder) and following batch normalization layers.

6.2.3 Applied Loss Function

The Generator has to fill the masked regions by exploiting the competitive role with respect to the Discriminator, preserving the statistics of the CMB. Therefore, to achieve better results, we have trained our generative model based on two loss functions: First $\mathcal{L}_1 = L_{GAN}$ where L_{GAN} is defined in Equation 6.1 and $\mathcal{L}_2 = MSE(\mathcal{P}_i, \mathcal{P}_g)$. Mean Square Error ($MSE$) is one of the most common loss functions for regression. We have exploited the $\mathcal{L}_2$ to apply the regression of ground truth for the masked regions and reconstructing the overall structure of the missing part, while $\mathcal{L}_1$ has the responsibility to create a fake image that looks real. In our architecture, these two loss functions are related to the α parameter.

Therefore applied loss function will be as follows:

$$\mathcal{L} = \alpha\mathcal{L}_1 + (1-\alpha)\mathcal{L}_2. \tag{6.5}$$

The control coefficient α changes from 0.01 to 0.2 while the model learns. At the beginning of the training, the value for α is low, which means the main goal is minimizing $\mathcal{L}_2$ that has the role of learning about filling the exact missing region. As the training goes on, we relax this condition through the training process and let $\mathcal{G}$ learn more about the statistics. In this way, in higher epochs, α value increases, and the Generator contributes more in filling the masked region. Also, we applied an adaptive learning rate for $\mathcal{G}$ and $\mathcal{D}$ during the training phase, considering the loss value in such a way that in each epoch, the $\mathcal{G}$ loss is compared with $\mathcal{D}$. In this method, if $\mathcal{G}$ loss is larger than $\mathcal{D}$ loss, the ratio of $\mathcal{G}$ loss to $\mathcal{D}$ loss: $n = \frac{\mathcal{L}^{\mathcal{G}}}{\mathcal{L}^{\mathcal{D}}}$ is calculated and Generator takes n times training more in epoch and vice versa. In this way, we prevent to reinforce one of two opponents, just for some epochs that this ratio might be extreme, we put a threshold $n < 20$, which means at maximum $\mathcal{G}$ or $\mathcal{D}$ can train 20 times more than the other one.

6.2.4 Working environment

In our work, we used *Keras* [1] package with *Tensorflow* backend. We trained different architectures for 70000 epochs using 64×64 pixels images and 32 as batch size. The number of layers, kernel size, the latent space dimension and number of filters are evaluated as different parameters of the architectures. The learning rate is initiated with 0.5 and decayed with the factor 0.9997. The α values we investigate are 0.01, 0.05, 0.1 and 0.2 where the masked areas are chosen from one of these ranges: $[10-30]$, $[80-100]$, $[150-170]$, $[220-240]$ and $[290-310]$ pixels. The model is trained on NVIDIA Tesla P100 and Quadro RTX 5000 GPUs and 30 GigaByte of memory.

[1] https://keras.io

6.3 Data set

In order to train the network and test the capability of generating the CMB masked part, we have used publicly available Planck simulated maps [2] and the corresponding mask. To avoid complications, we chose to work with "Spectral Matching Independent Component Analysis" (SMICA) component separation method and assumed that the component separation procedure works well except in the masked region. SMICA has been obtained by combining the multi-frequency Planck 2018 dataset to mitigate the foreground emission. SMICA [162] represents one of the four component separation approaches included in the Planck analysis [11]. We have also used the Planck intensity mask map, so the masked regions' shapes and distribution are exactly as they are in Planck observation.

The used simulated maps are noiseless but include all the systematic effects (effect of Time-ordered information and beams) entering the Planck observation and analysis pipeline. There are more details about the simulation pipeline and corresponding systematic effects in [163] for interested readers. We have tried different patch sizes and chosen 64×64 pixels due to the best result and computational costs. In fact, this choice prevents our generative model $\mathcal{G}$ from learning about larger scales, and the larger patches would need huge memory and is very time consuming. To train our network, we have used ten full-sky CMB simulations, and in total, more than 10^5 CMB patches as the training set, and five full-sky simulations are used for the test set.
In this work, we target to fill in the CMB missing regions, which are masked because of point sources, with different areas. In order to have a clear idea of the distribution of these masked regions in terms of size and area, Figure 6.2 shows that a large number of masked regions has an area less than 1000 pixels.

In order to have more quantitative statistics, in Table 6.1, by considering a threshold on the maximum area for masked regions (A_{masked}), we calculate which percentage of the total number of the masks with less than or equal to different limits, N_r, and the sky fraction associated to them, A_r. Moreover, A_r indicates the percentage of the masked sky by the masks with an area for the corresponding A_{masked}. We have used these values for testing our models and their statistics. For instance, by choosing $A_{masked} \leq 1500$ pixels, the GAN will fill in 96.08% of the total number of the masks and 2.97% of the whole masked area. In the typical

[2] http://pla.esac.esa.int

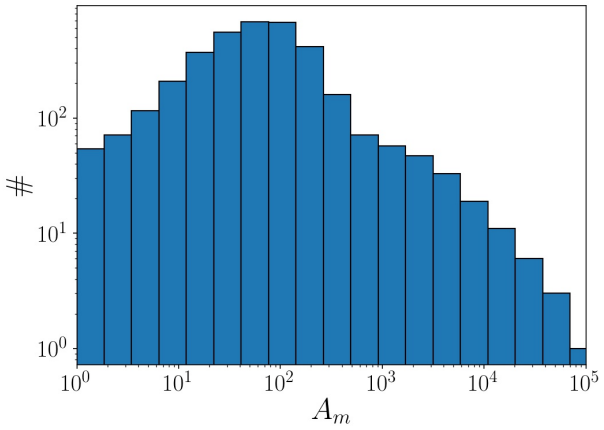

FIGURE 6.2: Masked regions area distribution for Planck intensity mask. The masked regions with area $A_m \approx 100\ pixel^2$ are the most probable.

$A_{masked} \leq [pixels]$	$A_{masked} \leq [arcmin^2]$	$N_r(\%)$	$A_r(\%)$
100	295	74.49	1.14
200	590	89.83	1.80
500	1475	93.11	2.10
1000	2951	94.79	2.51
1500	4426	96.08	2.97
2000	5901	96.81	3.37

TABLE 6.1: The percentage of masked regions in term of number, N_r, and fraction of masked regions respect to the whole sky (4π), A_r, for different A_{masked}.

masking procedure of an all sky CMB map, the majority of the excluded region is in the mask applied to the Galactic plane, needed in order to mask out the diffuse Galactic foreground signals. Still, a relevant fraction of point sources is masked out also at low Galactic latitudes.

6.4 Results and Discussion

We discuss the results of each architecture and case aforementioned in this Section. In order to do so, we need to define our methodology to compare different cases.

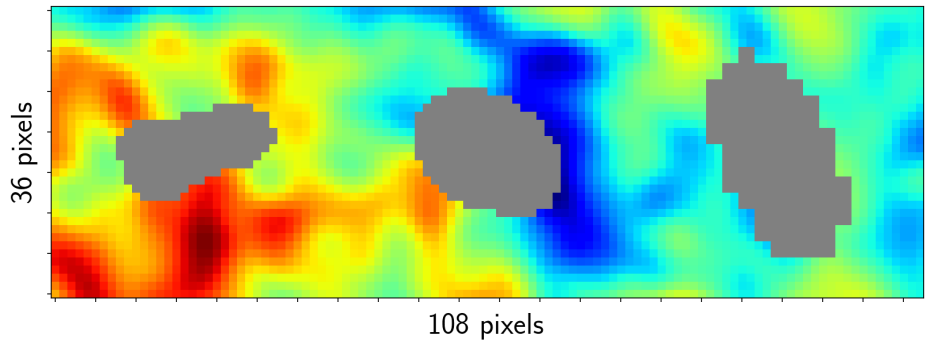

FIGURE 6.3: The different range of isolated masked areas which are used in our training. From left to right: $150 \leq A_{masked} \leq 170$, $220 \leq A_{masked} \leq 240$ and $290 \leq A_{masked} \leq 310$.

6.4.1 Methodology

It is common to visually compare the real and fake images provided by GAN to check the GAN performance. In the context of CMB analysis, though, we will consider the angular power spectrum as a diagnostics of the proposed algorithm's good functioning. Although, concerning the scientific goal, there is the possibility to apply another kind of statistical metric as a benchmark. This work will focus on the total intensity only, leaving the polarization to future works. That is also because it is expected that point source masking plays a less important role in polarization, with respect to total intensity [164].

In order to investigate how $\mathcal{G}$ can learn through different masked sizes, we have used two different masked area conditions for the training and test. The parameter A_{masked} sets the allowed range of masked regions. In our method, in the training phase, the A_{masked} has both upper and lower bands, limited to 20 pixels for all the models to focus on a specific range of areas. We have trained our model on masked regions with $A_{masked} = [10-30]$, $[80-100]$, $[150-170]$, $[220-240]$ and $[290-310]$ pixel area. The last three ones are reported in Table 6.2 and 6.3, resulting in a better performance for the current analysis. Figure 6.3 shows three samples of these applied masked areas. Instead, in the test set, we have just limited A_{masked} with an upper limit, and the maximum A_{masked} is equal and less than 2000, as reported in Table 6.4.

Our methodology is based on comparing the theoretical CMB intensity power spectrum with the inpainted one, for this purpose, we will use the same definition of the CMB power spectrum C_ℓ as Equation 2.18. For performance comparison and considering higher ℓs, we will plot the D_ℓ which is defined as $D_\ell = \frac{\ell(\ell+1)\,C_\ell}{2\pi}$.

In this analysis, we have used HEALPix *anafast* function for the power spectrum computations after the CMB patches are reassembled on the sky sphere. We cut and project back the CMB patches after the inpainting procedure using the *CCGPack*[3] package. The difference between the inpainted map power spectrum D_ℓ^{inp} and theoretical D_ℓ^{truth} one varies through different A_{masked} and g, so ΔD_ℓ can be written as:

$$\Delta D_\ell(A_{masked}, g) = D_\ell^{inp}(A_{masked}, g) - D_\ell^{truth}. \tag{6.6}$$

The variable g is referred to the inpainted CMB patch by the Generator. Now, we are ready to define our cost function, which will be used for testing the generative model performance:

$$\mathcal{C}(A_{masked}, g) = \sum_\ell \Big(\Delta D_\ell(A_{masked}, g)\Big)^2. \tag{6.7}$$

Since this is a function of A_{masked}, and A_{masked} might be different from a patch to another, we need to define a reference case in order to have a fair comparison for the large inpainted areas with respect to small ones. We defined a worst-case scenario for each chosen A_{masked} to compare how well the D_ℓ^{inp} is reconstructed. The worst-case scenario power spectrum D_ℓ^w is achieved if we assume that the masked regions are filled using the average of the rest of the map, which means that for each patch $\mathcal{P}$ the following relation will be valid:

$$\mathcal{P} = \begin{cases} \mathcal{P}_i & \mathcal{M}_i = 1 \\ \overline{T} & \mathcal{M}_i = 0 \end{cases}, \tag{6.8}$$

Here, $\overline{T}$ is the averaged intensity of the whole unmasked sky map, and i shows the pixel number. Then the cost of the worst-case scenario, $\mathcal{C}_w$, can be defined as:

$$\mathcal{C}_w(A_{masked}, g) = \sum_\ell \Big(D_\ell^w(A_{masked}, g) - D_\ell^{truth}\Big)^2. \tag{6.9}$$

Finally, we can define the relative cost, which from now on, we will indicate as $\mathcal{C}_r$ to compare different results for the different Generator architectures and A_{masked}.

[3]https://github.com/vafaei-ar/ccgpack

$$\mathcal{C}_r = \frac{\mathcal{C}(A_{masked}, g)}{\mathcal{C}_w}. \tag{6.10}$$

6.4.2 Results

We have applied our algorithm on two different types of CMB patches that henceforth we will refer to the hypothetical and the Planck mask. In the hypothetical mask, the generative model is asked to inpaint the same masked area size as it learns in the training phase, while in the Planck mask, the Generator should deal with any kind of mask sizes less than the specified A_{masked}.

The hypothetical mask is created assuming: one masked area exits within the intervals defined in Section 6.4.1 in each patch with 64×64 pixels. Of course, the number of masked regions within this range is much smaller in the Planck mask, but we investigate this case to evaluate model performance by training for a full sky mask. On the other hand, the different chosen maximum areas for the Planck mask are listed in Table 6.4 in order to check the model performance in case of facing larger masked regions.

We have probed various hyper-parameter spaces, including the appropriate depth of the $\mathcal{G}$ and $\mathcal{D}$ as well as different α parameters for the loss function and A_{masked}. In total, we have trained 60 various networks for 70000 epochs with different parameters. Here we are reporting the selected ones with the best results.

Table 6.2 and 6.3 show obtained $\mathcal{C}_r$ from inpainted CMB patches for different A_{masked} and α in the case of hypothetical sky by having 9 and 11 layers in $\mathcal{G}$. The number of $\mathcal{D}$ layers is always $l_{\mathcal{D}} = \frac{l_{\mathcal{G}}}{2}$, where $l_{\mathcal{G}}$ is the number of layers in the Generator, due to our trial and error that indicates $\mathcal{G}$ needs to be deeper than $\mathcal{D}$. Each specific network architecture is trained on 3 different range of areas $150 \leq A_{masked} \leq 170$, $220 \leq A_{masked} \leq 240$ and $290 \leq A_{masked} \leq 310$ pixels and for the test, a full-sky map with the masked area inside the threshold, is given to the network. Table 6.2 demonstrates that a GAN with Generator with nine layers as it is described in Section 6.2.2.2 and Discriminator with four layers, as Section 6.2.2.1 on the range of area = [150, 170] pixels has the best performance and least $\mathcal{C}_r$. Figure 6.4 shows a sample of 4 patches of ground truth CMB patches next to each other, masked and inpainted CMB, in the same architecture, for the visual comparison. From this Figure, one clearly can notice

the masked areas have different shapes but a size inside the range. Furthermore, we have computed the intensity power spectrum of our hypothetical CMB maps in all the cases and plotted the best case in Figure 6.5. In this Figure, for comparison, we have shown the observed CMB power spectrum as a baseline, D_ℓ, the worse, D_ℓ^w, and the best, D_ℓ^p, inpainted scenario. D_ℓ^p in this plot corresponds to the green cell in Table 6.2. Since the power spectrum itself is not very representative of the difference between them, we have plotted power spectrum residual, ΔD_ℓ, and error percentage in the middle and lower panels. In order to have more statistics, we have simulated 5 different hypothetical sky masks and done the same procedure. The shaded areas, which are 95% confidence level in Figure 6.5 come from these map realizations. We can see in the range ℓ less than 1700 the error is about 1%. Moreover, we followed the same procedure to plot the power spectrum for the cases, when $220 \leq A_{masked} \leq 240$ and $290 \leq A_{masked} \leq 310$, the results are demonstrated in the Figure 6.10, which are compatible with the yellow and orange highlighted cells in Table 6.3.

Now we would like to test our network with the same procedure on the real Planck 2018 intensity mask. In this step, we also drop the architecture with 11 layers since the result for the architecture with 9 layers turns out to be the best among the two. In addition, we picked the model trained on $290 \leq A_{masked} \leq 310$, which is favorable because it has less $\mathcal{C}_r$ for the larger masked area on the real Planck mask. Again, Figure 6.6 shows a sample of inpainted CMB patch compared to the input CMB. Our model is able to deal with different masked areas in terms of both size and shape. In Table 6.4, the $\mathcal{C}_r$ from inpainted CMB patches for different $A_{masked} \leq 2000$ pixels and α for the real sky, are reported. We would like to recall that our model, in this case, is just trained on masked areas with $A_{masked} \leq 310$, but it is able to predict and inpaint regions much larger on the real sky. For each different upper limit on A_{masked}, we have added the plot of the best predicted power spectra in comparison with the baseline in Figure 6.12. Still, owing to the fact $A_{masked} \leq 1500$ is the largest area in which the generative model can inpaint with minimum error. We have shown this case in Figure 6.7. As before, observed CMB D_ℓ, D_ℓ^w, D_ℓ^p, inpainted scenario power spectrum are compared. D_ℓ^p belongs to the cell with the blue highlight in Table 6.4 with $\alpha = 0.01$. We can notice that for $\ell < 1500$ the deviation of inpainted CMB map is negligible and around 1%.

Finally, we consider more examples of inpainted CMB patches as a visual comparison in Figure 6.8 and 6.9 for the hypothetical mask and in Figure 6.11 for the

A_{masked}[pix] \ α	0.01	0.05	0.1	0.2
150:170	0.64	1.10	0.56	0.98
220:240	1.29	1.06	0.85	0.77
290:310	1.81	1.87	1.64	1.58

TABLE 6.2: Obtained $\mathcal{C}_r$ for trained model with 9 layers architecture considering different α and A_{masked} on hypothetical sky mask. The highlighted cell shows the least $\mathcal{C}_r$ specifications. All the values are multiplied by 10^2.

A_{masked}[pix] \ α	0.01	0.05	0.1	0.2
150:170	2.76	2.55	2.43	4.22
220:240	1.53	0.99	0.81	0.77
290:310	1.71	1.53	1.02	1.53

TABLE 6.3: $\mathcal{C}_r$ for the trained model with 11 layers architecture considering different α and A_{masked} on hypothetical sky mask. The yellow and orange highlighted cells show the least $\mathcal{C}_r$ for $220 \leq A_{masked} \leq 240$ and $290 \leq A_{masked} \leq 310$ respectively, taking to account both this Table and Table 6.2. All the values are multiplied by 10^2 .

Planck intensity mask. Also the power spectra comparison in the cases of A_{masked} = 220 and 290 for hypothetical mask in Figure 6.10 is plotted. As the last plot also we have shown all the A_{masked} cases mentioned in Table 6.4 to check with the inpainted CMB maps with our baseline.

We wrap up the different cases in Figure 6.13. $\mathcal{C}_r$ versus different upper limits of A_{masked} for different α is plotted with 95% confidence level. We see that by enlarging the masked areas, $\mathcal{C}_r$ value gradually increases, but for $A_{masked} \geq 1500$, this growth is significant, so we rely on our generative model up to $A_{masked} \leq 1500$. Also, from this Figure, it is clear the change of α does not have a remarkable effect on $\mathcal{C}_r$ taking into account different statistical variations of inpainted maps.

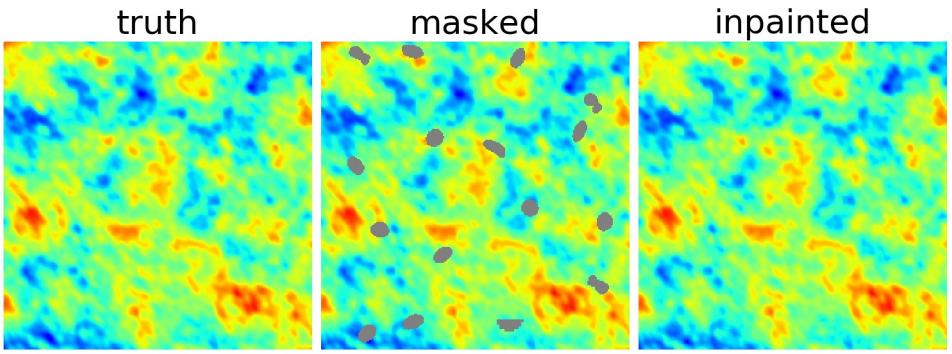

FIGURE 6.4: One sample including 4 patches of input CMB patch (left), masked patch in the middle and the prediction (right). The inpainted patches are produced using a hypothetical mask and $\alpha = 0.1$, 9 layers Generator and $150 \leq A_{masked} \leq 170$ model.

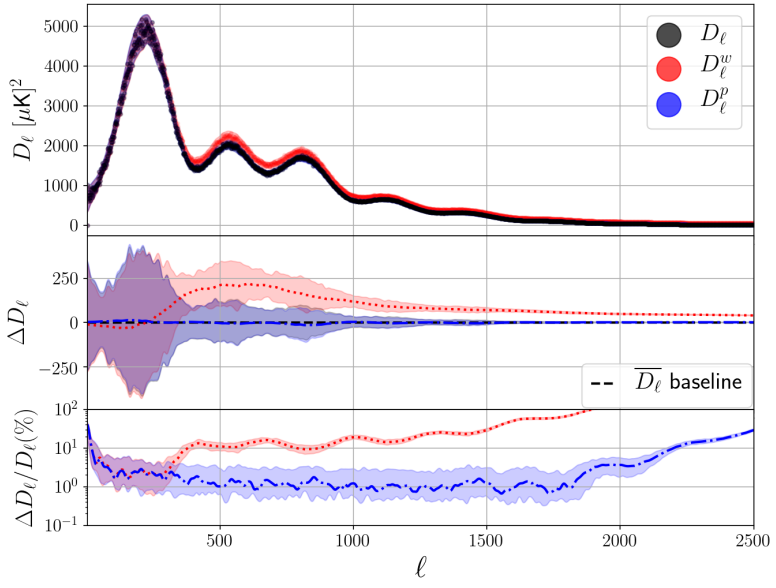

FIGURE 6.5: Upper panel demonstrates the comparison of CMB intensity power spectrum D_ℓ for the worse D_ℓ^w and best D_ℓ^p scenario which is the intensity power spectrum of inpainted CMB maps for the hypothetical full sky mask. The middle and lower panel show the deviation and residual percentage from the observed CMB power spectrum. The D_ℓ^p is the case with green highlight in Table 6.2. The shaded areas show the 2σ confidence level.

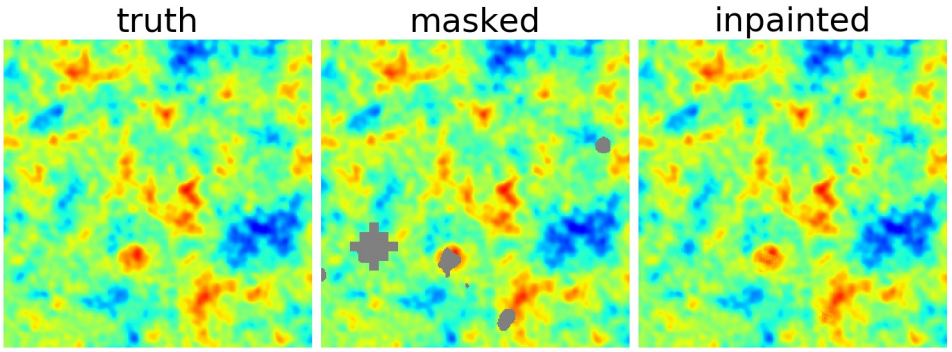

FIGURE 6.6: One sample of inpainted 64×64 pixels CMB patch. Input CMB patch (left), masked patch in the middle and the prediction on a large masked region (right). The masked areas come from Planck 2018 intensity mask.

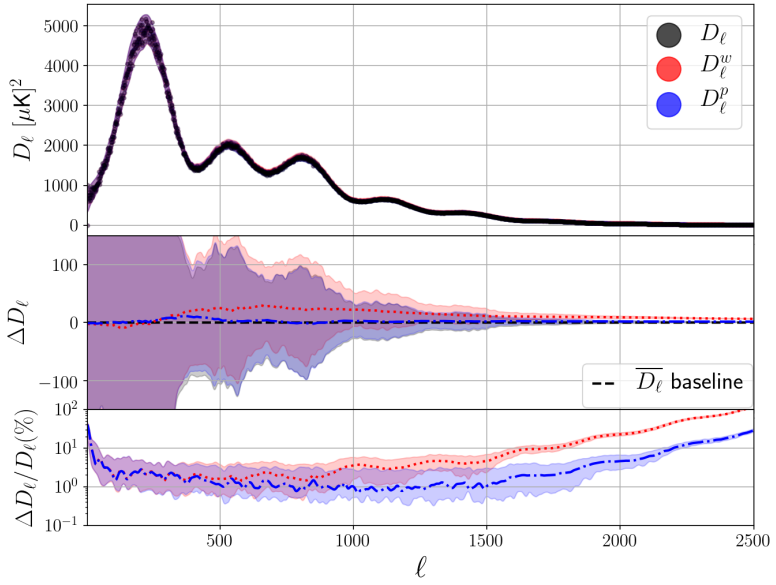

FIGURE 6.7: Upper panel demonstrates the comparison of CMB intensity power spectrum D_ℓ for the worse D_ℓ^w and best D_ℓ^p scenario inpainted CMB maps for the real full sky mask with $A_{masked} \leq 1500$. The middle and lower panel show the deviation and residual percentage from the observed CMB power spectrum. The D_ℓ^p is the case with the blue highlight in Table 6.4. The shaded areas show the 2σ confidence level.

A_{masked}[pix] \ α	0.01	0.05	0.1	0.2
≤ 100	2.64	2.92	2.56	2.64
≤ 200	3.66	3.82	3.66	3.74
≤ 500	5.26	4.98	5.47	5.25
≤ 1000	10.08	10.25	10.10	10.22
≤ 1500	10.95	20.35	10.98	20.22
≤ 2000	40.84	50.83	40.91	5.52

TABLE 6.4: $\mathcal{C}_r$ for trained model with 9 layers architecture and $A_{masked} = 290 : 310$ considering different α and A_{masked} on Planck mask. All the values are multiplied by 10^2. The highlighted blue cell shows the best performance of $A_{masked} \leq 1500$ which is our favourite model.

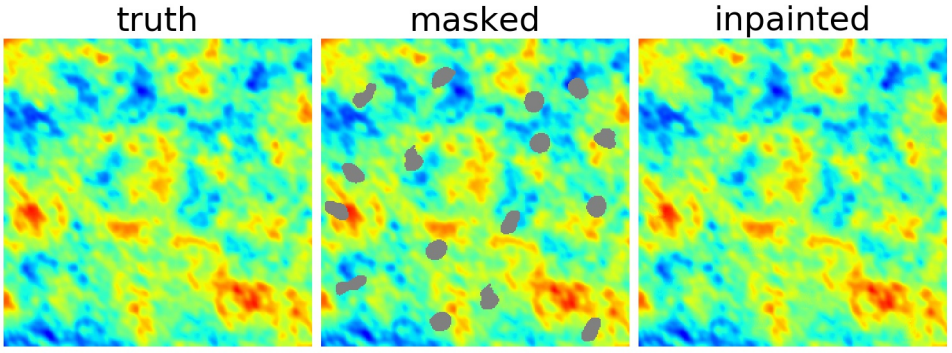

FIGURE 6.8: One sample including 16 patches of input CMB patch (left), masked patch in the middle and the prediction (right). The filled in patches are produced using a hypothetical mask and $\alpha = 0.2$, 11 layers Generator and $A_{masked} = 220$ model.

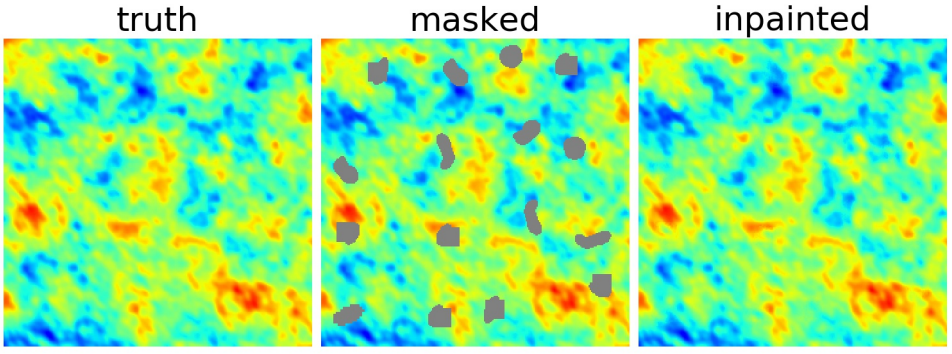

FIGURE 6.9: One sample including 4 patches of input CMB patch (left), masked patch in the middle and the prediction (right). The filled in patches are produced using a hypothetical mask and $\alpha = 0.1$, 11 layers Generator and $A_{masked} = 290$ model.

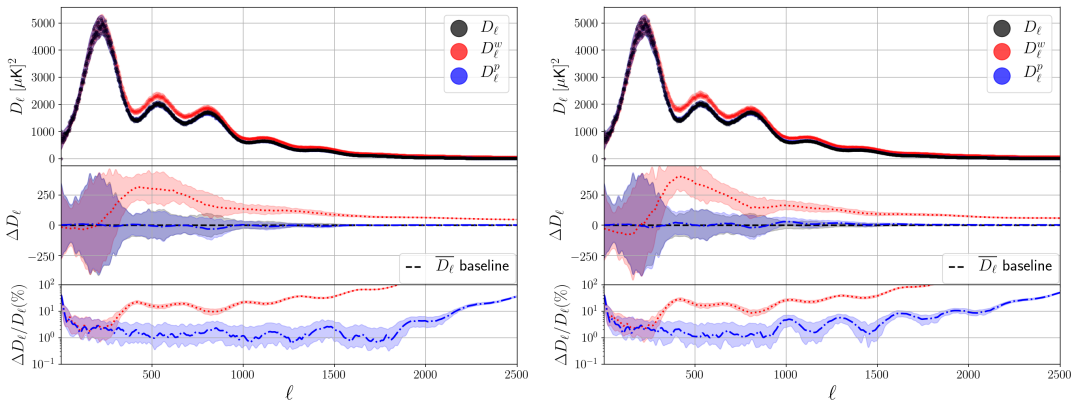

FIGURE 6.10: The caption is the same for Figure 6.5, but for $A_{masked} = 220$ and 290 pixels from left to right. The corresponding cases are yellow and orange highlighted cells in Table 6.3.

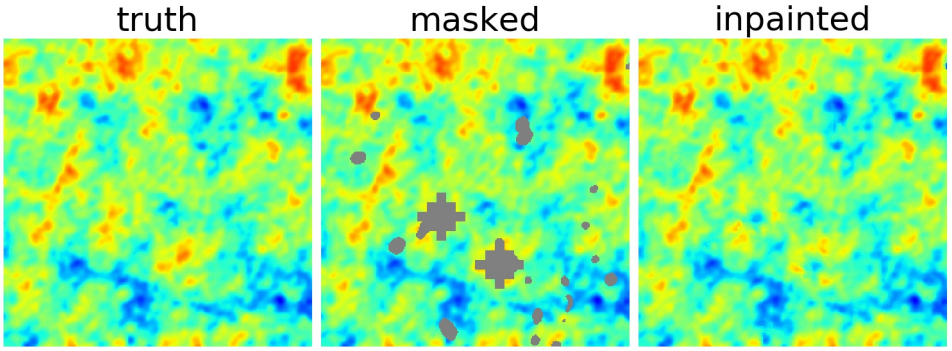

FIGURE 6.11: One sample of inpainted 64×64 pixels CMB patch. Input CMB patch (left), masked patch in the middle and the prediction on a large masked region (right). The masked areas come from Planck 2018 intensity mask.

6.5 Summary and outlook

In this chapter, we consider an GAN based approach to face the issue concerning the inpainting of the masked regions following point source removal in CMB maps, while keeping their statistics, such as the power spectrum, unchanged. We develop a modified generative model that is able to inpaint the CMB masked areas less than 1500 pixels with around 1% error on the CMB intensity power spectrum for $\ell < 1500$. Our model does not use any kind of prior and in the case of training on observed CMB patches, preserves the statistic, and therefore it is not limited to the reconstruction of Gaussian random fields.

Different setups are explored as well as diverse architectures. Considering the best results, we suggest a modified GAN architecture with 9 layers for Generator and 4 layers for Discriminator, trained on $290 \leq A_{masked} \leq 310$. Also, our network

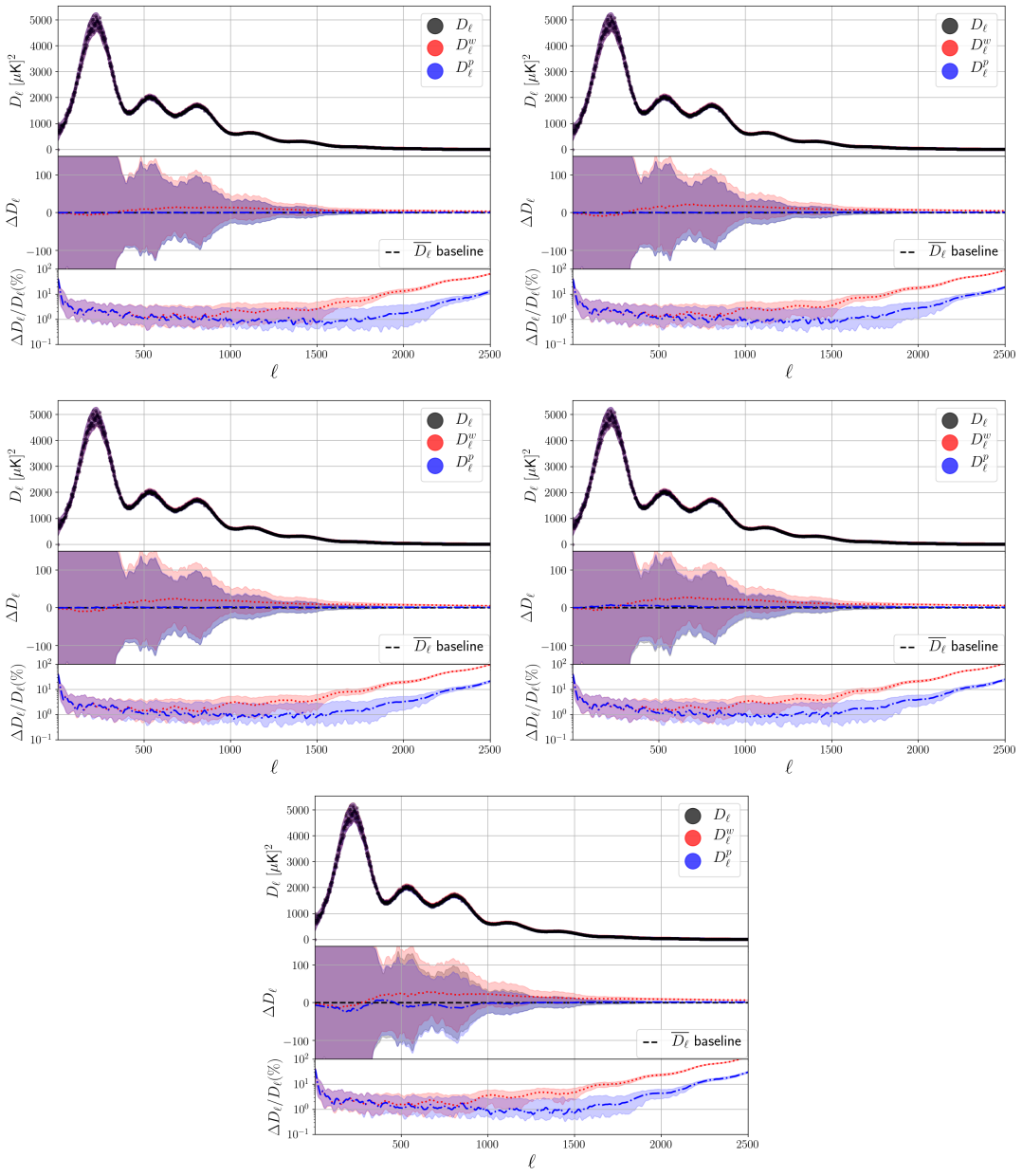

FIGURE 6.12: The caption is the same for Figure 6.7, but for A_{masked} less than 100, 200, 500, 1000, 2000 pixels sequentially, and α in these cases corresponds with the least $\mathcal{C}_r$ for each A_{masked} reported in Table 6.4.

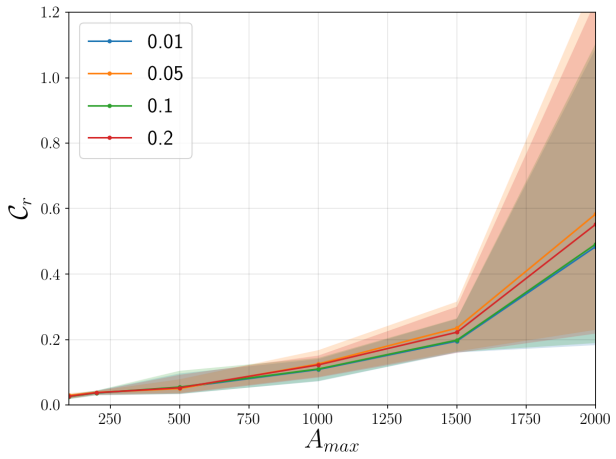

FIGURE 6.13: The summarizing $\mathcal{C}_r$ values for different A_{masked} upper limits and α, where the shaded areas show the 2σ confidence level.

takes advantage of using both MSE and GAN loss functions to learn the best strategy to inpaint different masked areas, and these two loss functions are related to each other by the α parameter. We have used a novel method in using a dynamic training rate of $\mathcal{G}$ and $\mathcal{D}$, calculated during each epoch, for the network. Furthermore, our model is not limited to a specific shape of the CMB patch, as well as a missing area smaller than 1500 pixels.

We have defined the $\mathcal{C}_r$ parameter, which is a measure of the network performance established on the power spectrum residual. The results of testing our model on both hypothetical and Planck 2018 intensity masks is reported in Table 6.2, 6.3 and 6.4. We have shown that our applied GAN architecture, in the best scenario, up to $A_{masked} \leq 1500$ and $\ell < 1500$, is able to inpaint the masked areas of the CMB map in such a way that the CMB intensity power spectrum is barely different, about 1%. Moreover, the generative model is almost insensitive to the choice of α between [0.01, 0.05, 0.1, 0.2] considering the statistical analysis.

We believe that the exploitation of the GAN and generative model as a part of the mapmaking pipeline in the next generation of CMB experiments might be relevant. In addition, as it is described, our generative model has the capability of not being biased to Gaussian fields; Since it does not have any Gaussian prior in the training phase, in case of having observed CMB as the training set. In future works, we will focus on larger masked regions. In that case, one needs more effective architecture to deal with very large inputs. Also, we'll be able to concentrate on either power spectrum or higher-order statistics optimization using multi-level CNN operations for the intensity maps as well as polarization. We will

keep discussing the future prospects concerning this line of investigation in the final chapter of this Thesis.

Chapter 7

Conclusions

In the forthcoming years, an incredible effort will take place towards detecting primordial GWs, which are the imprint of the inflationary era in the very early Universe. Detection and characterization of the amplitude of primordial gravitational waves or tensor-to-scalar ratio r will be possible by observing CMB polarization, specifically B-modes, with extremely high sensitivity. Over the last decade, operating CMB experiments looking at the CMB polarization made it clear that diffuse Galactic foregrounds are the main non-instrumental challenge for the detection of primordial GWs. The application of component separation techniques for extracting the Galactic foregrounds out of the multi-frequency CMB datasets turns out to be extremely important for the next generation of CMB experiments, in order to achieve the detection limit for r, down to 10^{-3}.

In this Thesis, we test the component separation on simulated data by one of the forthcoming CMB probes, the LSPE experiment. At the same time, we make the first steps into the implementation of novel, NN based algorithms, used for a physical understanding of the actual parameterization of foregrounds, prior to component separation, as well as for filling in regions where the foregrounds were so intense that the considered was masked out.

Our studies show that the LSPE will be able to set an upper limit for tensor-to-scalar ratio r at the level of 10^{-2} and detect a $r = 0.03$ with 95% confidence limits.

For this analysis, we assumed the common and basic models for the diffuse foreground emissions in polarization. On the other hand, as mentioned before, there is

strong evidence of the variability of foreground models in the sky. This variability makes a difference in the actual parameterization of foreground models that are needed to fit them out across the sky. We wanted to start considering methods that have to be applied before actual component separation, which are able to learn, from data, the actual parameterization of foregrounds, which is most convenient for their fitting and removal. Therefore, a new methodology is proposed and based on NNs, in Chapter 5. We have tested our model on different synchrotron models, AME, i.e. with different SEDs, presence of AME, and having thermal dust spectral index variable across the sky. The implemented NN has been check in two general ways: binary and Multi-class classifications. We have shown that the NN can reach an accuracy higher than 90%, meaning that the algorithm returns the correct foreground modeling in 93% of the considered sky area, in all the cases mentioned above when the test data set is with noise. Moreover, comparisons of the NN to a standard χ^2 information extracted from parametric component separation for the classification of foreground emission in the noisy case, shows the advantage of the NN in terms of accuracy. For the only case considered for this comparison, synchrotron with and without curvature in the SED, the accuracy reaches 97% for discerning synchrotron power law from curvature, including the noise, which is 20% more than the same accuracy estimated via χ^2 statistics.

In Chapter 6, we used NNs to face the issue of actual missing areas in CMB maps due to point source removal. Precise restoration of CMB masked regions is important in a number of post-processing of CMB maps, for example reducing the cosmological parameter estimation uncertainty. This operation, known in general as inpainting, consists in the reconstruction of lost, missed or damaged parts of an image. We have considered GANs algorithms to face this issue and we report different types of GAN architecture on various CMB masked areas. Moreover, we analyse the CMB power spectrum after the inpainting process for various cases and show that the inpainted maps, in the best case scenario, have a negligible deviation from the theoretical and observed intensity power spectrum. This deviation in the masked area less than 1500 pixels and $\ell < 1500$ is about 1%. Due to importance of CMB B-mode observations, we are going to apply the same method on the CMB polarization maps and power spectra reconstruction.

Each of the works presented in this Thesis is the beginning of investigation lines. On the LSPE front, one should realize that the experiment represents the first one dedicated to polarization, to observe large sky fraction for constraining B-modes,

over a large frequency interval. Therefore, the study would benefit from considering more complex foreground models, with enhanced variability of spectral indices, as well as new templates, including correlations between the main polarized foregrounds. On the NN side, the model recognition would also benefit from being applied to more complex skies, but also, its performance combined with component separation needs to be investigated further, beyond the promising comparison with the χ^2 statistics. In addition, one could go further from the Gaussian noise model on the noise simulation part and include other systematic effects. Finally, the NNs for inpainting should be used on real data analysis. Also, we plan to extend the analysis to the CMB polarization maps, where due to the fainter amplitude of the signal, the generative model's training is more challenging. On the other hand, all these new studies can be undertaken on the basis of the good results we got in the present Thesis, along the three considered lines of investigations, which we hope will be a stimulus for future ones. The studies, novel methodologies and tests presented in this Thesis represent the basis for new implementations, and applications to the future simulated and real CMB data from the B-mode probes that we discussed in Chapter 3.

Appendix A

Classification in presence of variant dust spectral index

As we mentioned in Chapter 3, Planck 2018 observations showed that thermal dust spectral index β_d is not constant all over the sky. In order to complete our analysis and get closer to the realistic case, we have also included a thermal dust model with a non-constant spectral index. All the results that we discussed in Chapter 5, are obtained based on simulations with a constant spectral index, as we described in Section 5.4. In this appendix, we explain the same analysis that we have done from Section 5.5 to 5.8, but this time in the presence of spatially variant dust spectral index.

For simulating the training and test maps, we have followed the same procedure as described in Section 5.4 with a difference for simulating the thermal dust emission model. Therefore the synchrotron, AME, and CMB radiations are untouched with respect to Section 5.4. For the thermal dust component, still we have used the PySM template, rescaled as a modified blackbody, as in Equation 3.3 but with spatially variant spectral index, considering a Gaussian distribution with mean value $\beta_d = 1.54$ and standard deviation equal to 0.1 which is motivated by Planck 2018 [11]. While the dust temperature remains constant $T_d = 20K$. The number of map components, training, test, and validation set vectors will be the same as before.

In order to avoid repetition, here we just bring the results of analysis and the maps concerning the NN predictions in the different cases. In case of changes in the NN's architecture, we have explained in each Section internally.

A.1 Binary classification

The architecture of the networks for the binary classification stays the same as before, meaning that NN for discerning the existence of synchrotron curvature of AME in the sky, has the same architecture of Section 5.3.1.

A.1.1 Synchrotron with and without curvature

We have applied the same procedure as Section 5.5.1 to distinguish synchrotron emission with curvature from the usual synchrotron power-law when there is a spatially variant dust spectral index. The number and resolution of simulated multi-frequency maps both for the training and test sets stay identical. In this case, the NN after 70 epochs get to 97% accuracy in the training phase, and the same accuracy can be obtained for the test set as well, which is reported in Table A.1. Figure A.1 shows the prediction of NN for the noiseless case. In general, we would expect by increasing the variables and adding complications to the foreground models, the accuracy drops. However, in this case, we see that accuracy decreases just by 1% with respect to the case of discerning simple power-low synchrotron from synchrotron with curvature when the thermal dust spectral index was constant.

Moreover, by investigation of the pixels where the NN associated the wrong model (white pixels in Figure A.1) and comparison to Figure 5.13, we realized the same physics properties apply here as well, which means that the NN has the tendency to misclassify the model when the relative amplitude of the synchrotron emission over dust is small. For example, in the region near Galactic coordinate (230°, +40°) where is well-known for the low amplitude of synchrotron, there are several misclassified pixels. In comparison to Figure 5.13, the misclassified pixels in this region slightly increased, which confirms our expectation.

A.1.2 Synchrotron and AME

We have followed a similar setup as mentioned in Section 5.5.2 to distinguish the presence of AME or not but this time, the dust spectral index is variant across the sky in our simulated maps. The NN with this setting, after 120 epochs is

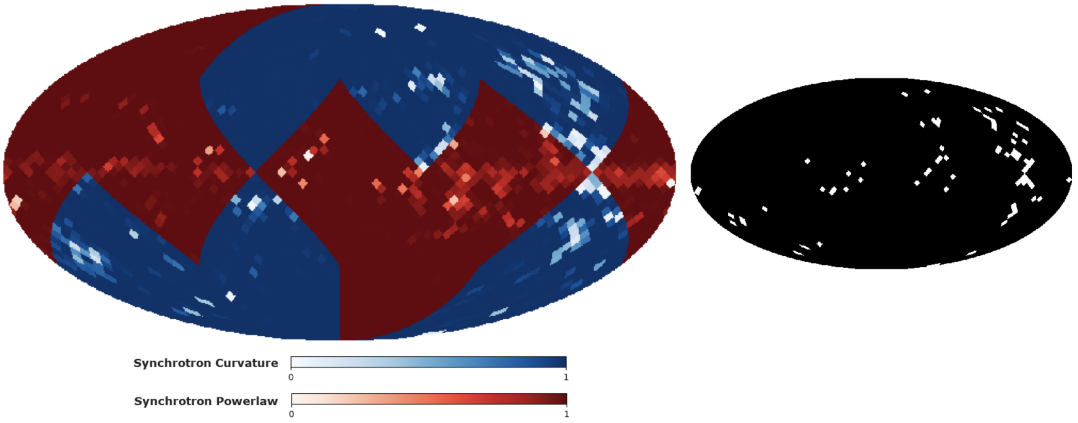

FIGURE A.1: NN prediction of synchrtron emission with and without curvature when dust spectral index is changing, the color scale and caption are the same as Figure 5.6

.

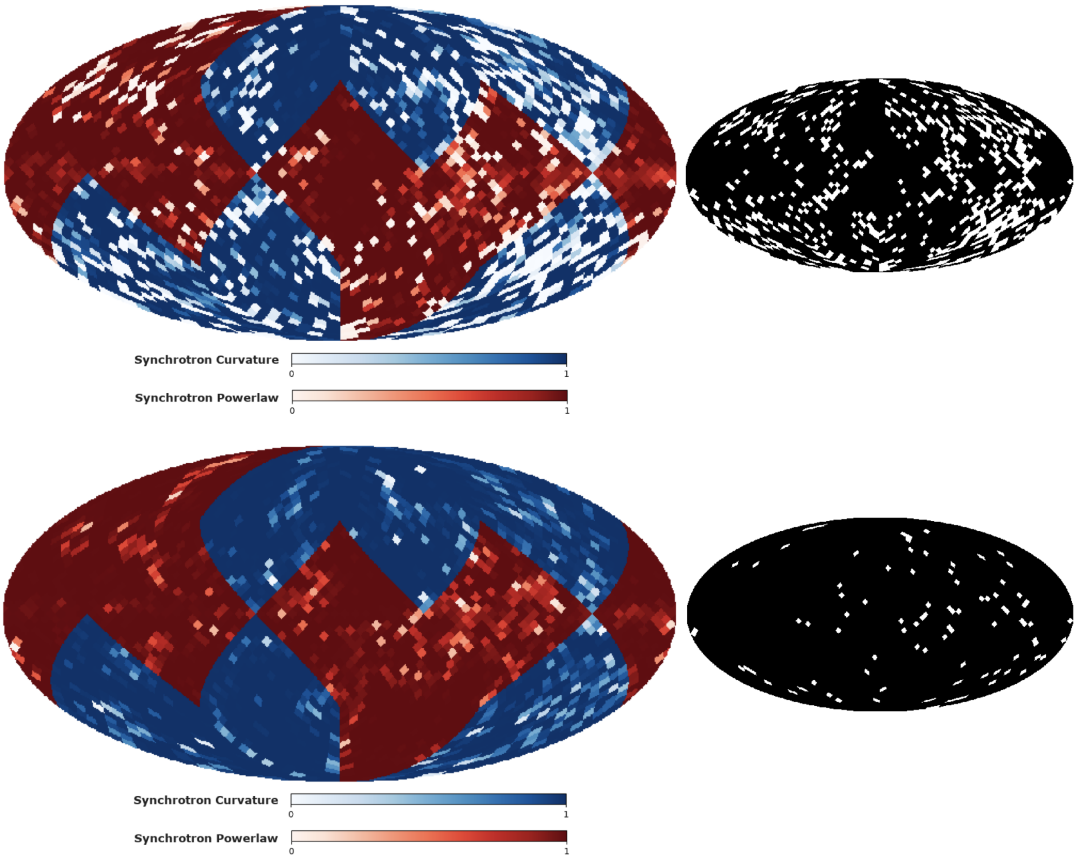

FIGURE A.2: The effect of including noise for the case of Figure A.1. Upper and lower panels shows NN prediction trained with training set without noise and with noise respectively.

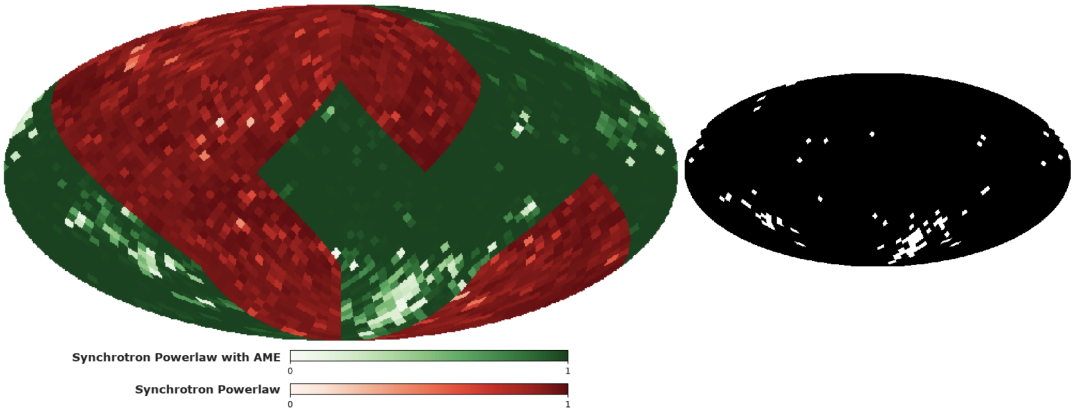

FIGURE A.3: NN prediction of synchrtron emission in presence of AME or not when dust spectral index is changing, the color scale and caption are the same as Figure 5.9

able to learn the difference of sky 90% in the training phase and 96% for the test set without noise (shown in Table A.1). Figure A.3 is showing the result of this analysis, which indicates some similarities in comparison with Figure 5.14. The same as the last Section, the physical interpretation of the misclassified pixels stays identical to Section 5.5.2. Therefore, it is hard for the NN to classify the model with or without AME where the AME emission is faint concerning the synchrotron one.

A.2 Multi classification

For this exercise, the number of simulated maps and vectors stay the same as Section 5.6. Although, the major difference of this analysis regarding the procedure done in Chapter 5 is changing the network's architecture in the case of Multi-class classification. Due to adding more complexity to the classification problem the architecture described in Section 5.3.2 was not able to reach the high accuracy mentioned in Table 5.3 and 5.4 for the different cases. We have added another layer with 272 neurons to the network mentioned in Section 5.3.2, so the applied architecture for classifying four foreground models has seven hidden layers with 272, 272, 136, 68, 34, and 17 neurons, with *tanh* activation function. Moreover, there are three dropout layers with a dropout rate = 0.2, applied to the first three hidden layers.

By changing the architecture as explained above, the NN after 171 epochs is able to learn four different combinations of the foreground models with almost the

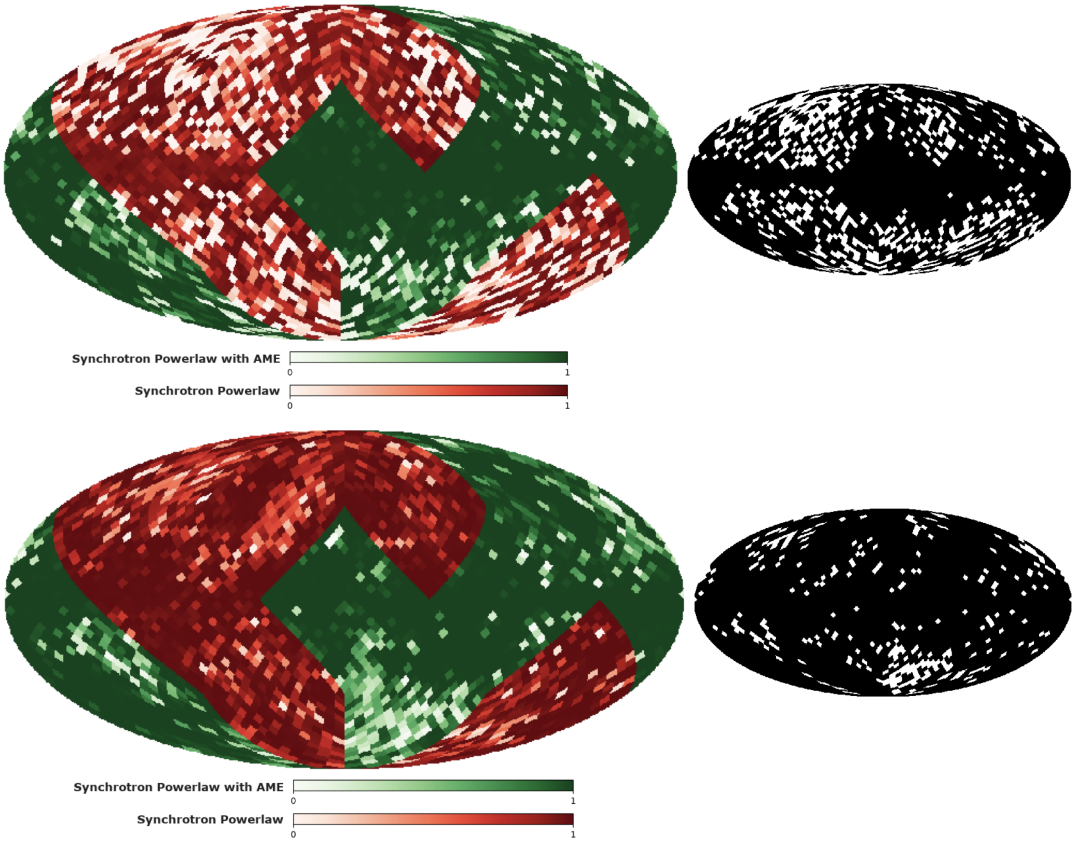

FIGURE A.4: The effect of including noise for the case of Figure A.3. Upper and lower panels shows NN prediction trained with training set without noise and with noise respectively.

Sky models	Accuracy on training set	Accuracy on test set
Pure power-law & Curvature	97%	97%
AME & Pure power-law	90%	96%
AME & Pure power-law & Curvature	82%	95%

TABLE A.1: Same as Table 5.3, accuracy on training and test sets of the NN for different sky models in the basic configuration without noise, in presence of dust spectral index variant.

same accuracy of Section 5.6. The accuracy in the training reaches 82% while for the test set it arrives to 95% (as mentioned in Table A.2). Figure A.3 shows the NN prediction of four models in the noiseless case when the dust spectral index is variant. By looking at this Figure, we could conclude that most misclassified pixels are in the regions where AME is present. This fact was quite predictable since AME emission is fainter than synchrotron and thermal dust emissions.

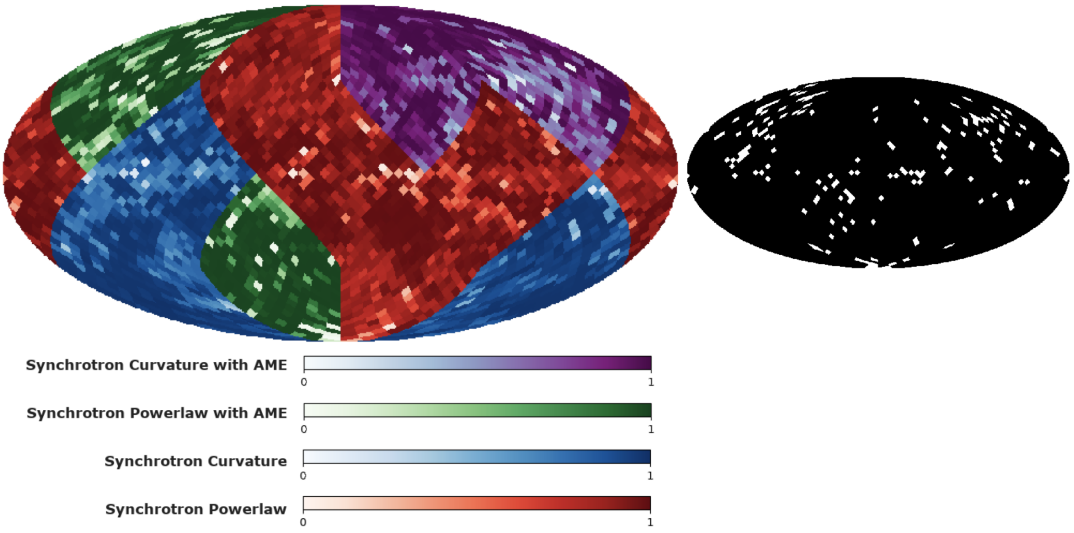

FIGURE A.5: NN prediction of multi foreground models classification when dust spectral index is changing, the color scale and caption are the same as Figure 5.12.

Sky models	Acc on test set for training with noiseless data	Acc for Re-Training with $N_{side} = 1024$, 1 noise realization	Acc for Re-Training with $N_{side} = 16$, 100 noise realizations
Pure power-law & Curvature	81%	96%	97%
AME & Pure power-law	78%	92%	91%
AME & Pure power-law & Curvature	63%	91%	90%

TABLE A.2: Same as Table 5.4, but in presence of variant dust spectral index.

A.3 Classification in presence of noise

We have tested our NN, also for the challenging case, in the presence of noise and variable thermal dust spectral index. The noise simulations and instrument sensitivities are the same as Section 5.7. We have applied two different approaches for testing the NN for the noisy data, as it is explained in Section 5.7. In the first approach, we checked the NN's performance when the NN had been trained only on noiseless data. In the second one, after the first step of training based on

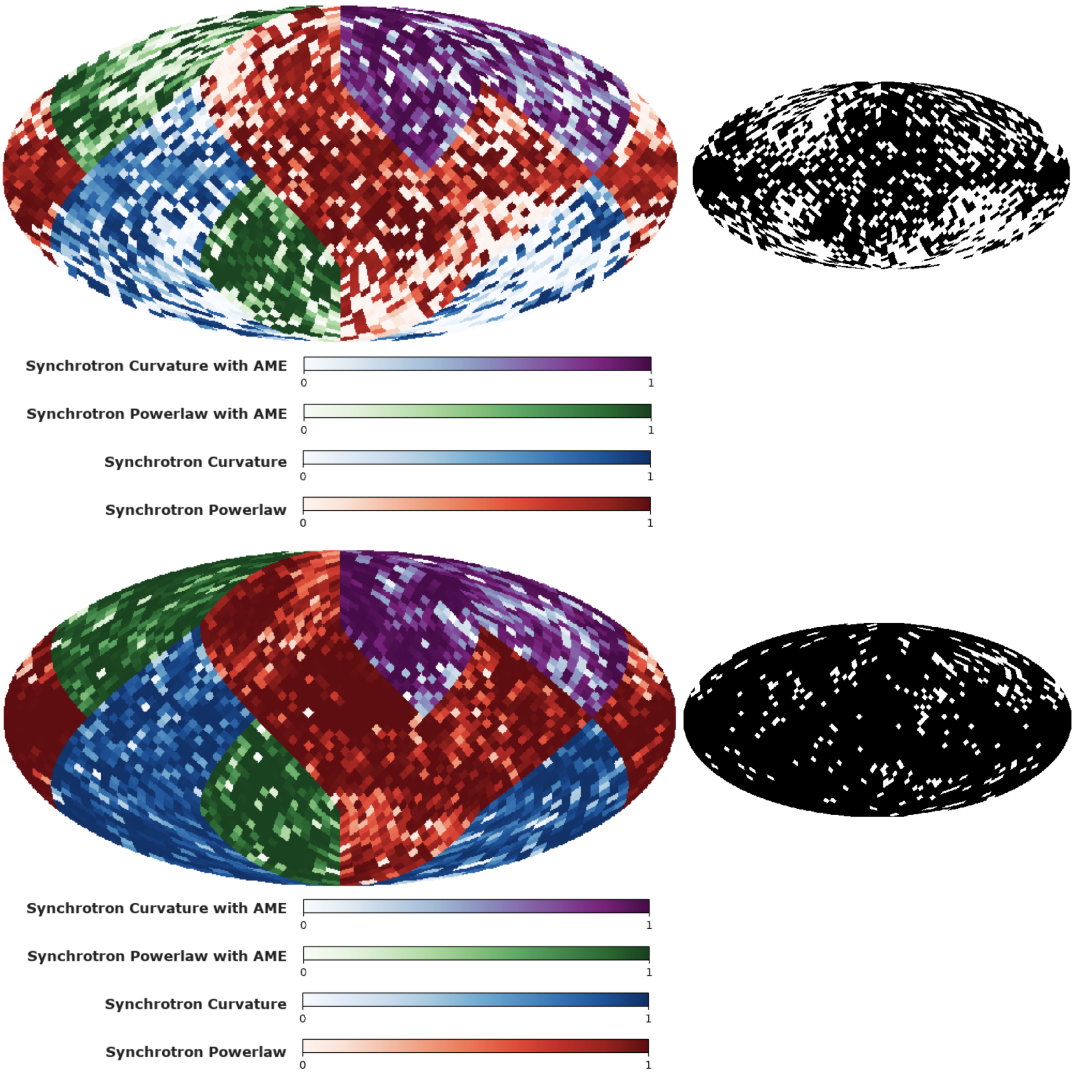

FIGURE A.6: The effect of including the noise in the training set in the case of having dust spectral index variety in the sky. The color scales for the considered models and panels explanations are the same as Figure 5.15.

noiseless data and saving the model, we have re-trained the model with fixed weights on the noisy data in two manners of high resolution ($N_{side} = 1024$) but one noise realization, and low resolution ($N_{side} = 16$) with one hundred noise realizations. These results and the NN predictions, including noisy and noiseless cases for different foreground components, are summarized in Table A.2.

In Figure A.2, the NN prediction for classifying synchrotron power-low from synchrotron curvature in the presence of noise and variable thermal dust spectral index has been shown. The Upper and lower panels of this Figure indicate that the NN has been trained without and with noisy data. In the upper panel, the

NN's prediction accuracy is 81% (first approach), but this accuracy increases up to 97% in the lower panel (second approach).

Figure A.4 carries the same information about the binary case for the discerning presence of AME in the sky or not. The upper panel of Figure A.4 shows 78% accuracy of the NN prediction for this case in the presence of noise and variable thermal dust spectral index. This accuracy can make progress to 91% for the NN re-trained with the noisy data. Figure A.6 shows the effect of including the noise in the Multi classification case. Upper and lower panels with 63% and 90% accuracy indicate the NN's improvement in classifying the correct model in the presence of noise. The important message taken from these figures is that the NN trained on only noiseless data is prone to misclassify the foreground model and has less accuracy while for the NN re-trained with noisy data, the accuracy increase significantly.

A.4 Comparison with chi-squared information

In this Section, we compare the chi-squared information with the NN prediction for discerning synchrotron with or without curvature when the dust spectral index is variant and in the presence of noise. In Figure A.7, we are showing this comparison where the upper panel indicates the preference of χ^2 information and lower panel is the NN prediction. The color scale and panel explanations are the same as Figure 5.16. Here, the upper panel that comes from chi-squared information show less accuracy (68%) with respect to Section 5.8 (73%). Due to adding another parameter (dust spectral index) to the solver of component separation in FGBuster while keeping the same signal-to-noise ratio, the numerical instabilities increases and the accuracy based on χ^2 information drops. The results shown in the lower panel are coming from the NN prediction when the model is re-trained with the low resolution ($N_{side} = 16$) but one hundred noise realizations. In this case, the NN distinguishes the foreground model with 97% accuracy correctly. This results indicate a significant improvement when the NN is applied to the noisy case in comparison to chi-squared information.

Figure A.8 shows the difference between χ^2 values in each pixel for the two different cases across the sky in the presence of variant dust spectral index. The same as

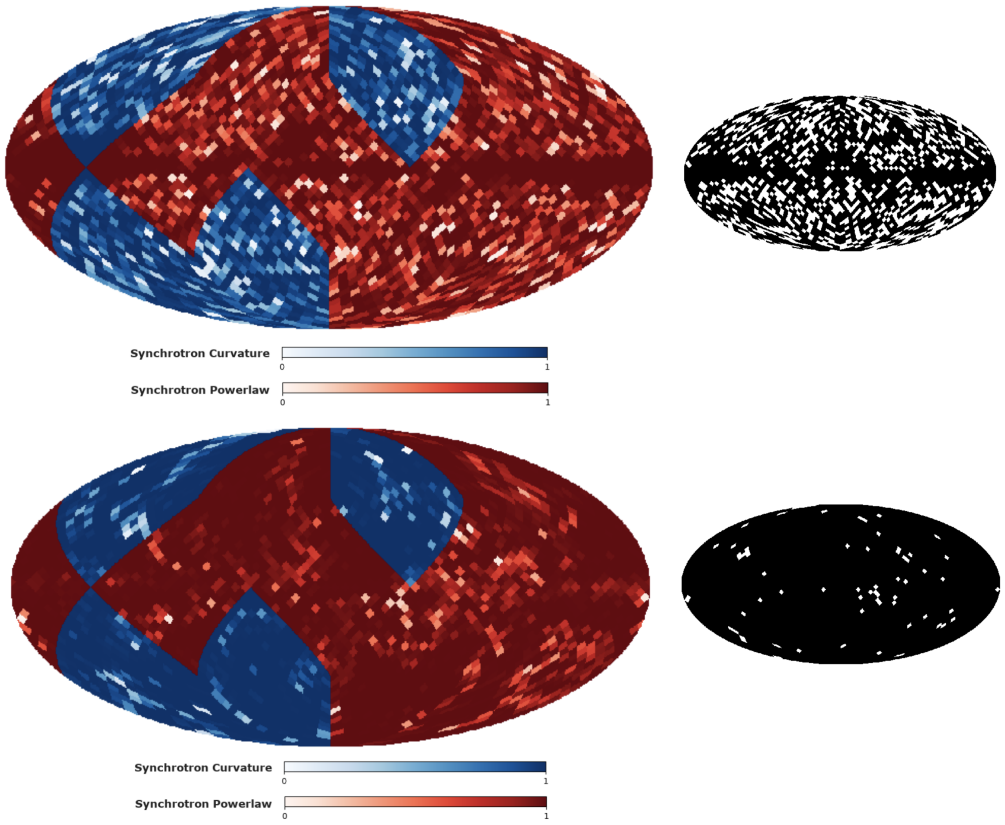

FIGURE A.7: Same as Figure 5.16, Comparison of χ^2 analysis with NN prediction in the presence of variant dust spectral index and noise on test maps. This comparison shows the advantage of using a NN approach, leading to a correct classification on about 97% of the pixels (lower panels) with respect to about 68% when the χ^2 information is used (upper panels).

Figure 5.17, in the region where the sky signal is low at intermediate and high galactic latitudes, the difference between the two models is very near zero.

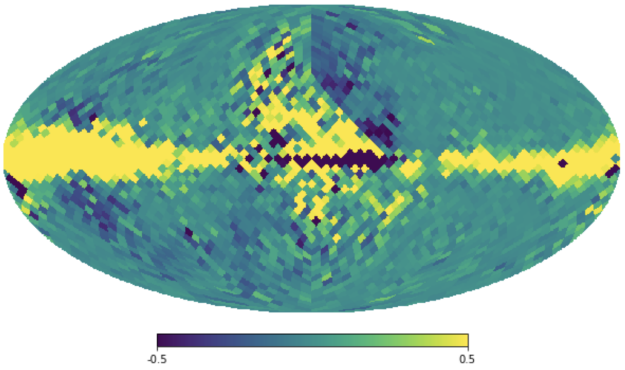

FIGURE A.8: Reduced χ^2 difference for each pixel, obtained when the fit is done considering pure power law SED for synchrotron and when curvature is included in presence of dust spectral index variant, meaning that in fitting procedure dust spectral index is another free parameter, The colorbar is the same as Figure 5.17.

Bibliography

[1] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. , 142:419–421, July 1965. doi: 10.1086/148307.

[2] J. C. Mather, E. S. Cheng, Jr. Eplee, R. E., R. B. Isaacman, S. S. Meyer, R. A. Shafer, R. Weiss, E. L. Wright, C. L. Bennett, N. W. Boggess, E. Dwek, S. Gulkis, M. G. Hauser, M. Janssen, T. Kelsall, P. M. Lubin, Jr. Moseley, S. H., T. L. Murdock, R. F. Silverberg, G. F. Smoot, and D. T. Wilkinson. A Preliminary Measurement of the Cosmic Microwave Background Spectrum by the Cosmic Background Explorer (COBE) Satellite. , 354:L37, May 1990. doi: 10.1086/185717.

[3] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, and et al. Nine-yearwilkinson microwave anisotropy probe(wmap) observations: Final maps and results. *The Astrophysical Journal Supplement Series*, 208(2):20, September 2013. ISSN 1538-4365. doi: 10.1088/0067-0049/208/2/20. URL http://dx.doi.org/10.1088/0067-0049/208/2/20.

[4] Y. Akrami et al. Planck 2018 results. I. Overview and the cosmological legacy of Planck. 7 2018.

[5] P. A. R. Ade et al. Measurements of Degree-Scale B-mode Polarization with the BICEP/Keck Experiments at South Pole. In *53rd Rencontres de Moriond on Cosmology La Thuile, Italy, March 17-24, 2018*, 2018.

[6] Wayne Hu and Martin J. White. A CMB polarization primer. *New Astron.*, 2:323, 1997. doi: 10.1016/S1384-1076(97)00022-5.

[7] P. A. R. Ade et al. Detection of B-Mode Polarization at Degree Angular Scales by BICEP2. *Phys. Rev. Lett.*, 112(24):241101, 2014. doi: 10.1103/PhysRevLett.112.241101.

[8] Marc Kamionkowski, Arthur Kosowsky, and Albert Stebbins. Statistics of cosmic microwave background polarization. *Phys. Rev.*, D55:7368–7388, 1997. doi: 10.1103/PhysRevD.55.7368.

[9] Paolo Campeti, Davide Poletti, and Carlo Baccigalupi. Principal component analysis of the primordial tensor power spectrum. *JCAP*, 1909(09):055, 2019. doi: 10.1088/1475-7516/2019/09/055.

[10] M. Tristram et al. Planck constraints on the tensor-to-scalar ratio. 10 2020.

[11] Y. Akrami et al. Planck 2018 results. IV. Diffuse component separation. 2018.

[12] N. Krachmalnicoff, C. Baccigalupi, J. Aumont, M. Bersanelli, and A. Mennella. Characterization of foreground emission on degree angular scales for CMB B-mode observations - Thermal dust and synchrotron signal from Planck and WMAP data. *Astron. Astrophys.*, 588:A65, 2016. doi: 10.1051/0004-6361/201527678.

[13] N. Krachmalnicoff et al. S-PASS view of polarized Galactic synchrotron at 2.3 GHz as a contaminant to CMB observations. *Astron. Astrophys.*, 618: A166, 2018. doi: 10.1051/0004-6361/201832768.

[14] E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173, March 1929. doi: 10.1073/pnas.15.3.168.

[15] Edwin Hubble and Milton L. Humason. The Velocity-Distance Relation among Extra-Galactic Nebulae. , 74:43, July 1931. doi: 10.1086/143323.

[16] H. P. Robertson. Kinematics and World-Structure. , 82:284, November 1935. doi: 10.1086/143681.

[17] Scott Dodelson. *Modern Cosmology*. Academic Press, Elsevier Science, 2003.

[18] Adam G. Riess et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.*, 116:1009–1038, 1998. doi: 10.1086/300499.

[19] S. Perlmutter et al. Measurements of Ω and Λ from 42 high redshift supernovae. *Astrophys. J.*, 517:565–586, 1999. doi: 10.1086/307221.

[20] Edvige Corbelli and Paolo Salucci. The Extended Rotation Curve and the Dark Matter Halo of M33. *Mon. Not. Roy. Astron. Soc.*, 311:441–447, 2000. doi: 10.1046/j.1365-8711.2000.03075.x.

[21] Douglas Clowe, Marusa Bradac, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *Astrophys. J. Lett.*, 648:L109–L113, 2006. doi: 10.1086/508162.

[22] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. 2018.

[23] Adam G. Riess, Stefano Casertano, Wenlong Yuan, Lucas M. Macri, and Dan Scolnic. Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM. *Astrophys. J.*, 876(1):85, 2019. doi: 10.3847/1538-4357/ab1422.

[24] Alan H. Guth and Erick J. Weinberg. Could the Universe Have Recovered from a Slow First Order Phase Transition? *Nucl. Phys.*, B212:321–364, 1983. doi: 10.1016/0550-3213(83)90307-3.

[25] Andrei D. Linde. Inflationary Cosmology. *Lect. Notes Phys.*, 738:1–54, 2008. doi: 10.1007/978-3-540-74353-8_1.

[26] Alan H. Guth and S. H. H. Tye. Phase Transitions and Magnetic Monopole Production in the Very Early Universe. *Phys. Rev. Lett.*, 44:631, 1980. doi: 10.1103/PhysRevLett.44.631,10.1103/PhysRevLett.44.963.2. [Erratum: Phys. Rev. Lett.44,963(1980)].

[27] A.D. Linde. A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, 108(6):389 – 393, 1982. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(82)91219-9. URL http://www.sciencedirect.com/science/article/pii/0370269382912199.

[28] Andrew R. Liddle. An Introduction to cosmological inflation. In *ICTP Summer School in High-Energy Physics and Cosmology*, pages 260–295, 1 1999.

[29] G. Gamow. The origin of elements and the separation of galaxies. *Phys. Rev.*, 74:505–506, August 1948. doi: 10.1103/PhysRev.74.505.2. URL https://link.aps.org/doi/10.1103/PhysRev.74.505.2.

[30] Ralph A. Alpher and Robert Herman. Evolution of the Universe. , 162 (4124):774–775, November 1948. doi: 10.1038/162774b0.

[31] J. C. Mather, E. S. Cheng, D. A. Cottingham, R. E. Eplee, Jr., D. J. Fixsen, T. Hewagama, R. B. Isaacman, K. A. Jensen, S. S. Meyer, P. D. Noerdlinger, S. M. Read, L. P. Rosen, R. A. Shafer, E. L. Wright, C. L. Bennett, N. W. Boggess, M. G. Hauser, T. Kelsall, S. H. Moseley, Jr., R. F. Silverberg, G. F. Smoot, R. Weiss, and D. T. Wilkinson. Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument. , 420: 439–444, January 1994. doi: 10.1086/173574.

[32] Y. Akrami et al. Planck 2018 results. I. Overview and the cosmological legacy of Planck. 2018.

[33] D. Fixsen, E. Cheng, D. Cottingham, Robert Eplee, D. Massa, John Mather, Stephan Meyer, R. Shafer, and E. Wright. The far infrared absolute spectrophotometer (firas): Cosmic background spectrum, dipole and distortions. 24:1149, 11 1992.

[34] Alexei A. Starobinsky. Spectrum of relict gravitational radiation and the early state of the universe. *JETP Lett.*, 30:682–685, 1979.

[35] Ruth Durrer. *The Cosmic Microwave Background*. Cambridge University Press, 2008. doi: 10.1017/CBO9780511817205.

[36] James M. Bardeen. Gauge-invariant cosmological perturbations. *Phys. Rev. D*, 22:1882–1905, October 1980. doi: 10.1103/PhysRevD.22.1882. URL https://link.aps.org/doi/10.1103/PhysRevD.22.1882.

[37] Viatcheslav F. Mukhanov and G.V. Chibisov. The Vacuum energy and large scale structure of the universe. *Sov. Phys. JETP*, 56:258–265, 1982.

[38] Jerome Martin, Christophe Ringeval, and Vincent Vennin. Encyclopædia Inflationaris. *Phys. Dark Univ.*, 5-6:75–235, 2014. doi: 10.1016/j.dark.2014.01.003.

[39] Y. Akrami et al. Planck 2018 results. IX. Constraints on primordial non-Gaussianity. 5 2019.

[40] Uros Seljak and Matias Zaldarriaga. A Line of sight integration approach to cosmic microwave background anisotropies. *Astrophys. J.*, 469:437–444, 1996. doi: 10.1086/177793.

[41] R. K. Sachs and A. M. Wolfe. Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. , 147:73, January 1967. doi: 10.1086/148982.

[42] Wayne Hu and Scott Dodelson. Cosmic Microwave Background Anisotropies. *Ann. Rev. Astron. Astrophys.*, 40:171–216, 2002. doi: 10.1146/annurev.astro.40.060401.093926.

[43] Anthony Challinor, Hiranya Peiris, Mario Novello, and Santiago Perez. Lecture notes on the physics of cosmic microwave background anisotropies. *AIP Conference Proceedings*, 2009. doi: 10.1063/1.3151849. URL http://dx.doi.org/10.1063/1.3151849.

[44] Uros Seljak and Matias Zaldarriaga. Signature of gravity waves in polarization of the microwave background. *Phys. Rev. Lett.*, 78:2054–2057, 1997. doi: 10.1103/PhysRevLett.78.2054.

[45] Matias Zaldarriaga. Nature of the E B decomposition of CMB polarization. *Phys. Rev. D*, 64:103001, 2001. doi: 10.1103/PhysRevD.64.103001.

[46] A. Blanchard and J. Schneider. Gravitational lensing effect on the fluctuations of the cosmic background radiation. , 184(1-2):1–6, October 1987.

[47] Antony Lewis and Anthony Challinor. Weak gravitational lensing of the CMB. *Phys. Rept.*, 429:1–65, 2006. doi: 10.1016/j.physrep.2006.03.002.

[48] Duncan Hanson, Anthony Challinor, and Antony Lewis. Weak lensing of the cmb. *General Relativity and Gravitation*, 42(9):2197–2218, June 2010. ISSN 1572-9532. doi: 10.1007/s10714-010-1036-y. URL http://dx.doi.org/10.1007/s10714-010-1036-y.

[49] N. Aghanim et al. Planck 2018 results. VIII. Gravitational lensing. 7 2018.

[50] Carlo Baccigalupi and Viviana Acquaviva. The CMB as a dark energy probe. *PoS*, CMB2006:008, 2006. doi: 10.22323/1.027.0008.

[51] Uros Seljak and Christopher M. Hirata. Gravitational lensing as a contaminant of the gravity wave signal in CMB. *Phys. Rev. D*, 69:043005, 2004. doi: 10.1103/PhysRevD.69.043005.

[52] P. A. R. Ade et al. BICEP2 / Keck Array x: Constraints on Primordial Gravitational Waves using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season. *Phys. Rev. Lett.*, 121:221301, 2018. doi: 10.1103/PhysRevLett.121.221301.

[53] S. P. Boughn and R. G. Crittenden. A detection of the integrated Sachs Wolfe effect. , 49(2-6):75–78, May 2005. doi: 10.1016/j.newar.2005.01.005.

[54] M.J. Rees and D.W. Sciama. Large scale Density Inhomogeneities in the Universe. *Nature*, 217:511–516, 1968. doi: 10.1038/217511a0.

[55] R.A. Sunyaev and Ya.B. Zeldovich. Small scale fluctuations of relic radiation. *Astrophys. Space Sci.*, 7:3–19, 1970.

[56] R. A. Sunyaev and Y. B. Zeldovich. The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4:173, November 1972.

[57] Nick Hand, Graeme E. Addison, Eric Aubourg, Nick Battaglia, Elia S. Battistelli, Dmitry Bizyaev, J. Richard Bond, Howard Brewington, Jon Brinkmann, Benjamin R. Brown, and et al. Evidence of galaxy cluster motions with the kinematic sunyaev-zel'dovich effect. *Physical Review Letters*, 109(4), July 2012. ISSN 1079-7114. doi: 10.1103/physrevlett.109.041101. URL http://dx.doi.org/10.1103/PhysRevLett.109.041101.

[58] M. Birkinshaw. The Sunyaev-Zel'dovich effect. *Physics Reports*, 310(2-3): 97–195, Mar 1999. doi: 10.1016/S0370-1573(98)00080-5.

[59] J. Chluba and K. Mannheim. Kinetic Sunyaev-Zeldovich effect from galaxy cluster rotation. *Astron. Astrophys.*, 396:419–428, 2002. doi: 10.1051/0004-6361:20021429.

[60] Kevork N. Abazajian et al. CMB-S4 Science Book, First Edition. 2016.

[61] George B. Rybicki and Alan P. Lightman. *Radiative Processes in Astrophysics*. 1986.

[62] U. Fuskeland, I.K. Wehus, H.K. Eriksen, and S.K. Naess. Spatial variations in the spectral index of polarized synchrotron emission in the 9-yr WMAP sky maps. *Astrophys. J.*, 790:104, 2014. doi: 10.1088/0004-637X/790/2/104.

[63] E. Carretti et al. S-band Polarization All Sky Survey (S-PASS): survey description and maps. *Mon. Not. Roy. Astron. Soc.*, 489(2):2330–2354, 2019. doi: 10.1093/mnras/stz806.

[64] A. Kogut. Synchrotron spectral curvature from 22 mhz to 23 ghz. *The Astrophysical Journal*, 753(2):110, June 2012. ISSN 1538-4357. doi: 10.1088/0004-637x/753/2/110. URL http://dx.doi.org/10.1088/0004-637X/753/2/110.

[65] Y. Akrami et al. Planck 2018 results. XI. Polarized dust foregrounds. 1 2018.

[66] A. Lazarian and D. Finkbeiner. Microwave emission from aligned dust. *New Astron. Rev.*, 47:1107–1116, 2003. doi: 10.1016/j.newar.2003.09.037.

[67] David J. Schlegel, Douglas P. Finkbeiner, and Marc Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. , 500(2):525–553, June 1998. doi: 10.1086/305772.

[68] A. Kogut et al. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Foreground Polarization. *Astrophys. J.*, 665:355–362, 2007. doi: 10.1086/519754.

[69] Steve K. Choi and Lyman A. Page. Polarized galactic synchrotron and dust emission and their correlation. *JCAP*, 12:020, 2015. doi: 10.1088/1475-7516/2015/12/020.

[70] P.A.R. Ade et al. Planck intermediate results. XXII. Frequency dependence of thermal emission from Galactic dust in intensity and polarization. *Astron. Astrophys.*, 576:A107, 2015. doi: 10.1051/0004-6361/201424088.

[71] Clive Dickinson, Yacine Ali-Haïmoud, Adam Barr, E. Battistelli, Aaron Bell, L. Bernstein, S. Casassus, K. Cleary, B. Draine, R. Génova-Santos, S. Harper, B. Hensley, Jaz Hill-Valler, Thiem Hoang, F. Israel, L. Jew, A. Lazarian, J. Leahy, J. Leech, and Matias Vidal. The state-of-play of

anomalous microwave emission (ame) research. *New Astronomy Reviews*, 80, 02 2018. doi: 10.1016/j.newar.2018.02.001.

[72] B. T. Draine and A. Lazarian. Electric dipole radiation from spinning dust grains. *Astrophys. J.*, 508:157–179, 1998. doi: 10.1086/306387.

[73] B. T. Draine and A. Lazarian. Magnetic dipole microwave emission from dust grains. *Astrophys. J.*, 512:740, 1999. doi: 10.1086/306809.

[74] R. Génova-Santos et al. QUIJOTE scientific results – I. Measurements of the intensity and polarisation of the anomalous microwave emission in the Perseus molecular complex. *Mon. Not. Roy. Astron. Soc.*, 452(4):4169–4182, 2015. doi: 10.1093/mnras/stv1405.

[75] R. Génova-Santos et al. QUIJOTE scientific results – II. Polarisation measurements of the microwave emission in the Galactic molecular complexes W43 and W47 and supernova remnant W44. *Mon. Not. Roy. Astron. Soc.*, 464(4):4107–4132, 2017. doi: 10.1093/mnras/stw2503.

[76] M. Remazeilles, C. Dickinson, H.K.K. Eriksen, and I.K. Wehus. Sensitivity and foreground modelling for large-scale cosmic microwave background B-mode polarization satellite missions. *Mon. Not. Roy. Astron. Soc.*, 458(2): 2032–2050, 2016. doi: 10.1093/mnras/stw441.

[77] Y. Ali-Haimoud, C. M. Hirata, and C. Dickinson. A refined model for spinning dust radiation. *Mon. Not. Roy. Astron. Soc.*, 395:1055, 2009. doi: 10.1111/j.1365-2966.2009.14599.x.

[78] C. L. Bennett, G. F. Smoot, G. Hinshaw, E. L. Wright, A. Kogut, G. de Amici, S. S. Meyer, R. Weiss, D. T. Wilkinson, S. Gulkis, M. Janssen, N. W. Boggess, E. S. Cheng, M. G. Hauser, T. Kelsall, J. C. Mather, Jr. Moseley, S. H., T. L. Murdock, and R. F. Silverberg. Preliminary Separation of Galactic and Cosmic Microwave Emission for the COBE Differential Microwave Radiometers. , 396:L7, September 1992. doi: 10.1086/186505.

[79] F. K. Hansen, A. J. Banday, H. K. Eriksen, K. M. Gorski, and P. B. Lilje. Foreground subtraction of cosmic microwave background maps using WI-FIT (wavelet-based high-resolution fitting of internal templates). *The Astrophysical Journal*, 648(2):784–796, September 2006. doi: 10.1086/506015. URL https://doi.org/10.1086%2F506015.

[80] Jacques Delabrouille, J.-F. Cardoso, and G. Patanchon. Multi-detector multi-component spectral matching and applications for CMB data analysis. *Mon. Not. Roy. Astron. Soc.*, 346:1089, 2003. doi: 10.1111/j.1365-2966.2003.07069.x.

[81] Davide Maino, S. Donzelli, A. Banday, F. Stivoli, and Carlo Baccigalupi. Cosmic microwave background signal in wilkinson microwave anisotropy probe three-year data with fastica. *Monthly Notices of the Royal Astronomical Society*, 374:1207 – 1215, 12 2006. doi: 10.1111/j.1365-2966.2006.11255.x.

[82] Vladislav Stolyarov, M.P. Hobson, A.N. Lasenby, and R.B. Barreiro. All-sky component separation in the presence of anisotropic noise and dust temperature variations. *Mon. Not. Roy. Astron. Soc.*, 357:145–155, 2005. doi: 10.1111/j.1365-2966.2005.08610.x.

[83] W. N. Brandt, C. R. Lawrence, A. C. S. Readhead, J. N. Pakianathan, and T. M. Fiola. Separation of Foreground Radiation from Cosmic Microwave Background Anisotropy Using Multifrequency Measurements. , 424:1, March 1994. doi: 10.1086/173867.

[84] H. K. Eriksen, C. Dickinson, C. R. Lawrence, C. Baccigalupi, A. J. Banday, K. M. Górski, F. K. Hansen, P. B. Lilje, E. Pierpaoli, M. D. Seiffert, K. M. Smith, and K. Vanderlinde. Cosmic Microwave Background Component Separation by Parameter Estimation. , 641(2):665–682, April 2006. doi: 10.1086/500499.

[85] Radek Stompor, Josquin Errard, and Davide Poletti. Forecasting performance of CMB experiments in the presence of complex foreground contaminations. *Phys. Rev. D*, 94(8):083526, 2016. doi: 10.1103/PhysRevD.94.083526.

[86] A. H. Jaffe, A. Balbi, J. R. Bond, J. Borrill, P. G. Ferreira, D. Finkbeiner, S. Hanany, A. T. Lee, B. Rabii, P. L. Richards, G. F. Smoot, R. Stompor, C. D. Winant, and J. H. P. Wu. Determining foreground contamination in cosmic microwave background observations: Diffuse galactic emission in the MAXIMA-i field. *The Astrophysical Journal*, 615(1):55–62, November 2004. doi: 10.1086/422635. URL https://doi.org/10.1086%2F422635.

[87] R. Stompor, Samuel M. Leach, F. Stivoli, and C. Baccigalupi. Maximum Likelihood algorithm for parametric component separation in CMB

experiments. *Mon. Not. Roy. Astron. Soc.*, 392:216, 2009. doi: 10.1111/j.1365-2966.2008.14023.x.

[88] J. Delabrouille, J. F. Cardoso, M. Le Jeune, M. Betoule, G. Fay, and F. Guilloux. A full sky, low foreground, high resolution CMB map from WMAP. , 493(3):835–857, January 2009. doi: 10.1051/0004-6361:200810514.

[89] Jason Dick, Mathieu Remazeilles, and Jacques Delabrouille. Impact of calibration errors on CMB component separation using FastICA and ILC. , 401 (3):1602–1612, January 2010. doi: 10.1111/j.1365-2966.2009.15798.x.

[90] Jean-Francois Cardoso, Maude Martin, Jacques Delabrouille, Marc Betoule, and Guillaume Patanchon. Component separation with flexible models. Application to the separation of astrophysical emissions. 3 2008.

[91] Nobuhiko Katayama and Eiichiro Komatsu. Simple foreground cleaning algorithm for detecting primordialb-mode polarization of the cosmic microwave background. *The Astrophysical Journal*, 737(2):78, August 2011. ISSN 1538-4357. doi: 10.1088/0004-637x/737/2/78. URL http://dx.doi.org/10.1088/0004-637X/737/2/78.

[92] E. Martinez-Gonzalez, J.M. Diego, P. Vielva, and J. Silk. CMB power spectrum estimation and map reconstruction with the expectation - Maximization algorithm. *Mon. Not. Roy. Astron. Soc.*, 345:1101, 2003. doi: 10.1046/j.1365-2966.2003.06885.x.

[93] D. Hanson et al. Detection of B-mode Polarization in the Cosmic Microwave Background with Data from the South Pole Telescope. *Phys. Rev. Lett.*, 111 (14):141301, 2013. doi: 10.1103/PhysRevLett.111.141301.

[94] P. A. R. Ade et al. A Measurement of the Cosmic Microwave Background B-Mode Polarization Power Spectrum at Sub-Degree Scales from 2 years of POLARBEAR Data. *Astrophys. J.*, 848(2):121, 2017. doi: 10.3847/1538-4357/aa8e9f.

[95] P. A. R. Ade et al. Planck intermediate results. XLI. A map of lensing-induced B-modes. *Astron. Astrophys.*, 596:A102, 2016. doi: 10.1051/0004-6361/201527932.

[96] Thibaut Louis et al. The Atacama Cosmology Telescope: Two-Season ACTPol Spectra and Parameters. *JCAP*, 1706(06):031, 2017. doi: 10.1088/1475-7516/2017/06/031.

[97] Luke Jew, Angela C Taylor, Michael E Jones, A Barr, H C Chiang, C Dickinson, R D P Grumitt, S E Harper, H M Heilgendorff, J Hill-Valler, and et al. The c-band all-sky survey (c-bass): Simulated parametric fitting in single pixels in total intensity and polarization. *Monthly Notices of the Royal Astronomical Society*, 490(2):2958–2975, October 2019. ISSN 1365-2966. doi: 10.1093/mnras/stz2697. URL http://dx.doi.org/10.1093/mnras/stz2697.

[98] J. A. Rubiño-Martín, R. Rebolo, M. Aguiar, R. Génova-Santos, F. Gómez-Reñasco, J. M. Herreros, R. J. Hoyland, C. López-Caraballo, A. E. Pelaez Santos, V. Sanchez de la Rosa, A. Vega-Moreno, T. Viera-Curbelo, E. Martínez-Gonzalez, R. B. Barreiro, F. J. Casas, J. M. Diego, R. Fernández-Cobos, D. Herranz, M. López-Caniego, D. Ortiz, P. Vielva, E. Artal, B. Aja, J. Cagigas, J. L. Cano, L. de la Fuente, A. Mediavilla, J. V. Terán, E. Villa, L. Piccirillo, R. Battye, E. Blackhurst, M. Brown, R. D. Davies, R. J. Davis, C. Dickinson, S. Harper, B. Maffei, M. McCulloch, S. Melhuish, G. Pisano, R. A. Watson, M. Hobson, K. Grainge, A. Lasenby, R. Saunders, and P. Scott. *The QUIJOTE-CMB experiment: studying the polarisation of the galactic and cosmological microwave emissions*, volume 8444 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 84442Y. 2012. doi: 10.1117/12.926581.

[99] G. Addamo et al. The large scale polarization explorer (LSPE) for CMB measurements: performance forecast. 8 2020.

[100] H. Sugai, P. A. R. Ade, Y. Akiba, D. Alonso, K. Arnold, J. Aumont, J. Austermann, C. Baccigalupi, A. J. Banday, R. Banerji, and et al. Updated design of the cmb polarization experiment satellite litebird. *Journal of Low Temperature Physics*, 199(3-4):1107–1117, January 2020. ISSN 1573-7357. doi: 10.1007/s10909-019-02329-w. URL http://dx.doi.org/10.1007/s10909-019-02329-w.

[101] M. Hazumi et al. LiteBIRD: A Satellite for the Studies of B-Mode Polarization and Inflation from Cosmic Background Radiation Detection. *J. Low. Temp. Phys.*, 194(5-6):443–452, 2019. doi: 10.1007/s10909-019-02150-5.

[102] A. Suzuki et al. The POLARBEAR-2 and the Simons Array Experiment. *J. Low. Temp. Phys.*, 184(3-4):805–810, 2016. doi: 10.1007/s10909-015-1425-4.

[103] Peter Ade et al. The Simons Observatory: Science goals and forecasts. *JCAP*, 1902:056, 2019. doi: 10.1088/1475-7516/2019/02/056.

[104] Juan I. Cayuso and Matthew C. Johnson. Towards testing CMB anomalies using the kinetic and polarized Sunyaev-Zel'dovich effects. *Phys. Rev. D*, 101(12):123508, 2020. doi: 10.1103/PhysRevD.101.123508.

[105] Amanda Yoho, Simone Aiola, Craig J. Copi, Arthur Kosowsky, and Glenn D. Starkman. Microwave Background Polarization as a Probe of Large-Angle Correlations. *Phys. Rev. D*, 91(12):123504, 2015. doi: 10.1103/PhysRevD.91.123504.

[106] Ben Thorne, Jo Dunkley, David Alonso, and Sigurd Naess. The Python Sky Model: software for simulating the Galactic microwave sky. *Mon. Not. Roy. Astron. Soc.*, 469(3):2821–2833, 2017. doi: 10.1093/mnras/stx949.

[107] J. A. Rubiño-Martín, Planck Collaboration, and QUIJOTE Collaboration. Cosmology with the Cosmic Microwave Background: Latest Results from the PLANCK satellite and the QUIJOTE experiment. In B. Montesinos, A. Asensio Ramos, F. Buitrago, R. Schödel, E. Villaver, S. Pérez-Hoyos, and I. Ordóñez-Etxeberria, editors, *Highlights on Spanish Astrophysics X*, pages 32–43, March 2019.

[108] K.M. Gorski, Eric Hivon, A.J. Banday, B.D. Wandelt, F.K. Hansen, M. Reinecke, and M. Bartelman. HEALPix - A Framework for high resolution discretization, and fast analysis of data distributed on the sphere. *Astrophys. J.*, 622:759–771, 2005. doi: 10.1086/427976.

[109] Farida Farsian, Nicoletta Krachmalnicoff, and Carlo Baccigalupi. Foreground model recognition through Neural Networks for CMB B-mode observations. *JCAP*, 07:017, 2020. doi: 10.1088/1475-7516/2020/07/017.

[110] A. Suzuki, P. A. R. Ade, Y. Akiba, D. Alonso, K. Arnold, J. Aumont, C. Baccigalupi, D. Barron, S. Basak, S. Beckman, and et al. The litebird satellite mission: Sub-kelvin instrument. *Journal of Low Temperature Physics*, 193(5-6):1048–1056, May 2018. ISSN 1573-7357. doi: 10.1007/s10909-018-1947-7. URL http://dx.doi.org/10.1007/s10909-018-1947-7.

[111] Ben Thorne, Jo Dunkley, David Alonso, Maximilian H. Abitbol, Josquin Errard, J. Colin Hill, Brian Keating, Grant Teply, and Edward J. Wollack. Removal of Galactic foregrounds for the Simons Observatory primordial gravitational wave search. 5 2019.

[112] M. Remazeilles et al. Exploring cosmic origins with CORE: B-mode component separation. *JCAP*, 04:023, 2018. doi: 10.1088/1475-7516/2018/04/023.

[113] Siamak Ravanbakhsh, Junier Oliva, Sebastien Fromenteau, Layne C. Price, Shirley Ho, Jeff Schneider, and Barnabas Poczos. Estimating Cosmological Parameters from the Dark Matter Distribution. 2017.

[114] Daniel George and E. A. Huerta. Deep Neural Networks to Enable Real-time Multimessenger Astrophysics. *Phys. Rev.*, D97(4):044039, 2018. doi: 10.1103/PhysRevD.97.044039.

[115] Arushi Gupta, José Manuel Zorrilla Matilla, Daniel Hsu, and Zoltán Haiman. Non-Gaussian information from weak lensing data via deep learning. *Phys. Rev.*, D97(10):103515, 2018. doi: 10.1103/PhysRevD.97.103515.

[116] H. U. Norgaard-Nielsen and H. E. Jorgensen. Foreground removal from CMB temperature maps using an MLP neural network. *Astrophys. Space Sci.*, 318: 195–206, 2008. doi: 10.1007/s10509-008-9912-6.

[117] Razvan Ciuca, Oscar Hernandez, and Michael Wolman. A convolutional neural network for cosmic string detection in cmb temperature maps. *Monthly Notices of the Royal Astronomical Society*, 08 2017. doi: 10.1093/mnras/stz491.

[118] J. Caldeira, W. L. K. Wu, B. Nord, C. Avestruz, S. Trivedi, and K. T. Story. DeepCMB: Lensing Reconstruction of the Cosmic Microwave Background with Deep Neural Networks. *Astron. Comput.*, 28:100307, 2019. doi: 10.1016/j.ascom.2019.100307.

[119] N. Krachmalnicoff and M. Tomasi. Convolutional neural networks on the healpix sphere: a pixel-based algorithm and its application to cmb data analysis. *Astronomy & Astrophysics*, 628:A129, August 2019. ISSN 1432-0746. doi: 10.1051/0004-6361/201935211. URL http://dx.doi.org/10.1051/0004-6361/201935211.

[120] Oludare Isaac Abiodun, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat AbdElatif Mohamed, and Humaira Arshad. State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4 (11):e00938, 2018. ISSN 2405-8440. doi: https://doi.org/10.1016/j.heliyon.2018.e00938. URL http://www.sciencedirect.com/science/article/pii/S2405844018332067.

[121] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, pages 65–386, 1958.

[122] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning, 2018.

[123] B. Widrow, Rodney Winter, and Robert Baxter. Layered neural nets for pattern recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36:1109 – 1118, 08 1988. doi: 10.1109/29.1638.

[124] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Training pruned neural networks. *CoRR*, abs/1803.03635, 2018. URL http://arxiv.org/abs/1803.03635.

[125] Adam Gaier and David Ha. Weight agnostic neural networks. *CoRR*, abs/1906.04358, 2019. URL http://arxiv.org/abs/1906.04358.

[126] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[127] Kurt Hornik, Maxwell Stinchcombe, Halbert White, et al. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[128] Tariq Rashid. *Make Your Own Neural Network*. CreateSpace Independent Publishing Platform, North Charleston, SC, USA, 1st edition, 2016. ISBN 1530826608.

[129] Tong Zhang. Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Proceedings of the twenty-first international conference on Machine learning*, page 116, 2004.

[130] Michael A. Nielsen. Neural networks and deep learning, 2018. URL http://neuralnetworksanddeeplearning.com/.

[131] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. ISSN 1476-4687. doi: 10.1038/323533a0. URL https://doi.org/10.1038/323533a0.

[132] Matthew D. Zeiler. ADADELTA: An Adaptive Learning Rate Method. *arXiv e-prints*, art. arXiv:1212.5701, December 2012.

[133] E. de la Hoz, P. Vielva, R.B. Barreiro, and E. Martínez-González. On the Detection of CMB B-modes from Ground at Low Frequency. 2 2020.

[134] R. Adam et al. Planck 2015 results. I. Overview of products and scientific results. *Astron. Astrophys.*, 594:A1, 2016. doi: 10.1051/0004-6361/201527101.

[135] Josquin Errard, Stephen M. Feeney, Hiranya V. Peiris, and Andrew H. Jaffe. Robust forecasts on fundamental physics from the foreground-obscured, gravitationally-lensed CMB polarization. *JCAP*, 1603(03):052, 2016. doi: 10.1088/1475-7516/2016/03/052.

[136] P. A. R. Ade et al. Planck 2013 results. XXVIII. The Planck Catalogue of Compact Sources. *Astron. Astrophys.*, 571:A28, 2014. doi: 10.1051/0004-6361/201321524.

[137] Alireza Vafaei Sadr and Farida Farsian. Inpainting via Generative Adversarial Networks for CMB data analysis. 4 2020.

[138] Yehuda Hoffman and E. Ribak. Constrained realizations of gaussian fields - a simple algorithm. *The Astrophysical Journal*, 380:L5–L8, 09 1991. doi: 10.1086/186160.

[139] Martin Bucher and Thibaut Louis. Filling in cosmic microwave background map missing data using constrained gaussian realizations. *Monthly Notices of the Royal Astronomical Society*, 424(3):1694–1713, June 2012. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2012.21138.x. URL http://dx.doi.org/10.1111/j.1365-2966.2012.21138.x.

[140] Jaiseung Kim, Pavel Naselsky, and Nazzareno Mandolesi. Harmonic inpainting of cosmic microwave background sky by constrained gaussian realization. *The Astrophysical Journal*, 750(1):L9, April 2012. ISSN 2041-8213. doi: 10.1088/2041-8205/750/1/l9. URL http://dx.doi.org/10.1088/2041-8205/750/1/L9.

[141] H. F. Gruetjen, J. R. Fergusson, M. Liguori, and E. P. S. Shellard. Using inpainting to construct accurate cut-sky CMB estimators. *Phys. Rev.*, D95 (4):043532, 2017. doi: 10.1103/PhysRevD.95.043532.

[142] Ben Hoyle. Measuring photometric redshifts using galaxy images and deep neural networks. *Astronomy and Computing*, 16:34–40, 2016.

[143] Daniel George and EA Huerta. Deep learning for real-time gravitational wave detection and parameter estimation: Results with advanced ligo data. *Physics Letters B*, 778:64–70, 2018.

[144] A Vafaei Sadr, Etienne E Vos, Bruce A Bassett, Zafiirah Hosenie, N Oozeer, and Michelle Lochner. Deepsource: point source detection using deep learning. *Monthly Notices of the Royal Astronomical Society*, 484(2):2793–2806, 2019.

[145] Siyu He, Yin Li, Yu Feng, Shirley Ho, Siamak Ravanbakhsh, Wei Chen, and Barnabás Póczos. Learning to predict the cosmological structure formation. *Proceedings of the National Academy of Sciences*, 116(28):13825–13832, 2019.

[146] Michelle Ntampaka, Camille Avestruz, Steven Boada, Joao Caldeira, Jessi Cisewski-Kehe, Rosanne Di Stefano, Cora Dvorkin, August E Evrard, Arya Farahi, Doug Finkbeiner, et al. The role of machine learning in the next decade of cosmology. *arXiv preprint arXiv:1902.10159*, 2019.

[147] A. Vafaei Sadr, M. Farhang, S. M. S. Movahed, B. Bassett, and M. Kunz. Cosmic string detection with tree-based machine learning. *Mon. Not. Roy. Astron. Soc.*, 478(1):1132–1140, 2018. doi: 10.1093/mnras/sty1055.

[148] Guangyu Zhang, Chi-Ting Chiang, Chris Sheehy, Anže Slosar, and Jian Wang. Predicting cmb dust foreground using galactic 21 cm data. *arXiv preprint arXiv:1904.13265*, 2019.

[149] Giuseppe Puglisi and Xiran Bai. Inpainting galactic foreground intensity and polarization maps using convolutional neural network, 2020.

[150] Kai Yi, Yi Guo, Yanan Fan, Jan Hamann, and Yu Guang Wang. CosmoVAE: Variational Autoencoder for CMB Image Inpainting. 2020.

[151] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications, 2020.

[152] Tilman Tröster, Cameron Ferguson, Joachim Harnois-Déraps, and Ian G. McCarthy. Painting with baryons: augmenting N-body simulations with gas using deep generative models. *Mon. Not. Roy. Astron. Soc.*, 487(1):L24–L29, 2019. doi: 10.1093/mnrasl/slz075.

[153] Juan Zamudio-Fernandez, Atakan Okan, Francisco Villaescusa-Navarro, Seda Bilaloglu, Asena Derin Cengiz, Siyu He, Laurence Perreault Levasseur, and Shirley Ho. HIGAN: Cosmic Neutral Hydrogen with Generative Adversarial Networks. 2019.

[154] Mustafa Mustafa, Deborah Bard, Wahid Bhimji, Zarija Lukić, Rami Al-Rfou, and Jan M. Kratochvil. Cosmogan: creating high-fidelity weak lensing convergence maps using generative adversarial networks. *Computational Astrophysics and Cosmology*, 6(1), May 2019. ISSN 2197-7909. doi: 10.1186/s40668-019-0029-9. URL http://dx.doi.org/10.1186/s40668-019-0029-9.

[155] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[156] Karol Kurach, Mario Lucic, Xiaohua Zhai, Marcin Michalski, and Sylvain Gelly. The GAN landscape: Losses, architectures, regularization, and normalization, 2019. URL https://openreview.net/forum?id=rkGG6s0qKQ.

[157] Charles Holt and Alvin Roth. The nash equilibrium: A perspective. *Proceedings of the National Academy of Sciences of the United States of America*, 101:3999–4002, 04 2004. doi: 10.1073/pnas.0308738101.

[158] Deepak Pathak, Philipp Krähenbühl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. Context encoders: Feature learning by inpainting. *CoRR*, abs/1604.07379, 2016. URL http://arxiv.org/abs/1604.07379.

[159] Andrew L. Maas. Rectifier nonlinearities improve neural network acoustic models. 2013.

[160] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.

[161] Alexey Dosovitskiy, Jost Tobias Springenberg, Maxim Tatarchenko, and Thomas Brox. Learning to generate chairs, tables and cars with convolutional networks, 2014.

[162] Jacques Delabrouille, J. F. Cardoso, and G. Patanchon. Multi-detector multi-component spectral matching and applications for CMB data analysis. *Mon. Not. Roy. Astron. Soc.*, 346:1089, 2003. doi: 10.1111/j.1365-2966.2003.07069.x.

[163] Y. Akrami et al. Planck 2018 results. VII. Isotropy and Statistics of the CMB. 2019.

[164] G. Puglisi, V. Galluzzi, L. Bonavera, J. Gonzalez-Nuevo, A. Lapi, M. Massardi, F. Perrotta, C. Baccigalupi, A. Celotti, and L. Danese. Forecasting the Contribution of Polarized Extragalactic Radio Sources in CMB Observations. *Astrophys. J.*, 858(2):85, 2018. doi: 10.3847/1538-4357/aab3c7.