Universiteit
Leiden
The Netherlands

Playing dice with the universe: Bayesian statistical analyses of cosmological models and new observables

Cañas Herrera, G.

Citation

Cañas Herrera, G. (2022, October 19). *Playing dice with the universe: Bayesian statistical analyses of cosmological models and new observables*. *Casimir PhD Series*. Retrieved from https://hdl.handle.net/1887/3483658

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3483658

Note: To cite this publication please use the final published version (if applicable).

Playing dice with the Universe: Bayesian Statistical analyses of cosmological models and new observables

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 19 oktober 2022
klokke 13:45 uur

door

Guadalupe Cañas Herrera

geboren te Griñón, Madrid, Spanje
in 1993

Promotor:	Prof. dr. A. Achucarro
Co-promotor:	Dr. A. Silvestri
Promotiecommissie:	Dr. M. Martinelli (Osservatorio Astronomico di Roma)
	Dr. A. Pourtsidou (Institute for Astronomy, Edinburgh, UK)
	Prof. dr. E. R. Eliel
	Prof. dr. H. Hoekstra
	Prof. dr. K. E. Schalm

Casimir PhD series, Delft-Leiden 2022-21
ISBN 978-90-8593-532-2
An electronic version of this thesis can be found at `https://openaccess.leidenuniv.nl`

This work is supported by the Delta Institute for Theoretical Physics (D-ITP consortium), a program of the Netherlands Organization for Scientific Research (NWO, OWC), and by the Leiden De Sitter programme.

The cover shows an illustration of the bright night sky. The painting was made by my father, *Miguel Cañas Muñoz*, who has a genuinely interest in astronomy and shared this passion with me when I was a little girl. I thank *Jaime Cañas Muñoz* for realizing the graphical design of the cover using the painting.

A mamá, papá y Juanchito

Glossary

Λ Cosmological constant.
ΛCDM Standard Cosmological model.
(C)DM (Cold) Dark Matter.

AIC Akaike Information Criterion.
AP Alcock-Pacyznski effect.

BAO Baryonic Acoustic Oscillations.
BBN Big Bang Nucleosynthesis.
BH Black Hole(s).
BIC Bayesian Information Criterion.
BICEP Background Imaging of Cosmic Extragalactic Polarization.
BM Baryonic Matter.

CAMB Code for Anisotropies in the Microwave Background.
CfA Center for Astrophysics Redshift Survey.
CLASS Cosmic Linear Anisotropy Solving System.
CMB Cosmic Microwave Background.
CPL Chevallier-Polarski-Linder parametrization.

DE Dark Energy.
DES Dark Energy Survey.

EC Euclid Consortium.
EFT Effective Field Theory.
EM Electromagnetic.
ESA European Space Agency.
ET Einstein Telescope.

FLRW Friedmann-Lemaitre-Robertson-Walker.

GC Galaxy Clustering.
GPs Gaussian Processes.
GR General Relativity.
GR Gravitational Lensing.
GW Gravitational-Waves.

IA Intrinsic Alignment.

IST:L Inter-Science Taskforce Likelihood.
IST:NL InterScience Taskforce Non-Linear.

LIGO Laser Interferometer Gravitational-Wave Observatory.
LISA Laser Interferometer Space Antenna.
LSS Large Scale Structure.

MCMC Monte Carlo Markov Chain.
ML Machine Learning.

NASA National Aeronautics and Space Administration.

SDSS Sloan Digital Sky Survey.
SHO Simple Harmonic Oscillator.
SKA Square Kilometre Array.
SNIa SuperNovae of type Ia.

WL Weak Lensing.
WMAP Wilkinson Microwave Anisotropy Probe.

XC Cross (X) - Correlations.

Contents

Dedication **iii**

Glossary **iv**

1 Introduction **1**
1.1 Evolution of a homogeneous and isotropic universe 3
1.2 A problem of initial conditions and inflation 8
1.3 Canonical single-field slow-roll inflation 10
1.3.1 Primordial perturbations 11
1.3.2 Primordial power spectrum 14
1.3.3 Beyond canonical single-field slow-roll inflation 15
1.4 The Large Scale Structure of the universe 16
1.4.1 The non-linear regime 19
1.5 Observations 20
1.5.1 Supernovae Type Ia (SNIa) 20
1.5.2 CMB angular power spectrum 21
1.5.3 Galaxy Clustering 22
1.5.4 Weak Lensing 24
1.6 The Standard Cosmological model 26
1.7 Data Analysis in Cosmology 27
1.7.1 Bayesian Statistics 28
1.7.2 Parameter Inference 30
1.7.3 Model Comparison 35
1.8 About this thesis 36

I Constraints on inflationary models **39**

2 Bayesian reconstruction of the inflaton's speed of sound using CMB data **41**
2.1 Introduction 41
2.2 Theoretical model 43
2.3 Methodology 45
2.3.1 Reconstruction model for the reduction in the speed of sound 45
2.3.2 Parameters and Priors 48

2.3.3 Data sets and sampler 50
2.4 Results 52
2.4.1 Consistency checks 52
2.4.2 Reconstruction of the inflaton's speed of sound profile $u(\tau)$ . . 54
2.5 Conclusions 56
2.6 Appendix: detailed results up to 4 nodes of $u(\tau)$ in log-log parameterization 58
2.6.1 One dip (*denoted by "A"*) 58
2.6.2 Two dips (*denoted by "B"*) 64
2.6.3 Three dips (*denoted by "C"*) 64
2.6.4 Four dips (*denoted by "D"*) 64
2.6.5 Reductions with substructure (*denoted by "S"*) 65

3 Constraints on α-attractor models 67
3.1 Introduction 67
3.2 Constraints from current CMB and LSS data 69
3.3 Forecast for future CMB-S4 observations 72
3.4 Conclusions 75
3.5 Appendix: from a two sigma bound to an upper limit 77

II Constraints on Gravitational Waves Physics 79

4 Cross-correlations of the anisotropies of the Astrophysical Gravitational-Wave Background with Galaxy Clustering 81
4.1 Introduction 81
4.2 Gravitational-Wave anisotropies 83
4.3 Cross-correlation with galaxy clustering 87
4.4 Information content 88
4.4.1 Model set-up 88
4.4.2 Behaviour of the cross-correlation 91
4.4.3 Constraining $\mathcal{K}(r)$ 92
4.5 Conclusions 95

5 Reconstruction of the propagation and origin of gravitational waves 99
5.1 Introduction 99
5.2 Gravitational wave propagation 102
5.3 Angular power-spectra 103
5.3.1 GW sources 103
5.3.2 Galaxies 105
5.4 Reconstructing GW physics 106
5.4.1 Mock data and fiducial model 106
5.4.2 Reconstruction Pipeline 108
5.5 Results 110
5.5.1 Reconstructions 110

5.5.2 Signal-to-Noise scaling . . . 112
5.6 Discussion and Conclusions . . . 115

III Data Science for the Euclid mission 117

6 Cosmological Likelihood for Observables in *Euclid* 119
6.1 The *Euclid* Mission . . . 120
6.1.1 Telescope, goals and the Euclid Consortium . . . 120
6.1.2 The role of IST:L and IST:Non-Linear . . . 122
6.2 Photometric and Spectroscopic surveys . . . 123
6.3 Theoretical predictions for *Euclid* primary observables . . . 125
6.3.1 Galaxy Clustering: Spectroscopic (GCsp) . . . 127
6.3.2 Weak Lensing (WL) . . . 129
6.3.3 Photometric Galaxy Clustering (GCph) . . . 132
6.3.4 Weak Lensing - Photometric Galaxy Clustering Cross-Correlation (XC) . . . 134
6.4 Benchmark data and covariance matrices . . . 135
6.4.1 Weak Lensing (WL) . . . 135
6.4.2 Photometric Galaxy Clustering (GCph) . . . 137
6.4.3 Weak Lensing - Photometric Galaxy Clustering Cross-Correlation (XC) . . . 137
6.4.4 Spectroscopic Galaxy Clustering (GCsp) . . . 138
6.5 Standard likelihood analysis . . . 139
6.6 `CLOE`: specifications and structure . . . 140
6.7 Validation of the benchmark data using `CLOE` . . . 143
6.8 The future of `CLOE` . . . 147
6.9 Appendix: fiducial and best fit values for the cosmological and nuisance parameters . . . 150

References 153

Summary 177

Samenvatting 181

List of Publications 185

Curriculum Vitae 187

Acknowledgments 189

Chapter 1

Introduction

Looking up at the night sky. A very small body movement that costs an insignificantly small number of calories. However, the consequences of bending the neck up 45 degrees are extraordinary. First, it triggers our innate human curiosity about the world we live in. Second, this curiosity activates our critical thinking by our wish to learn something new. All through history, physicists have tried to explain how nature works using this method of critical thinking, summarized in the scientific method. In the 20th century, a tiny neck movement translated into a series of remarkable discoveries in the field of Theoretical Physics, culminating in the work of A. Einstein about the Theory of General Relativity (Einstein, 1916). This milestone, together with discoveries in the atomic world, brought about the era of *Modern Physics*, establishing a revolution in our understanding of the universe.

During that century, humanity discovered that the universe is expanding (encoded in the Hubble parameter) (Hubble, 1929). Years later, the *Cosmic Microwave Background* radiation was predicted and detected (Penzias & Wilson, 1965). Furthermore, there appeared strong evidence that galaxies rotate faster in their outer regions pointing out the existence of an unknown type of matter called *Dark Matter* (Rubin, 1983), and ultimately, it was found out that the universe is currently undergoing an accelerated expansion thanks to the study of Supernovae type Ia observations (Schmidt et al., 1998).

These discoveries started the so-called *Modern Cosmology* era. Cosmologists aim to narrate the biography of the universe to understand its origin, structure and evolution, its properties, and, eventually, its final fate. Cosmological observations have increased our understanding of the universe's evolution from the time of the creation of the first nuclei up to the present. Yet, at earlier (and future) times the lack of cosmological information makes speculation dominate.

The ultimate goal of cosmologists is to find a cosmological model able to explain the current observational data, and so far, it seems that they have accomplished it. The *Standard Cosmological model*, also known as the ΛCDM model, establishes that our universe is mainly composed of two well-measured yet unknown components: a type of matter that is known to only interact through gravitation, Cold Dark Matter (CDM), and a substance responsible of the current accelerated expansion of the universe that can be modelled by a cosmological constant (Λ). The ordinary mat-

ter from which everything we observe is made of commonly called Baryonic Matter[1] (BM), only accounts for less than 5% of the total composition of the universe. Cosmologists have been able to test the robustness of the ΛCDM model mostly thanks to the high-accuracy measurements of the temperature anisotropies of the Cosmic Microwave Background (CMB) obtained by successful experiments such as the NASA WMAP mission (WMAP, 2003) and the ESA Planck mission (Planck Collaboration et al., 2014a).

The Standard Cosmological model relies on the assumption that the underlying distribution of the primordial density perturbations that seeded the early universe was Gaussian. These density fluctuations in the early universe are the origin of the current Large Scale Structure (LSS) of our universe that is observed today. The simplest model of inflation (an exponential expansion phase in the primordial universe) is compatible with this assumption, apart from explaining why the cosmological observations indicate that the universe is causally connected and flat.

Still, the ΛCDM model, though successful, fails to answer hot-burning questions in the field such as "what is the nature of Λ?" or "why is there a tension between measurements of the Hubble parameter using different observational probes?". For this reason, theoretical cosmologists focus on developing modifications of the Standard Cosmological model, and test them against astrophysical data to check whether alternative scenarios can provide a better explanation of the observations. Furthermore, cosmologists also work on constructing theoretical predictions of new observables that can be used to test a given model beyond ΛCDM. The key discipline used so far to compare theory with cosmological data is *Statistics*. In the last 20 years, not only the amount of cosmological and astrophysical data has increased significantly, but also our statistical analysis tools have evolved dramatically (Trotta, 2017).

In the coming decades, a substantial change in the type of cosmological observations used for statistical analysis will take place. Although the era of precision Cosmology based on the observations of the CMB has extensively advanced our understanding of the universe, it is close to an end[2]. In the next decade, the most precise constraints on the parameters of the ΛCDM model (or extensions of this model) will come from the information embedded in the large scale structure (LSS) of the universe, with several cosmological experiments and missions planned. The ESA mission *Euclid* (Laureijs et al., 2011), whose launch is expected in 2023, will precisely map the universe's structure by studying two observables: the Weak Lensing effect (WL), which studies the apparent change of the shapes of galaxies) and Galaxy Clustering (GC), which focuses on how well the distribution of galaxies in the universe can trace the underlying matter distribution.

Even though the goal of the *Euclid* mission is to shine some "light" on the origin and nature of Λ and Dark Matter, its data can also be used to explore the origin of

[1]Cosmologists usually misuse this term, as when they talk about "baryonic" matter, they also refer to leptons.

[2]The CMB still offers an incredible source of information encoded in other observables such as spectral distortions or the polarization signal. There are still efforts in proposing, designing, pursuing and realizing new experiments based on observations of the CMB. See for instance (Suzuki et al., 2018).

the primordial density fluctuations that seeded the very early universe and evolved up to the structure visible today. Therefore, the next generation of LSS surveys will have the power to indirectly constrain different inflationary models by studying the underlying spatial distribution of the cosmic web.

Furthermore, LSS surveys, together with observations of Gravitational-Wave (GW) events (Black Hole or Neutron Star mergers), lay the foundation for multimessenger Cosmology: cross-correlating the information contained in GW merger events and the LSS information of our universe opens a new window for research in Cosmology, in which the universe is observed not only through the electromagnetic spectrum but also through gravitational signals and the combination between these two.

In this self-contained chapter, we review the most important concepts of the Standard Cosmological Model following the reviews by (Baumann, 2009), (Carroll & Ostlie, 2014), (Dodelson, 2003) and (Mukhanov, 2005).

1.1 Evolution of a homogeneous and isotropic universe

Cosmological observations at sufficiently large scales (> 100 Mpc) support the idea that the universe is *isotropic* on large scales. If we also assume that we are not a special observer in the universe, we infer *homogeneity* of space. These two characteristics are the foundations of the *Cosmological Principle*. In the language of General Relativity, such a universe is described by the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric (Friedmann, 1924; Lemaître, 1931; Robertson, 1935; Walker, 1937):

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -dt^2 + a^2(t)\gamma_{ij}dx^i dx^j, \tag{1.1}$$

where:

- μ, ν, i, j take the values $\{\mu, \nu\} = \{0, 1, 2, 3\}$ and $\{i, j\} = \{1, 2, 3\}$.
- $a(t)$ is the scale factor, which accounts for the relative size of space-like hypersurfaces at a given time.
- x is the comoving coordinate, which defines a fixed set of points on a coordinate grid that grows with the expansion of the universe given by scale factor $a(t)$.
- t is the cosmic time, which is the time measured by comoving observers (observers who move with the space expansion determined by the rate of change of the scale factor a).
- γ_{ij} is the tensor corresponding to the spatial part of the metric, which in spherical coordinates (r, θ, ϕ) and together with the differentials $dx^i dx^j$ takes the form:

  $$\gamma_{ij}dx^i dx^j = \frac{dr^2}{1 - \kappa r^2} + r^2 d\phi^2 + r^2 \sin\theta^2 d\theta^2, \tag{1.2}$$

  where κ is the intrinsic curvature of 3-surfaces, and represents a *flat* ($\kappa = 0$), *positively curved* ($\kappa = +1$), or *negatively curved* ($\kappa = -1$) spatial slices.

For convenience, we define the *conformal time* τ:

$$d\tau = \frac{dt}{a} \rightarrow ds^2 = a^2(\tau)\left[-d\tau^2 + \gamma_{ij}dx^i dx^j\right]. \tag{1.3}$$

Light travels along null-geodesics ($ds^2 = 0$), and in the transformed FLRW metric (1.3), its propagation is the same as in Minkowski space when $\kappa = 0$. The scale factor $a(t)$ encodes the time evolution of the universe. Usually, it is normalised so that today t_0 it takes the value $a(t_0) = 1$. Due to the expansion of the universe, photons travelling from their place of emission towards us are redshifted. This redshift z is due to the stretching of electromagnetic waves along their path to us and it is quantified as

$$z + 1 = \frac{1}{a(t)}. \tag{1.4}$$

The FLRW metric is the solution to *Einstein Equations*:

$$G_{\mu\nu} = \frac{8\pi G}{c^2}T_{\mu\nu}, \tag{1.5}$$

where $G_{\mu\nu}$ is the Einstein tensor and $T_{\mu\nu}$ is the energy-momentum tensor of the matter components. If we choose to describe the matter present in the universe as a perfect fluid in the rest frame defined by its energy density ρ and pressure p, the tensor $T_{\mu\nu}$ takes the form:

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu}, \tag{1.6}$$

where $g_{\mu\nu}$ is the metric of equation (1.1) and u_μ is the four velocity of a perfect fluid. Solving equation (1.5) using natural units $c = \hbar = 1$, and defining the reduced Planck mass as $M_P = \sqrt{\frac{1}{8\pi G}}$, we obtain the well-known First and Second[3] *Friedmann* equations respectively (Baumann, 2009),

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{\rho}{3M_P^2} - \frac{\kappa}{a^2}, \tag{1.7}$$

$$\dot{H} + H^2 \equiv \frac{\ddot{a}}{a} = -\frac{1}{6M_P^2}(\rho + 3p). \tag{1.8}$$

where the derivative with respect to the cosmic time t is denoted with the dot ($\dot{} = d/dt$) and H is the *Hubble* parameter, a crucial function in Cosmology defined as:

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \tag{1.9}$$

Combining both equations (1.7) and (1.8) (or alternatively, calculating directly $\nabla_\mu T^{\mu\nu} = 0$) we obtain the continuity equation

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p). \tag{1.10}$$

[3]The Second Friedmann equation is usually called the acceleration equation.

The equation of state of a perfect fluid is often given in terms of the dimensionless state parameter, w, as $p = w\rho$. Solving equation (1.10) for each species i as a function of the scale factor $a(t)$,

$$\rho_i = \rho_{i,0}\,(a)^{-3(1+w_i)} \rightarrow \begin{cases} \rho_m \propto a^{-3} & (w=0) \quad \text{matter (Cold Dark Matter and baryons)} \\ \rho_r \propto a^{-4} & \left(w=\frac{1}{3}\right) \quad \text{radiation} \end{cases}, \tag{1.11}$$

where the sub-index 0 indicates a reference value, which we have taken at the present time. There is an extra component that must be taken into account to explain current cosmological observations: the cosmological constant Λ. This constant was first introduced in the Einstein Equations (1.5) as a term $\Lambda g_{\mu\nu}$ to obtain static solutions of the Friedmann equations, and was later reintroduced when the High-Z Supernova Search Team discovered that the expansion of the universe is accelerating thanks to the study of supernovae type Ia observations (Riess et al., 1998). Given equations (1.8) and (1.11), we observe that a late time acceleration of the expansion of the universe occurs when the universe starts to be dominated by an energy component with an equation of state $w \leq -1/3$; that is, a substance that violates the strong energy condition that is given by $w \geq -1/3$.

Why is this acceleration taking place? In quantum field theory there exists the prediction of an accelerated expansion of the universe due to the non-zero quantum mechanical vacuum energy. This vacuum energy behaves exactly like a cosmological constant with an equation of state $w_\Lambda = -1$, which results in a energy density

$$\rho_\Lambda(t) = \rho_\Lambda(t_0). \tag{1.12}$$

This constant energy density is described by the energy-momentum tensor

$$T^\Lambda_{\mu\nu} = -M_P^2 \Lambda g_{\mu\nu}. \tag{1.13}$$

Unfortunately, the theoretical predictions of the estimated value of the vacuum energy density do not agree with the cosmological observations. Still, current observations can be explained by a cosmological constant Λ behaving as a fluid-like energy component of the universe responsible of the accelerated late-time expansion. This approach receives the name of dynamical Dark Energy (DE), which characterizes the fluid by an equation of state $w_{\rm DE} = -1$. To infer information from the observational data, a phenomenological parametrization is often used, denominated CPL (Chevallier & Polarski, 2001)(Linder, 2003) defined as,

$$w_{\rm DE}(z) = w_0 + w_a(1 - a(z)). \tag{1.14}$$

We aim to provide DE or modified gravity (MG) models that are able to explain the current observational data to obtain some insight into the nature and origin of the current accelerated expansion. To find a balance between observations and plausible theories, it is interesting to study the Large Scale Structure of the universe from the predictions of the ΛCDM model in terms of the phenomenological functions[4]

[4]We are using the subscript "MG" to differentiate the symbol μ from other quantities that will be introduced later in this chapter.

Σ_{MG} and μ_{MG}, which parametrize modifications of the perturbed Einstein's equations relating the matter density contrast to the lensing of light and the Newtonian potential (Frusciante & Perenon, 2020). The function $\Sigma_{\rm MG}$ will be later used in chapter 6.

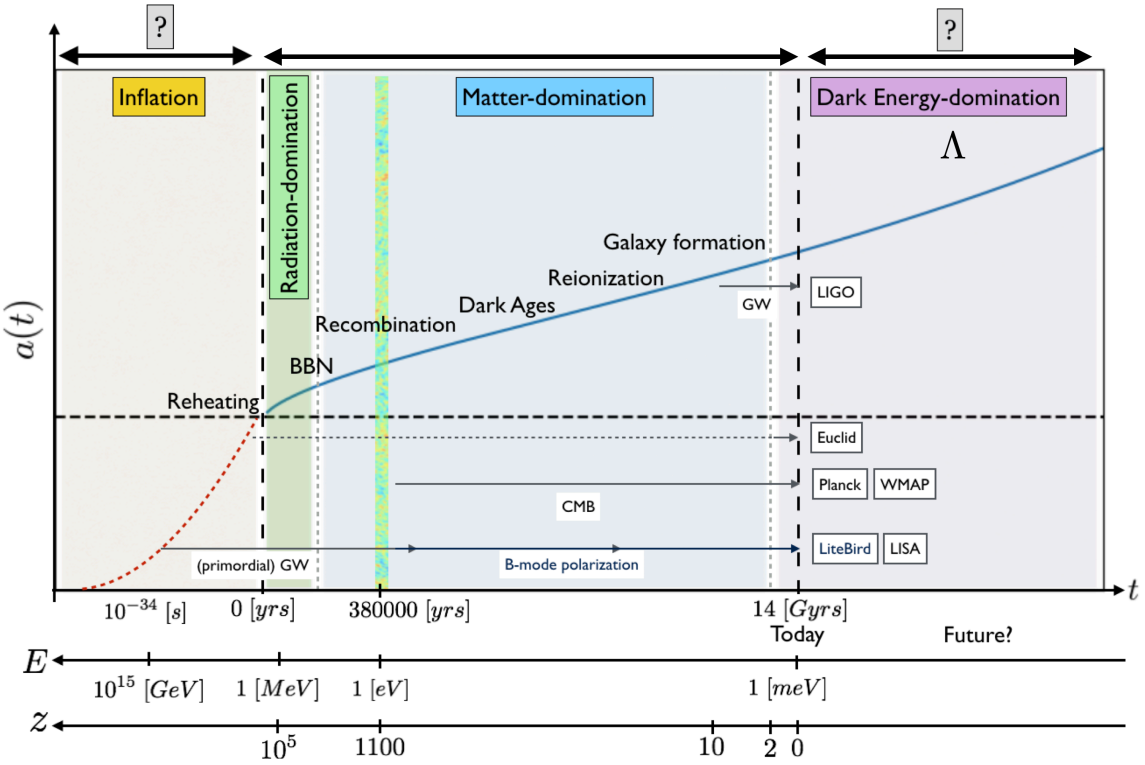

Figure 1.1: Evolution of the scale factor $a(t)$ as a function of the cosmic time t, redshift z and energy scale E. The main key events in the history of the universe are pointed out during the universe's timeline. We highlight speculative epochs of the universe: at earlier times, inflation and at later future times, Dark Energy-domination, where the current accelerated expansion of the universe is modelled in term of the cosmological constant Λ. Numerical integration of $a(t)$ has been done with (*Calculadora Cosmologica: code that solves Friedman Equations numerically given values for components of universe*, 2015). The time evolution of the scale factor during radiation-domination scales as $a(t) \propto t^{1/2}$ and during matter-domination as $a(t) \propto t^{2/3}$. Some cosmological space missions as well as experiments placed on Earth are included only for representative purposes. All acronyms can be found in the Glossary.

Given (1.11), we can determine the evolution of the scale factor, $a(t)$, as a function of time in the presence of these 3 different species. Defining the critical density $\rho_c \equiv 3M_p^2 H_0$ and the density parameter[5] $\Omega_{i,0} = \rho_i/\rho_c$ (H_0 is the Hubble parameter

[5]We are denoting the density parameters at present time with the sub-index 0. However, this sub-index is most of the times dropped, and we will follow this notation in the rest of the thesis.

evaluated today), we can re-write Friedmann equation (1.7) as,

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = H_0^2 \left[\Omega_{m,0}\left(\frac{1}{a}\right)^3 + \Omega_{r,0}\left(\frac{1}{a}\right)^4 - \frac{\kappa}{H_0^2}\left(\frac{1}{a}\right)^2 + \Omega_{\Lambda,0}\right]. \tag{1.15}$$

Given that observational data, in particular, very precise measurements of the Cosmic Microwave Background (CMB) radiation, support that the geometry of our universe is *flat* ($\kappa = 0$) (Planck Collaboration et al., 2014a), we can solve the differential equation of $a(t)$ as a function of time t ,

$$\dot{a} \sim a^{\frac{1}{2}(1+3w)} \rightarrow a(t) = \begin{cases} t^{2/(3+3w)} & \text{matter and radiation} \\ e^{Ht} & \text{Cosmological constant} \end{cases}. \tag{1.16}$$

The re-written Friedmann equation (1.15) shows how the three cosmological species evolve differently with respect to the scale factor, and how these species also scale in time according to equation (1.16). Knowing the values of the density parameters today (Ω_m, Ω_r and Ω_Λ), we can account for the expansion of the universe throughout its evolution. This difference in the scaling found in equation (1.11) shows how the universe progressed through different epochs, starting with a Radiation-domination era followed by Matter and, ultimately, by Dark Energy-domination (see Figure 1.1, where the evolution of the scale factor $a(t)$ as a function of time t is depicted).

Given Friedmann equations and equation (1.16), we can now explain the thermal history of the universe. At its very early stages, the universe behaved like a hot plasma made of relativistic particles forming a thermal batch. As a consequence of the universe's expansion, the universe's temperature cooled down, allowing the decoupling of various particles from the thermal bath. At a temperature $T \approx 10^{15}\ K$, it became more energetically favourable for quarks to exist in bound states such as protons and neutrons instead of being in a quark-gluon plasma. Later, it is believed that neutrinos decoupled from the cosmic thermal bath around $T \approx 10^{10}\ K$. Below that temperature, high-energy photons (gamma-rays) can no longer produce electron-positron pairs, and the populations of both electrons and positrons started to decrease steeply by particle-antiparticle annihilation.

When the universe was approximately 3 minutes old, it had reached the necessary low temperature to allow several light nuclei (Hydrogen, Helium and some Lithium) to form through a process called Big Bang Nucleosynthesis (BBN). This mechanism was responsible for the transition between a radiation-dominated universe and a matter-dominated universe ($z \approx 3600$). The predictions made in the context of BBN (such as the primordial abundances of those nuclei, which can be calculated by studying the corresponding cross-sections) are in agreement with the current experimental evidence (Cyburt, Fields, & Olive, 2003). The primordial abundance predictions are one of the two main successes of the Big Bang Theory.

At a temperature $T \approx 1$ eV (around $z \approx 1100$) the universe entered the so-called Recombination epoch. During this time, light nuclei began to bind with electrons to form neutral atoms. These bindings induced a drop in the number density of free electrons. Hence, the decoupling of photons from matter took place, as photons

could no longer scatter off of electrons, which is called *recombination*. These photons could finally travel freely through space, with an almost isotropic distribution. This primordial radiation is called the Cosmic Microwave Background (CMB). Its experimental detection by Penzias and Wilson (Penzias & Wilson, 1965) represents the second main success of the Big Bang Theory. More details about the modelling of the CMB observables are given in subsection 1.5.2.

Finally, after recombination, only the CMB radiation is emitted, introducing an epoch called *dark ages*. At around $z \approx 20$, the Large Scale Structure (LSS) of the universe started to form, where Dark Matter (DM) is thought to have a very important role. It is assumed that DM is composed of an "exotic" yet-unknown particle[6] that decoupled at earlier stages from the thermal bath and started to collapse gravitationally into halos. Thus, after recombination, stars began to form when Baryonic Matter (BM) collapsed to the centre of these pre-existing DM halos. Afterwards, due to gravitation, galaxies were formed. These objects emit enough light to re-ionize the medium by ripping electrons from the neutral atoms. At lower redshifts $z \approx 6$, clusters of galaxies began to form thanks to gravity, building the well-known web-like structure of the universe. At approximately $z \approx 0.55$, the contribution of DE is approximately equal to the contribution of matter, causing the universe to enter into the Dark Energy-domination era.

1.2 A problem of initial conditions and inflation

Despite cosmological observations having accounted for experimental evidence of the BBN as well as the CMB radiation, the detection of the CMB radiation also brought some of the biggest puzzles that Modern Cosmology had to face. First of all, it confirmed that our universe is currently dominated by Dark Energy and that we need close to 30% of Dark Matter to explain the observations of the angular power spectrum of the CMB temperature anisotropies. Furthermore, it showed the necessity of fine-tuning the initial conditions of the Big Bang. The fine-tuning of the initial conditions of the universe is usually shown in terms of the *horizon problem* and the *flatness problem*. To introduce these problems, we calculate the maximum comoving distance photons can travel in space-time from emission (t_e) to a detecting observer (t_d):

$$\chi_P = \int_{t_e}^{t_d} \frac{dt'}{a(t')} = \int_{a(t_d)}^{a(t_e)} \frac{1}{aH} d\ln a, \tag{1.17}$$

where the factor $1/aH$ is denominated the *comoving Hubble radius*[7]. The first problem is that not-causally connected patches in the CMB may seem to be in thermal equilibrium. Solving the integral in equation (1.17) for the different epochs in the universe, we observe that χ_P is an increasing function during Matter and Radiation-domination. If we calculate its size at the moment of the last scattering during the

[6]There are other alternative theories to DM that aim to explain the current behaviour observed in the observational data. See for instance (Verlinde, 2017)

[7]Cosmologists misuse this term often calling to the *comoving Hubble radius* with the word *horizon*.

CMB radiation emission, we discover that the CMB map is composed of at least $\sim 10^4$ causally-disconnected patches despite showing the same temperature up to 10^{-5} K.

On the other hand, the second problem exposes that, given today's observed values for the different components of the universe and the fact that our universe is *flat* ($\kappa = 0$), a *flatter* universe was required in the past due to the dependence of density with time. Rewriting equation (1.15) by absorbing the time dependencies into the density parameters Ω_i we see that

$$1 - \sum_{i\in\{m,\Lambda,r\}} \Omega_i(a) = \frac{-\kappa}{(aH)^2}. \tag{1.18}$$

According to the calculations, we need to fine-tune, at least, at the moment of BBN $\sum_{i\in\{m,\Lambda,r\}} \Omega_i(a_{\rm BBN}) - 1 \approx 10^{-16}$ to obtain a flat universe $\kappa = 0$ at present time.

An elegant solution to solve these two problems is a decreasing comoving Hubble radius in a very early stage of our universe. This accelerated expansion period in the early history of the universe is called inflation. Given Friedmann equations (1.7, 1.8):

$$\frac{d}{dt}\left(\frac{1}{aH}\right) < 0 \quad \longleftrightarrow \quad \frac{d^2a}{dt^2} > 0 \quad \longleftrightarrow \quad \rho + 3p < 0 \quad \longleftrightarrow \quad \omega < -\frac{1}{3}. \tag{1.19}$$

Re-writing the first term of equation (1.19), we obtain

$$\epsilon_1 \equiv -\frac{\dot{H}}{H^2} = -\frac{d\ln H}{dN} < 1, \tag{1.20}$$

where N is the *number of e-folds* (number of expansion times of the universe $a = \exp(N(t))$ so that $dN = Hdt$), ϵ_1 is usually called the *first slow-roll (kinematic) parameter*, and to ensure accelerated expansion $0 < \epsilon_1 < 1$. In the limit when $\epsilon_1 \to 0$, the Hubble parameter becomes constant and therefore, the space-time becomes de Sitter space. However, inflation needs to finish and last long enough to solve the *horizon problem* (approximately between 50 and 60 e-folds). This condition is studied by introducing a new parameter, ϵ_2, which relates the relative change of ϵ_1 during one e-fold as follows

$$\epsilon_2 \equiv -\frac{d\ln\epsilon}{dN} = \frac{\dot{\epsilon}}{H\epsilon}, \tag{1.21}$$

and is known as the *second slow-roll kinematic parameter*. The condition $|\epsilon_2| < 1$ ensures that the change of ϵ_1 is small and consequently, inflation can last. A shrinking Hubble radius (1.19) also provides a natural mechanism to explain the generation of the initial perturbations of the distribution of matter in the early universe that evolved up to the current Large Scale Structure visible today. To explain the mechanism, we use the Simple Harmonic Oscillator (SHO) as an example. Let's model the behaviour of small perturbations as damped SHOs due to the expanding space-time. Before inflation, all small-scale perturbations are inside the Hubble radius. When the comoving Hubble radius shrinks, the Hubble friction term starts to dominate and the scales become over-damped. If the scale is larger than the Hubble radius, they cannot longer move becoming *frozen*. After inflation, when the Hubble radius grows again, the friction term does not longer dominate the behaviour of the SHO, allowing the scales to *move* again. In the next section, we will introduce this mechanism using the simplest setting of single-field slow-roll inflation.

1.3 Canonical single-field slow-roll inflation

Inflation needed to stop at some point because we know, from cosmological observations, that the radiation and matter-domination epochs took place. For inflation to end, the space-time had to deviate from a perfect de Sitter space. Still, de Sitter space-times remains a good approximation for $\epsilon_1 \ll 1$. Together with $\epsilon_2 \ll 1$, they are the slow-roll conditions. We can discuss how microscopic physics can fulfil the slow-roll conditions by defining a new substance that violates the strong energy condition, similarly to the case of the cosmological constant. For simplicity, let us model this substance as a single scalar field ϕ that is homogeneous, so that $\phi(t, \mathbf{x}) = \phi(t)$. To describe the dynamics of the system, we introduce the second-order action for the field ϕ with a canonical kinetic term and a potential $V(\phi)$, minimally coupled to gravity, given by

$$S_2 = \int d^4x\sqrt{-g}\left[\frac{M_P^2}{2}R - \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right], \tag{1.22}$$

where R is the Ricci scalar curvature of the space-time and the first and second terms of the action describe the coupling of the *inflaton* ϕ to gravity. All this setting is called canonical single-field slow-roll inflation. Calculating $T_{\mu\nu}$ from the action (1.22) and comparing to the expression for a perfect fluid (1.6), we obtain the corresponding expressions for the density, ρ_ϕ, the pressure, p_ϕ, and the homogeneous field ϕ

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V, \qquad p_\phi = \frac{1}{2}\dot{\phi}^2 - V. \tag{1.23}$$

Introducing expressions (1.23) in (1.7) and (1.10), we can obtain the corresponding to the continuity equation as well as the Friedmann equations in terms of V and ϕ respectively:

$$\ddot{\phi} + 3H\dot{\phi} + \frac{dV}{d\phi} = 0, \qquad H^2 = \frac{1}{3M_P^2}\left(\frac{1}{2}\dot{\phi}^2 + V\right), \qquad \dot{H} = \frac{1}{M_P^2}\left(-\frac{1}{2}\dot{\phi}^2\right). \tag{1.24}$$

The equations in (1.24) can be used to express the slow-roll parameters ϵ_1 and ϵ_2 in terms of the potential V and the field ϕ. To fulfil the slow-roll conditions, the kinetic energy of the inflaton field has to be negligible with respect to the potential energy ($\dot{\phi} \ll V$), which makes the Hubble parameter to be nearly constant $\dot{H}/H \ll 1$, allowing an almost-exponential expansion with $a(t) \propto \exp Ht$. Also, the acceleration of the field has to be very small ($\ddot{\phi} \ll H\dot{\phi}$), ensuring inflation to last long enough. Moreover, when the slow-roll conditions are fulfilled, we can quickly see from equations (1.23) that the equation of state of this substance should be $w_\phi \approx -1$, as in the cosmological constant case. The slow-roll parameters can be also expressed in terms of the potential V:

$$\epsilon_V \equiv \frac{M_P^2}{2}\left(\frac{\partial_\phi V}{V}\right) \approx \epsilon_1 \qquad\qquad |\eta_V| \equiv \frac{M_P^2}{2}\frac{\partial_{\phi\phi}V}{V} \approx 2\epsilon_1 - \frac{1}{2}\epsilon_2, \tag{1.25}$$

and the slow-roll conditions imply that $\epsilon_V, |\eta_V| \ll 1$. The potential slow-roll parameters are commonly used in the literature and can be also related to the kinetic slow-roll parameters ϵ_1 and ϵ_2.

1.3.1 Primordial perturbations

Apart from solving ad-hoc the fine-tuning of initial conditions, inflation explains the origin of the primordial seeds that evolved to the current structure observed in the universe. The origin of these primordial density perturbations lies in the quantum fluctuations of the inflaton field with respect to the homogeneous background and, also, in how the evolution of these perturbations k behaves with respect to the comoving Hubble radius aH^{-1}. The overview is as follows (see Figure 1.2). When the universe underwent inflation, quantum fluctuations were stretched to a very large scale. When the physical wavelength[8] of this mode is larger than the scale of the Hubble radius, this mode is called *super-Hubble* ($k^{-1} \gg (aH)^{-1}$). These perturbations that exited the Hubble radius during this extreme expansion remained frozen. At later stages of the universe, when the comoving Hubble radius increased again during radiation and matter-domination (see equations (1.17) and (1.16)), these frozen modes re-entered the Hubble radius again, becoming *sub-Hubble* ($k^{-1} \ll (aH)^{-1}$) and forming the initial density fluctuations of the universe.

Following the discussion presented in (Baumann, 2009; Achúcarro & Welling, 2015), where the original references can be also found, we can illustrate this behaviour mathematically using the canonical single-field slow-roll inflation scenario as an example. For that, we use a perturbed expression of the FLRW metric in (1.1), $g_{\mu\nu}(t, \mathbf{x}) \equiv \bar{g}_{\mu\nu} + \delta g_{\mu\nu}(t, \mathbf{x})$, and of the field $\phi(t, \mathbf{x}) \equiv \phi(t) + \delta\phi(t, \mathbf{x})$. To parametrize $\delta g_{\mu\nu}$, a particular useful choice of coordinates is the *comoving gauge*,

$$\delta\phi = 0, \qquad \delta g_{ij}(t, \mathbf{x}) = a^2 \left[(1 - 2\mathcal{R}(t, \mathbf{x}))\delta_{ij} + h_{ij}(t, \mathbf{x})\right], \qquad \partial_i h_{ij} = 0, \tag{1.26}$$

where $\mathcal{R}$ denotes the scalar metric fluctuation (comoving curvature perturbation) and h_{ij} is the tensor part (gravitational waves) of the perturbed metric. In this gauge, the inflation field ϕ is unperturbed (the density of the fluid is constant) and the scalar degrees of freedom are parametrized by $\mathcal{R}$. Substituting equation (1.26) into the expression of the action for a single-field (1.22), and expanding in powers of $\mathcal{R}$, we find the quadratic action for the scalar metric fluctuation,

$$S_2 = \frac{1}{2} \int d^4x a^3 \frac{\dot{\phi}^2}{H^2} \left[\dot{\mathcal{R}}^2 - \frac{(\partial_i \mathcal{R})^2}{a^2}\right]. \tag{1.27}$$

Our goal is to derive the equation of motion for $\mathcal{R}$, show that it has a Simple Harmonic Oscillator (SHO) form, and promote the classical field $\mathcal{R}$ into a quantum operator. In order to solve the equations of motion derived from (1.27), the *Mukhanov-Sasaki variable* v is introduced (Mukhanov, 1988; Sasaki, 1983). It is defined as

$$v \equiv z\mathcal{R} \qquad\qquad z^2 \equiv 2a^2\epsilon. \tag{1.28}$$

Using the conformal time τ, defined in equation (1.3), and the Mukhanov-Sasaki variable v, we obtain the full quadratic action in perturbations with canonical kinetic

[8] A physical wavelength of a mode is proportional to k^{-1}. Larger physical wavelengths imply smaller values of k.

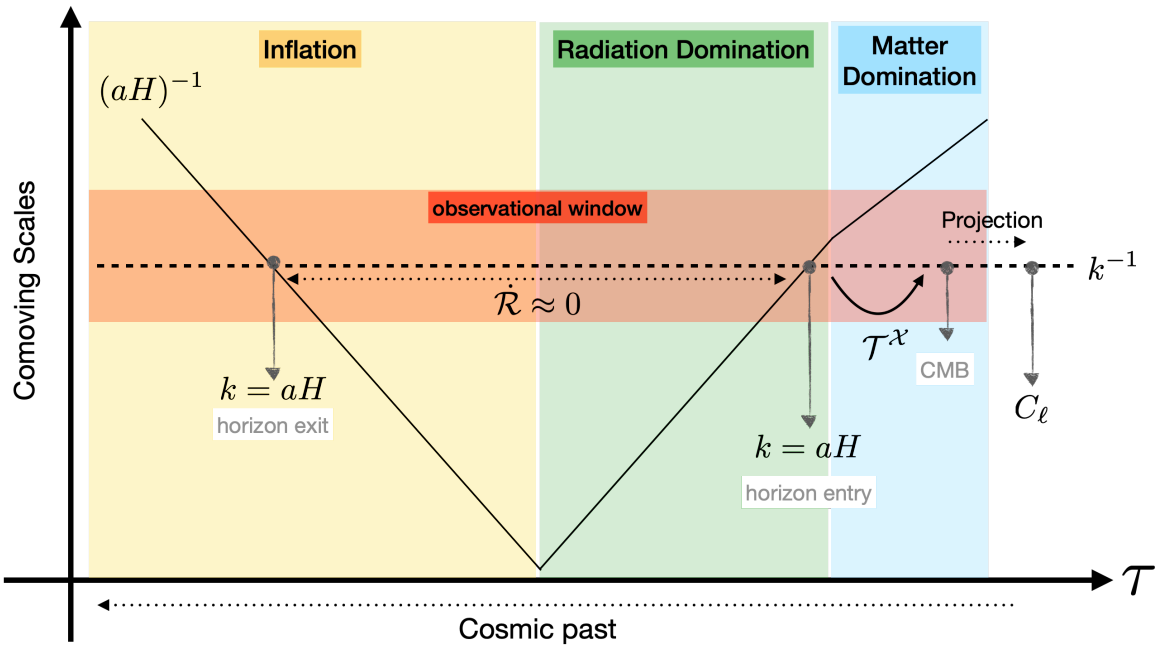

Figure 1.2: Graphical representation of the history of a particular comoving scale k^{-1} (dashed black line). This comoving scale is a sub-Hubble scale ($k^{-1} \ll (aH)^{-1}$) until it leaves the comoving Hubble radius $1/aH$ due to inflation, where it remains constant and super-Hubble ($k^{-1} \gg (aH)^{-1}$). It re-enters during the radiation-domination epoch, becoming sub-Hubble again ($k^{-1} \ll (aH)^{-1}$). The evolution of k^{-1} after re-entering is modelled according to the dynamics of the radiation and matter-domination epochs, and is encoded in the transfer function $\mathcal{T}^{\mathcal{X}}$, which is a mathematical tool that explains the physical phenomena in the period between inflation, radiation domination, matter domination, and the emission of the Cosmic Microwave Background. We can study the distribution of density perturbations in the angular power spectra of anisotropies of the Cosmic Microwave Background (CMB) C_ℓ. Adapted from (Baumann, 2009).

terms:

$$S_2 = \frac{1}{2}\int d\tau d^3x \left[(v')^2 - (\partial_i v)^2 + \frac{z''}{z}v^2\right], \tag{1.29}$$

where the derivative with respect to the conformal time τ is denoted as $df/d\tau = f'$. This action demonstrates that the *Mukhanov-Sasaki variable* v is the one we should use as the canonical quantization variable. In Fourier space, we obtain the equation of motion for v:

$$v_{\mathbf{k}}'' + \left(k^2 - \frac{z''}{z}\right) v_{\mathbf{k}} = 0. \tag{1.30}$$

where $v_{\mathbf{k}}$ is the amplitude of the Fourier mode and depends only on the magnitude of k. Equation (1.30) resembles the equation of a Simple Harmonic Oscillator (SHO) with a mass depending on the conformal time $m^2(\tau) \approx -z''/z$. Therefore, the quantization of the field v is obtained analogously to the treatment of the quantum SHO. We promote v and its corresponding conjugate momentum v' to a quantum operator $\hat{v}$,

$$\hat{v}_{\mathbf{k}} = v_{\mathbf{k}}\hat{a}_{\mathbf{k}} + v^*_{\mathbf{k}}\hat{a}^\dagger_{\mathbf{k}}, \tag{1.31}$$

where $\hat{a}_{\mathbf{k}}$ and $\hat{a}^\dagger_{\mathbf{k}}$ are the creation and annihilation operators, which satisfy the canonical commutation relation

$$[\hat{a}_{\mathbf{k}}, \hat{a}^\dagger_{\mathbf{k}'}] = (2\pi)^3\delta(\mathbf{k} - \mathbf{k}'). \tag{1.32}$$

Solving equation (1.30) is complicated as z depends on the background dynamics. However, some insight can be obtained if we study certain limits and constraints. However, an important constraint can be imposed at the earliest stage of our universe when $\tau \to -\infty$, which implies that all comoving scales were within the Hubble horizon (as $\tau \propto -(aH)^{-1}$, given equation (1.17)). In this limit, we choose the vacuum state for the fluctuation,

$$\hat{a}_{\mathbf{k}} |0\rangle = 0, \tag{1.33}$$

meaning that every mode k is assumed to have started its evolution in the vacuum state, such that there was no particle production. According to this, modes with high k do no feel the curvature of space-time, and the expectation value of the Hamiltonian in the minimal energy state corresponds to the one in flat (Minkowski) space. This boundary condition is known as the *Bunch-Davies* vacuum.

In the *quasi-De Sitter* regime, where H and ϵ are approximately constant, the full solution to equation (1.30) is,

$$v_{\mathbf{k}}(\tau) = \frac{e^{-ik\tau}}{\sqrt{2k}}\left(1 - \frac{i}{k\tau}\right). \tag{1.34}$$

Now, we can trace the history of a mode k mathematically by studying the asymptotic limits of the solution (1.34) in the quasi-De Sitter case ($H, \epsilon \approx$ constant):

- $|k\tau| \gg 1 \leftrightarrow k \gg aH \leftrightarrow k^{-1} \ll (aH)^{-1}$: the scale k is sub-Hubble, and the solution of $v_{\mathbf{k}}(\tau)$ is dominated by the oscillating exponential part $e^{-ik\tau}$. This is just the result of the above-mentioned Bunch-Davis vacuum.

- $|k\tau| \ll 1 \leftrightarrow k \ll aH \leftrightarrow k^{-1} \gg (aH)^{-1}$: the scale k is super-Hubble, and the dominant contribution to the solution is the divergent factor $1/\tau$. This means that, in this range, equation (1.34) is,

$$\lim_{\tau \to 0-} v_{\mathbf{k}}(\tau) = \frac{1}{\sqrt{2}k^{3/2}\tau} \propto \frac{aH}{k^{3/2}} \tag{1.35}$$

  where we have used that $\tau \sim (aH)^{-1}$. Recovering the definition of v from equation (1.28), we find that the curvature perturbation $\mathcal{R}$ is constant and

shows that super-Hubble scales remain frozen when they leave the horizon:

$$\mathcal{R} = \frac{v}{z} = \frac{v}{a\sqrt{2\epsilon}} \propto \frac{H}{k^{2/3}\sqrt{\epsilon}} \rightarrow \lim_{\tau \to 0^-} \dot{\mathcal{R}} = 0, \tag{1.36}$$

where this expression for $\mathcal{R}$ should be evaluated for each mode at the Hubble crossing radius $k = aH$.

In conclusion, during inflation, perturbation modes exit the horizon, becoming super-Hubble scales. Quantum fluctuations in the inflaton field lead to some parts of the universe being stretched for a longer period of time than others, which means that some sub-Hubble scales exit the Hubble radius due to the accelerated expansion of the universe later than other scales. All scales remained frozen until they re-enter at later stages of the universe when the comoving Hubble radius increases again. The before mentioned local time delay of the Hubble exit of different scales translates into density fluctuations, which follow their evolution during radiation-matter domination. These local differences in density become the seeds of all the current structures of the universe.

1.3.2 Primordial power spectrum

We can study the statistical probability distribution of the primordial fluctuations by using the power spectrum, $P_{\mathcal{R}}(k)$, which is the Fourier transform of two-point correlation function $\langle \mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\rangle$:

$$\langle \mathcal{R}(\mathbf{k}_1)\mathcal{R}(\mathbf{k}_2)\rangle \equiv (2\pi)^3 \delta^3(\mathbf{k}_1 - \mathbf{k}_2) P_{\mathcal{R}}(k), \tag{1.37}$$

where δ^3 is the three dimensional Dirac delta and $P_{\mathcal{R}}(k)$ is only dependent on the magnitude of the momentum scale $k = |\mathbf{k}|$, and not on the direction, due to isotropy. The primordial power spectrum is usually redefined as a dimensionless quantity, $\mathcal{P}_{\mathcal{R}}(k)$, as

$$\mathcal{P}_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2} P_{\mathcal{R}}(k). \tag{1.38}$$

Cosmological observations support statistically an almost scale invariant power spectrum, and this is why a phenomenological parametrization of $\mathcal{P}_{\mathcal{R}}(k)$ is broadly used. Using a Taylor-expansion up to the first order, the expression is

$$\log \frac{\mathcal{P}_{\mathcal{R}}(k)}{\mathcal{P}_{\mathcal{R}}(k_*)} \approx (n_s - 1)\left[\log k - \log k_*\right] + ... \rightarrow \mathcal{P}_{\mathcal{R}}(k) = A_s \left(\frac{k}{k_*}\right)^{n_s - 1}, \tag{1.39}$$

where k_* is a pivot scale, A_s is the scalar amplitude (the value of the spectrum at the pivot scale) and n_s is called the *spectral index*. The spectral index quantifies the scale dependence of $\mathcal{P}_{\mathcal{R}}(k)$. As mentioned above, current observational data are consistent with a power spectrum which is *nearly scale-invariant*. One can write the spectral index n_s and the scalar amplitude A_s as a function of the slow-roll parameters ϵ_1 and ϵ_2 for canonical single-field slow-roll inflation:

$$n_s = 1 - 2\epsilon_1 - \epsilon_2. \tag{1.40}$$

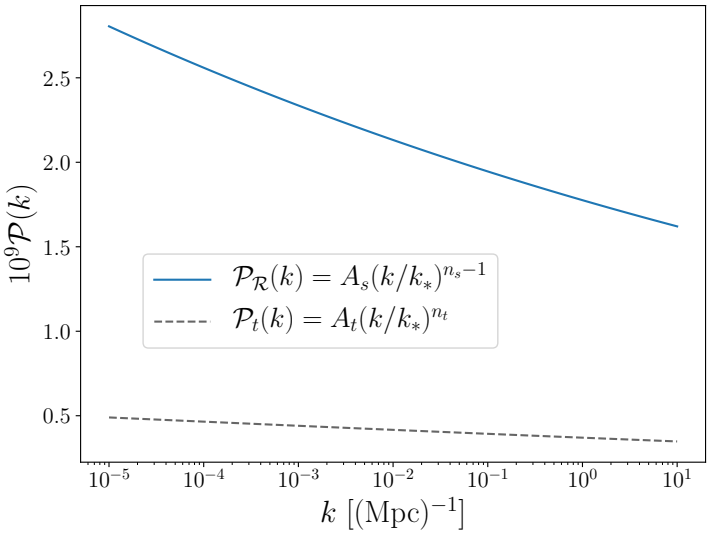

Figure 1.3: Tensor $\mathcal{P}_t(k)$ and scalar $\mathcal{P}_\mathcal{R}(k)$ primordial power spectra for the best fit values of ΛCDM taking the tensor-to-scalar ratio $r = 0.2$ for illustration purposes. Plot generated with `CAMB`.

$$A_s = \frac{H^2}{8\pi^2\epsilon_1}. \tag{1.41}$$

In the single-field slow-roll scenario, ϵ_1 and ϵ_2 are taken to be small, which indicates that n_s should slightly deviate from one, and this is why, by construction, single-field slow-roll inflation is consistent with current observations. If we assume that the distribution of the density perturbations is Gaussian, the power spectrum contains all information to characterize it. This time-dependence of the Hubble parameter (during inflation, the Hubble parameter decreases as ϵ increases) induces a slight increment of $\mathcal{P}_\mathcal{R}(k)$ for low values of k, producing a *red-tilted* primordial power spectrum. Moreover, the parametrization (1.39) is also used for the perturbations of the metric (tensor modes),

$$\mathcal{P}_t(k) = A_t \left(\frac{k}{k_*}\right)^{n_t}, \tag{1.42}$$

where A_t is the tensor amplitude and n_t is the *tensor tilt*. The ratio between both, the scalar and tensor amplitudes, is called the *tensor-to-scalar ratio*:

$$r = \frac{A_t}{A_s}. \tag{1.43}$$

1.3.3 Beyond canonical single-field slow-roll inflation

As seen in the previous section, the simplest inflationary scenario (the canonical single-field slow-roll case) gives by construction a nearly scale-invariant power

spectrum, which is the one supported by cosmological observations (Bennett et al., 2013)(Planck Collaboration et al., 2014b)(T. M. C. Abbott et al., 2018). If we are interested in studying deviations from this power-law primordial power spectrum (1.39), any alternative proposed approach has to predict a spectral tilt n_s smaller than one so that the power spectrum is still red-tilted.

In theoretical physics, it is customary to use a bottom-up approach to build an Effective Field Theory (EFT) to describe low-energy physical phenomena in a model-independent way. The idea behind EFT is that in nature there is often separation of scales: many aspects of physics at a given energy or length scale do not depend on detailed information about physical processes on much smaller length scales or much larger energies ("UV-scales"). EFT is a powerful framework to study the low-energy physics (with some limitations) by parametrizing the "unknowns" at UV-scales. In the case of inflation, we can have a single degree of freedom responsible for driving the perturbations and the background evolution, as in the canonical single field case we just discussed. However, in other scenarios, a single degree of freedom is responsible for the dynamics of the perturbations whereas some unknown effects on the background physics are encoded in other primordial functions. In this last case, the effective single field action for the perturbations up to second order, neglecting higher-order slow-roll corrections ($\sim \mathcal{O}(\epsilon^2)$) is given by,

$$S_2 = \int d^4x a^3 M_P^2 \epsilon_1 \left[\frac{\dot{\mathcal{R}}^2}{c_s^2} - \frac{(\partial_i \mathcal{R})^2}{a^2} \right], \tag{1.44}$$

where $c_s(t)$ is the curvature perturbation's speed of sound and encodes part of this unknown physics. This action has a single degree of freedom and reduces to the action (1.27) when $c_s = 1$. This action is constructed following a bottom-up approach by (Cheung, Fitzpatrick, Kaplan, Senatore, & Creminelli, 2008), where the action is written in terms of the Goldstone boson of time diffeomorphisms $\pi(t, \mathbf{x})$, which is related to the curvature perturbation $\mathcal{R}$ through the relation $\mathcal{R} = -H\pi$ and assuming $\dot{\pi}^3$ to be small and approximately constant. On the other hand, this action can be also obtained in a multifield setting by integrating out the heavy fields, as in reference (Achúcarro, Gong, Hardeman, Palma, & Patil, 2012). This action has been used in chapter 2 to study localized deviations from the power-law (1.39) in the primordial power spectrum, denominated in the literature as *features*.

1.4 The Large Scale Structure of the universe

The Friedmann-Lemaitre-Robertson-Walker (FLRW) metric provides a good description of the universe at large scales (those larger than 100 Mpc). Nevertheless, on smaller scales, the universe is no longer isotropic and homogeneous, and it begins to show a web of clustered matter, known as the Large Scale Structure (LSS) of the universe. The description of the physics taking place at smaller scales can be done by studying the perturbations around the FLRW cosmological background, assuming

that the perturbations remain small. Scales that can be treated linearly in perturbation theory are called *linear scales*. A detailed explanation of the theory of linear order perturbations can be found in (Dodelson, 2003).

We are interested in the time evolution of cosmological perturbations, which can be divided into three different epochs. At early times, all of the perturbation modes are outside the horizon and follow the same evolution. At intermediate times, the modes' wavelengths are within the horizon and the universe evolves from radiation to matter domination. The matter-radiation equality and the stage of the horizon crossing play an important role in the evolution of the modes. At late times, all the modes follow the same evolution again. We mainly observe the distribution of matter at late epochs, which corresponds to this latest stage.

Following the same perturbation theory approach explained above, let us start by defining the dimensionless matter density contrast δ_{m} in terms of the matter density ρ_{m} and the spatially constant density $\bar{\rho}_{\mathrm{m}}$:

$$\delta_{\mathrm{m}} = \frac{\delta\rho_{\mathrm{m}}}{\bar{\rho}_{\mathrm{m}}} = \frac{\rho_{\mathrm{m}} - \bar{\rho}_{\mathrm{m}}}{\bar{\rho}_{\mathrm{m}}}. \tag{1.45}$$

Poisson's equation, which is obtained by the theory of linear perturbations on the FLRW metric, relates the gravitational potential in Fourier space $\Phi(k, a)$ and the matter density contrast as:

$$\Phi = \frac{4\pi G\rho_{\mathrm{m}}a^2\delta_{\mathrm{m}}}{k^2} \tag{1.46}$$

We are mostly focused on studying the Fourier transform of δ_{m} because at linear scales, the different Fourier modes $\mathbf{k}$ are independent. Moreover, for an isotropic Gaussian field, the Fourier-transformed quantity is described by the power spectrum. In particular, the Fourier transformed matter density contrast $\delta_{\mathrm{m}}(\mathbf{k})$ can be described in terms of the matter power spectrum $P_{\delta\delta}(k, z)$ as

$$\langle\delta_{\mathrm{m}}(\mathbf{k}_1)\delta_{\mathrm{m}}(\mathbf{k}_2)\rangle = (2\pi)^3 P_{\delta\delta}(k, z)\delta^3(\mathbf{k}_1 - \mathbf{k}_2) \tag{1.47}$$

where $\delta^3(\mathbf{k}_1-\mathbf{k}_2)$ is the three-dimensional Dirac delta function, and assuming isotropy, the vector-dependency on $\mathbf{k}$ can be dropped. The matter power spectrum can be related to the primordial power spectrum $\mathcal{P}_{\mathcal{R}}(k)$ of curvature perturbations via the matter transfer function $\mathcal{T}_m$ through

$$P_{\delta\delta}(k, z) = 2\pi^2\mathcal{T}_m^2(k, z)\mathcal{P}_{\mathcal{R}}(k), \tag{1.48}$$

where the transfer function $\mathcal{T}_m$ is introduced to describe the combined effect of physical processes that affect the growth of our perturbations such as acoustic oscillations (see subsection 1.5.2), silk damping, radiation drag, free-streaming damping, among others. The transfer function $\mathcal{T}_m$ is usually computed by the numerical Boltzmann solvers like `CAMB` (Lewis, Challinor, & Lasenby, 2000; Howlett, Lewis, Hall, & Challinor, 2012) and `CLASS` (Essinger-Hileman et al., 2014). See Figure 1.4 for a visualization of the matter power spectrum. Combining the Boltzmann equations governing

the Dark Matter evolution and the Poisson equation described in equation (1.46) we can derive a second-order differential equation for the matter density perturbation $\delta_{\rm m}$ in Fourier Space:

$$\delta''_{\rm m}(z,k) + \left[\frac{H'(z)}{H(z)} - \frac{1}{1+z}\right]\delta'_{\rm m}(z,k) - \frac{3}{2}\frac{\Omega_{\rm m}(z)}{(1+z)^2}\delta_{\rm m}(z,k) = 0\,, \tag{1.49}$$

where, here, the prime denotes derivatives with respect to z, and the time-redshift evolution of $\Omega_{\rm m}(z)$ is given by equation (1.15). The second term of this equation can be seen as a friction term that shows how the growth of structures can be affected by the expansion of the universe. The last term is related to the enhancement of matter densities due to gravity. Note that this equation is only valid in the limit when $\delta_{\rm m} \ll 1$ (terms of order δ^2 are ignored).

The solution $\delta_{\rm m}(z,k)$ of equation (1.49) is scale-independent at later times; see e.g. (Dodelson, 2003). This scale-independence motivates the introduction of the growth factor $D(z)$ through

$$\delta_{\rm m}(z,k) = \delta_{\rm m}(z_\star,k)\frac{D(z)}{D(z_\star)}\,, \tag{1.50}$$

where $z_\star$ is an arbitrary reference redshift in the early stages of the matter-dominated era. Both $\delta_{\rm m}(z,k)$ and $D(z)$ can be used to define another important quantity, i.e. the growth rate $f(z)$:

$$f(z,k) = -\frac{{\rm d}\ln\delta_{\rm m}(z,\vec{k})}{{\rm d}\ln(1+z)}. \tag{1.51}$$

When $\delta_{\rm m}(z,k) \equiv \delta_{\rm m}(z)$ applies, the growth rate is itself a solution of a first-order differential equation:

$$f'(z) - \frac{f(z)^2}{1+z} - \left[\frac{2}{1+z} - \frac{H'(z)}{H(z)}\right]f(z) + \frac{3}{2}\frac{\Omega_{\rm m}(z)}{1+z} = 0\,, \tag{1.52}$$

with initial condition $f(z=z_\star)=1$. In the Large Scale Structure literature, the power spectrum of matter fluctuations $P_{\delta\delta}$ is usually normalised at present times by requiring that the r.m.s. variance on a sphere of radius $R_8 = 8\,h^{-1}\,$Mpc is equal to a normalisation factor dubbed σ_8, namely

$$\sigma_8^2 \equiv \frac{1}{2\pi^2}\int {\rm d}k\, P_{\delta\delta}(k,z=0)\,|W_{\rm TH}(kR_8)|^2\,k^2, \tag{1.53}$$

with $W_{\rm TH}(x) = 3(\sin x - x\cos x)/x^3$ a top-hat filter in Fourier space. It is also customary to introduce the parameter S_8 defined as,

$$S_8 = \sigma_8\sqrt{\frac{\Omega_{\rm m}}{0.3}}. \tag{1.54}$$

1.4.1 The non-linear regime

As mentioned above, equation (1.49) holds when $\delta \ll 1$. If we go to much smaller scales the perturbations become more important and the linear theory breaks down, which means that the growth of smaller massive perturbations cannot any longer be tracked with this equation. This breakdown corresponds approximately to a scale today of around $k \approx 0.1$ Mpc^{-1} and it is denominated the *non-linear regime*.

Assuming the Standard Cosmological Model, structures are considered to be formed hierarchically. This means that small structures are formed first around the primordial over-densities of the matter density field. Due to gravity, these structures collapse into dark matter halos. Larger structures form later by the accretion of mass nearby or by the merging of different halos. Beyond the scales of a galactic halo or inside it, it is very difficult to study the evolution of structures.

There exist different theoretical approaches to extend the model predictions to smaller scales; for instance, the standard perturbation theory (SPT), see the review by (Bernardeau, Colombi, Gaztañaga, & Scoccimarro, 2002). However, we need to rely on prescriptions obtained from the fitting of cosmological simulations to go to smaller scales. A widely-known used prescription to model the non-linear part of the matter power spectrum $P_{\delta\delta}$ is `halofit` (R. E. Smith et al., 2003), where the authors obtained the fitting prescription from a library of N-body cosmological simulations (see Figure 1.4 for the `halofit` correction to $P_{\delta\delta}$). Although the code `halofit` is able to predict correctly the non-linear matter power spectrum up to very small scales, the fitting prescription was obtained from Cold Dark Matter simulations. This means that `halofit` may provide biased results for the non-linear modelling if massive neutrinos are taken into account or the non-linear modelling requires to go to even smaller scales where the impact of baryonic interactions is non-negligible. Because of this reason, other corrections are needed, the most famous being the Bird and Takahashi corrections (Bird, Viel, & Haehnelt, 2012) (Takahashi, Sato, Nishimichi, Taruya, & Oguri, 2012).

Recently, other tools have been developed to predict the non-linear corrections to the matter power spectrum. For example, the code `hmcode` (Mead et al., 2016) introduced physically-motivated free parameters for the halo model formalism that were fitted using high-resolution N-body simulations for a variety of cosmological models. Moreover, another mechanism to make predictions for the non-linear regime is based on *emulators* (see for instance (Heitmann, Lawrence, Kwan, Habib, & Higdon, 2014)). The idea behind emulators is based on running high-resolution simulations at key points in the cosmological parameter space so that the full parameter space can be recovered evenly. Then, an emulator is constructed to interpolate between the values of the grid so that it gives predictions for any set of parameters within the space and corrections for the matter power spectrum are computed.

1.5 Observations

The number of precise and new observational probes used to constrain our understanding of the underlying physics that explain the universe has increased significantly in the last decades. In this section, we detail the most important ones. Some of these observational probes rely on definitions of *distances* in our universe. Cosmologists work mainly with three different types of distances:

- **Comoving distance χ:** measures the distance travelled by a light ray between an object at redshift z and us in a coordinate system that expands with the scale factor $a(t)$. The comoving distance has already been introduced in equation (1.17), and rewriting it in terms of the Hubble parameter $H(z)$ we obtain:

$$\chi = \int_0^z \frac{dz\prime}{H(z\prime)} \tag{1.55}$$

  where z is the redshift and $z = 0$ is usually taken at present time.

- **Angular diameter distance $D_{\rm A}$:** measures the distance to a distant object of proper size l subtended by an angle θ as

$$D_{\rm A} = \frac{l}{\theta}. \tag{1.56}$$

  For flat cosmologies (i.e: ΛCDM), relating the angle θ to the comoving distance χ, we arrive at the redshift z dependent expression:

$$D_{\rm A} = \frac{\chi}{1+z}. \tag{1.57}$$

- **Luminosity distance $D_{\rm L}$:** measures the distance to a source by means of studying the flux F of an object with known luminosity L:

$$D_{\rm L}^2 = \frac{L}{4\pi F} \tag{1.58}$$

  For flat cosmologies (i.e: ΛCDM), we can relate $D_{\rm L}$ with the angular diameter distance $D_{\rm A}$:

$$D_{\rm L} = (1+z)^2 D_{\rm A}. \tag{1.59}$$

1.5.1 Supernovae Type Ia (SNIa)

Some of the best cosmological constraints so far have been obtained using observations of Supernovae Type Ia (SNIa): a type of supernovae that does not have any presence of hydrogen but does show strong Si II absorption lines in their spectra. This type of supernovae is believed to originate from the thermonuclear disruption of carbon-oxygen white dwarfs. Due to their characteristics (they are very bright, and can be spotted at high redshift), they are used as distance indicators in Cosmology.

In particular, the main observable used in SNIa observations is the distance modulus[9] $\mu_{\rm D}(z)$:

$$\mu_{\rm D} = 5\log_{10}\left(\frac{H_0}{c}D_{\rm L}\right) \tag{1.60}$$

where $D_{\rm L}$ is the luminosity distance defined above. Note that the speed of light factor c was recovered in this equation.

1.5.2 CMB angular power spectrum

As we mentioned in the last section, the inflationary hypothesis is not only used to explain the fine-tuning of the initial conditions issues but it provides an elegant mechanism of production of the density fluctuations that seeded the early universe. These primordial density fluctuations have been seen in the temperature anisotropies $\Delta T/T_0$ of the Cosmic Microwave Background (CMB). The CMB is a relic radiation in our universe. After Big Bang nucleosynthesis, the universe was filled with a very dense baryon plasma, where baryons tend to cluster together due to gravitation. However, the pressure created by photons was so large that it could stop the clustering. We say that the baryonic plasma was suffering *acoustic oscillations*. When the universe was approximately 380000 years old (at a redshift $z_{\rm dec} \approx 1090$), it cooled down to a temperature of approximately 1 eV, allowing light nuclei begin to bind with electrons to form neutral atoms. These bindings induced a drop in the number density of free electrons. Hence, the decoupling of photons from matter took place, as photons could no longer scatter with electrons. This is the so-called *recombination* epoch. These photons could finally travel freely through space, with an almost isotropic distribution. The relic temperature of this radiation is now $T_0 \approx 2.7$ K and makes up the CMB. The acoustic oscillations got imprinted in a characteristic scale r_s defined as

$$r_s(z_{\rm dec}) = \int_{z_{\rm dec}}^{\infty} d\tilde{z}\frac{c_s(\tilde{z})}{H(\tilde{z})}, \tag{1.61}$$

where c_s is the sound speed of the photon-baryon fluid.

The temperature anisotropies $\Delta T/T_0$ can be related to the primordial power spectrum $\mathcal{P}_{\mathcal{R}}(k)$ through the angular power spectrum C_ℓ^{TT}:

$$C_\ell^{TT} = \frac{2}{\pi}\int k^2 dk \mathcal{P}_{\mathcal{R}}(k)\mathcal{T}_\ell^T(k)\mathcal{T}_\ell^T(k), \tag{1.62}$$

where $\mathcal{T}_\ell^T(k)$ is the corresponding transfer function. The angular power spectrum of temperature anisotropies C_ℓ^{TT} can be seen in Figure 1.4. In general, the CMB background provides more information than the one encoded in the temperature anisotropies. In fact, in the CMB map, we can also study the polarization modes E and B of photons. Thus, the general expression for the angular power spectra of the CMB is

$$C_\ell^{XY} = \frac{2}{\pi}\int k^2 dk P(k)\mathcal{T}_\ell^X(k)\mathcal{T}_\ell^Y(k), \tag{1.63}$$

[9]We have decided to use the subscript D for the distance modulus μ to differentiate this variable from the cosine of the angle in the expression for the galaxy power spectra in subsection 1.5.3.

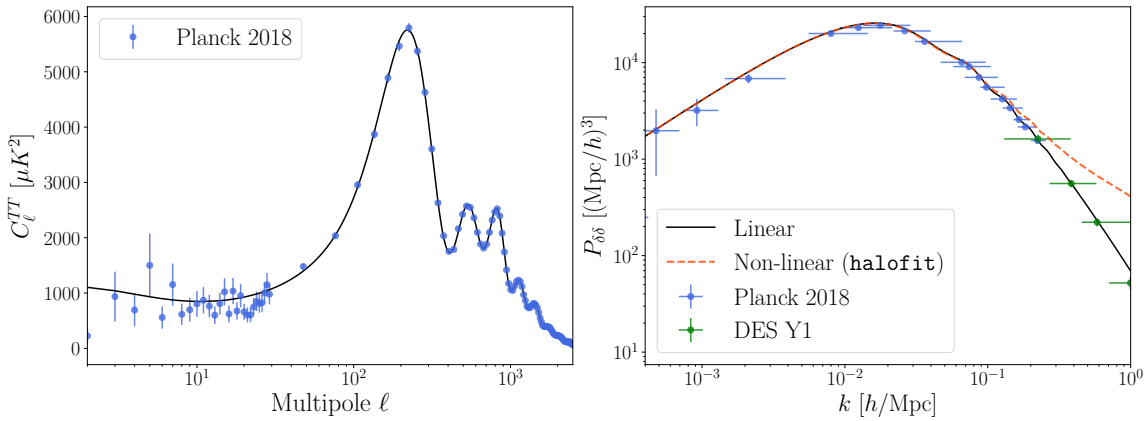

Figure 1.4: **Left:** CMB angular power spectrum of temperature anisotropies C_ℓ^{TT} with the experimental data of Planck 2018. **Right:** matter power spectra $P_{\delta\delta}$ (linear and non-linear) with Planck 2018 and DES Y1 data. The plot has been generated with `CAMB` and `matplotlib` using the available data provided by (Chabanier et al., 2019) and (Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020b). Solid line represents the best theoretical fit according to Planck 2018 data using the linear prediction for the matter power spectrum. The amplitude of the non-linear prediction (dashed red line), using `halofit`, has been increased for illustration purposes.

where X and Y may refer either to the temperature T or the polarization modes E and B. Furthermore, $\mathcal{T}_\ell^X$ are the transfer functions, which are usually written as a line-of-sight integration in conformal time that contains the source factor $S^X(k,\tau)$ and the geometric projection $P_\ell^X(k|\tau_0-\tau|)$ based on Bessel functions (see Figure 1.2):

$$\mathcal{T}_\ell^X(k) = \int_0^{\tau_0} d\tau S^X(k,\tau) P_\ell^X(k|\tau_0-\tau|). \tag{1.64}$$

The computation of the transfer functions is usually done numerically. There are several codes that are designed for this goal. The most well-known ones are `CLASS` (Essinger-Hileman et al., 2014) and `CAMB` (Lewis et al., 2000; Howlett et al., 2012). The Planck mission (Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020b) has measured successfully C_ℓ^{TT}, the cross-correlation spectrum C_ℓ^{TE} and C_ℓ^{EE}.

1.5.3 Galaxy Clustering

We have seen in section 1.4 that we can model the distribution of matter in our universe by the matter power spectrum $P_{\delta\delta}$. The key question is how we can infer from observations the matter distribution by looking at the bright distant objects present in the night sky. Galaxy Clustering (GC) describes how well the distribution of galaxies in the universe traces the underlying matter distribution. The theoretical prediction for the positions of galaxies in the universe depends on the cosmological model, so measuring the position from galaxies is a powerful proof to constrain cosmological parameters. For that, we need a robust modelling for the relation between the galaxy

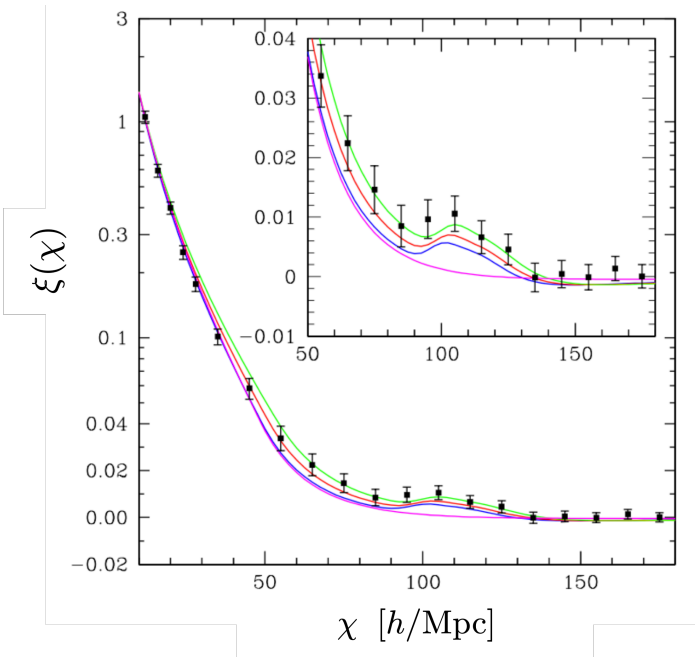

Figure 1.5: Redshift-space correlation function compared to different cosmological models. The data points are taken from the Sloan Digital Sky Survey (SDSS). The magenta line shows the prediction of ΛCDM without baryons ($\Omega_b = 0$), which demonstrates that the BAO peak is statistically significant. The plot is original from (Eisenstein et al., 2005).

and the matter distributions. This relation is encoded in the so-called *galaxy bias* b. One of the simplest models relating the matter and galaxy densities is given by a constant factor (Peacock & Dodds, 1994):

$$\delta_g = b\delta_m. \tag{1.65}$$

This model is known to be very simplistic but is still used for forecasting. In chapter 6 different bias models are shown.

When galaxy clustering is used as an observational probe to constrain cosmological models, other observational probes can be taken into account to increase in accuracy. One of these probes is the Baryonic Acoustic Oscillations (BAO), which do not only affect photons (as we have seen in the case of the CMB in subsection 1.5.2) but also baryons. The characteristic radius scale formed when the waves froze during electron-photon decoupling (at around $z_{\text{drag}} \approx 1020$ during the so-called drag epoch) is imprinted as an excess of power on the distribution of baryons. Since baryons and dark matter interact gravitationally, the latter would tend to cluster on this particular scale. Consequently, galaxies likely form in the higher density regions corresponding to the BAO scale $r_{\text{s}}(z_{\text{drag}})$ (see Figure 1.5).

Another probe is Redshift Space Distortions (RSD). When we measure galaxy redshifts, we need to take into account both the peculiar velocities of galaxies[10] and the Hubble flow. If only the Hubble recession is considered when we convert redshift measurements to distances, we will recover a distorted field showing RSD. We can write down the relation between the redshift space galaxy power spectrum $P_{\rm gg}$ and the real-space matter power spectrum $P_{\delta\delta}$ as (Kaiser, 1987)(Hamilton, 1998)

$$P_{\rm gg}(k, \mu_k, z) = P_{\delta\delta}(k)(b + b_\nu f(z)\mu_k^2)^2 \tag{1.66}$$

where b is the bias introduced in equation (1.65), b_ν is the bias between galaxy and matter velocity distributions (which is usually neglected), f is the growth rate introduced in section 1.4 and μ_k is the cosine of the angle to the line-of-sight. The RSD also carry cosmological information in the growth rate f.

Finally, redshift surveys such as the future *Euclid* mission (see chapter 6) will measure the positions of galaxies and their redshift. Still, to compare this information to theoretical predictions based on cosmological models, we need to translate the position and redshift of galaxies to comoving coordinates. To do this process, we need to assume a fiducial cosmological model. If the used fiducial model does not agree with the real cosmological model of our universe, some extra distortions similar to the RSD (Ballinger, Peacock, & Heavens, 1996) will be introduced. This is called the Alcock-Pacyznski (AP) effect. This effect is taken into account when the recipe of the *Euclid* mission observables is introduced in chapter 6.

1.5.4 Weak Lensing

Another approach to studying the formation of structures and constraining the underlying cosmological model is not to look only at the position of galaxies in the sky but also at their shapes. One of the most fascinating predictions of Einstein's Theory of General Relativity[11] (GR) is that the light from distant objects will be deflected if there is a massive concentration along its way, inducing, for instance, distortions in the shape of galaxies. This effect is called Gravitational Lensing (GL) and we can use it to infer the mass distribution in the universe. There are different regimes of gravitational lensing. When the deflection of light coming from the source is large, we are in the so-called strong lensing limit, where even multiple images of the same object or giant arcs are seen on the sky (see Figure 1.6). Another example is when we study the light coming from background distant galaxies. The magnitude of the distortions is smaller than in the strong lensing regime, where the effect induced in the shapes of galaxies is very subtle. However, we can statistically average all these small distortions to infer the mass distribution in the LSS of the universe (Kaiser, 1998). We call this limit the Weak Lensing (WL) regime, and we can relate the distortion of the shapes of galaxies with the underlying mass power spectrum.

[10]The gravitational force exerted on galaxies by their neighbouring mass can induce peculiar velocities deviating from the Hubble flow.

[11]In fact, General Relativity was widely accepted after the detection of the deflection of a starlight ray during a solar eclipse, whose magnitude was well in agreement with that predicted by Einstein's theory (Dyson, Eddington, & Davidson, 1920).

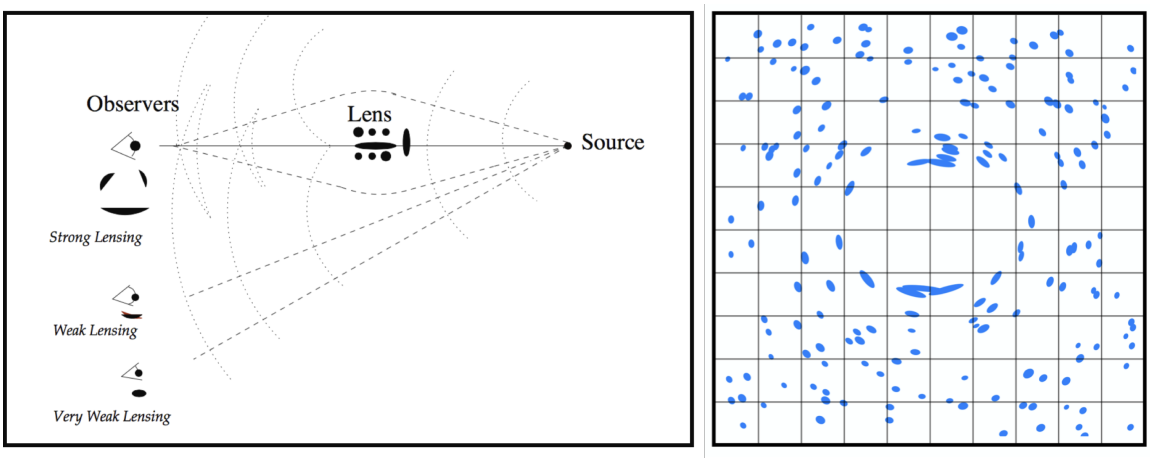

Figure 1.6: **Left**: schematic illustration of the light wavefront and the different regimes of gravitational lensing depending on the position of the observers. Deflection arcs and double images, for instance, fall in the strong lensing limit (Courbin et al., 2002). **Right**: illustration of the distortions caused by strong gravitational lensing (centre of the image) and by weak lensing (the shape of the galaxies shows small variations as we move away from the centre of the image) (Peirone, 2020).

Mathematically, the distortion of light that induces the apparent change in the intrinsic ellipticities of galaxies is described using GR. Deriving the geodesic equation for the transverse motion of photons travelling from the source towards the observer and integrating twice, we can obtain the position of the imaged shape as a function of the source position:

$$\frac{\partial\theta^i_{\text{source}}}{\partial\theta^j} - \delta^{\text{K}}_{ij} \equiv \int_0^{\chi_{\text{source}}} d\tilde{\chi} \left(1 - \frac{\tilde{\chi}}{\chi_{\text{source}}}\right) \tilde{\chi}(\Phi_{,ij}), \tag{1.67}$$

where χ_{source} is the comoving distance to the source, θ^j is the angle under the given light ray between the source and the observer, θ^i_{source} is the unlensed angle, i, j represent the direction on the sky, δ^{K}_{ij} is the Kronecker delta and Φ is the gravitational potential introduced in equation (1.46). We usually describe the change between the source position and the observed one using the two-dimensional symmetric matrix:

$$\frac{\partial\theta^i_{\text{source}}}{\partial\theta^j} - \delta_{ij} = \begin{pmatrix} 1 - \kappa_{\text{GL}} - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa_{\text{GL}} + \gamma_1 \end{pmatrix} \tag{1.68}$$

where κ_{GL} is denominated *convergence*[12], which describes the magnification of an image, and γ_1, γ_2 are the two components of the so-called *shear*, which describe the distortions and is the most important quantity for weak lensing studies.

Therefore, weak lensing makes circular images look like elliptical, and this effect can be studied statistically to obtain valuable cosmological information using current

[12]We have added the subscript GL to differentiate this κ from the one appearing in the FLRW metric.

and future galaxy surveys. However, the high-quality of incoming surveys also requires accurate modelling of the systematic effects of weak lensing measurements. In this respect, it is crucial to model the ellipticity of the galaxy shapes robustly. Furthermore, the Intrinsic Alignment (IA) of galaxies must also be taken into account so that the weak lensing signal does not become contaminated. During their formation and evolution, galaxies tend to align with the matter distribution and neighbouring galaxies due to tidal forces induced by the surrounding large scale structure. These intrinsic alignments, if not taken into account, can mimic the weak lensing signal and produce a systematic signal larger than the experimental errors. In chapter 6, the modelling for the *Euclid* weak lensing observable, including IA corrections, is explained in detail.

1.6 The Standard Cosmological model

Current cosmological observations described in section 1.5 support statistically the Standard Cosmological Model, also known as the Concordance model or the ΛCDM model. This model concludes that we live in a spatially flat universe, which contains less than 5% of Baryonic Matter (BM) and approximately 26% of Cold Dark Matter (CDM). The other main contribution comes from Dark Energy (DE), which behaves like vacuum energy and is modelled with the cosmological constant Λ. The percentage of radiation is negligible today. The ΛCDM model assumes that the underlying distribution of the primordial density perturbations is Gaussian, and that the corresponding primordial power spectrum of curvature perturbations is almost-scale invariant.

This model is able to explain most of the current cosmological observations with only six free parameters, whose values have been obtained with high accuracy (see Table 1.1). The procedure and techniques to obtain the values of these parameters using cosmological data coming from the CMB and LSS are explained in detail in section 1.7.

In the last years, data analysis of the ΛCDM model using different observational probes has given values for the ΛCDM parameters that are in tension[13]. One example is the case is the value of the Hubble constant H_0, whose value obtained by direct measurements from SNIa is significantly higher than the indirect value obtained using CMB data (Riess et al., 1998). In the regime of LSS, other tensions are found. This is the case for the amplitude σ_8, whose value obtained from CMB measurements is in tension with the one obtained from cluster abundance. This tension in σ_8 induces another tension of 3σ on the value of S_8 measured from WL when compared to the CMB Planck results. A detailed review about the current state of cosmological tensions can be found in reference (Abdalla et al., 2022).

Different explanations aim to bring some light on these tensions. Some scientists strongly believe that the tensions may hint at new physics beyond the Concordance

[13]A parameter value given two data sets is referred to be in tension when their mean values and confidence intervals overlap partially or slightly.

Symbol	Description	Value
$\Omega_b h^2$	Density parameter of Baryonic Matter	0.02242 ± 0.00014
$\Omega_c h^2$	Density parameter of Cold Dark Matter	0.11933 ± 0.00091
$100\theta_{MC}$	Measure of the sound horizon at last scattering. It can be used to infer a value for H_0	1.04101 ± 0.00029
τ_{reion}	Thomson scattering optical depth due to reionization	0.0561 ± 0.0071
$\ln(10^{10} A_s)$	Scalar amplitude of the primordial power spectrum	3.047 ± 0.014
n_s	Spectral index of the primordial power spectrum	0.9665 ± 0.0038

Table 1.1: Mean values and 68% intervals for the base-ΛCDM model parameters from Planck CMB power spectra, in combination with CMB lensing and BAO, obtained by the Planck Collaboration. The decimal notation used in this table agrees with that of (Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020b). The parameter h represents the reduced Hubble constant defined as $h \equiv H_0/(100\,\mathrm{km\,s^{-1}Mpc^{-1}})$. The scale amplitude A_s is fitted using a redefinition of it as $\ln(10^{10} A_s)$ by the Planck collaboration. The inflationary parameters n_s and A_s were evaluated at the pivot scale $k_* = 0.005$ Mpc^{-1}.

model. Others support the assumption that the data analysis is contaminated by observational systematical errors that were beyond the control of the analysis methodology. In any case, the effort of the cosmological community is undeniable, either by proposing extensions of the ΛCDM model that could explain the tensions, or by further reviewing any possible systematics introduced on the measurements or on the data analysis.

1.7 Data Analysis in Cosmology

Statistics are crucial to evaluate the fit of cosmological observations to any theoretical model. The starting point is to define the *scientific question*. For example, you may be interested in finding the parameters that fit best a chosen theoretical model, or you may want to discover if a model, given some parameters, is statistically preferred than another one. These questions in Cosmology are usually answered using a Bayesian Statistical framework. This section aims to be a practical guide on how Bayesian statistics are currently used in the context of Cosmology and represents the main methodology followed in the scientific content of this thesis.

1.7.1 Bayesian Statistics

Bayesian statistics assume that the parameters of interest, given that the model parameters are treated as random variables, have probability distributions (Ivezic, Connolly, VanderPlas, & Gray, 2014). Bayesian statistical analyses rely on *Bayes' Theorem*, which provides the probability distribution of the parameters θ given a model M and the observed data $\vec{d}$. This probability distribution, $P(\theta|\vec{d}, M)$, called *posterior distribution*, is defined as

$$P(\theta|\vec{d}, M) = \frac{\mathcal{L}(\vec{d}|\theta, M)\Pi(\theta|M)}{\mathcal{Z}(\vec{d}|M)}. \tag{1.69}$$

In equation (1.69), we find several probability distributions defined below:

- $\mathcal{L}(d|\theta, M)$ is the so-called *likelihood*, which gives the probability of observing the data d given the model M and the parameter values θ. It measures the compatibility of the data with the hypothesis.
- $\Pi(\theta|M)$ is the *prior distribution*, which is the probability distribution of the parameters θ taking into account all available external information. For instance, this information may come from previously collected data, limits imposed by theory, results of previous experiments, etc. This prior should not take into account the actual data d.
- $\mathcal{Z}(d|M)$ is the *evidence* (also called *marginal likelihood*), which gives the probability of observing the data given the external information and the chosen model.

Likelihood

The likelihood $\mathcal{L}$ is a probability distribution that determines how well a model M with a given set of parameter values θ matches the given experimental data d. The likelihood considers the assumptions of the model as well. In Cosmology, we are especially interested in testing whether a cosmological model M, such as ΛCDM or extensions of it, with its parameter values θ, can fit current data. Therefore, a relation between the experimental data, d, and the theoretical predictions for cosmological observables given by the model is needed. In many scenarios, the likelihood $\mathcal{L}$ is modeled as a multivariate Gaussian distribution function of dimension i:

$$\mathcal{L} = \frac{1}{(2\pi)^i\sqrt{|C|}} e^{-\frac{1}{2}(\vec{d}-\vec{T}(\theta))^t C^{-1}(\vec{d}-\vec{T}(\theta))}, \tag{1.70}$$

where $\vec{d}$ is the data vector, $\vec{T}(\theta)$ is the theory vector, C is the covariance matrix of the data $\vec{d}$, i is the number of variables of the data set $\vec{d}$ and $|C|$ is the determinant of the covariance matrix. This approach of using a multivariate Gaussian distribution to model the likelihood is applied in CMB experiments, for instance, or in the *Euclid* mission (see chapter 6, in particular section 6.5).

Prior

The already known information about the parameters of interest is summarized in a probability distribution denominated prior. The information coming from previous experiments can be used as prior distributions. Moreover, it is customary in Cosmology to use Gaussian priors for the parameters θ if the results from previous experiments show that the posterior distributions of those parameters θ resemble a Gaussian distribution. In general, to define a prior for the set of parameters θ, we need to set the probability distribution and the hyper-parameters[14] of the prior distribution.

In our research, in some cases, we restrict the hyper-parameters of the prior distribution according to the theory. In these cases, for each θ, we assign a uniform distribution, which is defined within the limits allowed by the theory, to ensure that the probability area in the parameters space is equal for all the parameter values. These priors are sometimes called uninformative priors, although there is a notorious debate in the field of Statistics about whether uniform probability distributions are indeed the less informative distributions. In other situations, we define more complex prior distributions, according to the theory, so that we have an efficient sampling (for instance, a Beta function, described in chapter 2).

We distinguish between two types of prior distributions: the conjugate and non-conjugate priors. In the case of using a conjugate prior, the posterior distribution is in the same probability distribution family as the prior distribution (for example, if the prior is a Gausiaan distribution and the likelihood is defined as in equation (1.70), the posterior distribution is also Gaussian). Otherwise, for non-conjugate priors, the posterior distribution is not part of the same family as the prior and this usually increases the difficulty of the sampling procedure (see subsection 1.7.2).

Bayesian Statistics vs. Frequentist Statistics

A different approach to statistical problems is the *Frequentist approach*. This approach is very common in particle physics and other fields of science. Working with an example, this approach is, in general, based on defining a *null hypothesis* such as "the Primordial Power Spectrum does not deviate from an almost-scale invariant power law", and an *alternative hypothesis*, for instance, "the Primordial Power Spectrum has some features described by the model M" (Chluba, Hamann, & Patil, 2015). Based on the results, the null hypothesis is either rejected or retained.

In the Frequentist approach, the most important output is not a probability density distribution of the parameters, but a numerical value called *statistic*, which is used to determine whether the null hypothesis can be rejected or not. Associated with the statistic, the assumed underlying probability distribution, and the degrees of freedom we have a p-value (Robert, 2011). This value is defined as the probability of obtaining a statistic value equal to or more extreme than the observed one under

[14]We call hyper-parameters the parameters of the prior distributions. Examples of hyper-parameters are, for instance, the bounds of a uniform distribution. More examples of hyper-parameters can be seen in chapter 2

the assumption that the null hypothesis is correct. If the associated p-value is smaller than the probability of rejecting the null hypothesis when it is true[15], α, we reject the null hypothesis in favour of the alternative one. In general, α is set to 0.05.

There is a notorious debate regarding the choice of a Bayesian or Frequentist approach to tackle scientific problems. Both of them offer pros and cons: many researchers criticise the use of a prior in Bayesian Statistics; in particular, whether the selection of the prior distribution may be misleading. On the other hand, Frequentist statistics is considered less flexible. Besides, the publication bias has induced a tendency on obtaining p-values smaller than 0.05 (Viechtbauer, 2007). In some exceptional cases, there has been a malpractice called *p-hacking*, which consists on changing models and hypothesis on the fly during research and/or modifying the data sets to obtain a p-value below 0.05 (Reyes, Dieste, C., & Juristo, 2020; Head, Holman, Lanfear, Kahn, & Jennions, 2015). In any case, the use of Bayesian statistics in infering the parameters distributions of a cosmological model or in the research of extended cosmological models is more popular, due to the demanding numerical simulations and other procedures required to analyse the data in the Frequentist approach (Schafer & Stark, 2003). One of these requirements is the large number of experiments (or repetitions of measurements), such as in particle physics, where a collision can be repeated numerous times. Unfortunately, in Cosmology, we cannot perform several experiments for the same research question and this is one of the main reasons behind the application of Bayesian statistical analysis (Schafer, 2015).

1.7.2 Parameter Inference

When only interested in the estimation of the values of the set of parameters θ that best fit a model M, the normalization $\mathcal{Z}(d|M)$ of the Bayes' Theorem (Equation 1.69) is usually omitted, because it provides only a re-scaling of the normalization of the distribution. Thus

$$P(\theta|d, M) \propto \mathcal{L}(d|\theta, M)\Pi(\theta|M). \tag{1.71}$$

To infer the best value of the parameters θ we need to explore the parameter space and test if the model fits the data well for a wide range of parameter values within this space. In Cosmology, we usually deal with a very large number of parameters with non-conjugate prior distributions, which complicate the analytical evaluation of the posterior distribution $P(\theta|d, M)$ (Equation 1.71). Therefore, we need to use numerical tools to sample computationally the posterior distribution. Usually, this procedure is done by evaluating possible parameter values using a grid. These tools are based on a random sample drawn from the real posterior distribution. The most common ones are *Monte-Carlo Markov Chain* (MCMC) methods (see (Eckhardt, 1987) for a review). The name Monte Carlo simply means obtaining a representation of the distribution by sampling it randomly, while the Markov Chain improves the sampling efficiency.

[15]This is also known as type-I error.

A Markov Chain is defined by a series of random variables where the probability of the outcome of a random variable in the current step depends only on the outcome of the random variable in the previous step (Markov, 2006). We are interested in two main properties of the Markov Chains. First, we need the chain to be *stationary*: the distribution should not depend on time (meaning, it should not depend on the number of steps). Second, we need the chain to reach a state where the next elements of the chain are picked from the high-density regions of the posterior distribution (that is, the elements are selected from the region close to the likelihood maximum[16]).

In the simplest scenario for an MCMC, we compute the set of parameter values iteratively. At each iteration, we check whether the new set of values fits better than the previous one; this means that the set of values sampled from the posterior distribution is closer to the maximum of the likelihood. This step is done by checking the values given a selection criterion. Different sampling algorithms use different selection criterion. If the selection criterion is fulfilled, the algorithm selects the new set of values. Otherwise, it keeps the previous one. At the end of each iteration, the set of parameter values is saved. The ordered set of saved iterations forms what we call a *chain*. The density of points in the chain gives the posterior distributions of the parameters.

Once the posterior distribution is sampled, we usually marginalise it to show the results of a multi-parameter fit. Marginalization involves projecting the distributions to all other dimensions, and it is usually achieved by summing or integrating over the unwanted distributions of the parameters as

$$P_{\Theta_1}(\theta_1) = \int_{\theta_2} P_{\Theta_1,\Theta_2}(\theta_1, \theta_2) d\theta_2 \tag{1.72}$$

where Θ_1 are the set of random variables with values θ_1, which are the variables that we want to keep, and Θ_2 are the set of random variables with values θ_2 that we aim to get integrated (or "marginalized" over). In Cosmology, it is customary to use already available codes to post-process the MCMC chains, being one of the most popular codes `GetDist` (Lewis, 2019).

There are several sampling algorithms for the MCMC process. In this thesis, we focus on Metropolis-Hastings (MH) and Nested Sampling algorithms. In Cosmology, there are several Bayesian Analysis Framework codes designed to carry out statistical analyses of parameter inference of cosmological models against data. The most famous ones are `MontePython` (Audren, Lesgourgues, Benabed, & Prunet, 2013), `CosmoMC` (Lewis & Bridle, 2002) (and its new `python` version `Cobaya` (Torrado & Lewis, 2021, 2019)) and `CosmoSIS` (Zuntz et al., 2015). These frameworks usually contain a so-called *theory code* (for instance `CAMB`, `CLASS` or both) to provide theoretical predictions of the cosmological observables, different data sets and several *samplers* (such as the nested samplers `PolyChord` (W. J. Handley, Hobson, & Lasenby, 2015a, 2015b) or `MultiNest` (Feroz & Hobson, 2008; Feroz, Hobson, & Bridges, 2009; Feroz, Hobson, Cameron, & Pettitt, 2019), and Metropolis-Hastings MCMC samplers).

[16]Maximize the likelihood $\mathcal{L}$ is equivalent to minimize the $-\log \mathcal{L}$.

Metropolis-Hastings algorithm

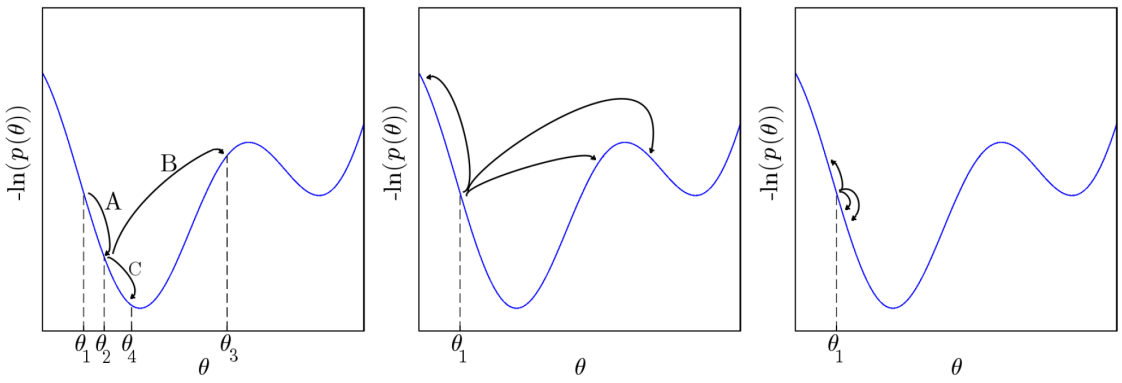

Figure 1.7: **Left**: an example of a Markov Chain constructed by the Metropolis-Hastings algorithm. It starts at 1, then 2 is proposed and accepted (step A), 3 is proposed and refused (step B), and finally, 4 is proposed and accepted (step C). The resulting chain is $(\theta_1, \theta_2, \theta_2, \theta_4)$. **Central**: an example of what happens when the proposal step is too broad: the different proposed new set of values are further away from the minimum of the log-posterior, and therefore, they get rejected. As a result, the chain lacks mobility because all the proposals are unlikely and it may get stuck. **Right**: an example of what happens with a proposal step that is too narrow: all the proposed new set of values are accepted, making the chain to move towards the minimum of the log-posterior but it samples the parameter space very slowly and the convergence takes longer. Figure from (Leclercq et al., 2014).

This algorithm (Hastings, 1970) is based on a particular choice of the selection criterion to decide whether to keep the set of values of the parameter space during the current step θ', or to come back to the previous step θ. This criterion is based on calculating a ratio, called *acceptance ratio*, a, defined as[17]

$$a = \frac{p(\theta')q(\theta|\theta')}{p(\theta)q(\theta'|\theta)}, \tag{1.73}$$

where $p(\theta')$ is the posterior distribution and, the proposal distribution $q(\theta'|\theta)$, is the conditional probability of proposing a new θ' given the previous θ values. In general, q is proposed to be a Gaussian distribution. At each step, a sample of θ' from $q(\theta'|\theta)$, and a random number v from a uniform distribution between 0 and 1, are drawn. If $a \geq v$, we accept θ', and it becomes a new state of the chain. Otherwise, we reject θ' and the new state of the chain is θ again. This is shown in figure 1.7. In this figure, it can be observed that each accepted step only depends on the previous one and that, eventually, it is independent of the number of steps, making the chain a Markov Chain. To speed up the algorithm, a covariance matrix C of previous sampling runs can be used in the likelihood calculation so that the algorithm samples

[17]Do not confuse with the scale factor $a(t)$ used in previous sections.

more efficiently the parameter space with prior information obtained from another statistical study.

Still, the major disadvantage of this algorithm is that the sampled sets of parameters are correlated because we only use the previous sample to obtain the current one. This means that, if we start the algorithm with a value within the parameter space that has a low probability, the chain will not reflect the underlying distribution very well until a higher likelihood region is reached. Therefore, a small percentage of the initial values of the chain are usually discarded (the so-called *burn-in* phase), in order to achieve the stationary property. Sometimes, the chains may also get stuck in a lower probability region and not be able to escape from it to further explore the parameter space. This is why, when this algorithm is used, several chains with different initial values are usually run. In any case, this algorithm also suffers in multi-modal posterior distributions where the chain can get stuck in relative maxima of the likelihood (see for instance multi-modal posterior distribution at Chapter 2). A similar problem will arise if there are high correlations or degeneracies among the parameters (showing correlated peaks in the posterior and "*banana*" contour plots).

The Metropolis-Hastings algorithm allows us to construct a Markov Chain containing as many values as parameters we are trying to infer per step. Once that the chain is stationary, we need to decide if we can stop the sampling. Therefore, a question arises: *how do we decide this?* A simple answer is not available, as no test can affirm whether a chain has converged or not. Still, some convergence diagnostics may point out some necessary conditions for a chain to show some convergence, although they are not sufficient conditions. Some diagnostics include:

- Individual segments of the chain show similar results, as long as the chain is much longer (sufficient large number of steps in the chain) than any obvious correlation.
- A reasonable number of accepted proposed steps. If the acceptance is too high, it may point out that the chain is slowly converging; however, if the acceptance is too small, it may indicate that the chain is locally stuck. An optimal acceptance rate for a Metropolis-Hastings MCMC is considered to be around 25-30% (Gelman, Gilks, & Roberts, 1997).
- As a result of running several chains at the same time with different starting points to avoid them getting stuck, we can apply a different stopping criterion: the *Gelman-Rubin* statistical criterion R (Gelman & Rubin, 1992). This test consists of comparing the variance within individual chains to the variance between chains. Their ratio, R, should be close to 1 when the chains have converged. In Cosmology, the usual convergence ratio is R-1<0.01.

Nested Sampling

To circumvent the issues presented by the Metropolis-Hastings algorithm, an alternative way of sampling the posterior was developed by J. Skilling (Skilling, 2006):

the *Nested Sampling* algorithm. This algorithm is based on calculating the evidence Z (marginalized likelihood). From the Bayes' theorem (1.69):

$$Z(d|M) \times P(\theta|d, M) = \mathcal{L}(d|\theta, M) \times \Pi(\theta|M) \rightarrow \tag{1.74}$$

$$\int Z(\theta) \times p(\theta) d\theta = \int \mathcal{L}(\theta) \times \Pi(\theta) d\theta \quad \rightarrow \quad Z = \int \mathcal{L} dX. \tag{1.75}$$

where we have defined dX as $dX = \Pi(\theta)d\theta$ and the prior and posterior are normalised to the unit total. From this equation (1.75), we observe that Z, which is the normalization in equation (1.69), is defined as the area below the curve $\mathcal{L}dX$ (see right panel of Figure 1.8). Therefore, the algorithm developed by Skilling set up a genuine way to estimate Z doing a numerical integration, whereas the sampling of the posterior p is a by-product of this calculation.

In 1-dimension, the algorithm works as follows: in the first step, we generate a starting set S_0 of n samples uniformly distributed over the space and allowed by the prior Π. Next, we delete the lowest likelihood sample $\mathcal{L}_0$ in S_0, and replace it with a new uniform sample with a higher likelihood $\mathcal{L}_1 > \mathcal{L}_0$ (this is called a hard constraint on likelihood value), moving to step S_1.

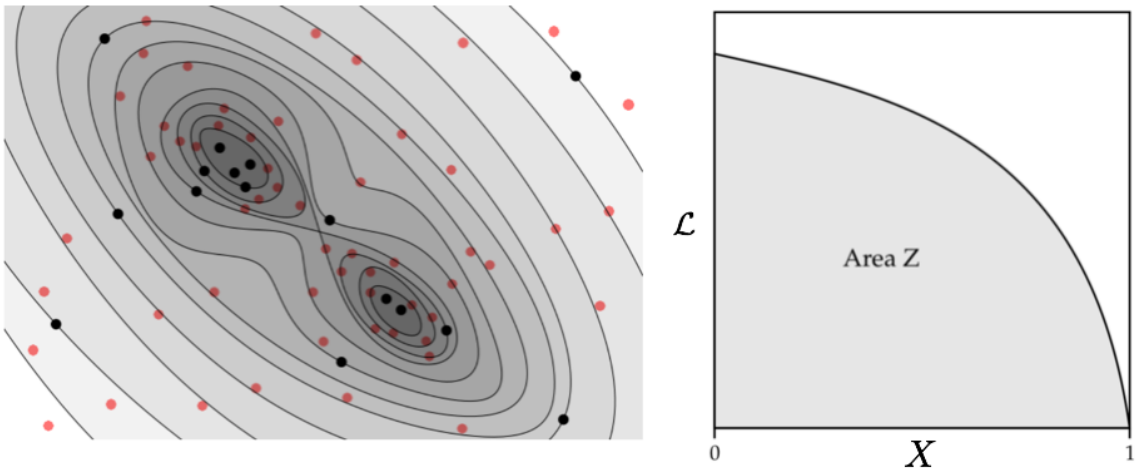

Figure 1.8: *Left*: Graphical interpretation of the Nested Sampling algorithm by looking at the parameter space: the contours specify the areas of the search after each step, where the point with the lowest likelihood was discarded. In every step, the contours are reduced, moving closer to the regions of higher likelihood and finding different likelihood clusters (separated contour regions with high likelihood). This figure is the result of an animation from *W. Handley* lecture at (Lesgourgues, 2018). *Right*: Likelihood function $\mathcal{L}$ as a function of the parameter $X = \int d\theta \Pi(\theta)$, showing how the area below the curve is the evidence Z. Adapted from (Skilling, 2006).

The live-evidence, Z_{live}, is related to the live points X_{live} and their corresponding likelihood $\mathcal{L}_{live}$. The live points X_{live} are the survivor points of the n+1 sample S_{n+1} (see left panel of figure 1.8, black points). The live evidence is approximately equal

to $Z_{live} \approx \langle \mathcal{L}_{live} \rangle X_{live}$ (the value of the likelihood is similar when we are reaching convergence, so we can approximate the integral as the average value of the likelihood of every point times the number of points). The last process is repeated until the live evidence, Z_{live} is a small fraction of the total evidence Z. The algorithm needs a precise value for this fraction, given by the user, to determine if convergence is achieved or not. The value of this fraction is what we call the "stopping criterion". The set S of n samples is constantly updated after every step and the set of dead points (the erased points with low likelihood) with an appropriate weighting factor are the posterior samples (see left panel of figure 1.8, red points). This procedure can be generalized to multiple dimensions (Skilling, 2006). The points chosen randomly from the region $\mathcal{L}(X)$ are representative of the posterior, and therefore, the samples of the posterior distribution p are obtained by assigning weights w_j to the discarded points so that:

$$p(\theta) \approx \frac{w_j \mathcal{L}(\theta)}{Z}. \tag{1.76}$$

1.7.3 Model Comparison

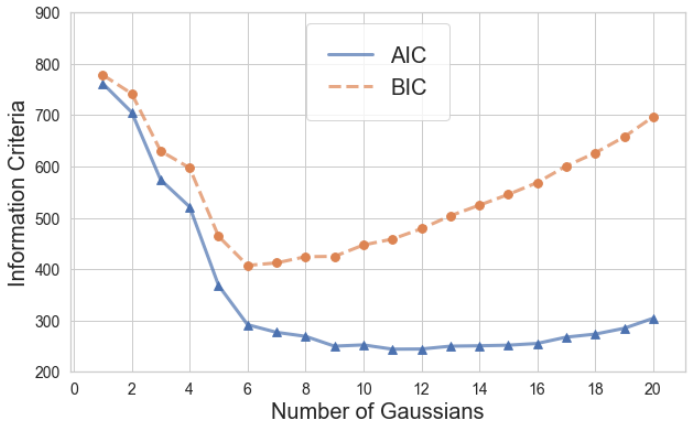

Figure 1.9: Bayesian Information Criteria: proof-of-concept. The data set `make_moons` from the `python` package `sklearn` has been fitted using the Gaussian Mixture model, also available in `sklearn`. The horizontal axis represents the number of Gaussian functions used to fit the data (more Gaussian functions represent an increment in the number of parameters k). The vertical axis shows both the *Akaike Information Criterion* (AIC) and the *Bayesian Information Criterion* (BIC). This example shows the asymptotic behaviour of the AIC (blue triangles) and the growing tendency for the BIC (orange circles), after some local minima. We conclude that the most statistical convenient number of Gaussian functions is 6 (according to the BIC) and 9 (according to the AIC, although after 6 the gain is not really significant) because, after those local minima, both criteria remain either stable (AIC) or increase (BIC). Example inspired by the documentation of `sklearn` (*Scikit-learn 1.1.1 documentation: Gaussian Mixture Model Selection*, 2022).

In many different scenarios, we are interested in knowing whether a model M_1 is statistically favoured with respect to a different model M_2. In Bayesian statistics, this is usually computed by calculating the *Bayes Factor*, which is inferred from applying model comparison to Bayes' theorem (Skilling, 2006):

$$B = \frac{P(M_1|d)}{P(M_2|d)} = \frac{Z(d|M_1)\Pi(M_1)}{Z(d|M_2)\Pi(M_2)}, \tag{1.77}$$

where $P(M|d)$ are the posterior probability distributions integrated over all the parameter space Θ. If our sampling algorithm allows us to calculate evidences Z, for example with the Nested Sampling algorithm, we can compute B directly by taking the ratio of the evidences and priors (in some circumstances when the priors are uniform, the ratio of priors $\Pi(M)$ is assumed to be close to unity). If B is larger than one, this means that the data support the model M_1 statistically better in comparison to model M_2.

The numerical integration to obtain $P(M|d)$ may be infeasible, because, in general, we might need a large number of samples. For this reason, alternatives to the Bayes Factor are sometimes used, which are defined in terms of the likelihood $\mathcal{L}$ and the complexity of the model, and are called *information criteria*. In general, these methods do not simply choose one model versus a different one by analysing which likelihood is the highest, because a model with more parameters mostly leads to a higher likelihood. In most cases, the information criteria penalize a model according to its number of parameters k. The most popular information criteria are the *Akaike Information Criterion* (AIC) (Akaike, 1974) and the *Bayesian Information Criterion* (BIC) (Stoica & Selen, 2004). The AIC and BIC are defined respectively as

$$AIC = -2\log\mathcal{L} + 2k, \tag{1.78}$$

$$BIC = -2\log\mathcal{L} + k\ln N, \tag{1.79}$$

where N is the number of data points used in the fit. In principle, the model with lower BIC or AIC is favoured. It is worth mentioning that the BIC and AIC tend to show an asymptotic behaviour beyond a certain number of parameters (see figure 1.9), implying that no gaining is obtained when more parameters are included in the model. The quantity $-2\log\mathcal{L}$ is known as the *deviance*[18].

1.8 About this thesis

Since the last decade, Modern Cosmology is increasingly becoming heavily-based on data science, in particular, on Bayesian Statistics. This is the primary methodology and common motto of all the chapters of this thesis. The thesis is divided into three different parts:

[18]In Cosmology, we denote the deviance $-2\log\mathcal{L}$ with the symbol χ^2. This is a language misuse, probably because the deviance follow a χ^2 probability distribution.

- The first part is focused on data science and inflation, pursuing to constrain inflationary models using advanced inference techniques and forecasting tools. **Chapter 2** shows the first-ever results of the reconstruction of the inflaton's speed of sound using the latest Cosmic Microwave Background (CMB) data from Planck 2018 and modern algorithms (Gaussian Processes). **Chapter 3** is dedicated to the forecast of a particular class of single-field inflation models, known as α-attractor, for a future CMB stage-IV experiment using a model-dependent alternative approach for the sampling of the inflationary parameters based on current constraints obtained by CMB and LSS data. These two chapters are based on references (Cañas-Herrera, Torrado, & Achúcarro, 2021) and (Cañas-Herrera & Renzi, 2021), respectively.

- The second part of the thesis is dedicated to the novel interest in the cross-correlations of Gravitational-Wave (GW) physics and Large Scale Structure observables, concretely, Galaxy Clustering. In particular, we study how we could exploit the information contained in these new observables by forecasting their behaviour and possible detection using future experimental set-ups and statistics. **Chapter 4** focuses on the study of unresolved GW events that form the Astrophysical Gravitational-Wave background, and how we can use the cross-correlation of the anisotropies of that background with Galaxy Clustering to extract both astrophysical and cosmological information. This work is based on (Cañas-Herrera, Contigiani, & Vardanyan, 2020). On the other hand, in **Chapter 5** we investigate how machine learning techniques can be used to reconstruct the propagation of tensor perturbations by combining the spatial correlation between resolved GW mergers and galaxies. This chapter is based on (Cañas-Herrera, Contigiani, & Vardanyan, 2021).

- Finally, the third part of this thesis is dedicated to the *Euclid* mission and the current work developed by the author of this thesis as a member of the *Euclid Consortium*. **Chapter 6** focuses on a crucial data science tool for the mission: the code *Cosmological Likelihood for Observables in Euclid* (`CLOE`). In this chapter, the implemented cosmological recipe, as well as the description of the *Euclid* likelihood, are depicted. The results concerning the structure of `CLOE` and its performance to constrain cosmological parameters are a novel preview of the content of the *Key Project papers* of the *Euclid* Science Working Group Inter-Science Taskforce Likelihood (IST:L).

Part I

Constraints on inflationary models

Chapter 2

Bayesian reconstruction of the inflaton's speed of sound using CMB data

This chapter is based on:
Bayesian reconstruction of the inflaton's speed of sound using CMB data
Guadalupe Cañas-Herrera, Jesús Torrado Cacho and Ana Achúcarro
(2021, June), Physical Review D 103, 12, 123531, arXiv:2012.04640.

2.1 Introduction

The standard cosmological model (ΛCDM) is currently favored by the available data. It assumes that primordial fluctuations are Gaussian and defined by an almost scale-invariant primordial power spectrum (see Chapter 1). These assumptions do not point to any particular origin, although the simplest inflationary model, canonical slow-roll single-field inflation, naturally predicts them. By contrast, other models of inflation predict deviations from the near scale-invariant spectrum in the form of features. If ever detected, they would open a new window of research in the field of primordial dynamics. See i.e: (Chluba et al., 2015; X. Chen, 2010; Slosar et al., 2019; Palma, 2015).

The study of features of primordial origin can be done within an Effective Field Theory approach. Within this scenario, features can be produced by the time dependence of primordial functions such as the slow-roll parameters or the speed of sound of the effective inflaton (the adiabatic mode). In particular, small, soft and transient reductions in the inflaton's speed of sound produce such correlated localized oscillatory features in the n-point correlation functions. In the 3-point function (or bispectrum), these localized oscillations present a distinct difference in phase between the squeezed and equilateral configurations (Achúcarro, Gong, Palma, & Patil, 2013).

The Planck Collaboration (Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020b) searched for deviations of the canonical scenario in its last release of data. Nevertheless, they did not find strong evidence in the context of features in

the primordial power spectrum (Planck Collaboration, Akrami, et al., 2020). They included, for the first time, a joint search of correlated simple features in the primordial power spectrum and in the bispectrum, also without significant results. However, the Planck Collaboration has not studied in detail different feature templates such as the above mentioned ones due to small and transient reductions of the inflaton's speed of sound. This motivates us to continue our previous study (Achúcarro, Atal, Ortiz, & Torrado, 2014; Achúcarro, Atal, Hu, Ortiz, & Torrado, 2014; Hu & Torrado, 2015; Torrado, Hu, & Achúcarro, 2017), in preparation for a future release of the Planck bispectrum likelihood or for future investigation in light of incoming Large Scale Structure surveys.

Most of the time, the study of features in both observables (primordial power spectrum and higher correlation functions) are model-dependent, both regarding their physical origin and the ansatz used. In our latest paper in this series (Torrado et al., 2017), we already pointed out the need for testing more flexible feature templates to mitigate the dependence on the ansatz. Within this approach, we can test whether multiple and consecutive reductions of the inflaton's speed of sound can take place consecutively, a possibility already pointed out in the previous work (Torrado et al., 2017). Furthermore, reconstructing the inflaton's speed of sound allows us to test more complex feature templates with variable amplitude and oscillation frequency, which implies more possibilities to fit well-motivated deviations from ΛCDM beyond those that only used a pre-defined ansatz for the features.

Reconstructions at the level of the primordial power spectrum have already been attempted (Planck Collaboration et al., 2016; Hazra, Shafieloo, & Souradeep, 2013; Hunt & Sarkar, 2014; Ravenni, Verde, & Cuesta, 2016; Durakovic, Hunt, Mukherjee, Sarkar, & Souradeep, 2018; Ballardini, 2019; Durakovic, Hunt, Patil, & Sarkar, 2019; W. J. Handley, Lasenby, Peiris, & Hobson, 2019). However, there is not enough constraining power in Planck's power spectrum alone to decide on a particular model for the features. Model-informed reconstructions have the advantage of increasing the constraining power by adding the information contained in higher-order correlation functions; but this is only possible if the constraints of the theoretical model are properly imposed on the reconstructed spectrum, so that it will always lead to a consistent prediction. The task of imposing these physical constraints along the reconstruction is non trivial (Appleby, Gong, Hazra, Shafieloo, & Sypsas, 2016). It is advisable, instead, to reconstruct the primordial dynamics directly. In our case, we reconstruct the inflaton's speed of sound: the timing, intensity and rate of its reduction. Since we are reconstructing the underlying function leading to the correlated features, it is not only guaranteed that we will obtain a consistent bispectrum feature prediction using power spectrum data alone, but we will also be able to use both data sets simultaneously to get a more stringent reconstruction once a bispectrum likelihood has been released.

In this chapter, we develop a new analysis pipeline that uses Gaussian Processes (GPs), a hyper-parametric regression technique, to model the inflaton's speed of sound profile. The analytic nature of GPs makes easy to impose the constraints of the theoretical model, which involve derivatives of the reconstructed function. For a given number of nodes in the GP, we construct a prior on the hyper-parameters of

the GP model (the position of the nodes and the correlation length), that maximizes entropy with respect to the bare physical constraints. In this way, we verify that nodes are not placed wherever they would lead to an unphysical reconstruction, and that the density with which the hyper-parameters are explored reproduces the measure of the physical prior (W. Handley & Millea, 2019; Gariazzo & Mena, 2019; Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020b).

We test our new pipeline against Planck 2018 temperature, polarization and lensing CMB angular power spectrum data, obtaining corresponding posteriors of the parameters of interest and several *maxima a posteriori*. Our results do not only reproduce our previous findings (Torrado et al., 2017), but allow for combinations of multiple consecutive reductions as well as more complex shapes.

This chapter is organized as follows. In section 2.2 we review the theoretical framework for inflationary correlated features in the primordial power spectrum due to transient reductions in the speed of sound. In section 2.3, we explain the methodology used to generate features in the primordial power spectrum: the parametrization for the reduction in the speed of sound (2.3.1), the chosen priors for the the different parameters (2.3.2) and the computational procedure (2.3.3). In section 5.5, we present the results corresponding to the fitting of features using the CMB angular power spectrum. Finally, we discuss the results, draw our conclusion and show prospective work for the future in section 2.5.

2.2 Theoretical model

We follow the Effective Field Theory (EFT) of inflationary perturbations (Cheung et al., 2008) to characterize the fluctuations of comoving curvature perturbations around an inflating cosmological background. It starts with an effective action for the Goldstone boson of cosmic time diffeomorphisms $\pi(t, \mathbf{x})$. This Goldstone boson is related to the comoving curvature perturbation $\mathcal{R}(t, \mathbf{x})$ through the relation $\mathcal{R} \approx -H(t)\pi(t, \mathbf{x})$, with the Hubble parameter $H(t) \equiv \dot{a}/a$, with a being the scale factor (where the dot denotes derivatives with respect to cosmic time t). The effective single field action for π up to second order is given by

$$S_2 = \int d^4x a^3 M_P^2 \epsilon_1 H^2 \left[-\frac{\dot{\pi}^2}{c_s^2} + \frac{(\partial_i \pi)^2}{a^2} \right], \tag{2.1}$$

where $M_P = 1/\sqrt{8\pi G}$ is the reduced Planck Mass in natural units $c = \hbar = 1$, $\epsilon_1 \equiv -\dot{H}/H^2$ is the first slow-roll parameter and $c_s \equiv c_s(t)$ is the time-dependent speed of sound.

The effective single field action up to third order, neglecting higher order slow-roll corrections ($\sim \mathcal{O}(\epsilon_1^2)$) and assuming $\dot{\pi}^3$ to be small and approximately constant reads

(Achúcarro, Gong, et al., 2012),

$$S_3 = \int \Big[d^4x a^3 M_P^2 \epsilon_1 H^2 \\ - 2Hsc_s^{-2}\pi\dot{\pi}^2 - (1-c_s^{-2})\dot{\pi}\left(\frac{\dot{\pi}^2}{c_s^2} - \frac{(\partial_i\pi)^2}{a^2}\right)\Big], \tag{2.2}$$

where $s \equiv s(t)$ parameterizes the change in the speed of sound $c_s(t)$ defined as

$$s \equiv \frac{\dot{c_s}(t)}{(c_s(t)H)}. \tag{2.3}$$

The physical details of the theory are encoded in the speed of sound c_s and in its corresponding rate of change given by s. The speed of sound c_s accounts for the effects of integrating out the heavy fields within the effective action. To get an insight of what this variable $c_s(t)$ means, we look at the particular case of an effective theory for the comoving curvature perturbation $\mathcal{R}$, when a strong turn in the inflationary trajectory in multifield space is supported by a heavy field $\mathcal{F}$ with "effective mass" $M_{\rm eff}$. In this case, the curvature perturbation $\mathcal{R}$ is kinetically coupled to the heavy field $\mathcal{F}$. This effective action is similar to the EFT of inflation equation (2.1), with the speed of sound c_s of the adiabatic perturbation $\mathcal{R}$ given by (Achúcarro, Gong, Hardeman, Palma, & Patil, 2011a, 2011b)

$$c_s^{-2} = 1 + \frac{4\Omega^2}{k^2/a^2 + M_{\rm eff}}, \tag{2.4}$$

where Ω is the the angular velocity when there is a turn in the inflationary trajectory, inducing a momentary reduction on the speed of sound c_s (Achúcarro, Gong, et al., 2012). The effect of this variable speed of sound c_s can be seen in the primordial power spectrum $\mathcal{P}_\mathcal{R}$, in the bispectrum $\mathcal{B}_\mathcal{R}$ and in higher-order correlation functions. In particular, transient variations of c_s produce localized oscillatory and correlated features in both $\mathcal{P}_\mathcal{R}$ and $\mathcal{B}_\mathcal{R}$ (Achúcarro, Atal, et al., 2012). Generally, $c_s(t)$ encodes the effect of derivative interactions.

Under the assumption of small, mild and transient reductions of the speed of sound c_s, the modifications in the primordial power spectrum of curvature perturbations with respect to the almost-invariant power law spectrum, $\Delta\mathcal{P}_\mathcal{R}/\mathcal{P}_{\mathcal{R}0}$, were already calculated (Achúcarro et al., 2013). The quadratic action of EFT of inflation, equation (2.1), is divided into a free part (resembling single field inflation, with $c_s = 1$) and a small perturbation:

$$S_2 = \int \mathrm{d}^4x\, a^3 M_{\rm P}^2 \epsilon H^2 \left(\dot{\pi}^2 - \frac{(\partial_i\pi)^2}{a^2}\right) \\ - \int \mathrm{d}^4x\, a^3 M_{\rm P}^2 \epsilon H^2 \Big(\left(1 - c_s^{-2}\right)\dot{\pi}^2 \Big), \tag{2.5}$$

Transitioning from cosmic time t to the conformal time τ, so that $d\tau = dt/a(t)$, using the *in-in* formalism (Weinberg, 2005) and the following definition of the variable u,

$$u(\tau) \equiv (1 - c_s^{-2}(\tau)), \tag{2.6}$$

the change in the primordial power spectrum $\Delta\mathcal{P}_\mathcal{R}$ is given by the Fourier transform of the reduction in the speed of sound c_s:

$$\frac{\Delta\mathcal{P}_\mathcal{R}}{\mathcal{P}_{\mathcal{R}0}} = k\int_{-\infty}^{0} d\tau u(\tau)\sin(2k\tau), \tag{2.7}$$

where $\mathcal{P}_{\mathcal{R}0}$ is the power law expressed in equation (1.39).

2.3 Methodology

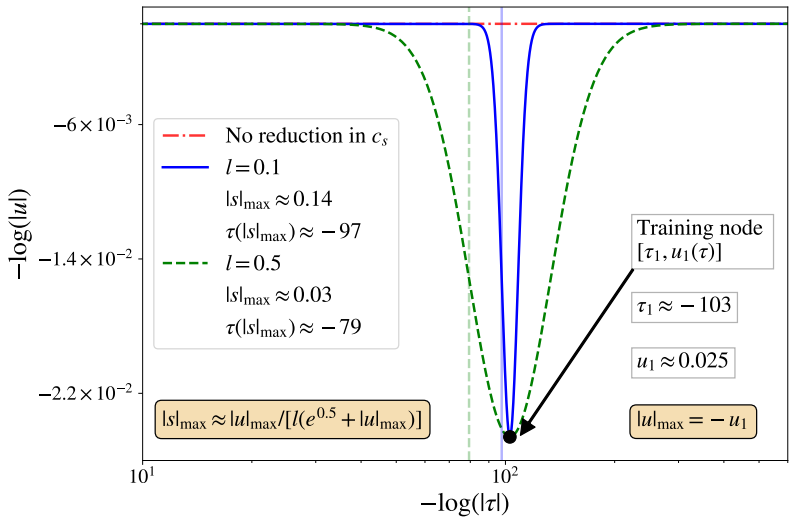

Figure 2.1: Example of reductions of the inflaton's speed of sound $u(\tau)$ in logarithmic space GP with a single node at (τ_1, u_1). The reductions peak at $\tau_1 \approx 103$ with a maximum reduction value $-u_1 = |u|_{\max}\approx 0.025$. The width of each of the reductions (given by the correlation length l, which, for a single GP node, plays the role of the standard deviation) is different, being the green-dashed parametrization ($l = 0.5$) milder than the solid blue one ($l = 0.1$). The rate of change of $u(\tau)$, see equation (2.3), can be approximated as $|s|_{\max}= |u|_{\max}/[l(e^{0.5} + |u|_{\max})]$ using this parametrization. The vertical lines indicate the values of τ at which $|s|_{\max}$ is reached in each case.

The reduction of the speed of sound and its rate of change are encoded in $u(\tau)$ and $s(\tau)$ respectively. We aim to use current cosmological data (the temperature, polarization and lensing power spectrum of Planck 2018) to estimate them given the theoretical framework presented in section 2.2. To do that, we use *Bayesian inference*, as explained Chapter 1.

2.3.1 Reconstruction model for the reduction in the speed of sound

We use Gaussian Processes (GPs) (Rasmussen & Williams, 2005) as an interpolator for reconstructing the speed of sound of the inflaton. The *mean* curve of a GP,

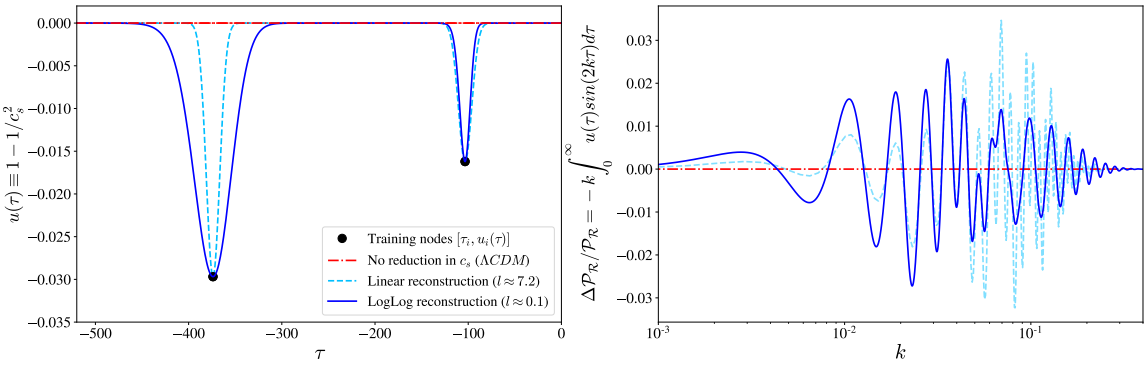

Figure 2.2: **Left panel**: Reconstruction of $u(\tau)$ using a GP on $\log|u(\log|\tau|)|$ (continuous line) and on $u(\tau)$ (dashed line), both using as training nodes $(\tau, u) \in \{(-100, -0.016), (-375, -0.03)\}$, and correlation length for each model such that the width of the mode corresponding to the first node is similar (notice the difference in width of the mode of the second one). As the reconstruction is done in conformal time τ, the x-axis is always negative and $\tau = 0$ indicates the end of inflation. **Right panel**: Corresponding feature in the primordial power spectrum. Notice how, despite the similar position of the training nodes, their features look quite different: similar in the leftmost oscillations (corresponding to the node at $\tau = -100$), but very different after that, due to the broader width of the first mode.

which we use to represent the speed of sound evolution, is smooth by construction (given our choice of kernel, see below) and naturally returns to a baseline value away from the *nodes* of the interpolator, which is useful for representing the transient character of the speed of sound reductions. The length scale over which the return to the baseline happens is called *correlation length*, and it is the same for all individual nodes in a particular realisation. The particular properties of the correlation between nodes is given by the *kernel* of the GP, which we choose to be a *squared exponential* kernel. This means that when the interpolator is defined by a single node placed at (x_1, y_1), the interpolating curve looks like a (non-normalised) Gaussian peaking at the node's position, with standard deviation equal to the correlation length l, i.e. $y(x) = y_1 \ \exp\left[-1/2 \ (x - x_1)^2/l^2\right]$.

We aim to reconstruct $u(\tau) \equiv (1 - c_s^{-2}(\tau))$. Since $u(\tau)$ is a negative quantity, it makes sense to reconstruct the logarithm of $-u = |u|$, to guarantee that the GP interpolator, once exponentiated, conserves sign. On the other hand, there is a choice to be made about the scale of the conformal time axis: whether to reconstruct $\log|u(\tau)|$ or $\log|u(\log|\tau|)|$. We show results for the latter case in this section for illustration purposes.

As explained above, in the case where a single node is placed at $(\tau_1, \log(-u_1))$ the reconstruction of $\log(-u(\tau))$ corresponds to a single transient reduction given by a log-normal function of conformal time, whose maximum occurs exactly at the node. The parameters that we would try to infer from the data would be the position of the node $(\tau_1, \log(-u_1))$ and the correlation length l representing the standard deviation of the log-normal. The rate of change of the reduction, of interest in our

theoretical framework, would peak approximately at $|s|_{\max} \approx 0.5l|u_1|/(\exp(0.5)+|u_1|)$. See Figure 2.1 for an example.

To reconstruct $\log|u(\log|\tau|)|$ as a generalization of the previous case using GPs, we choose a number i of training nodes[1] $(\log|\tau_i|, \log|u_i|)$ where $\log|u_i| = \log|u(\log|\tau_i|)|$ (see Figure 2.2), to which we fit a GP with kernel function

$$\kappa(\log|\tau_i|, \log|\tau_{i+1}|;\, l) = c^2 \exp\left\{-\frac{1}{2}\left(\frac{\log|\tau_i| - \log|\tau_{i+1}|}{l}\right)^2\right\}, \tag{2.8}$$

where c is the *output scale*, and l the *correlation length*. The output scale c plays no role in this approach, and can be fitted using *maximum likelihood* and then ignored. The correlation length will be sampled together with the position of the nodes. To compute the GPs, we use the Python package `sklearn` (*Scikit-learn 0.19.1 documentation: Gaussian Processes*, 2018). The *mean* of the GP is used as an interpolator for $\log|u(\log|\tau|)|$, and reads, in terms of the training nodes, as the matrix product:

$$\log|u(\log|\tau|)| = \kappa(\log|\tau|, \log|\tau_i|;\, l) \times [\kappa(\log|\tau_i|, \log|\tau_j|;\, l)]^{-1} \log|u_j|, \tag{2.9}$$

where the first kernel function κ is a vector of evaluations at the requested $\log|\tau|$ combined with each of the training $\log|\tau_i|$, the second one is the matrix of evaluations of κ for each pair of training nodes (i, j), and the final term is the vector of training $\log|u_j|$. Once $u(\tau)$ is generated, we calculate $s(\tau)$ from equation (2.3) numerically, which we can rewrite more conveniently as

$$\begin{aligned} s(\tau) &= \frac{1}{2}\frac{u}{1-u}\frac{d\log u}{d\log|\tau|} \\ &= \frac{1}{2}\frac{u}{u-1}\frac{1}{l^2}\left[(\log|\tau| - \log|\tau_i|)\,\kappa(\log|\tau|, \log|\tau_i|;\, l)\right] \times \\ &\quad [\kappa(\log|\tau_i|, \log|\tau_j|;\, l)]^{-1} \log|u_j|, \end{aligned} \tag{2.10}$$

where we have taken the derivative after substituting u by the mean of the GP defined in equation (2.9). Notice that this reproduces the matrix product in equation (2.9), just changing the first vector. Finally, we compute the power spectrum feature of equation (2.7) from a fine sampling of the GP using the `FFTLog` algorithm (Hamilton, 2000; *fftlog - python wrapper FFTLog*, 2016). The density and limits of the $\log|\tau|$ sampling for the `FFTLog` are chosen adaptively to minimise computational costs and guarantee the accurate computation of the transform.

The most consequential difference of the choice between linear and logarithmic τ in the GP will show up whenever we have nodes separated by a distance much larger than the correlation length, appearing as isolated (log)Gaussians: in the linear case, their width in τ will be similar, whereas for the logarithmic one, the width will scale logarithmically (see Figure 2.2). Looking at the first equality in equation (2.10), and seeing how s depends on the logarithmic derivative on τ, it is easy to see that the

[1] Notice that our use of GPs as interpolators does not involve machine-learning, but we are borrowing the term *training* from its literature.

linear parameterization is going to struggle to place two or more nodes away from each other, since $s(\tau)$ will peak at highly different values in each of them, making it hard not to violate the perturbativity bounds on s (see subsection 2.3.2).

Thus for the primary results in this chapter, we model $\log|u(\log|\tau|)|$. To mitigate excessive sensitivity to the prior of our results, we also perform a reconstruction in $u(\tau)$. Notice that by modelling u and not $\log|u|$ we need to deal with cases in which $u(\tau)$ goes positive, by assigning it null prior density. However, those are generally disfavored by the data (require large l compared to the distance between nodes), and the large difference between the $\log|u(\log|\tau|)|$ and the $u(\tau)$ reconstruction is useful for assessing prior sensitivity.

2.3.2 Parameters and Priors

The action described in equations (2.1) and (2.2) is perturbative in terms of $(1 - 1/c_s^2)$. It implies that the reduction in the speed of sound, c_s, cannot be too big ($|u| \ll 1$) and the rate of change in the reduction cannot be too fast ($|s| \ll 1$). Also, the contributions of the slow-roll corrections ϵ_1, ϵ_2 have to be smaller than those of the variable speed of sound c_s. We need to impose these conditions for all values of τ, but it is enough to restrict to the point where $u(\tau)$ and $s(\tau)$ take their maximum value $(|u|_{\max}, |s|_{\max})$. Note that imposing the perturbative limit on $|s|_{\max}$ satisfies the consistency conditions in (Céspedes, Atal, & Palma, 2012; Adshead & Hu, 2014; Cannone, Bartolo, & Matarrese, 2014). In short:

$$\max(\epsilon_1, \epsilon_2) \ll \max(|u|_{\max}, |s|_{\max}) \ll 1. \tag{2.11}$$

In (Torrado et al., 2017) we argued that this condition could be naturally imposed by a prior Beta(5, 5) on $\max(\log_{10}|u|_{\max}, \log_{10}|s|_{\max})$ between the extremes in equation (2.11) (see left part of Figure 2.3), the logarithm coming from the difference in order of magnitude between both bounds.

Contrary to (Torrado et al., 2017), in this work $|u|_{\max}$ and $|s|_{\max}$ are not sampled directly. Instead, the parameter space for the feature consists of the position of the nodes $\{(\tau_i, u_i = u(\tau_i))\}$ and the correlation length l. Thus, the total number of feature parameters is $2N + 1$ for a number N of nodes. Imposing the Beta prior described above is not as simple as sampling the GP parameters from some prior, computing $|u|_{\max}$ and $|s|_{\max}$ along the reconstruction, and multiplying by the Beta density. That procedure will likely introduce undesired information in the shape of under- or overdensities in the probability induced on $(|u|_{\max}, |s|_{\max})$, which would finally diverge significantly from a Beta. The correct way to proceed so that the induced probability on the physical parameters is the desired one is by constructing a distribution on the parameters of the nodes that *maximises entropy* with respect to the desired one, which can be computed, according to (W. Handley & Millea, 2019), as

$$\pi_{\text{pert, maxent}}(\tau_i, u_i, l) = \frac{\pi_0(\tau_i, u_i, l)\pi_{\text{pert}}(|u|_{\max}, |s|_{\max})}{P(|u|_{\max}, |s|_{\max} \mid \pi_0)}, \tag{2.12}$$

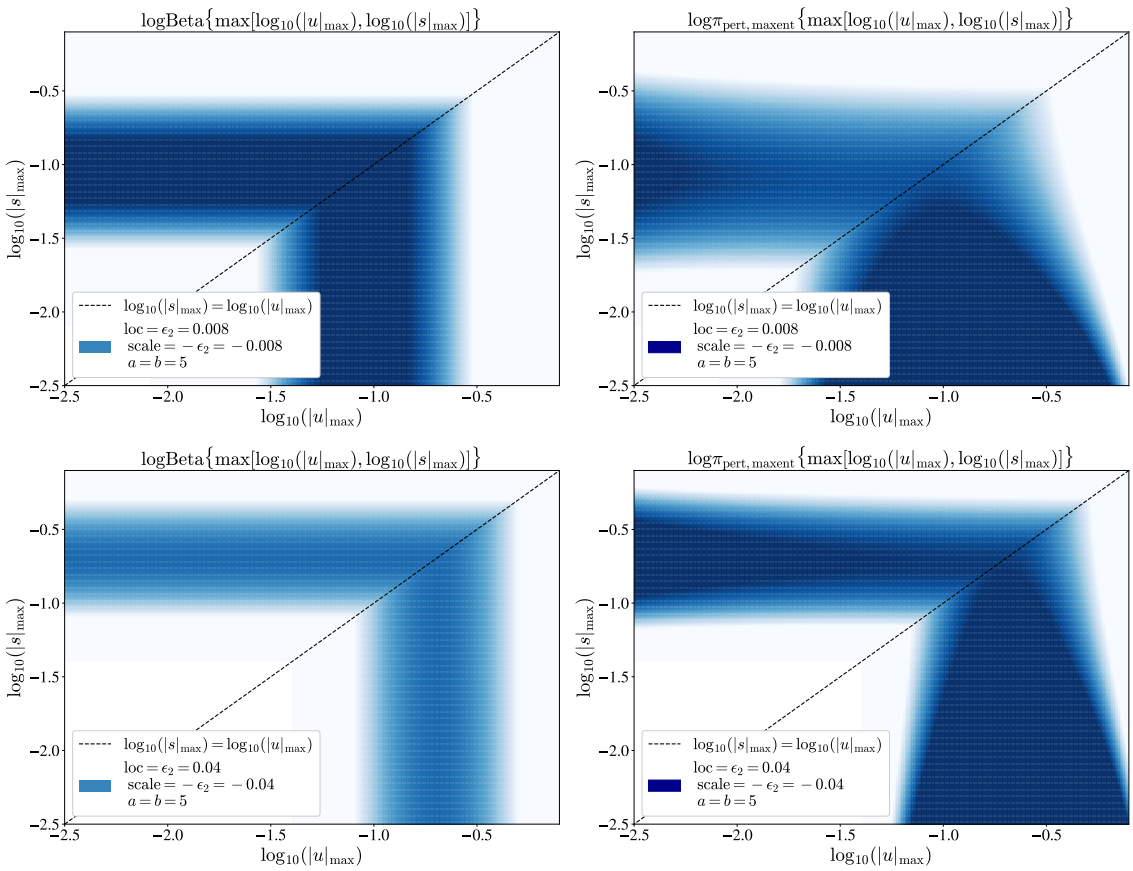

Figure 2.3: A description of the Bayesian priors adopted in this study. **Left panels**: Desired Beta density distribution for the parameters $log_{10}(|u|_{\max})$ and $log_{10}(|s|_{\max})$, with $\epsilon_2 = 0.008$ and $\epsilon_2 = 0.04$, respectively. **Right panels**: probability density distribution induced by π_0 on $log_{10}(|s|_{\max})$ and $log_{10}(|u|_{\max})$ (defined as P in the denominator of equation (2.12)), for a Gaussian-Processes reconstruction in $\log|u(\log|\tau|)|$ using 2 training nodes. The initial prior $\pi_0(\tau_i, u_i, l)$ induces overdensities at low $log_{10}(|s|_{\max})$ due to this quantity being doubly-correlated to early-time position of the nodes and large correlation lengths; equation (2.12) corrects for this effect.

where $\pi_0(\tau_i, u_i, l)$ is some initial prior on the node parameters and $\pi_{\rm pert}(|u|_{\max}, |s|_{\max})$ the Beta prior described above. The term in the denominator is the probability density induced by π_o on $(|u|_{\max}, |s|_{\max})$, which we compute from a Monte Carlo sample from π_0 using `PolyChord` (W. J. Handley et al., 2015a, 2015b). The Monte Carlo sample is fed to `GetDist` (Lewis, 2019) to construct a density estimator. Both the Beta prior and the maxent prior can be seen in Figure 2.3.

For π_0, we choose log-uniform prior distributions on the correlation length l and on the training node location (τ_i, u_i)[2]. The bounds for the time-positions τ_i are chosen so that the feature falls in the CMB window function (features running from scales

[2]Note that we are sampling the training nodes and the correlation length in a logarithmic scale as we expect them to vary several orders of magnitude.

$k \approx 10^{-3}$ to $k \approx 3 \times 10^{-1}$, though a larger region has been scanned as a consistency check (see subsection 2.4.1). The bounds for the amplitude of the reductions at the nodes, u_i, are chosen to generously fulfill the EFT condition in equation (2.11). Summarizing:

$$\begin{aligned}\pi_0(\log_{10}(|\tau_1|), \ldots, \log_{10}(|\tau_n|), \log_{10}(|u_1|), \ldots, \\ \log_{10}(|u_n|), \log_{10} l) = \mathcal{U}(1.8 < \log_{10}(|\tau_n|) < 3.3)\times \\ \prod_{i=n}^{2} \mathcal{U}(\log_{10}(|\tau_i|) < \log_{10}(|\tau_{i-1}|) < 3.3)\times \\ \prod_{i=1}^{n} \mathcal{U}(-4 < \log_{10}(|u_i|) < 0) \times \pi(l)\end{aligned} \tag{2.13}$$

where $\mathcal{U}$ means a uniform distribution, and the prior on the time positions of the nodes includes sorting so that $\tau_i < \tau_{i-1}$ (i runs from $1 \ldots n$). The prior on l is chosen so that it produces reasonable values of $|s|_{\max}$. For each of the two reconstructions studied here, $\log|u(\log|\tau|)|$ and $u(\tau)$, the boundaries can be chosen as[3]

$$\begin{aligned}&\pi_{\log|u(\log|\tau|)|}(l) = \mathcal{U}(-2 < \log_{10} l < 2) \qquad \text{and} \\ &\pi_{u(\tau)}(l) = \mathcal{U}(-2 < \log_{10} l < 3.3).\end{aligned} \tag{2.14}$$

As an improvement on (Torrado et al., 2017), in this work we do not fix the value of the slow-roll parameters ϵ_1 and ϵ_2 as bounds of the perturbativity condition equation (2.11). Instead we let the bounds of the Beta distribution run dynamically, marginalizing over the slow-roll parameters. We use as priors uniform distributions $\mathcal{U}(0.0001 < \epsilon_1 < 0.05)$ and $\mathcal{U}(-0.06 < \epsilon_2 < 0.06)$, which encompass the ΛCDM posterior found for them in Planck 2018 (Planck Collaboration et al., 2016).

A diagram showing the different assumptions that enter the final prior is shown in Figure 2.4.

2.3.3 Data sets and sampler

To constrain the reduction of the speed of sound, we use the Planck 2018 polarized CMB and lensing data. In particular we use the product of the low multipole likelihoods `lowT` and `lowE`, the *unbinned* high-ℓ likelihood `plik_TTTEEE` and the `lensing` likelihood. We use the unbinned likelihoods because of the fast frequency of oscillations in the features, as was already pointed out in (Torrado et al., 2017).

We compute the changes to the CMB power spectra C_ℓ using the Boltzmann code `CAMB` (Lewis, 2013), modified accordingly to account for the increased sampling in k needed by the oscillatory features $\Delta\mathcal{P}_\mathcal{R}/\mathcal{P}_{\mathcal{R}0}$ in the primordial power spectrum. We sample over the parameter space described in section 2.3.2, i.e. the positions of the

[3]Notice that, while smaller values of l will result in sharper reductions with too high, forbidden $|s|_{\max}$ values, larger values of l would result in small $|s|_{\max}$ values which are actually allowed as long as $|u|_{\max}$ fulfills equation (2.11). In any case, we are imposing these upper l boundaries for the main runs, since reductions with very small $|s|_{\max}$ tend not to be easily distinguishable from changes in the background cosmological model (see subsection 2.4.1).

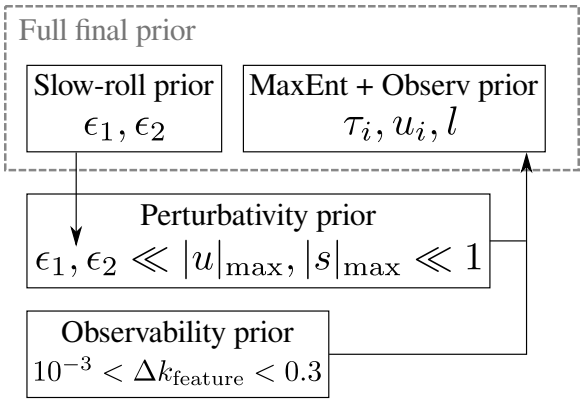

Figure 2.4: Diagram showing the structure of the prior. The conformal time τ_i of the nodes must fulfil that the feature happens within the observable CMB window. The physically-motivated perturbativity condition of equation (2.11) is imposed using *maximum entropy* (see text) on the position of the nodes and the correlation length. Since the value of the slow-roll parameters influences the prior on the position of the nodes via the lower bound of the perturbativity condition, the full prior (dashed grey box) is non-separable.

training nodes $\{(\tau_i, u_i = u(\tau_i))\}$, the correlation length l of the GP, and the kinetic slow-roll parameters ϵ_1 and ϵ_2. We also allow for the possibility of tensor modes, as changes in the Sachs-Wolfe plateau caused by them could possibly be correlated with features at very large scales. We track as derived parameters the scalar tilt n_s, the tensor-to-scalar ratio r and the EFT parameters $(|u|_{\max}, |s|_{\max})$. We fix the rest of cosmological parameters of ΛCDM to the best fit of Planck 2018 with the present likelihoods, as well as the *nuisance* parameters of the likelihoods. Fixing the ΛCDM parameters is justified by previous sensitivity analyses in (Torrado et al., 2017), that we repeat here for the background ΛCDM parameters by exploring a broader range of τ and l than the one indicated above (see subsection 2.4.1, where we have also assessed the impact of fixing the nuisance parameters of the Planck likelihoods).

We obtain the posterior distribution of the parameters using the sampler `PolyChord` (W. J. Handley et al., 2015a, 2015b). We use this nested sampler since, from previous searches, we expect the posterior distributions of $u_{\max}$ and τ_i to be multi-modal. The handling of the priors, likelihoods, Boltzmann code and sampler is managed by the Bayesian framework `Cobaya` (Torrado & Lewis, 2021, 2019). The analysis of the posterior distributions is carried out using `GetDist` (Lewis, 2019).

We sample the posterior of two different parameterizations of the GP sound speed reconstruction: $\log|u|(\log|\tau|)$ and $u(\tau)$, in the following called simply *logarithmic* and *linear* parameterizations, respectively. We know the logarithmic parameterization is more stable numerically, as it consistently makes the reconstruction of $u(\tau)$ negative. However, we still use the runs in the linear parameterization for the purposes of assessing prior sensitivity. For the first sampling processes (up to three GPs nodes), we run `Cobaya` in parallel launching 8 MPI processes, each allowed to thread across 3 CPU cores. In the case of 4 nodes, we run `Cobaya` with 32 MPI processes, each

allowed to thread across only one single CPU core. The nested sampler `PolyChord` has been run with 1000 live points (which is far above the requirements for the current number of dimensions in the parameter space) and a stopping criterion of 0.01. The computation time varies depending on the number of training nodes in the GPs: from a few days with only 1 node, up to several weeks with 4 nodes.

All the *maxima a posteriori* (MAP) presented in the next section have been obtained running `Py-BOBYQA` (Cartis, Fiala, Marteau, & Roberts, 2018; Cartis, Roberts, & Sheridan-Methven, 2018) (a Python implementation of the `BOBYQA` algorithm (Powell, 2009), available via `Cobaya`), initialized on the relevant local maxima of the `PolyChord` samples.

2.4 Results

2.4.1 Consistency checks

Before presenting our results, we shortly discuss whether the assumptions made in previous sections were justified. In particular, we have tested whether we find clear posterior modes outside the (τ_i, l) prior region described section 2.3.2 (the *CMB window* prior), and whether in posterior modes either in our initial prior or in the broader region, the assumption of no-correlation with background cosmological parameters is fulfilled.

To do that, we produced a 1-node posterior sample in the logarithmic parameterization in the enlarged prior region $0 < \log_{10}(|\tau_i|) < 4.3$ and $2 < \log_{10} l < 10$, and let the background ΛCDM parameters vary. No significant modes were found outside the original, reduced prior region. We found mild modes in the region $1.8 < \log_{10}(|\tau_i|) < 3.3$ and $2 < \log_{10} l < 3.3$, which presented some degeneracy between Ω_m, n_s and the reconstruction parameters ($\rho \approx 0.17$), due to the fact that these features can be confused with the shape of the first and second acoustic peaks (already observed in (Torrado et al., 2017)). This justifies restricting ourselves to the prior described in 2.3.2, since any mode found outside of it would not be distinguishable from background cosmology.

The check for degeneracies between the c_s reconstruction parameters and the slow-roll parameters is of particular importance, since the latter determine the perturbative prior limits on the former (see equation 2.11 and Figure 2.3). We have found no significant degeneracies, neither in the tests described above nor in the final runs. We have reproduced the Planck ΛCDM posterior on the slow-roll parameters in all cases (see Figure 2.5).

Most of the results below have been run both in the linear and logarithmic parameterizations for the Gaussian Process (GP) reconstruction of the speed of sound profile $u(\tau)$. The results agree with each other, in particular, for the maxima a posteriori found at late conformal time (i.e. towards the end of inflation, with $-\tau_i$ of a few hundreds, where both parameterizations look similar). However, the logarithmic parameterization differs from the linear one when training nodes are thrown at early values of conformal time (i.e. $-\tau_i$ over 800), see Figure 2.2). This is due to modes

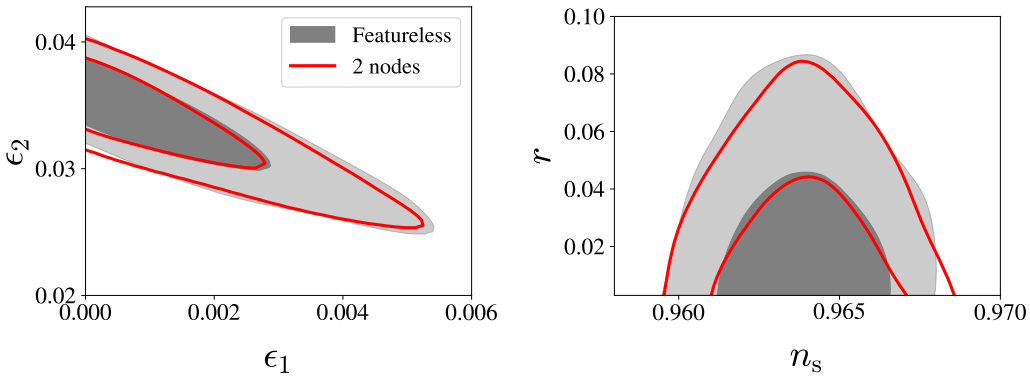

Figure 2.5: Posterior distributions of the primordial parameters: the kinetic slow-roll parameters ϵ_1 and ϵ_2, and the derived spectral index n_s and tensor-to-scalar ratio r, for a 2-node reconstruction in the logarithmic parameterization (red line). The grey contours correspond to the featureless ΛCDM scenario. The correspondence between both posteriors is due to the absence of degeneracies between the c_s reconstruction parameters and the slow-roll parameters. Similar results are found for 1, 3 and 4 nodes.

of constant width in logarithmic scale getting broader the further we go along the axis of conformal time. In the 1-node case, the linear parameterization reproduces the results in (Torrado et al., 2017) (which uses a Gaussian ansatz in $u(\tau)$), whereas the logarithmic parameterization produces different 1-node posterior modes (see Sec. 2.6.1 in appendix A).

It is worth remarking that the use of the logarithmic parametrization does not compromise the flexibility of our reconstruction of $u(\tau)$. Even though the logarithmic parameterization reconstructs naturally profiles of $u(\tau)$ with broad reductions at earlier conformal times and narrower reductions at later times (which is preferred so that $|s|_{\rm max}$ is not violated), narrow reductions at early conformal times can always be achieved by adding further nodes that would force the profile to return to zero. If the data and EFT conditions did allow for a narrow reduction at earlier conformal times, we would have seen it during the analysis of the posterior distributions when more than one training node was used.

Moreover, we have assessed the effect of fixing the nuisance parameters of the Planck likelihoods by running a minimizer around both the baseline, featureless ΛCDM model and the MAP features presented below, now letting the nuisance parameters vary. We find that this choice has almost no impact in our analysis, changing the $\Delta\chi^2$ values by less an unit.

Finally, we have assessed the impact of using separately each of the high-ℓ *unbinned* TT and EE Planck 2018 data sets, in order to check which subset dominates the posterior around each of the fits. To do that, we have combined each of these subsets with low-ℓ temperature and polarisation, and lensing data, fixing LCDM and nuisance parameters, and assuming a single dip. Using high-ℓ TT data alone, we recover the main single-dip maxima a posteriori described in the next section and appendix 2.6

(i.e: $\tau_1 \approx -100, -200, -400, -1000$). When using high-$\ell$ EE data alone, we clearly recover the dip at $\tau_1 \approx -100$, whereas just barely the peaks at $\tau_1 \approx -200$ and $\tau_1 \approx -400$, and none of the earlier-time dips found in combination with TT data. This shows that our posterior is temperature-dominated in most of the parameter range. Concordance at late times is encouraging, but it needs to be explored further with future less noisy polarised data from space- and ground-based surveys.

2.4.2 Reconstruction of the inflaton's speed of sound profile $u(\tau)$

In this section, we present the results of the GP reconstruction in the logarithmic parameterization using the Planck data as described above, and imposing the Maximum-Entropy prior described in 2.3.2 for the derived quantities $|u|_{\max}$ and $|s|_{\max}$.

When presenting our results, we use an effective $\Delta\chi^2$ where we have subtracted the χ^2 of the MAP of the featureless ΛCDM (obtained by using a minimizer) for the same likelihood combination (see subsection 2.3.3). Note that this effective $\Delta\chi^2$ is not meant for model selection purposes and it is used for illustration only. As an example, a triangle plot of the posterior distribution for the 2-nodes case can be seen in Figure 2.6.

We have reconstructed the inflaton's speed of sound profile $u(\tau)$ using up to four training nodes. We have stopped there after checking that the Akaike Information Criterion (AIC) (Akaike, 1974) has a minimum for three training nodes and stabilizes after that. The profile $u(\tau)$ shows different patterns depending on how many training nodes are used in the GP reconstruction. We have decided to classify all possible profiles $u(\tau)$ based on whether they show differentiated and non-overlapping reductions (that we denominate *dips*) or they present some kind of substructure:

- **One single dip**: usually present at either late values of conformal time (-100, -200, -400), or at earlier times (-800, ~ -1000). Early-time dips produce features in the CMB power spectra localised in ℓ's up to the first acoustic peak, whereas features from late-time dips affect the power spectra along the full ℓ range. Similar profiles were already found in previous studies (see subsection 2.4.1). Details on this posterior modes can be found in appendix 2.6.1.

- **Combination of non-overlapping reductions (2, 3 and 4 dips)**: appearing when more than one training node is used, they consist of consecutive, isolated reductions in the speed of sound.[4] Details can be found in appendices 2.6.2, 2.6.3 and 2.6.4. These combinations can be classified as (for details see appendix 2.6):

  - *All dips at late conformal times*: when at least two training nodes are considered, there is a preference for two of the possible dips remaining at

[4]Notice that the number of training nodes is not always equal to the number of dips: reconstructions with m dips found with m GP nodes usually re-appear as posterior modes in the $m+1$ GP nodes case, where one of the nodes is placed at $u_i \approx 0$.

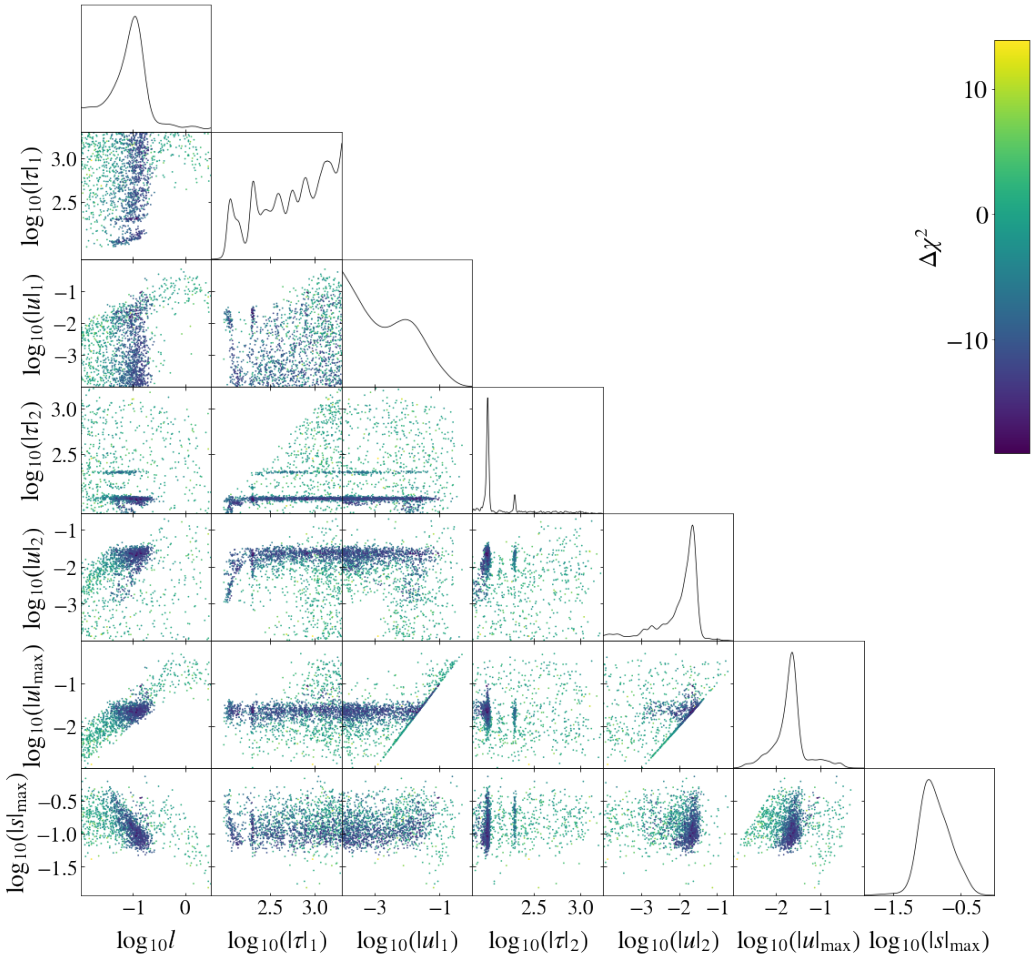

Figure 2.6: Posterior distribution for the reconstruction of the speed of sound's profile $u(\tau)$ using 2 training nodes and the logarithmic parametrization shown in Figure 2.2. We use $\Delta\chi^2 = \chi^2_{\rm model} - \chi^2_{\rm baseline}$ as the variable for the scatter plot's colour scale, the reference $\chi^2_{\rm baseline}$ corresponding to the MAP of the baseline ΛCDM model to the same datasets. We show the parameters of the training nodes (τ_i, u_i) and the correlation length l (described in 2.3.1; priors in 2.3.2). We also show the posteriors of the EFT parameters $(|u|_{\max}, |s|_{\max})$ (described in section 2.2, and not sampled directly, but derived from the nodes parameters). It can be seen how longer correlation lengths (broader reductions) lead to lower values of $|s|_{\max}$, and vice-versa. The posterior distributions for different numbers of nodes display similar patterns. For all cases, the posterior distributions are clearly multi-modal.

late-time values of τ_i, combining either -100 and -200, or -400 and -200. Their effect in the CMB power spectrum overlap each other along a large range of ℓ's.

– *Combination of early- and late-time dips:* these appear typically as a combination of features at both low ℓ's (from the early-time dips) and high ℓ's

(from the late time ones), e.g. from the presence of dips both at -800 and -100.

- **Dips with substructure:** We have found some *maxima a posteriori* where the reconstructed $u(\tau)$ does not show clearly separated reductions, but a more complex profile with some degree of substructure. These substructures are presented either at early and late τ_i, trying to fit some of the characteristic features of the CMB angular power spectrum (i.e: $\ell \approx 20 - 40$ feature). The fits are presented in Appendix A, in subsection 2.6.5.

As noticed in previous works (Torrado et al., 2017), we do not have a highly predictive posterior of the maximum of the rate of change of the sound speed, $|s|_{\rm max}$, whose value is mostly constrained by the prior information. By contrast, the positions of the nodes (the oscillation frequency of the features in the power spectrum) are tightly constrained within each of the multiple posterior modes, specially for nodes at late conformal time.

Using the sampling results of the profile $u(\tau)$ with four training nodes in the GPs, we have reconstructed the allowed confidence contours for $u(\tau)$ given Planck 2018 data. The result can be seen in Figure 5.3. As expected from the 1-D marginalized posterior distributions, the confidence contours are narrower around $\tau_i = [-100, 200]$. These modes were found in every single reconstruction of the inflaton's speed of sound independently of the number of training nodes (and were also observed in previous studies (Torrado et al., 2017)), and usually show the highest individual dip $\Delta\chi^2$ with respect to ΛCDM (since they produce features at a long range of ℓ for which Planck has low error bars). On the other hand, the confidence contours are broader for earlier conformal times $\tau_i < -400$. This is the range of τ where we have found the modes at $\tau_i = [-800, -1000]$ and some degree of substructure. In this range, the posterior distributions are not very predictive (see again Figure 2.6, where the posterior peaks are small for $\tau_i < -800$), since they produce low-multipole features hidden by cosmic variance.

2.5 Conclusions

We have searched for features in the primordial power spectrum as given by the last release of Planck 2018 data. Following an EFT of inflation approach, we have focused our search on features coming from reductions of the sound speed of the inflaton, assuming these reductions to be small, mild and transient. These feature templates were not tested by the Planck Collaboration.

We have improved over previous studies (which used a single-reduction Gaussian ansatz) by developing a reconstruction technique for the speed of sound's profile based on Gaussian Processes. We have also marginalised over the slow roll parameters to allow for a dynamical prior. In this new pipeline, the parameters of the reconstruction (the position of the training nodes and the correlation length) are fitted to the Planck 2018 data. The physical constraints of the model are imposed on the reconstruction parameters by means of a Maximum-Entropy prior defined on the EFT quantities

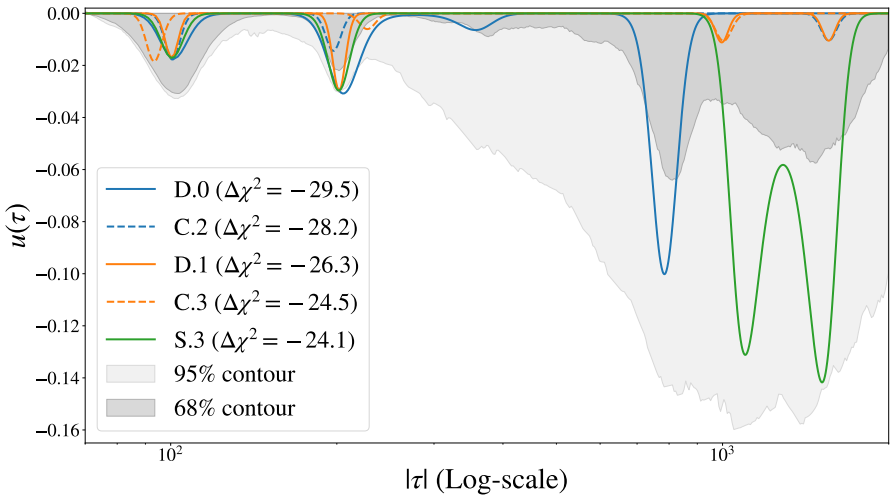

Figure 2.7: Reconstruction of the inflaton's speed of sound profile $u(\tau)$ based a 4-nodes GP, where the confidence contours (68% and 95%) are shown. We are able to constrain the shape the inflaton's speed of sound more stringently at late times (up to ~ -200), whereas the confidence intervals get larger at earlier times (i.e: starting from -800). This difference in the constraining power between early and late conformal times is mostly due to early-time reductions being associated to low-multipole features where cosmic variance is largest. The best *maxima a posteriori* are also plotted on top of the confidence contours: two 3-dipped cases (labelled C.2, dashed blue, and C.3, dashed orange), two 4-dipped cases (labelled D.0, solid blue, and D.1, solid orange) and a 3-dipped case in which one of the dips possesses some substructure (labelled S.3, solid green). For additional fits and a more detailed presentation of them see appendix 2.6.

$(|u|_{\max}, |s|_{\max})$, which define the consistency bounds of the model. We have also tracked as derived parameters n_s and r.

This template-free reconstruction of $u(\tau)$ has allowed us to make an exhaustive search of more flexible features' templates, constrained only by EFT conditions. The analysis of the result of Bayesian parameter inference on the Planck 2018 data has demonstrated that there are many possible different and complex $u(\tau)$ profiles which are consistent with Planck's CMB power spectra. As expected, none of these fits is preferred with respect to ΛCDM (their $\Delta\chi^2$'s are not significant), although show some interesting results in terms of new feature templates. First, we have argued that there is a strong preference for two consecutive reductions of the speed of sound to coexist at late times around $\tau_i \approx -200$ and $\tau_i \approx -100$. Also, combinations of modes at late conformal time $\tau_i \approx -100$ and early conformal time $\tau_i \approx -800$ are also possible. Second, we have found certain profiles which show some degree of sub-structure at early and late conformal times. Finally, we have been able to obtain reconstruction

confidence contours for the $u(\tau)$ profile given the results obtained with four training nodes.

In the future, we plan to exploit this robust and novel pipeline in the search of features using new sets of data (in particular, Large Scale Structure surveys or the CMB bispectrum). Furthermore, the improvement of current data (for example, the polarization of the CMB) will also help to reduce the noise and, therefore, the uncertainty we have at large scales. If the noise is reduced, we could discern how realistic the reductions at earlier conformal times are. Moreover, we also consider introducing new features coming from a variable first slow-roll parameter (Durakovic et al., 2019) to perform a joint search of both patterns: features induced by a variable $c_s(\tau)$ and $\epsilon(\tau)$.

2.6 Appendix: detailed results up to 4 nodes of $u(\tau)$ in log-log parameterization

In this appendix, we explain in detail the several *maxima a posteriori* found during the sampling runs when the profile of the inflaton's speed of sound $u(\tau)$ was reconstructed using Gaussian Processes up to 4 training nodes.

2.6.1 One dip (*denoted by "A"*)

These profiles of $u(\tau)$ show only one single reduction of the inflaton's speed of sound or *one dip*. These reductions can be found using just one training node in the Gaussian Process. The modes at late conformal time τ_i (see the first row of Figure 2.8) present a well defined oscillation frequency τ_i (at -100, -200 and -400). The value of $|u|_{\rm max}$ is around 0.02 and the rate of change in the speed of sound $|s|_{\rm max} \gg |u|_{\rm max}$. These dips are exactly reproduced with the linear parametrization of the reconstruction of $u(\tau)$. They were already listed during previous searches using Planck 2013 and 2015 data (Hu & Torrado, 2015; Torrado et al., 2017). In particular, the mode corresponding at $\tau_i \approx -400$ was identified faintly in (Torrado et al., 2017). These modes are present in a broad multipole ℓ range ($\ell \approx [100 - 2000]$), fitting some structures in the temperature and polarization data.

Modes at early conformal time τ_i (-800, -1000) (see the second row of Figure 2.8) are found, but are more poorly constrained and with worse $\Delta\chi^2$ with respect to ΛCDM. Mode A.3 at $\tau_i \approx -800$ shows similar characteristics to the modes at a late conformal time (small amplitude and same behaviour in the EFT parameter s). It fits an apparent oscillating structure of the temperature C_ℓ at the first acoustic peak. This mode is also found with the linear parametrization and in previous searches in (Hu & Torrado, 2015; Torrado et al., 2017). The modes at -800 and -1000, with a larger amplitude, have $|s|_{\rm max} \approx |u|_{\rm max}$. They slightly differ from the modes found in previous studies. The main reason is that $u(\tau)$ reconstructed using the logarithmic parametrization differs from the linear one at high values of τ_i. The modes at -1000 try to fit the characteristic $\ell \approx 20 - 40$ structure of the CMB temperature angular power spectrum. As identified in (Torrado et al., 2017), this kind of features impose

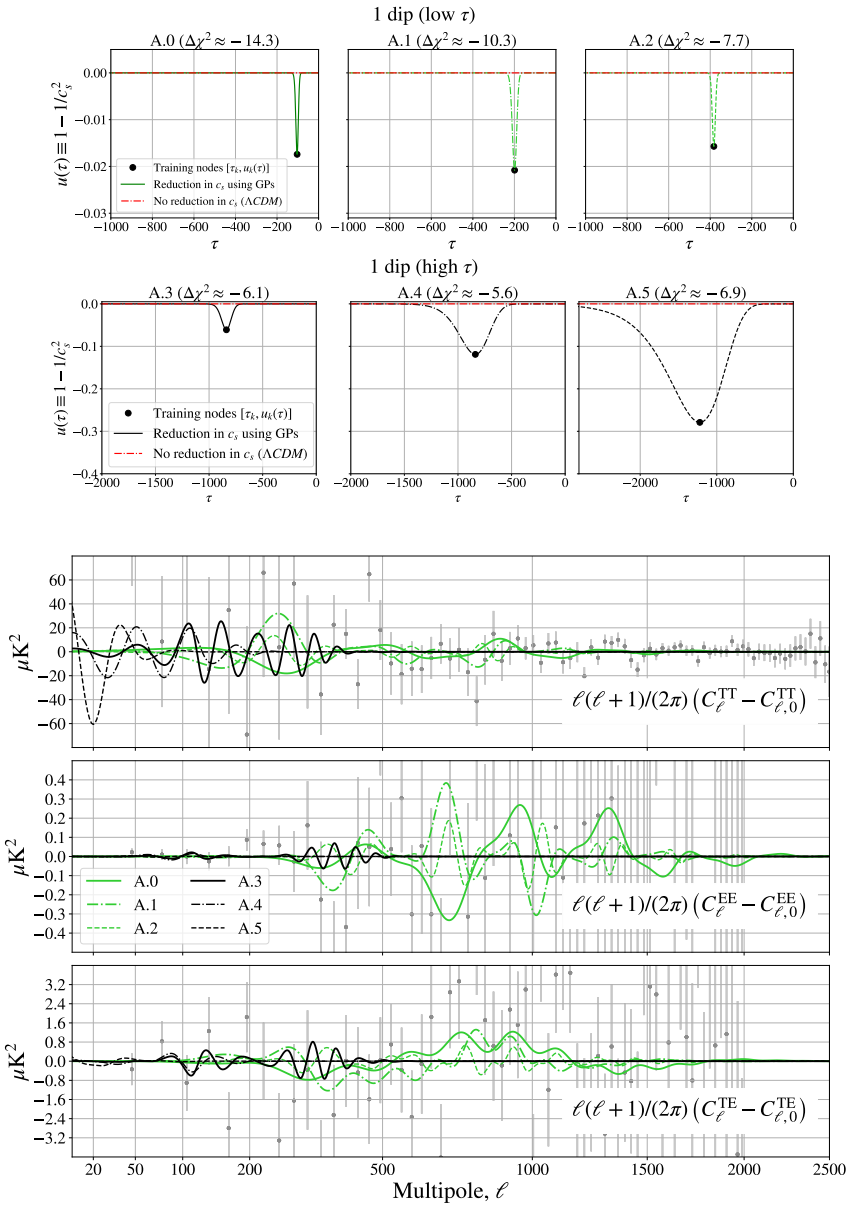

Figure 2.8: One single dip, one training node. **Top:** Different profiles $u(\tau)$ for the 6 *maxima a posteriori* when only 1 training node is used (and consequently only one dip is visible). The reconstruction is done following the logarithmic parametrization explained in subsection 2.3.1. We found a principal MAP and 5 other fits when the multimodal posterior distribution is further analysed (see, for example, Figure 2.6, where other peaks in the posterior distribution are visible). **Bottom**: Differences in the CMB temperature (TT), E-polarization (EE) and cross-correlated power spectra (TE) between the MAP to the Planck 2018 data and the featureless ΛCDM baseline model for the reconstructed speed of sound profiles $u(\tau)$ A.0 - A.5 shown above. Notice how these profiles fit small deviations from ΛCDM at low and high multipoles ℓ. The same color and line-style correspondence between the $u(\tau)$ profiles and the differences in the CMB spectra has been used.

2.6. APPENDIX: DETAILED RESULTS UP TO 4 NODES OF $U(\tau)$ IN LOG-LOG PARAMETERIZATION

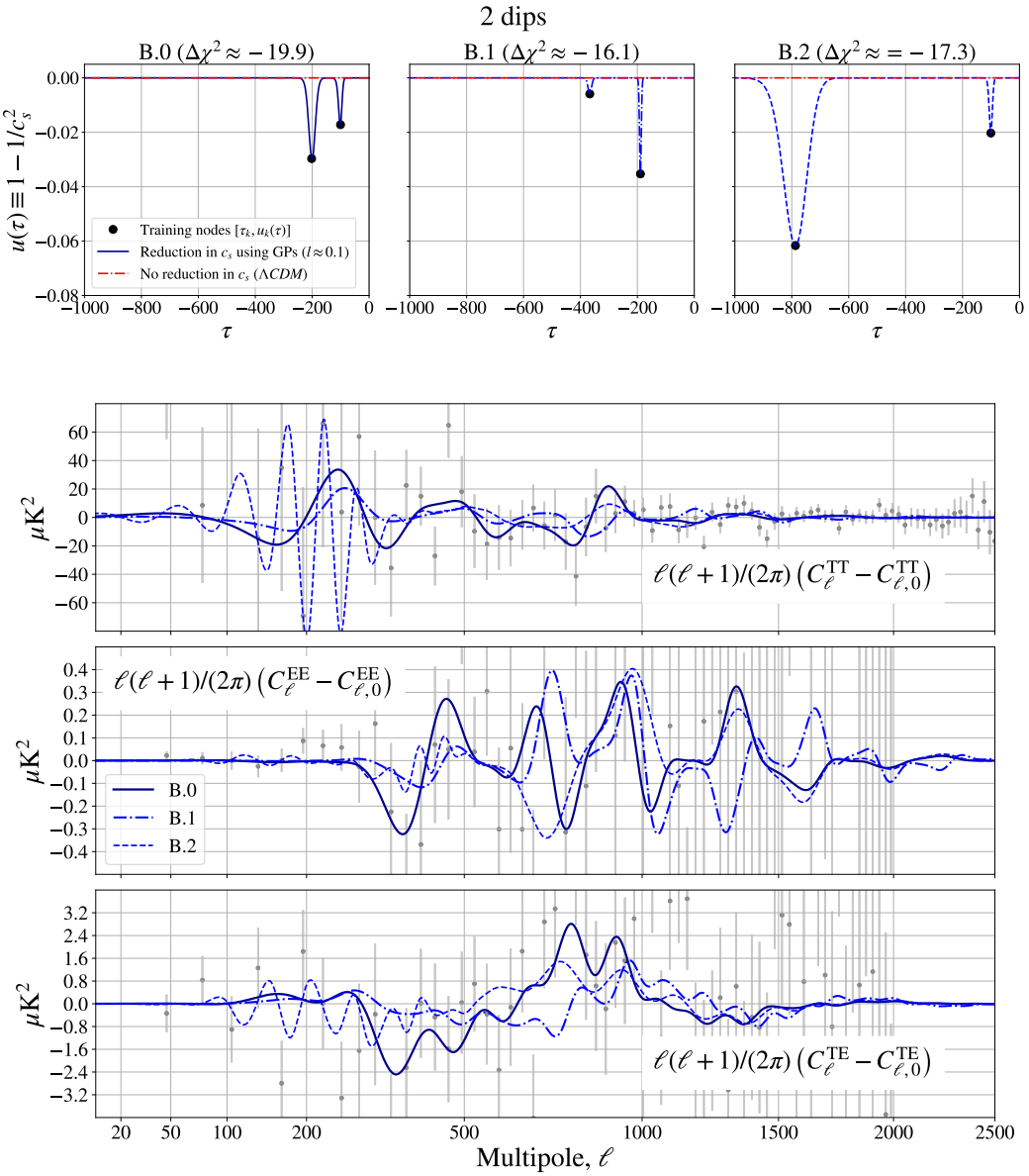

Figure 2.9: 2 non-overlapping dips, 2 training nodes. **Top:** Different profiles $u(\tau)$ for the 3 *maxima a posteriori* when only 2 training nodes are used and only two clearly different dips are observed. The reconstruction is done following the logarithmic parametrization explained in subsection 2.3.1. We found a principal best fit and 2 other fits when the multimodal posterior distribution is further studied (see, for example, Figure 2.6, where other peaks in the posterior distribution are visible). **Bottom**: Differences in the CMB temperature (TT), E-polarization (EE) and cross-correlated power spectra (TE) between the best fit to the Planck 2018 data and the featureless ΛCDM baseline model for the reconstructed speed's of sound profile $u(\tau)$ shown above. Notice how these profiles fit small deviations from ΛCDM at low and high multipoles ℓ. The same color and line-style correspondence between the $u(\tau)$ profiles and the differences in the CMB spectra has been used.

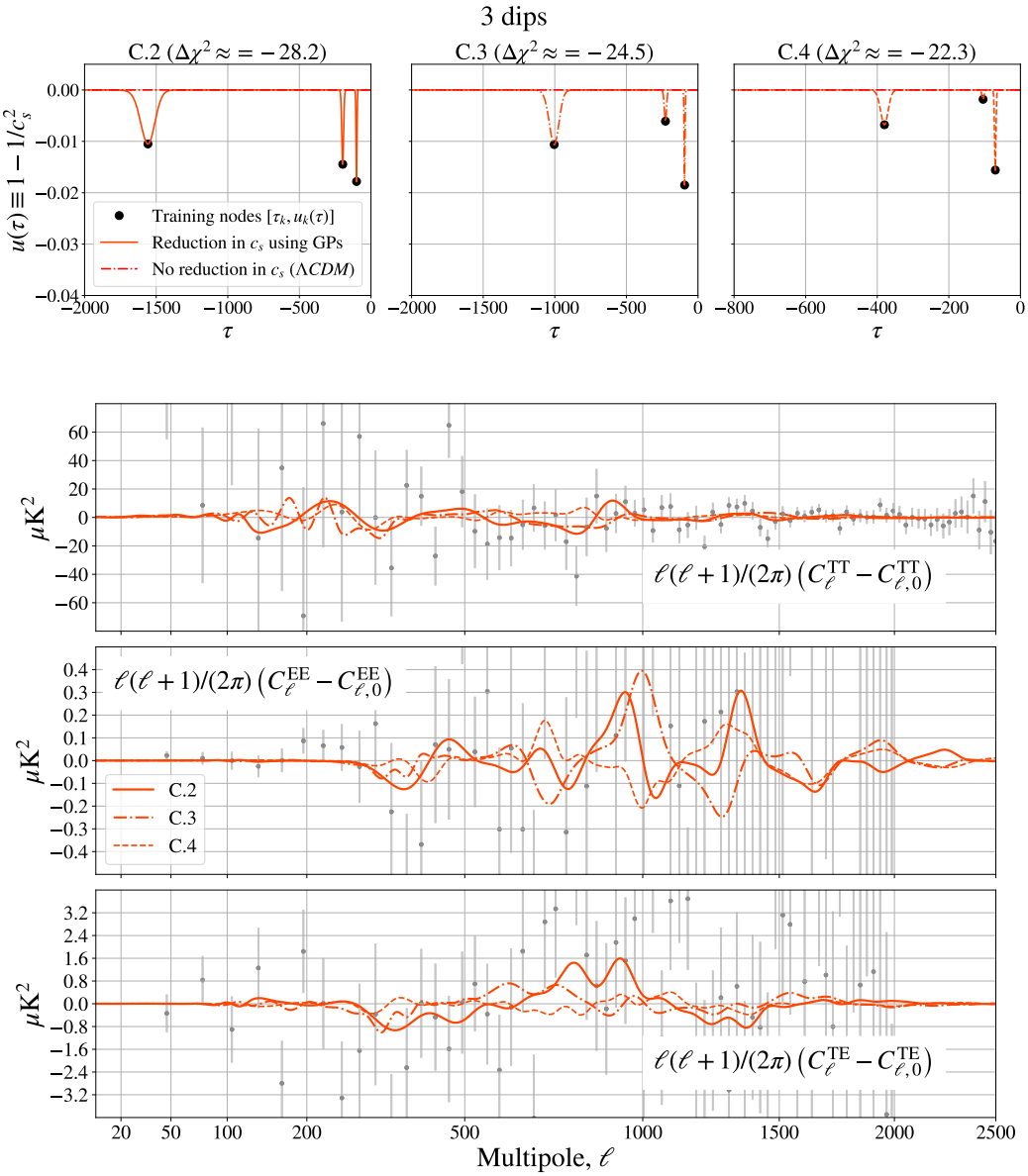

Figure 2.10: 3 non-overlapping dips, 3 training nodes. **Top:** Different profiles $u(\tau)$ for the 3 *maxima a posteriori* when only 3 training nodes are used and three differentiated dips are observed. The reconstruction is done following the logarithmic parametrization explained in subsection 2.3.1. We found 3 MAP when the corresponding multimodal posterior distribution is further studied (see, for example, Figure 2.6, where other peaks in the posterior distribution are visible). **Bottom**: Differences in the CMB temperature (TT), E-polarization (EE) and cross-correlated power spectra (TE) between the best fit to the Planck 2018 data and the featureless ΛCDM baseline model for the reconstructed speed's of sound profile $u(\tau)$ shown above. Notice how these profiles fit small deviations from ΛCDM at low and high multipoles ℓ. The same color and line-style correspondence between the $u(\tau)$ profiles and the differences in the CMB spectra has been used.

2.6. APPENDIX: DETAILED RESULTS UP TO 4 NODES OF $U(\tau)$ IN LOG-LOG PARAMETERIZATION

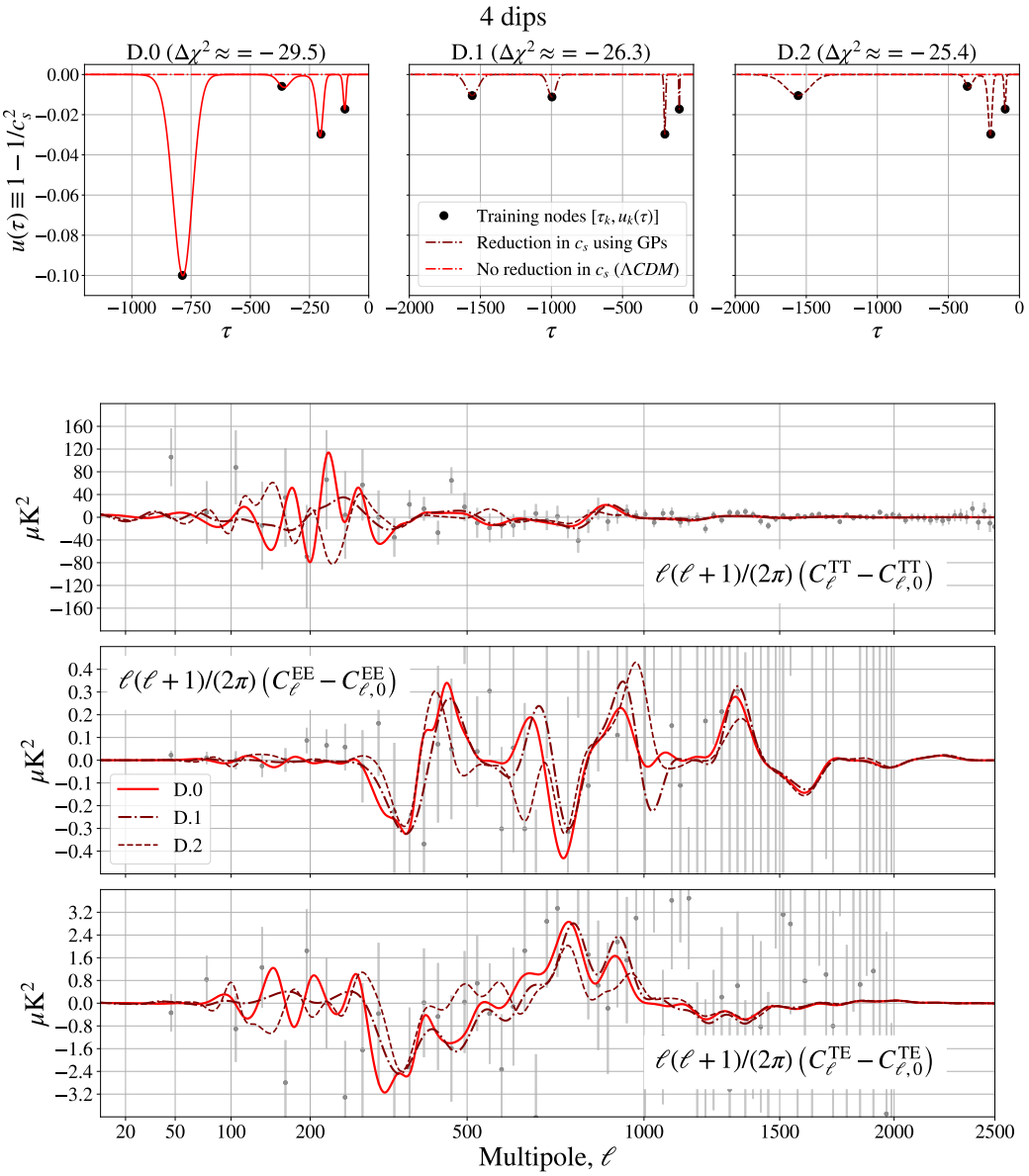

Figure 2.11: 4 non-overlapping dips, 4 training nodes. **Top:** Different profiles $u(\tau)$ for the 3 *maxima a posteriori* when only 4 training nodes are used and four differentiated dips are observed. The reconstruction is done following the logarithmic parametrization explained in subsection 2.3.1. In this case, we observe how possible reductions at $\tau \approx -100, -200, -400$ and -800 can consecutively take place. **Bottom**: Differences in the CMB temperature (TT), E-polarization (EE) and cross-correlated power spectra (TE) between the best fit to the Planck 2018 data and the featureless ΛCDM baseline model for the reconstructed speed's of sound profile $u(\tau)$ shown above. Notice how these profiles fit small deviations from ΛCDM at low and high multipoles ℓ. The same color and line-style correspondence between the $u(\tau)$ profiles and the differences in the CMB spectra has been used.

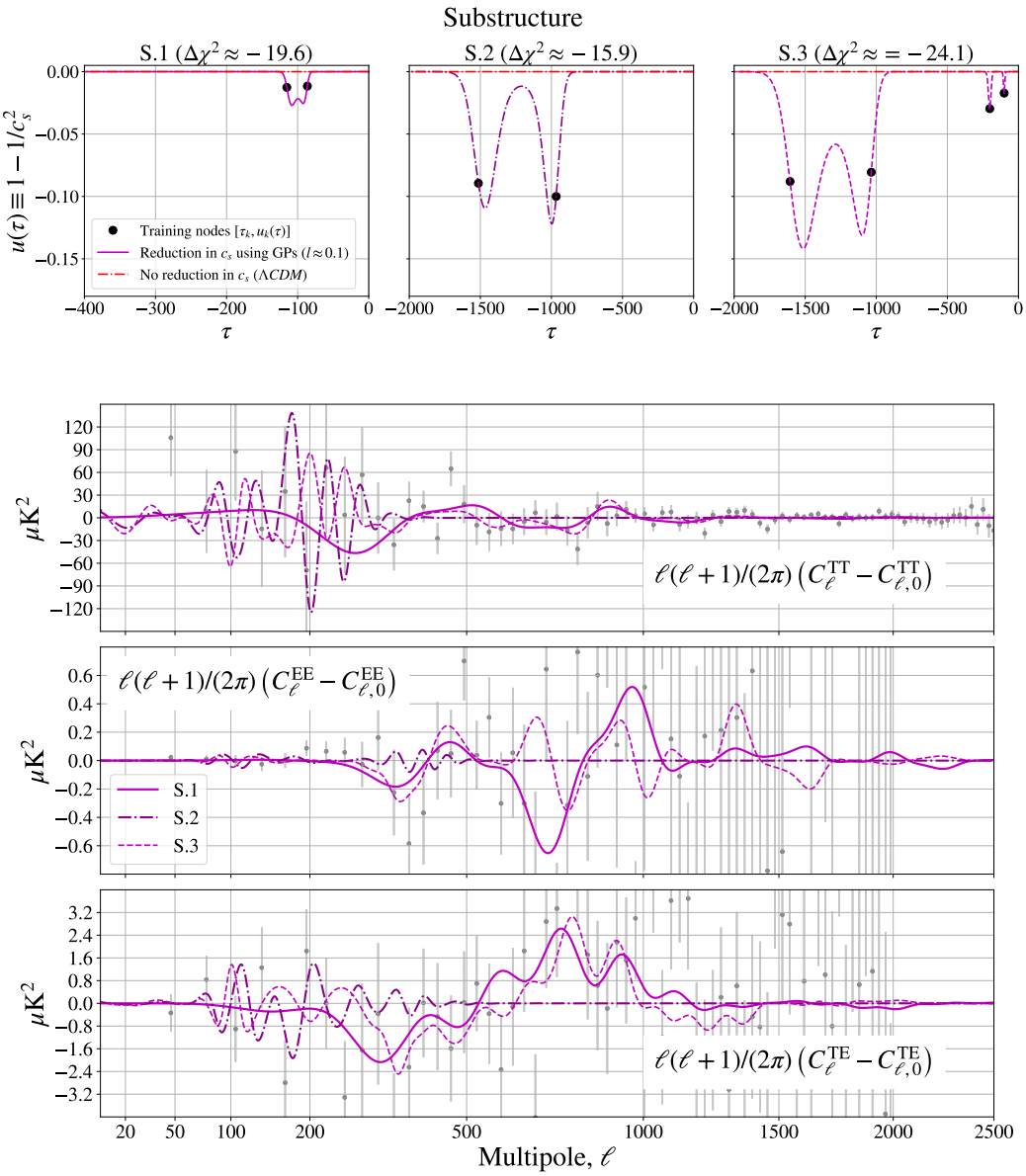

Figure 2.12: Profiles with substructure (2 training nodes and 4 training nodes). **Top:** Different profiles $u(\tau)$ for 2 *maxima a posteriori* when only 2 or training nodes are used and the profiles of $u(\tau)$ show some grade of substructure. The reconstruction is done following the logarithmic parametrization explained in subsection 2.3.1. We found 2 fits (one an earlier conformal time and a another one at late conformal time), when the corresponding multimodal posterior distributiosn are further studied (see, for example, Figure 2.6, where other peaks in the posterior distribution are visible). **Bottom**: Differences in the CMB temperature (TT), E-polarization (EE) and cross-correlated power spectra (TE) between the best fit to the Planck 2018 data and the featureless ΛCDM baseline model for the reconstructed speed's of sound profile $u(\tau)$ shown above. Notice how these profiles fit small deviations from ΛCDM at low and high multipoles ℓ. The same color and line-style correspondence between the $u(\tau)$ profiles and the differences in the CMB spectra has been used.

a tighter upper limit on the scalar-to-tensor ratio r, although these are still within the analogous bounds in ΛCDM.

2.6.2 Two dips (*denoted by "B"*)

In this case, the inflaton would suffer two consecutive reductions of the speed of sound (due to, for instance, two consecutive turns in the field space). In the previous study using Planck 2015 (Torrado et al., 2017), it was pointed out that, a priori, the features due to single reductions of the sound speed that do not overlap can in principle co-exist. These combinations would be modes at early τ_i with another late mode (i.e: -1000 and -100). However, the results of the reconstruction using two training nodes show a richer picture (see Figure 2.9). We have identified that the modes at -100 and -200 can result from an overlapping feature that is preferred by the data, and thus, it is the overall MAP for the 2-nodes reconstruction. The dips at -200 and -400 can also co-exist (mode B.1), with a worse $\Delta\chi^2$. These two overlapping features fit TT, TE and EE structure across a large range of ℓ. On the other hand, there is a possible combination of the modes at -800 and -100 (see the first row of Figure 2.9). This $u(\tau)$ profile includes the fitting of the apparent oscillations around the first acoustic peak and small deviations across the rest of the multipole scale. All of these combinations of modes fulfill $|s|_{\rm max} \gg |u|_{\rm max}$.

2.6.3 Three dips (*denoted by "C"*)

When the profile of the inflaton's reduction of the speed of sound $u(\tau)$ is reconstructed using 3 training nodes, we find more complex profiles. The usual dips at late τ_i around -100 and -200 combine to mild and small reductions at early conformal times around -1000 and -1500 (see modes C.2 and C.3 of subsection 2.6.3, respectively). The combinations that are preferred by the data are those whose earlier training node τ_i is placed around $u_i \approx -0.01$. These small dips at early τ_i were not found alone during the search using one training node (modes A shown in Figure 2.8). However, these modes at early conformal times are loosely constrained and the confidence intervals are large. Overall, these profiles fits the data very similarly as the single standing modes A.0 and A.1, as well as the combination B.0. Furthermore, we have also verified that the modes A.0 and A.2 can co-exist (contrary to the case of reconstruction with 2 training nodes, where the combination -100 and -400 was not found) if a very small mode is added close to -100. All these profiles presented in Figure 2.10 are in the limit $|s|_{\rm max} \gg |u|_{\rm max}$. Finally, the 2-dip profiles explained above are also reproduced when we run with 3 training nodes.

2.6.4 Four dips (*denoted by "D"*)

We find similar profiles for $u(\tau)$ as the ones for 3 training nodes, adding one extra dip and finding the remaining possible combination of nodes at late τ_i around -100, -200 and -400 with modes at earlier times at -800, -1000 or -1500 (see modes D.0, D.1 and D.3 of Figure 2.11). Thus, we have corroborated that the modes A.0, A.1, A.2

and A.4 can, in principle, co-exist. Still, the small modes at early $\tau_i \approx -1000, -1500$ are less likely to show up in the posterior, as they are poorly constrained given the data. It is worth mentioning that we can mostly reproduce all the different profiles found during the search using up to three nodes when four training nodes are used. In this case, one, two or three training nodes are placed in such a way the corresponding profile looks very similar to the cases A, B or C (either the training node is placed close to $u_i \approx 0$ or close to the previous training mode itself).

2.6.5 Reductions with substructure (*denoted by "S"*)

Apart from concatenations of transient reductions in the speed of sound, we have also observed some possible fits which show more complicated feature patterns according to the data. These are profiles of $u(\tau)$ that do not clearly show full dips but have some kind of substructure (see the upper row of Figure 2.12, profile S.1). We find a sub-structured maxima a posteriori at a late conformal time (centred around $\tau_i \approx -100$), which resembles the mode A.0 but with two small sub-reductions. Similarly to A.0, the limit of the EFT functions is $|s|_{\rm max} \gg |u|_{\rm max}$.

Motivated by the loose constraints of the training nodes in the range of early conformal times $-800 < \tau_i < -3000$, we have launched a GPs reconstruction using two training nodes, which are restricted to remain in the range $-3500 < \tau_i < -990$, obtaining the profile S.2 of Figure 2.12. This profile have the particularity that $|s|_{\rm max} \ll |u|_{\rm max}$. This profile tries to fit not only the structure of the CMB TT angular power spectrum around $\ell \approx 20 - 40$ but also the apparent structure of the first acoustic peak in the TT and TE data. To answer the question if it is possible the substructure mode S.2 to co-exist with any of the modes at late conformal time τ_i, we have relaunched the GPs reconstruction with 4 training nodes, constraining the two early ones in the range $-3500 < \tau_1 < \tau_2 < -990$. When this constraint is imposed, the later training nodes τ_3 and τ_4 are placed clearly around -100 and -200. The resulting profile (see mode S.3 in Figure 2.12) can fit the CMB data at low ℓ but also in a broader range similarly to the case of the mode A.0, increasing the statistical significance $\Delta\chi^2$. In this case, the EFT limit is reverted to $|s|_{\rm max} \gg |u|_{\rm max}$ due to the narrow mode at -100.

Chapter 3

Constraints on α-attractor models

This chapter is based on:
Current and future constraints on single-field α-attractor models
Guadalupe Cañas-Herrera, Fabrizio Renzi
(November, 2021), Physical Review D 104, 10, 103512, arXiv:2104.06398.

3.1 Introduction

In this chapter, we forecast the possible constraints that a future CMB StageIV (CMB-S4 hereafter) experiment may impose on inflationary observables in the optimistic scenario of a detection of non-vanishing tensor anisotropies in the Cosmic Microwave Background (CMB) polarization and temperature data. In general, the approach followed within the community (see e.g. (Planck Collaboration et al., 2014b, 2016; Chiang et al., 2010; Ade et al., 2016; Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020b)) is to sample the inflationary parameters without assuming any specific inflationary model a priori. While this approach has the advantage of exploring the inflationary sector model-independently, it does not allow for a complete sampling study of the parameter space in a specific model. Moreover, the assumption that the inflationary observables are independent of one another is in contrast with the prediction of any theory of inflation, which, for instance, assumes the validity of the slow-roll conditions (see e.g. (Renzi, Shokri, & Melchiorri, 2020; Shokri, Renzi, & Melchiorri, 2019; Giarè, Di Valentino, & Melchiorri, 2019)). In this work, conversely to the current literature on the subject, we follow a model-dependent approach imposing a specific model *a priori* and calculate the inflationary observables directly imposing the slow-roll conditions on the inflationary potential.

As mentioned in Chapter 1, the ΛCDM model is based on the simplest inflationary paradigm: canonical slow-roll single-field inflation. Within this approach, the power spectra of scalar and tensor comoving curvature perturbations are parametrised as

power laws:

$$\mathcal{P}_{\mathcal{R}}(k) = A_s\Big(\frac{k}{k_\star^S}\Big)^{n_s-1+\frac{\alpha_S}{2}\log k/k_\star^s} \tag{3.1}$$

$$P_T(k) = rA_s\Big(\frac{k}{k_\star^t}\Big)^{n_t} \tag{3.2}$$

where $k_\star^s = 0.05$ Mpc^{-1} and $k_\star^t = 0.002$ Mpc^{-1} and the subscripts stand for scalar and tensor perturbations respectively. The powers of the parametrizations are the scalar and tensor indices (n_s and n_t) and the scalar power spectrum have been further expanded in terms of the running of the spectral index α_S. Statistical analysis of recent cosmological observations (Planck observations of the Cosmic Microwave Background (CMB) (Planck Collaboration, Akrami, et al., 2020; Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020c, 2020b) and Large Scale Structure (LSS) surveys (To et al., 2021; Heymans et al., 2020)) support this parametrisation for the scalar fluctuations with $10^9 A_S \approx 2.1$ and $n_s \approx 0.965$ (Planck Collaboration, Akrami, et al., 2020; Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020c, 2020b).

In the last decade, the bound on the amplitude of PGWs (parametrized typically with the tensor-to-scalar ratio, r) has not yet seen significant improvement, where only an upper limit $r_{0.002} < 0.056$ at 95% C.L. has been provided in the last data release of the Planck Collaboration (Planck Collaboration, Akrami, et al., 2020) combining Planck and BICEP2/Keck array (BK15) data (Ade et al., 2018). Detecting those PGWs would give a direct measurement of the energy scale during inflation, as well as a clear distinguishable signature of the quantum origin of primordial fluctuations. In the upcoming decade, a new generation of CMB experiments (e.g. BICEP3 (Grayson et al., 2016), CLASS (Essinger-Hileman et al., 2014) , SPT-3G (Benson et al., 2014), Advanced ACTPol (Henderson et al., 2016), LBIRD (Suzuki et al., 2018) and CMB-S4 (K. N. Abazajian et al., 2016)) are expected to strongly improve the sensitivity on the B-modes polarization in the Cosmic Microwave Background (CMB), possibly revealing first evidences for inflationary tensor modes with amplitudes $r \sim 0.01 - 0.001$. That range is precisely expected in many well-motivated models, such as the Starobinsky inflation, which is considered the benchmark of future CMB experiments. However, while a measure of a non-vanishing r would be of key importance for inflationary theories, it will not allow understanding the inflationary mechanism in detail but only its energy scale. It is therefore timely to investigate, given future CMB experiments, what would be the freedom in a generic inflationary framework that is left in case of the optimistic scenario of a non-vanishing tensor-to-scalar ratio measure.

In particular, there is a general class of models called α-attractors, that has gained lots of popularity because of their agreement with observational constraints and the universality of their predictions for the inflationary observables (Iarygina, Sfakianakis, Wang, & Achucarro, 2019; Iarygina, Sfakianakis, Wang, & Achúcarro, 2020; Aresté Saló, Benisty, Guendelman, & Haro, 2021b, 2021a; Rodrigues, Santos da Costa, & Alcaniz, 2021). Recently, α-attractors have also been of the interest to study a possible connection between early and late physics in the context of Dark Energy

(Akrami, Kallosh, Linde, & Vardanyan, 2018; Akrami, Casas, Deng, & Vardanyan, 2021; Miranda, Fabris, & Piattella, 2017; Pozdeeva, 2020; Pozdeeva & Vernov, 2021). This set of models have also been embedded in a more general multi-field inflationary scenario and in $\mathcal{N} = 1$ supergravity. In the context of supergravity, the α-attractor can be represented by a potential of the form:

$$\frac{V(\varphi)}{V_0} = (\tanh(\beta\varphi/2))^{2n} \tag{3.3}$$

where $\beta^2 = 2/3\alpha$ and n is an arbitrary value. It is important to note that the "attractor behaviour" of this potential sits on the fact that the observable predictions are the same up to leading order regardless of the value of n while they differ only in sub-leading corrections. Assuming slow-roll inflation and the α-attractor form of the inflationary potential the observational predictions for the inflationary observables can be written as:

$$r = \frac{12\alpha}{N^2} \tag{3.4a}$$

$$n_s = 1 - \frac{2}{N} = 1 - \sqrt{\frac{r}{3\alpha}}, \tag{3.4b}$$

$$\alpha_S = -\frac{2}{N^2} = -\frac{r}{6\alpha}. \tag{3.4c}$$

where N is the number of e-folds to inflation to last. These definitions in terms of parameter α encompass several inflationary models and clearly reduce to the well-known Starobinsky inflation for $\alpha = 1$ (Kallosh, Linde, & Roest, 2013; Kallosh, Linde, Roest, & Yamada, 2017; Carrasco, Kallosh, Linde, & Roest, 2015). Moreover, for a broad class of potentials V, as long as $\alpha \ll O(1)$, the scalar spectral index n_s, its running α_S and the tensor-to-scalar ratio r converge to the functional form of equation (3.4) regardless of the kinectic terms of the theory. It has also been showed that this statement holds true in some multi-field inflation regimes (Achúcarro, Kallosh, Linde, Wang, & Welling, 2018), where the conditions that guarantee the universality of the observational predictions for the inflationary parameters are derived by imposing constraints on the potential. The universality of the observational constraints is one of the most important features of single-field α-attractor models.

3.2 Constraints from current CMB and LSS data

Current CMB (Planck Collaboration et al., 2014b, 2016; Chiang et al., 2010; Ade et al., 2016; Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020b) and LSS (To et al., 2021) data are unable to constrain the tensor-to-scalar ratio if r is sampled independently from the scalar index n_s. However, by imposing the α-attractor model a priori, we force a specific functional relation between n_s and r that allows us to translate the sub-percentage constraints on n_s from current data into a constraint on r in the context of α-attractors.

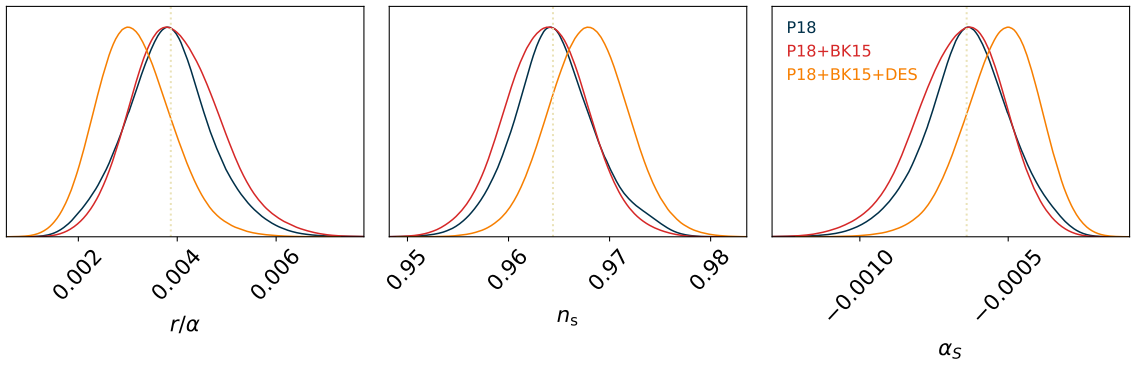

Figure 3.1: Posterior distribution for Planck 2018 data alone and combined with the Biceps/Keck 2015 B-mode data and with Large Scale Structure DES data. The dotted lines denote the expected values for Staronbinsky inflation ($\alpha = 1$ and $N \approx 60$).

In fact, by imposing an inflationary model we are selecting a subset of the parameter space allowed by the data when r and n_s are considered independently in cosmological parameter estimations. This is particularly evident if one considers the relation between r and n_s in the α-attractor model given by equation (3.4b).

Current data cannot break the $r - \alpha$ degeneracy, and therefore, sampling it in our analysis would give no insights on the α-attractor models. For this reason, instead of sampling r and α independently, we use the ratio r/α as a parameter for our MCMC analysis.

Along with the ratio r/α, we consider as independent parameters the other five standard ΛCDM ones: the baryon $\omega_b = \Omega_b h^2$ and the CDM $\omega_c = \Omega_c h^2$ densities, the Hubble constant H_0, the optical depth τ and the amplitude of scalar perturbations A_s. We also let free the running of the spectral index α_S. As α-attractors satisfy the usual inflationary consistency relation, we fix the index of tensor modes to $n_T = -r/8$. The uniform prior distribution imposed on these parameters are reported in Table 3.1.

Parameter	Prior range
$\Omega_b h^2$	$[0.005, 0.1]$
$\Omega_c h^2$	$[0.001, 0.99]$
H_0	$[20, 100]$
τ	$[0.01, 0.8]$
$r_{0.002}/\alpha \cdot 10^3$	$[0.5, 8]$
$\log(10^{10} A_s)$	$[1.61, 3.91]$
$\sum m_\nu$	$[0, 1]$
$N_{\rm eff}$	$[2, 5]$

Table 3.1: Range of uniform priors distributions imposed on the sampled parameters during the analysis.

The predictions of the theoretical observational probes are calculated using the latest version of the cosmological Boltzmann integrator code `CAMB` (Lewis et al., 2000; Howlett et al., 2012). To compare our theoretical predictions with data, we use the full 2018 Planck temperature and polarization datasets which also includes multipoles $\ell < 30$ (Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020a). We combine the Planck likelihood with the Biceps/Keck 2015 B-mode data (Ade et al., 2018) and the combination of galaxy clustering and weak lensing data from the first year of the Dark Energy Survey (DES Y1) (T. M. C. Abbott et al., 2018). The posterior distributions of the cosmological parameters have been explored using the publicly available version of the Bayesian analysis tool `Cobaya` (Torrado & Lewis, 2021). In particular, the posteriors have been sampled using the MCMC algorithm developed for CosmoMC (Lewis & Bridle, 2002; Lewis, 2013) and tailored for parameter spaces with a speed hierarchy.

The 1D posterior of r/α, n_s and α_S resulting from our Bayesian statistical analysis employing Planck 2018 data in combination with DES and BK15 data are reported in Figure 3.1. Given the sub-percentage constraints on the scalar index, we found a sub-percentage constraint on the tensor amplitude *i.e.* $r/\alpha = 0.00387^{+0.00078}_{-0.00094}$ for Planck 2018 data alone. Due to this correlation, we see also that there is virtually no difference between the results using the Planck 2018 data and combining them with the Biceps/Keck 2015 data ($r/\alpha = 0.00400^{+0.00076}_{-0.00095}$) as the constraint on the scalar index is unchanged (if not for a statistically insignificant shift in the posterior mean between the two runs). When LSS data (i.e: DES) are included in the analysis, a shift in the spectral index n_s with respect to CMB data is found. DES data prefer a slightly higher value for n_s, shifting accordingly the running of the spectral index α_S and the tensor-to-scalar ratio r. However, the results remain consistent with that from PK18 and PK18+BK15 within 1σ (see also Figure 3.1).

Incidentally, we also obtain a constraint on the running of the scalar index α_S (related to r/α by equation (3.4c)) away from zero at 4 standard deviation *i.e.* $\alpha_S = -6.4^{+1.6}_{-1.3} \cdot 10^{-4}$. It is worth noting that (as for r) this is due to the specific correlation which arises in α-attractor inflation between the parameters of the scalar and tensor spectrum. This result, however, points out that future measurements of $r_{0.002}$ and $n_{\rm nrun}$ could potentially rule out the α-attractor model: they are key parameters in studying the viability of an inflationary model and should be considered in the future analysis of CMB and LSS data.

We conclude this section with the following two considerations:

- Given $n_s \approx 0.965$ and $r \to 0$, the best-fit model for CMB and LSS data is the case of $\alpha = 1$, which corresponds to Starobinsky inflation, and this cannot be distinguished from a generic model with $\alpha \neq 1$ unless future experiments provides a measure of either polarization B-modes or α_S.

- Current data are consistent with $r \to 0$. Nevertheless, not all values of α are allowed. Instead, they set an upper limit on the value of the tensor-to-scalar ratio and this knowledge can be used to obtain an upper limit for α:

$$\alpha \lesssim \frac{r_{\rm lim}}{r_0} \equiv \alpha_{\rm lim}, \tag{3.5}$$

where $r_{\rm lim}$ is the experimental threshold for a given experimental configuration and r_0 is the mean of the r/α posterior. Being $r_{\rm lim} \approx 0.1$ for P18 data and $r_{\rm lim} \approx 0.06$ for P18+BK15 data combined (Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020b), we correspondingly find $\alpha_{\rm lim} \approx 25$ and $\alpha_{\rm lim} \approx 15$ for P18 and P18+BK15 respectively. These upper limits would be the same that one would obtain running an MCMC analysis with r and α considered independently (see Appendix 3.5) .

3.3 Forecast for future CMB-S4 observations

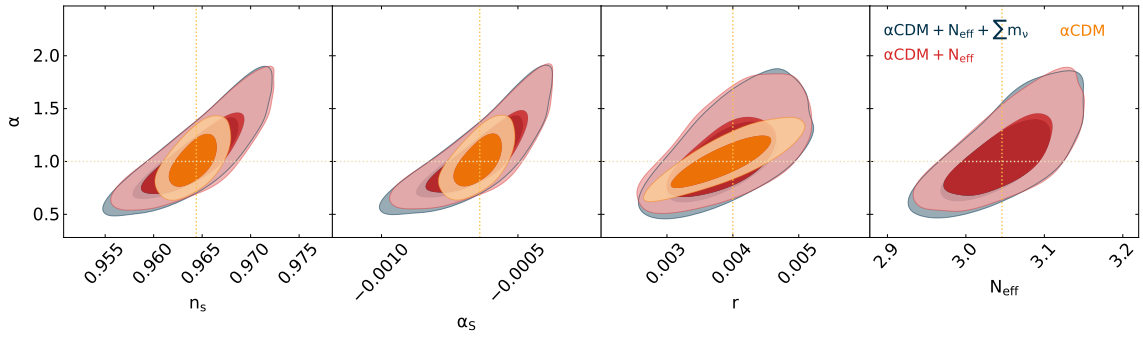

Figure 3.2: Forecasted 2D contours at 68% and 95% C.L. for α attractor inflationary parameters for a CMB-S4 experiments with and without allowing the neutrino sector to vary (αCDM, αCDM+$N_{\rm eff}$ and αCDM+$N_{\rm eff} + \sum m_\nu$ respectively).

While current data are unable to constraint the value of α given the current experimental sensitivity, future generation CMB experiments are expected to strongly improve the sensitivity on the B-mode polarization signal of the CMB, possibly discovering evidence of a primordial tensor mode with amplitude in the range of $r \sim 0.01 - 0.001$ (Ade et al., 2019; Lee et al., 2019; K. Abazajian et al., 2022; Hazumi et al., 2020). In particular, this is the range of predictions for many well-motivated inflationary models, such as Starobinsky inflation, considered the benchmark for future CMB observations (Ade et al., 2019; Lee et al., 2019; K. Abazajian et al., 2022; Hazumi et al., 2020). In this section, we study the optimistic scenario of a future detection in the CMB anisotropies of non-vanishing tensor amplitude and we forecast the constraints achievable with a CMBS4-like experiment on the parameters of the α-attractor model.

We consider as a baseline model a minimal extended ΛCDM cosmology with the inclusion of non-vanishing tensor-to-scalar ratio, r and α. This extended model constitutes our simulated data sets. The value of r is chosen correspondingly to the best-fit value obtained with a Starobinsky model using only Planck 2018 data *i.e.* $r = 0.00387$, while we fix $\alpha = 1$. The value of the scalar index and its running are also fixed to $n_s = 0.964$ and $\alpha_S = 0.0006$. The remaining ΛCDM parameters values are: $\omega_b = 0.0221$, $\omega_c = 0.12$, $H_0 = 67.3$, $\tau = 0.06$ and $\ln(10^{10} A_s) = 3.05$. As the new generation of CMB experiment also expects to set some light in the neutrino sector,

we also explore the number of effective degrees of freedom of relativistic species $N_{\rm eff}$ and the sum of the neutrino masses $\sum m_\nu$ in the forecast.

Both simulated data and theoretical models are computed with the latest version of the Boltzmann code `CAMB` (Lewis et al., 2000; Howlett et al., 2012). To extract constraints on cosmological parameters, we make use of the Monte Carlo Markov Chain (MCMC) code CosmoMC (Lewis & Bridle, 2002; Lewis, 2013) which compares theory with a simulated dataset using a given likelihood.

As in (Di Valentino, Holz, Melchiorri, & Renzi, 2018; Renzi, Hogg, Martinelli, & Nesseris, 2021; Renzi, Cabass, Di Valentino, Melchiorri, & Pagano, 2018; Cabass et al., 2016), we built our forecasts for future CMB experiments following a well-established and common method. Using the set of fiducial parameters described above, we compute the angular power spectra of temperature C_ℓ^{TT}, E and B polarization $C_\ell^{EE,BB}$ and cross temperature-polarization C_ℓ^{TE} anisotropies. We produce synthetic realization of future data adding to the theoretical power spectra, an exponential noise of the form (Perotto, Lesgourgues, Hannestad, Tu, & Y Y Wong, 2006):

$$N_\ell = w^{-1} \exp(\ell(\ell+1)\theta^2/8\ln 2) \tag{3.6}$$

where θ is the FWHM angular resolution and w^{-1} is the experimental sensitivity expressed in μK arcmin. The polarization noise is derived equivalently assuming $w_p^{-1} = 2w^{-1}$ since one detector measures two polarization states. The simulated spectra, realized accordingly to the previous discussion, are compared with theoretical ones using a "CMB-like" likelihood as in (Perotto et al., 2006; Audren, Lesgourgues, Bird, Haehnelt, & Viel, 2013)

For this chapter, we have constructed synthetic realizations of CMB data for only one experimental configuration, namely CMB-S4 (see e.g. (K. N. Abazajian et al., 2016)). The CMB-S4 dataset is constructed using $\theta = 3'$ and $w = 1\,^-$K arcmin, and it operates over the range of multipoles $5 \leq \ell \leq 3000$, with a sky coverage of the 40%. Furthermore CMB-S4 is expected to reach a target sensitivity on the tensor-to-scalar ratio of $\Delta r \sim 0.0006$, whose goal is to provide a 95% upper limit of $r < 0.001$. Therefore the value chosen for our fiducial model is well within the scope of an experiment like CMB-S4. However, the corresponding sensitivity on the value of the running of the scalar index, α_S would be only $\Delta\alpha_S = 0.002$ which would clearly not be enough for a joint detection of r and α_S assuming Starobinsky inflation. Thus, it may not be possible to distinguish between a generic α-attractor model with $r \sim 0.004$ and Starobinsky inflation despite a future detection of a non-vanishing tensor amplitude.

In α-attractors, however, the uncertainties about the correct shape of the inflationary potential, defining the value of r, n_s and α_S, are parameterized with the α parameter. Therefore a measure of the value of α would also give us insights about the correct shape of the inflationary potential and correspondingly on the correct theory of inflation. A CMBS4-like experiment will be able to give such insights provided a detection of a non-vanishing tensor amplitude.

Current data only place a loose upper bound $0 \leq \alpha \lesssim 15$ (Kallosh & Linde, 2019) correspondingly to P18+BK15 upper limit on the tensor amplitude $r < 0.056$

at 95%C.L. (Planck Collaboration, Akrami, et al., 2020). To correctly explore the available parameter space for α we therefore employ a logarithmic prior on its value $-6 \leq \log_{10}\alpha \leq 1$ while we keep the priors on the other parameters as of Table 3.1. We refer to this model as αCDM. From our CMB-S4 forecasts we obtain a 15% bound on the parameter $\alpha = 1.01^{+0.14}_{-0.18}$, clearly showing the ability of future CMB experiments of bounding single-field slow-roll inflationary models. Models with $\alpha \geq 2$ and $\alpha \leq 0.5$ would be potentially excluded at more than 2 standard deviations in the optimistic scenario of a PGWs detection with amplitude in the range of the Starobinsky model. We eventually extend this baseline model including the number of relativistic neutrino species $N_{\rm eff}$, (αCDM$+N_{\rm eff}$). When $N_{\rm eff}$ is varied, we find a 5% reduction of the accuracy with which α is measured *i.e.* $\alpha = 1.07^{+0.18}_{-0.23}$ while the bound on the tensor-to-scalar ratio is basically the same in the two cases, *i.e.* $\sigma(r) = 0.00050$. Conversely we found an increase in the error budget of the scalar index and running, passing from $\sigma(n_s) = 0.0016$ and $\sigma(\alpha_S) = 0.00006$ (αCDM) to $\sigma(n_s) = 0.0035$ and $\sigma(\alpha_S) = 0.0001$ (αCDM $+ N_{\rm eff}$) an worsening of a factor around two in both cases. It is worth stressing that Primordial Gravitational waves may also contribute to the number of relativistic species being themselves relativistic degrees of freedom (Cabass et al., 2016; T. L. Smith, Pierpaoli, & Kamionkowski, 2006; Clarke, Copeland, & Moss, 2020). This contribution can be calculated analytically to be :

$$N_{\rm eff,GW} \sim \frac{r A_s}{n_T}(A^{n_T} - B^{n_T}) \tag{3.7}$$

where A and B are two real numbers and $A, B \gg 1$. This contribution is clearly extremely small for red spectra ($n_T \leq 0$) but may be important in inflationary theories where blue spectra ($n_T > 0$) can be produced (see e.g. (Mukohyama, Namba, Peloso, & Shiu, 2014; Namba, Peloso, Shiraishi, Sorbo, & Unal, 2016; Stewart & Brandenberger, 2008; Hebecker, Jaeckel, Rompineve, & Witkowski, 2016; Peloso, Sorbo, & Unal, 2016; Giarè & Melchiorri, 2021; Giarè & Renzi, 2020)). Consequently the only interaction between PGWs and neutrinos considered in this work is the one arising from neutrino anisotropic stress after neutrino decoupling at $T \lesssim 1$ MeV (Kojima, Kajino, & Mathews, 2010). These constraints are virtually unmodified when we further extend our baseline model, allowing the whole neutrino sector to vary *i.e.* $N_{\rm eff}+\sum m_\nu$. The 2D contours for both our forecasts are reported in Figure 3.2. A strong correlation now arises between α and the other inflationary parameters conversely to what we found with the Planck data. This is due to the power of CMB-S4 of resolving the B-mode spectrum, consequently breaking the degeneracy between r and n_s. Nevertheless, the situation is unchanged for the scalar running. The strong bound we find on the scalar running is in fact due to imposing the α-model a priori. Even a StageIV experiment would not have the required accuracy to measure the tiny scalar running predicted by α-attractor inflation. When r, n_s and α_S are independently varied (*i.e.* neglecting the consistency relation in equation (3.4)) the running is fixed only with an error $\sigma(\alpha_S) = 0.0029$ at 68% C.L., an order of magnitude higher than when the α-model is imposed *a priori* and in good agreement with the expected sensitivity for the CMB-S4 experiment (K. N. Abazajian et al., 2016).

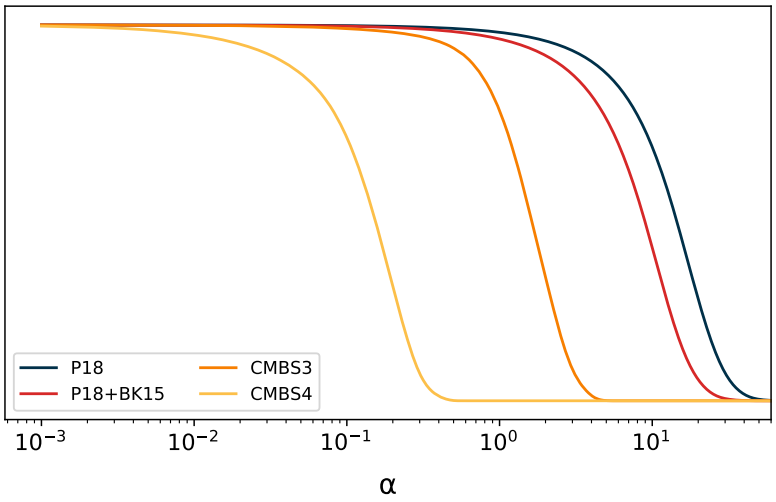

Figure 3.3: 1D posterior for the parameter α for several experimental configurations. These posterior distributions are obtained with the method described in Appendix 3.5. The CMB Stage-III (CMBS3) constraint is obtained assuming a target sensitivity of $r_{\rm lim} = 0.01$ corresponding to $\alpha_{\rm lim} \approx 3$. This sensitivity would be achievable by a stage-III experiment such as SPT-3G (Benson et al., 2014) or BICEP3 (Grayson et al., 2016).

We conclude noting that, as well as for current data, one can forecast the corresponding upper limit on α from equation (3.5) in the pessimist scenario which CMBS4 will not be able to detect a tensor-to-scalar ratio above the target sensitivity. Assuming $r_{\rm lim} = 0.001$ (K. Abazajian et al., 2022), one finds $\alpha_{\rm lim} \approx 0.26$ which would exclude Starobinsky inflation at ten standard deviations. With respect to Planck data, CMBS4 will provide an improvement on the measure of $\alpha_{\rm lim}$ of two orders of magnitude even in the pessimistic case of not detecting any B-mode polarization signal. The constraints on r/α will be instead improved only by a factor of four leading to $r/\alpha = 0.00386 \pm 0.00035$ when neutrinos parameters are fixed to their ΛCDM values. We show in Figure 3.3 a comparison of the upper bounds on α achievable by the experimental configurations considered in this work.

3.4 Conclusions

We have carried out a Bayesian analysis with current CMB and LSS data to constrain inflationary observables (the scalar spectral index n_s, its running α_S and the tensor-to-scalar ratio r). With the current constraining power on n_s and imposing the α-attractor model a priori in our analysis, the possible values of the ratio r/α are narrowed in a band of around 0.004. However current data do not have enough sensitivity to break the degeneracy between r and α and consequently to constrain any deviation from the Starobinsk inflationary model due to the fact the predicted tensor-to-scalar ratio is much smaller than the current upper limit of LSS and CMB

data. Consequently, we focused our attention on forecasting the constraints achievable by a future CMBS4 experiment assuming a tensor-to-scalar ratio corresponding to the value obtained from Planck data imposing the α-attractor model a priori, see section 3.2.

The forecast is performed from a Bayesian statistical approach, where α is let free to be sampled from a logarithmic prior distribution. Future CMB-S4 experiments will then be able to constrain α as long as the value of r is above the target sensitivity expected from such experiment *i.e.* $r > 0.001$ (K. N. Abazajian et al., 2016; K. Abazajian et al., 2022). Conversely, in the pessimist scenario that even in future CMB-S4 data will not measure a tensor amplitude above the target sensitivity, the situation will be exactly as for current data and only an upper limit on the value of α could be placed. We forecasted the corresponding limit on α to be $\alpha_{\rm lim} \approx 0.26$, an improvement of two orders of magnitude with respect to Planck data alone.

In conclusion, a future CMB-S4 experiment will have enough sensitivity to significantly constrain single-field slow-roll inflationary models. In the case of an optimistic detection of a non-vanishing tensor amplitude, it would be able to shed light on both the energy scale and the shape of the inflationary potential, while in the pessimistic scenario of a non-detection of tensor modes it would still be able to place a tight upper limit on the value of α and exclude Starobinsky inflation at 10σ . We underline that, when the running of the spectral index α_S is free to vary, it is always different from zero as expected from the inflationary consistency relation of the α attractor model. However, we show that the value expected for the scalar running given the current constraints on the scalar index is so small that it will not be detectable by a future CMB-S4 experiment (with an expected sensitivity of $\Delta\alpha_S \sim 0.003$), but it may be reachable when information from future weak lensing and galaxy clustering measurements will be included (Euclid Collaboration et al., 2020; Font-Ribera et al., 2014; LSST Science Collaboration, Abell, et al., 2009). The combination of future weak-lensing surveys and CMBS4 would possibly reach a target sensitivity on $\Delta\alpha_S \sim 0.001$, a factor of three better than CMBS4 alone[1]. This is enough to constrain α_S at a level compatible with the value expected from α-attractor models. Note that a measure of α_S would constitute a smoking-gun for inflation as well as a measure of a non-zero tensor amplitude. Therefore, future LSS surveys and CMB experiments will either give us a measure of both r and α_S in the most optimistic scenario or they will be able to significantly reduce the available parameter space for single-field slow-roll inflation in the most pessimistic one.

[1]This is derived assuming an improvement of a factor $\sigma(\alpha_S)_{\rm Planck}/\sigma(\alpha_S)_{\rm CMBS4} \sim 3$ of the forecasted constraints on α_S with respect to the combination of weak-lensing and Planck data from Tab.21 of (Amendola, Appleby, et al., 2018)

3.5 Appendix: from a two sigma bound to an upper limit

In this section we briefly describe the procedure used to convert the two sigma bound on r/α into an upper limit on α assuming an experimental threshold $r_{\rm lim}$. Let us start noting that an upper limit on the tensor-to-scalar ratio, $r < r_{\rm lim}$ at 95% C.L. can be represented by an half-normal distribution with standard deviation, σ, given by the following equation:

$$\int_{-r_{\rm lim}}^{r_{\rm lim}} \mathcal{N}(x \,|\, 0, \sigma) dx = 0.95 \tag{3.8}$$

Solving for σ and applying an inverse transform sampling technique, we can extract samples from the half normal distribution. Then, for each samples of the half Gaussian of r/α, we can use equation (3.5) to calculate a sample of the distribution of α. This procedure allows to reconstruct the posterior of α starting from the bound on r/α and it is equivalent to perform a full MCMC analysis with an experimental configuration that can reveal tensor modes with amplitude $r > r_{\rm lim}$ at 95% C.L. . As shown in Fig.(3.3), the results on α agrees almost perfectly with the approximate results obtained considering a delta distribution for r_0 and $r_{\rm lim}$. Thus, we conclude that the uncertainties in the measure of r/α can be negligible in deriving an upper limit for the values of α.

Part II

Constraints on Gravitational Waves Physics

Chapter 4

Cross-correlations of the anisotropies of the Astrophysical Gravitational-Wave Background with Galaxy Clustering

This chapter is based on:
Cross-correlation of the astrophysical gravitational-wave background with galaxy clustering
Guadalupe Cañas-Herrera, Omar Contigiani and Valeri Vardanyan.
(August, 2020), Physical Review D 102, 4, 043513, arXiv:1910.08353.

4.1 Introduction

Gravitational waves (GWs) are one of the striking predictions of the General Theory of Relativity (Einstein, 1916, 1918). The first indirect detection was obtained by measuring the orbital decay of a pulsar binary system by Hulse and Taylor (Hulse & Taylor, 1975) and, a century after they were conjectured, the GW signal of a merging black hole binary was detected by the Laser Interferometer Gravitational-Wave Observatory (LIGO) (LIGO Scientific collaboration and VIRGO collaboration, 2016b). Because the strain of GWs is less affected by distance compared to electromagnetic radiation, they potentially contain important information about sources which would be otherwise too dim to be observable. This discovery paved the way for a new multi-messenger era in cosmology and opened a new window into the physics of compact objects and gravity (Ezquiaga & Zumalacárregui, 2018).

Every GW signal observed so far has been emitted from bright sources resolved as distinct events, such as low-redshift black hole (LIGO Scientific collaboration and VIRGO collaboration, 2016a, 2017a, 2017b, 2017c) and neutron star binary mergers (LIGO Scientific collaboration and VIRGO collaboration, 2017d). However, in addition to resolved events, one can expect the presence of a GW background (GWB) produced by the superposition of unresolved compact binaries that are either too far

away or too faint to be detected individually. In practical terms, these unresolved sources form stochastic GWBs, which may differ in spectral shape and frequency depending on the source population (Thrane et al., 2009).

For instance, supermassive black hole binaries form a stochastic background in the nHz band, which is expected to soon be detected by the Pulsar Timing Array (PTA) (S. R. Taylor et al., 2016; Arzoumanian et al., 2018; Mannerkoski, Johansson, Pihajoki, Rantala, & Naab, 2019). While in the mHz band, the mergers of a similar population of massive binaries are expected to be detected as resolved events by the Laser Interferometer Space Antenna (LISA) (Wyithe & Loeb, 2003).

GWBs might also have a cosmological origin. Examples of such backgrounds are those produced in the early Universe, such as during inflation (T. L. Smith et al., 2006), or a phase transitions (Blanco-Pillado, Olum, & Siemens, 2018). Moreover, a hypothesized primordial black hole population (Sasaki, Suyama, Tanaka, & Yokoyama, 2018) might also contribute to the total number of compact binaries in the Universe. Many of these cosmological backgrounds are predicted to be isotropic and they can extend over multiple frequency bands, from nHz up to GHz (Sousa, Avelino, & Guedes, 2020; Wang, Huang, Li, & Liao, 2020).

In this chapter, we discuss the background due to solar-mass sized stellar remnants (black hole or neutron star binaries). The astrophysical GWB resulting from their inspiral and coalescence should be detectable not only in mHz band (Z.-C. Chen, Huang, & Huang, 2019), but also in the Hz to kHz band. In this range, LIGO searches of this background have already been performed (LIGO Scientific collaboration and VIRGO collaboration, 2019).

While the experimental challenges associated with the detection of this GWB are not the focus of this work, it is worth pointing out that fundamental obstacles persist in both frequency ranges. In the mHz band, the reconstruction is hindered by the presence of an additional low-frequency background induced by Galactic white dwarf binaries (Amaro-Seoane et al., 2012). To address this complication, previous works have shown that this background can be removed by exploiting the yearly modulation of space-based GW observatories (Adams & Cornish, 2014). On the other hand, the main obstacle in the Hz-kHz is represented by the large shot noise contribution. Because the astrophysical GWB in this band is comprised of multiple unresolved transient events, a low event rate induces a large theoretical uncertainty in the total expected energy density. In particular, the contribution of this effect to the scale-dependence of the signal has a divergent formal expression (Jenkins & Sakellariadou, 2019; Cusin, Dvorkin, Pitrou, & Uzan, 2018a).

None of these GWBs have been detected yet. Still, if ever observed, they would be the direct analogues of electromagnetic backgrounds formed by the superposition of multiple astronomical signals. Examples of this type of backgrounds are the cosmic infrared background (CIB) (Hazumi et al., 2020), produced by stellar dust, and the cosmic X-ray background (CXB) (Fabian & Barcons, 1992), formed by numerous extragalactic X-ray sources.

The anisotropies of the astrophysical GWB have been extensively studied for years (Thrane et al., 2009) and, more recently, two independent groups *Cusin et. al.* (Cusin, Dvorkin, Pitrou, & Uzan, 2018b; Cusin, Pitrou, & Uzan, 2017) and *Jenkins et. al.*

(Jenkins & Sakellariadou, 2018; Jenkins, Sakellariadou, Regimbau, & Slezak, 2018) obtained discrepant predictions for the scale-dependent signal (Cusin et al., 2018a; Jenkins, Sakellariadou, et al., 2019). The main disagreements are related to the shape of the angular power spectrum as well as the overall amplitude of the signal. The difference in shape seems to be related to the treatment of non-linear scales (see also section 4.2 of this chapter), whereas the difference in amplitude is due to the chosen normalization. Here, let us mention that the main focus of their investigations so far has been the study of the autocorrelation signal and its shot-noise component, with further studies in this field being carried out also in (Cusin, Pitrou, & Uzan, 2018; Cusin, Dvorkin, Pitrou, & Uzan, 2020; Pitrou, Cusin, & Uzan, 2020; Cusin, Durrer, & Ferreira, 2019). It is, however, worth pointing out that signals beyond autocorrelation, such as the cross-correlation between GWB and galaxy clustering or weak lensing convergence, have also been modelled to some extent (see *e.g.* (Cusin, Dvorkin, Pitrou, & Uzan, 2019)).

Here, we study the cross-correlation between the anisotropies of the astrophysical GWB and galaxy clustering (GC), and argue why it represents the ideal observable to detect the background and measure its properties. There are three main reasons for this choice. First, the distribution of compact mergers forming the GWB is determined by the distribution of their host galaxies. This means that one should expect a relatively large correlation between the two signals. Second, the cross-correlation signal for diffuse backgrounds is expected to have a larger signal-to-noise ratio compared to the autocorrelation signal, hence the former is likely to be detected earlier (Ando, Benoit-Lévy, & Komatsu, 2014). Third, as presented in the next section, our investigation shows that the autocorrelation signal of the astrophysical GWB is very sensitive to small-scale structure, while the cross-correlation signal is free from this problem. In a somewhat similar spirit, Refs. (S. Mukherjee, Wandelt, & Silk, 2020b, 2020a; Calore, Cuoco, Regimbau, Sachdev, & Serpico, 2020) have recently studied the cross-correlation of *resolved* GW sources with large scale structure and lensed cosmic microwave background.

Our chapter is organized as follows. In section 4.2 we review the main aspects of the GWB autocorrelation signal and highlight its limitations. In section 4.3, we present the angular power spectrum of the cross-correlation signal and calculate the expected shot-noise contamination (section 4.5). In section 4.4 we demonstrate how the cross-correlation can be used to constrain the average power emitted by unresolved GW sources as a function of redshift, and quantify the required signal-to-noise ratio and angular resolution. To do this, we use a fiducial cosmology based on the best-fit results of Planck 2018 (Planck Collaboration, Aghanim, Akrami, Ashdown, et al., 2020b) and a toy model for the GWB. Finally, we present our conclusions in section 4.5.

4.2 Gravitational-Wave anisotropies

In this section, we discuss the autocorrelation signal of the anisotropic GWB. This signal, as well as the shot-noise contamination, have been extensively studied in

previous works (Jenkins & Sakellariadou, 2019; Jenkins, Romano, & Sakellariadou, 2019; Cusin, Dvorkin, et al., 2019). Here, we review the main aspects of modelling these and describe some particularities.

Our starting point is the definition of the dimensionless energy density of GWs from a given direction of the sky $\hat{\mathbf{r}}$, per unit solid angle:

$$\Omega_{\rm GW}(\nu_0, \hat{\mathbf{r}}) = \frac{\nu_0}{\rho_c} \frac{\mathrm{d}\rho_{\rm GW}(\nu_0, \hat{\mathbf{r}})}{\mathrm{d}\nu_0 \mathrm{d}^2\hat{\mathbf{r}}}, \tag{4.1}$$

where $\rho_{\rm GW}(\nu_0, \hat{\mathbf{r}})$ is the present-day energy density in GWs, ν_0 is the observed frequency and $\rho_c = 3H_0^2/8\pi G$ is the critical density of the Universe. Note that, from now on, we suppress the frequency dependence. We model this signal as

$$\Omega_{\rm GW}(\hat{\mathbf{r}}) \equiv \int dr\ r^2 \mathcal{K}(r) n(\vec{\mathbf{r}}), \tag{4.2}$$

where $n(\vec{\mathbf{r}})$ is the galaxy density field in comoving coordinates $\vec{\mathbf{r}}$, and $\mathcal{K}$ is the GW kernel that encodes the average contribution of a galaxy to Ω_{GW} as a function of comoving distance r. In practice, this includes information about the star formation history of the Universe and the properties of the emitting binary population. It is instructive to rewrite equation (4.2) in terms of the galaxy overdensity $\delta_{\rm g}(\vec{\mathbf{r}}) \equiv n(\vec{\mathbf{r}})/\bar{n}(r) - 1$, with $\bar{n}(r)$ being the average number density of galaxies, defined as $\bar{n}(r) \equiv \int \mathrm{d}^2\hat{\mathbf{r}} n(\vec{\mathbf{r}})/4\pi$. With this notation we have

$$\Omega_{\rm GW}(\hat{\mathbf{r}}) = \int dr\ r^2 \mathcal{K}(r) \bar{n}(r) \left(\delta_{\rm g}(\vec{\mathbf{r}}) + 1\right). \tag{4.3}$$

From this point, the angular power spectrum of the anisotropic GWB $C_\ell^{\rm GW}$ can be calculated to be

$$C_\ell^{\rm GW} = 4\pi \int_{k_{\rm min}}^{k_{\rm max}} \frac{dk}{k} |\delta\Omega_\ell|^2 \mathcal{P}(k) + B_\ell^{\rm GW}. \tag{4.4}$$

Here $\delta\Omega_\ell(k)$ is given by

$$\delta\Omega_\ell(k) = \int dr\ r^2 \mathcal{K}(r) \bar{n}(r) T_g(k, r) j_\ell\left(kr\right), \tag{4.5}$$

where T_g is the synchronous gauge transfer function relating the galaxy power spectrum to the primordial one $\mathcal{P}(k) = A_s \left(k/k_*\right)^{n_{\rm s}-1}$, and j_ℓ is the spherical Bessel function of order ℓ. Note that the galaxy bias is implicitly absorbed in T_g. Note also that in equation (4.5) we neglect relativistic corrections, as they are generally found to be below cosmic variance (Bertacca et al., 2019).

The term $B_\ell^{\rm GW}$ in the power spectrum is the shot-noise bias term introduced by the spatial and temporal shot-noise in the distribution of the individual events forming the GWB. Following (Jenkins & Sakellariadou, 2019), we write the shot-noise contribution in the kHz band as

$$B_\ell^{\rm GW} = \int dr\ \mathcal{K}^2(r) \bar{n}(r) r^2 \left[1 + \frac{1 + z(r)}{R(r) T_O}\right]. \tag{4.6}$$

Because of the low event rate in this frequency range, this noise contribution is inversely proportional to the average number of events per galaxy, written as the average redshifted event rate $(1+z)/R(r)$ multiplied by the observing time T_O. However, because the duration of the inspiral phase in the mHz band is much larger than any reasonable observing time, the contribution of the term $1/(R(r)T_0)$ is negligible in this case.

The GWB discussed here is an integrated signal. Because of this, the low-redshift objects might significantly contribute to the GWB. Indeed, the astrophysical models of (Cusin, Dvorkin, et al., 2019) suggest that the combination

$$\tilde{\mathcal{K}}(r) = \mathcal{K}(r)\bar{n}(r)r^2 \tag{4.7}$$

is not decaying to negligible values close to redshifts $z \sim 0$. This introduces two complications in the modelling.

The first is connected to the shot noise. To highlight this, we rewrite equation (4.6) as

$$B_\ell^{\rm GW} = \int dr\, \frac{\tilde{\mathcal{K}}^2(r)}{\bar{n}(r)r^2}\left[1 + \frac{1+z(r)}{R(r)T_O}\right]. \tag{4.8}$$

From this expression, it is clear that the shot-noise has a divergent expression due to low-redshift (low-r) contributions. To obtain a well-behaved prediction for the autocorrelation signal, this divergence can be suppressed if local events are excluded from the background. This is equivalent to setting a lower limit in the integral above different from zero.

Second, there exist a complication derived from the scale dependent part of the angular power spectrum (the first term in equation (4.4)), which is expected to receive non-negligible contributions from small, highly non-linear scales. To get some intuition about this feature, let us simplify our expression for the GWB angular power spectrum by using the so-called Limber approximation

$$j_\ell(x) \to \sqrt{\frac{\pi}{2\alpha}}\delta_{\rm D}\left(\alpha - x\right), \tag{4.9}$$

where $\delta_{\rm D}$ is the Dirac delta-function and $\alpha \equiv \ell + 1/2$. Using this in Eq. (4.5) and neglecting the bias term we obtain

$$C_\ell^{\rm GW} \approx \frac{2\pi^2}{\alpha}\int_{k_{\rm min}}^{k_{\rm max}} \frac{{\rm d}k}{k^3}\tilde{\mathcal{K}}^2\left(\frac{\alpha}{k}\right)\mathcal{S}^2\left(k, \frac{\alpha}{k}\right), \tag{4.10}$$

$$\mathcal{S}(k,r) \equiv T_g(k,r)\mathcal{P}(k)^{1/2}. \tag{4.11}$$

What equation (4.10) demonstrates is that $\tilde{\mathcal{K}}(r)$ acts as a modified kernel and selects a particular domain in the k-integral. This causes small scales to contribute significantly to $C_\ell^{\rm GW}$, unless $\tilde{\mathcal{K}}$ is vanishing at the lower end of its argument or $\tilde{\mathcal{S}}^2/k^3$ is falling fast enough at large values of k. As the modelling of the galaxy power spectrum at non-linear scales is highly uncertain, this feature is signalling a potential danger of using the autocorrelation signal as a probe of GW merger history or cosmology.

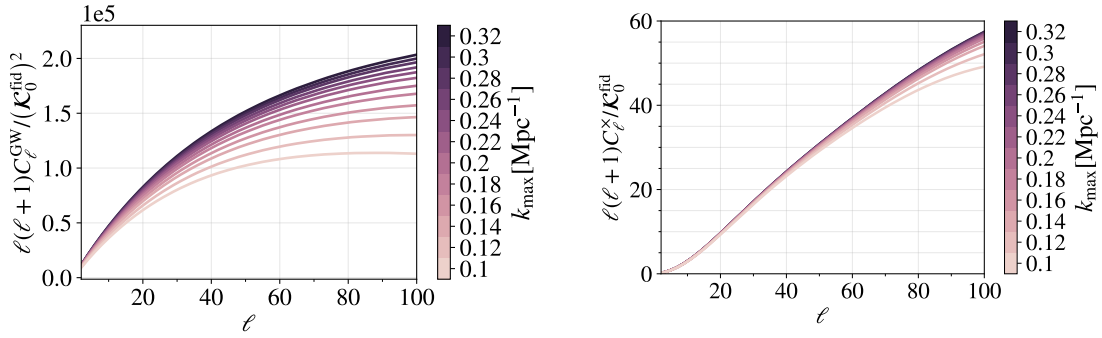

Figure 4.1: ***Left panel:*** Linear autocorrelation power spectra C_ℓ^{GW} of the GWB of a constant $\tilde{K}(r)$ for a set of upper limits of the integral in equation (4.4), in units of Mpc^{-1}. ***Right panel:*** The same as in the left panel, but for the cross-correlation between a galaxy sample (centered at $z = 0.5$) and the GWB, $C_\ell^{\times}$. Both of the panels are supposed to be understood as normalized with respect to the amplitude of the fiducial GWB model, to be discussed in detail later.

To accurately assess the impact of the issue mentioned above, let us turn to the results of exact numerical computations which do not rely on the Limber approximation. Having in mind the speed requirements of our later parameter analysis, we have developed a fast numerical procedure[1] to compute the integrals in equations (4.4) and (4.5), given the dark matter transfer function $T_{{\rm m}(k,r)}$ calculated using an Einstein-Boltzmann solver [2].

A technical remark is in order here. Given the rapidly-oscillatory nature of the spherical Bessel functions in equation (4.5), we have precomputed the line-of-sight integrals over these Bessel functions on bins of a fine r-grid. On the speed grounds, the source terms are then inserted only on a much coarser grid, which is only justified if these source functions do not vary significantly between two coarse-grid points. While this assumption is well justified for the transfer functions, we can only use our integrator if the kernel $\mathcal{K}(r)$ does not have rapid changes. In this chapter, we consider only such smooth-enough kernels (and window functions – see the next sections). We have verified the reliability of our integration procedure against a modified version of the latest public version of `CAMB` (Lewis et al., 2000; Howlett et al., 2012).

Our results are illustrated in the left panel of Figure 4.1, where we have chosen several values of $k_{\rm max}$, the upper limit of the integral in equation (4.4), and calculated the corresponding angular power spectra for the multipoles in the range $\ell = [2, 100]$. Note in particular that the magnitude of the signal changes drastically with $k_{\rm max}$, meaning that the autocorrelation signal depends heavily on the shape of the low redshift power spectrum on non-linear scales. This is likely one of the causes behind the

[1]The codes used in this chapter are publicly available at `https://github.com/valerivardanyan/GW-GC-CrossCorr`.

[2]In this chapter, we use the ΛCDM limit of the `EFTCAMB` code (Hu, Raveri, Frusciante, & Silvestri, 2014; Raveri, Hu, Frusciante, & Silvestri, 2014) for simplicity, as it is easier to output the required transfer functions as a function of redshift.

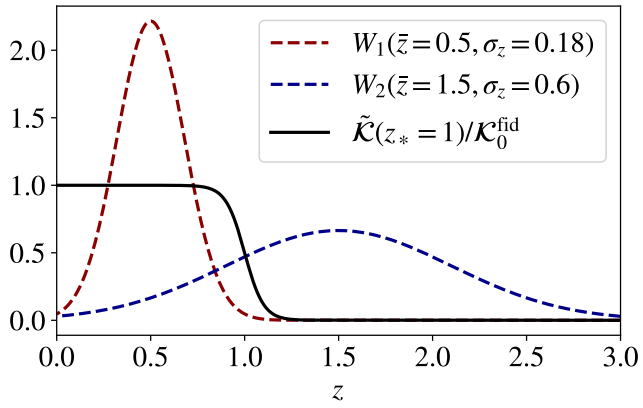

Figure 4.2: Fiducial model as a function of redshift z, of the GW source kernel $\tilde{\mathcal{K}}(r(z))$ in Eq. (4.7). In practice, we cut off the low-redshift sources with comoving distances smaller than 150 Mpc (see the text for details). The galaxy clustering window functions W_1 and W_2 are assumed to be Gaussian.

discrepancy between *Jenkins et. al.* and *Cusin et. al.* and suggests that an accurate prediction of the autocorrelation signal should take into account not only the shot-noise contribution (Jenkins, Romano, & Sakellariadou, 2019; Cusin, Dvorkin, et al., 2019), but also the uncertainties due to baryonic effects in the matter distribution at small scales (Debackere, Schaye, & Hoekstra, 2020; Schneider et al., 2019). We point out, in particular, that the galaxy catalogue based on dark-matter-only simulations of (Blaizot et al., 2005) and the `halofit` model of (Takahashi et al., 2012) are not designed to consistently or accurately model this uncertainty. While not shown, we point out that this problem is even more noticeable at high ℓ, where a larger value of $k_{\rm max} \sim 5$ Mpc^{-1} is required for the integrals to converge (as highlighted in (Cusin et al., 2018a)).

4.3 Cross-correlation with galaxy clustering

In this section, we introduce the main concepts necessary for modelling the cross-correlation signal and discuss its advantages.

First of all, we define the observed overdensity of galaxies in the given direction $\hat{\mathbf{r}}$ per unit sold angle as

$$\Delta(\hat{\mathbf{r}}) = \int dr\ W_i(r)\delta_g(\vec{\mathbf{r}}), \tag{4.12}$$

where $W_i(r)$ is the probability density function of the galaxies' comoving distances (also referred to as the *GC window function*) and $\delta_g(\vec{\mathbf{r}})$ is the galaxy overdensity defined earlier. Using equation (4.12), the angular power spectrum of GC, $C_\ell^{\rm GC}$, can be shown to be

$$C_\ell^{\rm GC} = 4\pi \int \frac{dk}{k} |\Delta_\ell(k)|^2 \mathcal{P}(k) + \frac{1}{n_i}, \tag{4.13}$$

where $\Delta_\ell(k)$ is given by

$$\Delta_\ell(k) = \int dr\ W_i(r) T_i(k,r) j_\ell(kr). \tag{4.14}$$

$T_i(k,r)$ is the transfer function for the galaxy overdensity in the selected redshift range $W_i(r)$, $j_\ell(kr)$ is the spherical Bessel function of order ℓ and n_i is the average number of galaxies per steradian, also dependent on the specific redshift selection $W_i(r)$. This final quantity appears in the second term in equation (4.13) and dictates the size of the shot-noise component of the power-spectrum.

Using equations (4.5) and (4.14), one can derive the angular power spectrum of the cross-correlation $C_\ell^\times$ of the GWB and the GC maps, given by equation (4.2) and (4.12). This is

$$C_\ell^\times = 4\pi \int \frac{dk}{k}\ \delta\Omega_\ell^*(k) \Delta_\ell(k) \mathcal{P}(k) + B_\ell, \tag{4.15}$$

where the shot-noise contribution B_ℓ, derived in Appendix 4.5, can be shown to be

$$B_\ell = \int dr\ W_i(r) \mathcal{K}(r). \tag{4.16}$$

With these expressions in mind, we can now discuss how the cross-correlation signal can be used to address the modelling challenges we have presented in the previous section.

To address the first one, we notice that, while the $1/r^2$ divergence is still present in the integral in equation (4.16), this integral is generally well behaved if the window function $W_i(r)$ decays fast enough at small redshifts. Notice that this is impossible to do in the equivalent expression for the autocorrelation in equation (4.6).

With respect to the second issue, we compare in Figure 4.1 the effects of the small-scale power spectrum on both the auto and cross-correlation. To explain the different behaviour, we note that the equivalent of equation (4.10) for the cross-correlation is

$$C_\ell^\times \approx \frac{2\pi^2}{\alpha} \int_{k_{\min}}^{k_{\max}} \frac{\mathrm{d}k}{k^3} W_i\left(\frac{\alpha}{k}\right) \tilde{\mathcal{K}}\left(\frac{\alpha}{k}\right) \mathcal{S}^2\left(k, \frac{\alpha}{k}\right). \tag{4.17}$$

Because GC surveys allow for redshift-selection of the sources, the GC window function $W_i(r)$ can be taken to be peaked at some non-zero redshift and quickly decaying for larger or smaller values of r. equation (4.17) shows that this behaviour cuts off the contribution from very large and very small scales, as shown in the right panel of Figure 4.1.

4.4 Information content

4.4.1 Model set-up

In this section, our primary goal is to explore the sensitivity of the cross-correlation signal to various parameters and estimate its information content. To this end, we model the signal using simple, but representative assumptions about the GW and GC

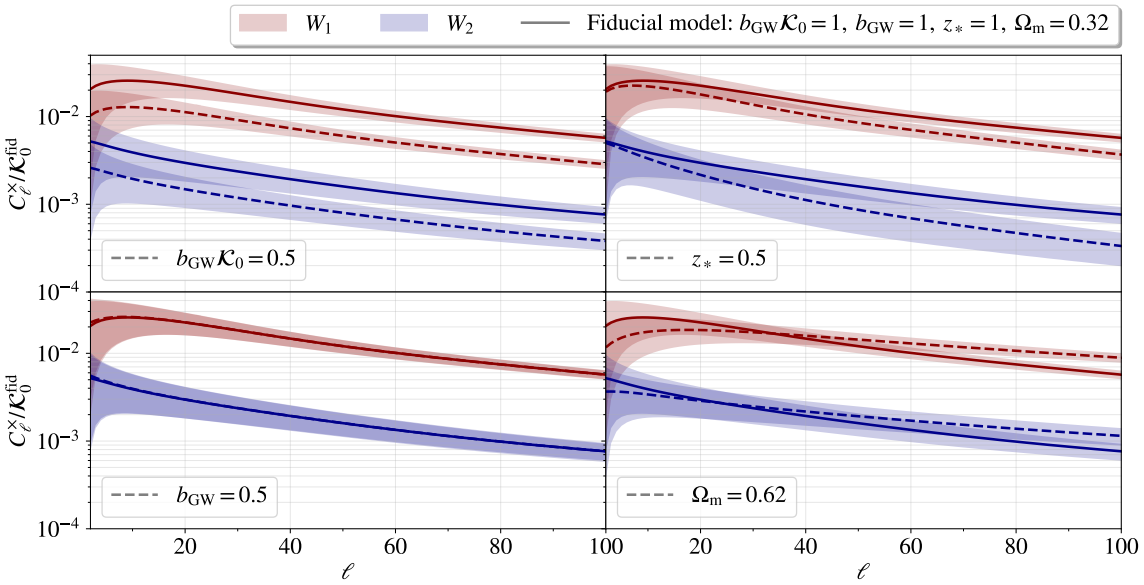

Figure 4.3: Effects of the model parameters $b_{\rm GW}\mathcal{K}_0, z_*, \Omega_{\rm m}$ and $b_{\rm GW}$ on the cross-correlation signal. The uncertainties are the cosmic variance defined in Appendix 4.5. Note particularly that in the case of both of the window functions W_1 and W_2 the change in $b_{\rm GW}\mathcal{K}_0$ induces a significant change in the amplitude of the signal (***upper left panel***), while when the combination $b_{\rm GW}\mathcal{K}_0$ is fixed, the signal is not sensitive to the value of the GW bias $b_{\rm GW}$ (***lower left panel***). Note that mostly the high-ℓ multipoles are sensitive to changes in z_* (***upper right panel***). Note also that the change on $\Omega_{\rm m}$ modifies the tilt of the signal, without altering its overall amplitude (***lower right panel***). All of the panels are supposed to be understood as normalized with respect to the amplitude of the fiducial GWB model.

maps. This allows us to derive an upper limit on the constraining power by assuming the theoretical minimum uncertainty due to cosmic variance.

We base our model for $\tilde{\mathcal{K}}(r)$ on the physically motivated one of *Cusin et. al.* (Cusin, Dvorkin, et al., 2019), by noting that their function $\mathcal{A}(z)$ is the analogue of our $\tilde{\mathcal{K}}(r)$ in redshift space. In this reference, in particular, it is shown that $\mathcal{A}(r)$ is a slowly-evolving function of redshift, and has a similar shape over a wide range of frequencies and assumptions about the source population (see their figures 19 and 13). Thus, we model the kernel as

$$\mathcal{K}(r) = \frac{\mathcal{K}_0}{2\bar{n}(r)r^2}\left\{\tanh\left[10(z_*(r) - z(r))\right] + 1\right\}, \tag{4.18}$$

where $\mathcal{K}_0$ is the amplitude of the kernel, z_* is a cut-off redshift, and $\bar{n}(r) \approx 10^{-1}$ Mpc^{-3} is the average comoving galaxy number density estimated using Figure 4 of (Schaye et al., 2015). We do not implement a redshift dependence for this quantity because its value is relevant only for the shot-noise component of the cross-correlation, found to be negligible in the cases considered here. In our fiducial model, we assume $z_* = 1$ (see Figure 4.2), as it is known by *Cusin et al.* that the astrophysical kernel

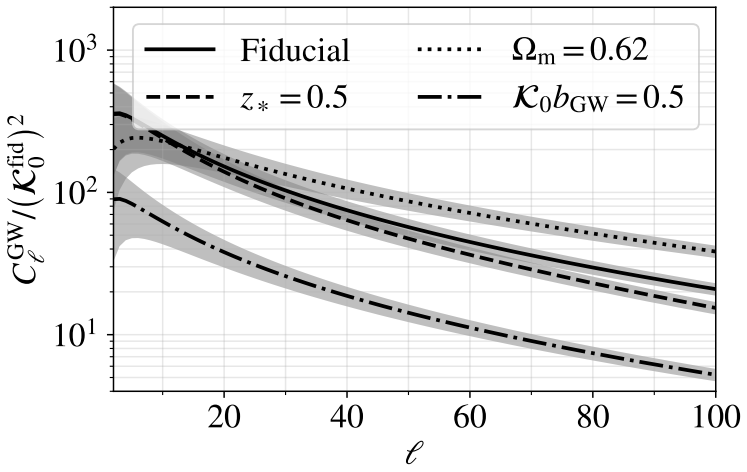

Figure 4.4: Effects of the model parameters $b_{\rm GW}\mathcal{K}_0, z_*, \Omega_{\rm m}$ and $b_{\rm GW}$ on the autocorrelation signal. The uncertainties are defined as in Figure 4.3. The curves should be understood as normalized with respect to the amplitude of the fiducial GWB model.

$\mathcal{K}(r)\bar{n}(r)r^2$ is expected to decay around that value in redshift. Notice that, while $\mathcal{K}_0$ should be dimensionful, its units are irrelevant to us because the cross-correlation signal is proportional to its value. For the rest of the chapter, we call $K_0^{\rm fid}$ the fiducial value of this quantity.

In the next subsections, we study the cross-correlation between the GWB modelled above and two galaxy catalogues centred at different redshifts. The two window functions, W_1 and W_2, are assumed to be Gaussian distributions centered at $\bar{z} = \{0.5, 1.5\}$ and with widths of $\sigma_z = \{0.18, 0.6\}$. These values are picked so that the two selections overlap with the constant portions of $\tilde{\mathcal{K}}(r)$.

Moreover, we model the transfer functions in equations (4.14) and (4.5) by using a linear bias approximation (valid for large scales):

$$T_i(k) = b_i T_{\rm m}(k, r), \tag{4.19}$$

and

$$T_g(k, r) = b_{\rm GW} T_{\rm m}(k, r), \tag{4.20}$$

where $T_{\rm m}(k, r)$ is the transfer function for cold dark matter and the b_X are known as bias parameters. When varying our model, we freeze the bias of both galaxy catalogues since it can be extracted from the clustering autocorrelation signal alone. On the contrary, we treat the GW bias $b_{\rm GW}$ as a free parameter and we assume it to be a constant over redshift. While this is not necessarily true, in the absence of shot-noise, only the combination $b_{\rm GW}\tilde{\mathcal{K}}(r)$ appears in the signal. This implies that a more complex model can always capture any redshift dependence through the function $\tilde{\mathcal{K}}(r)$. Note, however, that breaking the degeneracy between the linear bias of the GW population and the amplitude of the astrophysical kernel $\mathcal{K}(r)$ requires a full

understanding of the GWB kernel and all ingredients (Scelfo, Bellomo, Raccanelli, Matarrese, & Verde, 2018).

For the rest of the analysis, we focus on the mHz frequency band, and assume that low-redshift events (below $r = 150$ Mpc) can be filtered. In our modelling, as discussed in the previous sections, these assumptions are essential to obtain a well-behaved signal which is not overwhelmed by noise. For reference, under these assumptions we get the following relative noise values at $\hat{\ell} = 10$:

$$\frac{B_{\hat{\ell}}^{\mathrm{GW}}}{C_{\hat{\ell}}^{\mathrm{GW}}} \approx \frac{B_{\hat{\ell}}}{C_{\hat{\ell}}^{\times}} \approx 10^{-4}. \tag{4.21}$$

The first value is derived using the inspiral time of a solar mass black hole binary starting from 1 mHz (Blanchet, Damour, Iyer, Will, & Wiseman, 1995), an observing time of 1 year and a merger rate of 10^{-5} per year (B. P. Abbott et al., 2019).

As a summary of our model, Figure 4.2 contains the two window functions W_1, W_2 and the kernel $\tilde{\mathcal{K}}(r)$.

4.4.2 Behaviour of the cross-correlation

Before attempting to reconstruct the parameters of our model from mock data, let us gain some insights into the response of the cross-correlation signal on various parameters.

First, we explore the dependence of the signal on the kernel amplitude $\mathcal{K}_0$, or, more precisely, the combination $b_{\mathrm{GW}}\mathcal{K}_0$. In the upper left panel of Figure 4.3 we can see that in the case of both of the window functions W_1 and W_2 the change of the amplitude induces a significant change in the signal. Note that here the bias itself is fixed. In reality, the kernel amplitude $\mathcal{K}_0$ and the bias are perfectly degenerate with each other since the two appear as proportionality constants to both cross-correlation signals. To see this, in the lower left panel of Figure 4.3 we demonstrate the impact of varying b_{GW} on the signal when $b_{\mathrm{GW}}\mathcal{K}_0$ is held fixed. Note that a similar scaling with the kernel amplitude is present also for the autocorrelation signal shown in Figure 4.4, which is proportional to $(b_{\mathrm{GW}}\mathcal{K}_0)^2$.

Second, we turn our attention to the dependence of the signal on the turnover redshift z_*. In the upper right panel of Figure 4.3 we see that the change of z_* induces a change in the shape of the signal. The signal with W_2 is sensitive to z_*, while in the case of W_1 the signal is practically independent of it. A similarly small effect is also visible in the autocorrelation signal in Figure 4.4.

Third, it is interesting to show the effect of Ω_{m} on the signal. Specifically, in the lower left panel of Figure 4.3, it is demonstrated that the effects of Ω_{m} and $\mathcal{K}_0$ are qualitatively different from each other. Indeed, changing Ω_{m} rotates the signal, while $\mathcal{K}_0$ affects the amplitude of the signal. This rotation effect due to varying Ω_{m} is expected, as a similar effect is observed in the galaxy clustering autocorrelation signal. Indeed, such a behaviour in the signal allows galaxy clustering to constrain both Ω_{m} and the normalization of the matter power spectrum σ_8 (see e.g. (Euclid Collaboration et al., 2020)).

Finally, we point out that the scale-dependent power spectra discussed in this sections do not have a clear peak for any value of ℓ and practically do not show any sign of decaying power for small scales. This is in contrast to the naive expectations based on galaxy clustering result. This difference is due to the interplay between projected scales and redshift selection described in section 4.2, together with the use of relatively wide effective window functions ($\tilde{K}(r)$, W_1 and W_2).

4.4.3 Constraining $\mathcal{K}(r)$

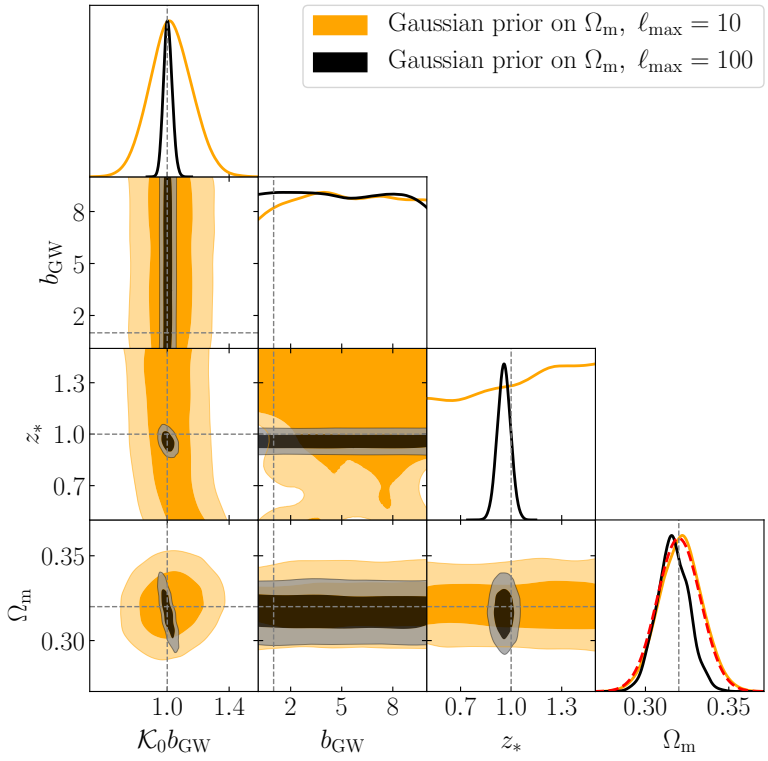

Figure 4.5: Posterior distributions for the cases of $\ell_{max} = 10$ (orange) and $\ell_{max} = 100$ (black), with Planck-2018-like Gaussian prior on Ω_m (shown in red dashed line). Contours represent the 68% and 95% confidence regions. We can clearly see that $\mathcal{K}_0 b_{GW}$ is constrained even in the case of the limited angular resolution, while b_{GW} is never separately constrained. The turn-over redshift z_* is unconstrained for the low-resolution case, while it is tightly constrained for the case of $\ell_{max} = 100$. Finally, Ω_m is prior dominated for the low-resolution case, while it beats the prior in the high-resolution scenario. Also noteworthy are the degeneracies between Ω_m and $\mathcal{K}_0 b_{GW}$, as well as between z_* and $\mathcal{K}_0 b_{GW}$.

The goal of this section is to understand the constraining power of the cross-correlation signal by studying how precisely the astrophysical model can be inferred from a noisy C_ℓ measurement.

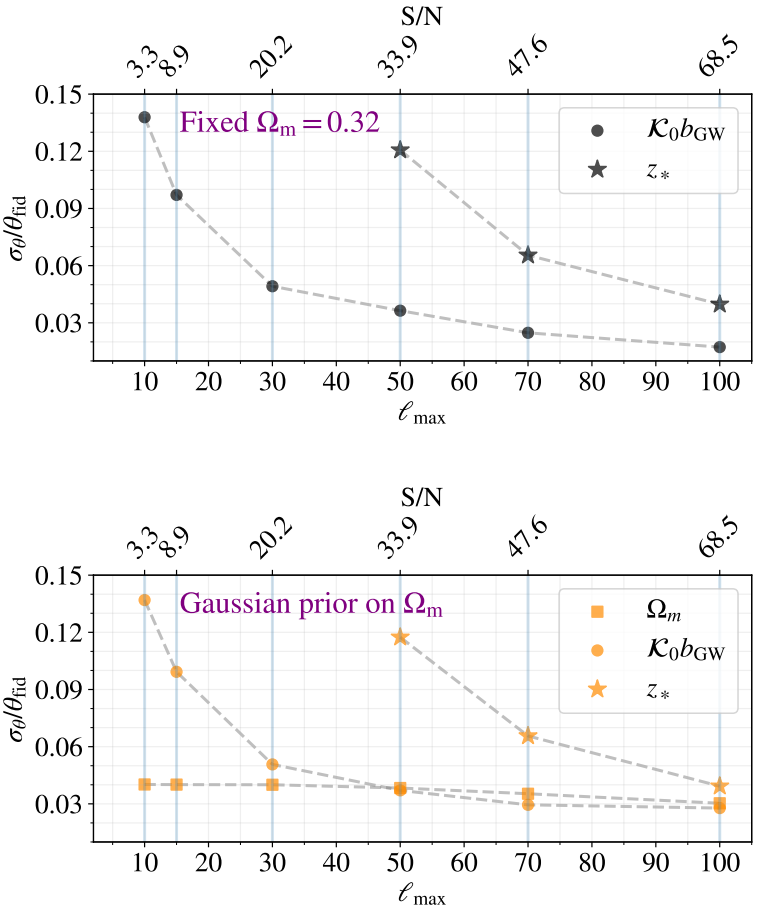

Figure 4.6: Constraints on the GWB parameters $(b_{\rm GW}, \mathcal{K}_0, z_*)$ and cosmology $(\Omega_{\rm m})$ obtained using the cross-correlation signal with two window functions as a function of the maximum multipole included in the analysis. Cosmic-variance limited measurements are assumed for all the constraints, so these should be understood as the best-case scenario results. Larger values of the signal-to-noise ratio (S/N) correspond to better angular resolution (see Eq. 4.22). We have explored the effect of $\Omega_{\rm m}$ on these constraints by either fixing its value (***left panel***), or setting a Planck-2018-like Gaussian prior (***right panel***). Remarkably, the combination $b_{\rm GW}\mathcal{K}_0$ can be constrained even with very limited angular sensitivity. The turnover location z_* is practically unconstrained for $\ell_{\rm max} \lesssim 50$, and $\Omega_{\rm m}$ is prior dominated for these multipoles. In case of $\ell_{\rm max} \gtrsim 50$ all the relevant parameters are tightly constrained, and for $\ell_{\rm max} \sim 100$ the constraints are at the level of a few percent. Notably, the cosmology (mimicked by varying $\Omega_{\rm m}$ in our analysis) can match and surpass the CMB results only in case of high angular resolution/signal-to-noise. For reference, $\ell_{\rm max} = 100$ roughly corresponds to 2 degrees.

In our analysis, we focus on the best-case scenario of cosmic-variance limited uncertainties as derived in Appendix 4.5 and use a simple proxy for the overall signal-to-noise ratio of the cross-correlation, defined as

$$\left(\frac{\mathrm{S}}{\mathrm{N}}\right)^2 \equiv \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \frac{\left(C_\ell^\times\right)^2}{\mathrm{Var}C_\ell^\times}. \tag{4.22}$$

Let us note that in our setup the GC signal dominates over the GC shot noise, implying that equation (4.22) is indeed the theoretical limit for uncertainties. In the presence of multiple, independent window functions, we simply sum the relative signal-to-noise expressions in quadrature.

We compute the cross-correlation power spectra, given in equation (4.15), using the model presented in subsection 4.4.1, and attempt to recover the model parameters from a noisy realization. To explore the inferred constraints as a function of angular resolution and S/N levels, we do this in several multipole ranges of ℓ with $\ell_{\min} = 2$ and varying $\ell_{\max}$.

The parameters of interest in our analysis are the amplitude of the GWB kernel $\mathcal{K}_0$ and the turnover redshift z_*. In addition to these, we also explore the bias b_{GW} and Ω_{m} to see if variations in $T_g(k, r)$ can affect the inferred $\mathcal{K}(r)$, and to explore the possible degeneracies between the GWB model and cosmology. To include the effects of varying Ω_{m} we have precomputed the dark matter transfer functions for a grid of Ω_{m} values, and have inferred the results for the intermediate values through nearest neighbour interpolation.

The exploration of the parameter space is carried out using the MCMC `python` code `emcee` (Foreman-Mackey, Hogg, Lang, & Goodman, 2013). We have employed a Gaussian likelihood function on C_ℓ with diagonal covariance matrix given through equation (4.35), and the prior ranges given in Table 4.1. Note that since we expect $\mathcal{K}_0$ to be degenerated with b_{GW}, we do not vary $\mathcal{K}_0$ itself, but rather vary the combination $b_{\mathrm{GW}}\mathcal{K}_0$.

Parameter	Fiducial value	Prior
$b_{\mathrm{GW}}\mathcal{K}_0$	1	$[0.01, 100]$
b_{GW}	1	$[0.1, 10]$
z_*	1	$[0.5, 1.5]$
Ω_{m}	0.32	$\mathcal{G}(0.32, 0.013)$

Table 4.1: Prior ranges of the sampled parameters. For Ω_{m} we use a Planck-2018 inspired Gaussian prior.

The main results of the analysis are summarised in Figure 4.6, where we show the expected constraints on the parameters of interest as a function of the maximum multipole included in the analysis. We also show the corresponding cosmic-variance-only signal-to-noise ratios.

Let us first have a look at the left panel of the figure, which corresponds to a fixed Ω_{m} value. As we see, $b_{\mathrm{GW}}\mathcal{K}_0$ is constrained and, notably, this is true even in

the limited multipole range corresponding to $\ell_{\rm max} = 10$. This is expected, as a clear detection of the signal is associated with a measurement of its amplitude. On the other hand, less encouraging are the results for the turnover redshift z_*, which can be constrained only for $\ell_{\rm max} \gtrsim 50$ or, equivalently, a S/N of ~ 33.

As we will prove shortly, the poor constraint on the amplitude is because, in the small multipole range, $\Omega_{\rm m}$ is very poorly constrained. At the same time, it is strongly degenerate with the amplitude.

In the right panel of the figure, we now impose a Gaussian prior on $\Omega_{\rm m}$, with its variance being comparable to the Planck-2018 constraint on $\Omega_{\rm m}$. While the z_* results are not affected, the uncertainties on the amplitude are now slightly inflated, due to a degeneracy between $\Omega_{\rm m}$ and $b_{\rm GW}\mathcal{K}_0$. This is also visible in the signal responses plotted in Figure 4.3.

Let us now fully concentrate on the two limiting angular sensitivities in our analysis. We take a LIGO-like angular sensitivity limited to the multipole range of $\ell \in [2, 10]$, as well as an angular sensitivity of a hypothetical high-resolution GW detector corresponding to $\ell \in [2, 100]$. The full constraints, for the case of Gaussian priors on $\Omega_{\rm m}$, are presented in Figure 4.5.

We can clearly see that $\mathcal{K}_0 b_{\rm GW}$ is constrained even in the case of the limited angular resolution, while $b_{\rm GW}$ is never separately constrained. We have checked that the latter feature is also present in all the other runs presented in this Section. This justifies our choice to vary the combination $b_{\rm GW}\mathcal{K}_0$ instead of varying $b_{\rm GW}$ and $\mathcal{K}_0$ separately.

The turn-over redshift z_* is unconstrained for the low-resolution case, while it is tightly constrained for the case of $\ell_{\rm max} = 100$. The dark matter abundance $\Omega_{\rm m}$ is prior dominated for the low-resolution case, while it beats the prior in the high-resolution scenario. Also noteworthy are the degeneracies between $\Omega_{\rm m}$ and $\mathcal{K}_0 b_{\rm GW}$, as well as between z_* and $\mathcal{K}_0 b_{\rm GW}$. These can be easily understood by inspecting the combined behaviours presented in Figure 4.3.

Before turning to our conclusions let us mention that the results presented in this section depend on the precise details of the GC window functions and GWB detection and more precise results can only be obtained by performing a realistic forecast with exact survey/detector specifications. While we leave a more detailed investigation for future research, our results suggest that a cosmic-variance limited measurement of the GWB anisotropies down to $\ell \sim 100$ is able to tightly constrain the redshift evolution of the GW kernel $\tilde{\mathcal{K}}$, as well as to provide Planck-like constraints on cosmological parameters.

4.5 Conclusions

In this chapter, we have discussed in detail the angular power spectrum of the cross-correlation between the GWB of astrophysical origin and GC.

We have shown that, contrary to the autocorrelation signal, the cross-correlation signal does not depend heavily on the small-scale galaxy power spectrum and hence is a more robust observational probe. To this point, we have also shown that the shot-

noise associated with this signal is small for realistic choices of the window functions W_i.

Then, armed with these results, we studied in detail the properties of the angular power spectra for a range of model parameters. In particular, we have shown how the signal is sensitive to the turnover redshift z_* of the GWB kernel, a combination of its amplitude and the bias $b_{\rm GW}\mathcal{K}_0$, as well as the dark matter abundance $\Omega_{\rm m}$. We have also shown that the signal is not separately sensitive to $b_{\rm GW}$ and $\mathcal{K}_0$. A summary of all of this is presented in Figure 4.3.

As one of the main goals of this chapter, we have performed a Bayesian parameter estimation using an MCMC sampling based on mock data with cosmic-variance-limited uncertainties. This choice allows us to provide an upper limit on the constraining power of this new observational probe (Figure 4.6). In particular, we have demonstrated that the cross-correlation signal is a powerful tool to constrain the properties of the GWB kernel $\mathcal{K}(r)$ if appropriate GC window functions are used. This is true even when marginalizing over uncertainties in the cosmology gravitational-wave bias.

We have quantified for the first time the need of high-resolution GW detectors in order to extract the full information content of the GWB of astrophysical origin. In particular, we have shown that both a high angular resolution and a high signal-to-noise ratio ($\ell \sim 100$, S/N ~ 70) are required to recover both the matter abundance $\Omega_{\rm m}$ and features of the kernel $\mathcal{K}(r)$ as a function of redshift. Note, in particular, that these requirements are far above the angular resolution of present-day and near-future detectors (roughly $\ell \lesssim 10$, and even $\ell \lesssim 4$ for LISA (Ungarelli & Vecchio, 2001; LIGO Scientific collaboration and VIRGO collaboration, 2019)). While this is not the priority of currently proposed third-generation detectors (Maggiore et al., 2020), it is worth noting that the advantages of high-resolution gravitational-wave astronomy are numerous and not limited to the study of this anisotropic background (J. Baker et al., 2019).

The case for studying the cross-correlation is strengthened by noticing that the anisotropies of the GWB in kHz band will most probably first be measured through cross-correlation with galaxy surveys, as the latter will provide a guiding pattern to be looked at in the noisy GW data. Given the promising nature of our results regarding the constraints of the GW kernel parameters and $\Omega_{\rm m}$, we believe that the cross-correlation between GW and GC has the potential to be a robust observational probe in the era of multi-messenger cosmology.

Appendix 1: Shot-noise for the cross-correlation signal

We follow (Jenkins & Sakellariadou, 2019) and evaluate the shot-noise contribution to the observed cross-correlation signal $C_\ell^\times$ in terms of the shot-noise contribution to the covariance between the observed maps $\Omega(\hat{\mathbf{r}})$ and $\Delta(\hat{\mathbf{r}}')$. Our starting point is

$$B_\ell = \int d^2\hat{\mathbf{r}} P_\ell(\hat{\mathbf{r}}\cdot\hat{\mathbf{r}}')\mathrm{Cov}[\Omega(\hat{\mathbf{r}}),\Delta(\hat{\mathbf{r}}')]_{\rm SN}. \tag{4.23}$$

By keeping in mind that $\tilde{\mathcal{K}}(r) = r^2\mathcal{K}(r)\bar{n}(r)$ and that $\delta_g(\vec{\mathbf{r}}) = (n(\vec{\mathbf{r}}) - \bar{n}(r))/\bar{n}$ we use the definitions in equations (4.2), (4.12) to write:

$$\begin{aligned}&\mathrm{Cov}[\Omega(\hat{\mathbf{r}}), \Delta(\hat{\mathbf{r}}')]_{\mathrm{SN}} = \\ &\int dr \int dr' \frac{r^2}{\bar{n}} \times \mathrm{Cov}[\mathcal{K}(r)n(\vec{\mathbf{r}}), W_i(r')n(\vec{\mathbf{r}}')]_{\mathrm{SN}}.\end{aligned} \tag{4.24}$$

As a side note, we point out that this expression is a stretch of notation since, formally, the quantities $\mathcal{K}(r)n(\vec{\mathbf{r}})$ and $W(r)n(\vec{\mathbf{r}})$ represent the mean values of the variables that we are trying to correlate. To proceed, we notice that $W(r)n(\vec{\mathbf{r}})$ is proportional to the number density of galaxies visible in the galaxy survey and that $\mathcal{K}(r)n(\vec{\mathbf{r}})$ is proportional to the number density of GW events around an infinitesimal volume centred in $\vec{\mathbf{r}}$. This is confirmed by the formalism used in the aforementioned references (Jenkins & Sakellariadou, 2019) and (Cusin et al., 2018b) to predict a realistic $\mathcal{K}(r)$.

In a finite volume δV_i we write down the number of GW mergers as

$$\Lambda_i = \sum_k^{N_i} \lambda_k, \tag{4.25}$$

where N is the number of galaxies present in this volume and the λ_j-s are the number of events for each galaxy. If we assume that N and λ_k are Poisson distributed, Λ_i follows a compound Poisson distribution with variance

$$\mathrm{Var}[\Lambda_i] = \langle\Lambda_i^2\rangle - \langle\Lambda_i\rangle^2 = \langle N_i\rangle \left(\langle\lambda\rangle + \langle\lambda\rangle^2\right). \tag{4.26}$$

If we call f the fraction of galaxies in the volume δV_j visible in the galaxy survey we also derive:

$$\mathrm{Cov}[fN_j, \Lambda_i] = f\langle N\rangle\langle\lambda\rangle\delta_{ij}, \tag{4.27}$$

where δ_{ij} is the Kronecker delta. By going back to the continuous case, we obtain the following result:

$$\mathrm{Cov}[\mathcal{K}(r)n(\vec{\mathbf{r}}), W_i(r')n(\vec{\mathbf{r}}')]_{\mathrm{SN}} = \bar{n}(r)W_i(r)\mathcal{K}(r)\delta^3(\vec{\mathbf{r}} - \vec{\mathbf{r}}'). \tag{4.28}$$

Finally, by plugging everything into equation (4.23) we obtain the result shown in the main text:

$$B_\ell = \int dr\ W_i(r)\mathcal{K}(r). \tag{4.29}$$

Appendix 2: Cosmic variance of the cross-correlation signal

Assume we have two maps on the sky, corresponding to the GWB and GC anisotropies. The angular decomposition coefficients $a_{\ell m}^{\mathrm{GW}}$ and $a_{\ell m}^{\mathrm{GC}}$ are assumed to

be Gaussian random variables with zero mean, and each m-mode is drawn from the same distribution. The relevant angular power spectra are defined as

$$C_\ell^\times \equiv \mathrm{Cov}\left[a_{\ell m}^{\mathrm{GW}}, a_{\ell m}^{\mathrm{GC}}\right], \tag{4.30}$$

$$C_\ell^{\mathrm{GW}} \equiv \mathrm{Var}\left[a_{\ell m}^{\mathrm{GW}}\right], \tag{4.31}$$

$$C_\ell^{\mathrm{GC}} \equiv \mathrm{Var}\left[a_{\ell m}^{\mathrm{GC}}\right]. \tag{4.32}$$

It is then trivial to construct an unbiased estimator of the cross-correlation power spectrum as

$$\widehat{C_\ell^\times} = \frac{1}{2\ell+1}\sum_{m=-\ell}^{+\ell} a_{\ell m}^{\mathrm{GW}} a_{\ell m}^{\mathrm{GC}}. \tag{4.33}$$

The variance of this estimator can then be shown to be

$$\begin{aligned}\mathrm{Var}C_\ell^\times = \frac{1}{(2\ell+1)^2}\sum_{m=-\ell}^{+\ell}\mathrm{Var}\left[a_{\ell m}^{\mathrm{GW}} a_{\ell m}^{\mathrm{GC}}\right] = \\ \frac{1}{(2\ell+1)^2}\sum_{m=-\ell}^{+\ell} C_\ell^{\mathrm{GW}} C_\ell^{\mathrm{GC}} + \\ \mathrm{Cov}\left[\left(a_{\ell m}^{\mathrm{GW}}\right)^2, \left(a_{\ell m}^{\mathrm{GC}}\right)^2\right] - \mathrm{Cov}\left[a_{\ell m}^{\mathrm{GW}}, a_{\ell m}^{\mathrm{GC}}\right]^2.\end{aligned} \tag{4.34}$$

In summary, we have

$$\mathrm{Var}C_\ell^\times = \frac{C_\ell^{\mathrm{GW}} C_\ell^{\mathrm{GC}} + \left(C_\ell^\times\right)^2}{2\ell+1}, \tag{4.35}$$

where we have used the Gaussianity of $a_{\ell m}$'s. Making the $a_{\ell m}^{\mathrm{GC}} \to a_{\ell m}^{\mathrm{GW}}$ replacement turns this expression into

$$\mathrm{Var}C_\ell^{\mathrm{GW}} = \frac{2\left(C_\ell^{\mathrm{GW}}\right)^2}{2\ell+1}, \tag{4.36}$$

which, of course, recovers the usual cosmic variance result.

Chapter 5

Reconstruction of the propagation and origin of gravitational waves

This chapter is based on:
Learning how to surf: Reconstructing the propagation and origin of gravitational waves with Gaussian Processes
Guadalupe Cañas-Herrera, Omar Contigiani and Valeri Vardanyan.
(September, 2021), The Astrophysical Journal, 918, 20, arXiv:2105.04262.

5.1 Introduction

The first direct detection of gravitational waves (GWs) by (B. P. Abbott et al., 2016) triggered a rapidly increasing interest in exploiting this new field for cosmological information. GWs alone are not particularly useful because the data provides only a sky position and a measure of the luminosity distance to the source. However, GW sources can serve as powerful cosmological probes when combined with electromagnetic (EM) data, from which redshifts can be extracted. This idea dates back to (Schutz, 1986) and has two main variations.

The first, simplest, possibility is the observation of so-called *standard sirens* (Holz & Hughes, 2005), GW sources with EM counterparts from which a redshift can be observed. As an example, a population of binary neutron stars, such as the already observed GW170817 (B. P. Abbott et al., 2017c), can be used to reconstruct the luminosity-redshift function and constrain cosmological observables. In this context, the aforementioned observation has already lead to promising results in constraining the Hubble constant $H_0 = 100h$ km/s/Mpc (B. P. Abbott et al., 2017b; H.-Y. Chen, Fishbach, & Holz, 2018).

The second possibility is to use *dark sirens* with a statistical counterpart. The likely counterpart is identified using the rough sky-localization offered by current GW detectors in combination with deep galaxy catalogs. This technique, sometimes combined with the first, has also been used to extract a measurement of H_0 (Soares-Santos et al., 2019; B. P. Abbott et al., 2021) and a refined version of it is the focus of this work. Assuming that both galaxies and the hosts of GW mergers are biased

tracers of the same cosmological structure, it is possible to measure a non-zero cross-correlation signal between the two (see section 5.3). At the linear level, in particular, the description of this signal is especially simple. It should be noted that this idea has been explored by others before, sometimes using a different formalism (Namikawa, Nishizawa, & Taruya, 2016; S. Mukherjee & Wandelt, 2018; Scelfo et al., 2018) and that similar methods have already been employed to provide a measurement of H_0 (Finke, Foffa, Iacovelli, Maggiore, & Mancarella, 2021).

In addition to these two methods, modelling of higher-order effects in the merger dynamics or knowledge of intrinsic source distribution can also be used to extract the luminosity distance-redshift relation from GW observations alone. Because GW signals provide a measurement of the redshifted chirp-mass of the system, features dependent only on the mass of the system (e.g. tidal effects) can be used to extract the cosmological redshift present in the signal (Messenger & Read, 2012). Furthermore, expected features in the source distribution, such as, e.g. the pair instability mass gap (Farr, Fishbach, Ye, & Holz, 2019) or the peak of the star formation rate (Ye & Fishbach, 2021) can also be used to extract this relation from catalogs of GW events.

We also mention that, in principle, the same idea can be used in the absence of resolved events. Previous works, however, have shown that a detection of this effect from a stochastic background signal is unlikely to happen soon due to the low signal-to-noise ratio (SNR) attainable by current and proposed experiments covering the optimal wavelength range (Cañas-Herrera et al., 2020; Alonso, Cusin, Ferreira, & Pitrou, 2020).

In this chapter, we discuss the possibility of using the non-zero correlation between the distribution of EM galaxies and resolved GW mergers to jointly extract information about the two main quantities in the field of GW cosmology: the luminosity distance as a function of redshift, describing the propagation of gravitational waves across cosmic time, and the linear bias of GW sources with respect to the underlying dark matter distribution of the Universe, describing their clustering properties. The linear bias, in this context, is a multiplicative factor relating the spatial correlation function of a source population to the correlation function of the matter distribution of the Universe.

It is already established that GWs carry the potential of constraining the fundamental laws of gravity. This is the case because propagation of GWs in modified gravity scenarios differs from predictions of the general theory of relativity (GR) in multiple ways (Deffayet & Menou, 2007; Garoffolo, Tasinato, Carbone, Bertacca, & Matarrese, 2020; Ezquiaga & Zumalacárregui, 2020). One of the clear signatures is the speed of tensor modes which can be both sub- and superluminal as opposed to the GR case, where GWs propagate at the speed of light. The tight constraints on speed deviations imposed by the multimessenger observations of GW170817 (B. P. Abbott et al., 2017a), for example, have ruled out a wide parameter space of otherwise viable scalar-tensor theories of gravity (Pettorino & Amendola, 2015; Lombriser & Taylor, 2016; Sakstein & Jain, 2017; Ezquiaga & Zumalacárregui, 2017; Creminelli & Vernizzi, 2017; T. Baker et al., 2017). Similarly, implications for bi-gravity models have also been demonstrated in the literature (Amendola, Könnig, Martinelli, Pet-

torino, & Zumalacarregui, 2015; Max, Platscher, & Smirnov, 2017, 2018; Akrami, Brax, Davis, & Vardanyan, 2018, June; Belgacem et al., 2019).

Another striking difference between modified gravity and GR is the modified friction of GWs (Amendola, Sawicki, Kunz, & Saltas, 2018; Belgacem, Dirian, Foffa, & Maggiore, 2018). This feature arises in models with non-minimal coupling of a scalar field and gravity which manifests itself as a redshift-dependent gravitational coupling. As a result, the inferred luminosity distance to GW sources differs from the corresponding EM luminosity distance. This interesting phenomenon has already been constrained using the aforementioned multimessenger detection of GW170817 (Arai & Nishizawa, 2018; Lagos, Fishbach, Landry, & Holz, 2019) and the mass distribution of existing GW catalogs (María Ezquiaga, 2021). In this work, we will investigate the possibility of testing this hypothesis using the spatial clustering of GW sources and galaxies.

The discovery of the first LIGO-Virgo binary black hole has been used to motivate alternative scenarios where the binary did not represent the endpoint of stellar evolution, but originated either as a pair of primordial black holes (PBH) or some exotic compact objects (Bird et al., 2016; Sasaki, Suyama, Tanaka, & Yokoyama, 2016; Bustillo et al., 2021). In particular, the last half-decade has seen a resurgence in interest for PBHs (Clesse & García-Bellido, 2017; Sasaki et al., 2018; Raccanelli, Kovetz, Bird, Cholis, & Munoz, 2016; Raccanelli, Vidotto, & Verde, 2018). The main difference between the PBH and stellar evolution scenarios is the spatial distribution of GW mergers, measurable both in the redshift and sky distribution of the sources. In subsection 5.3.1 we expand on how to model this difference through the linear bias and present our model.

In addition to presenting our formalism, we investigate a possible method to precisely reconstruct the redshift evolution of clustering and modified gravity effects in the upcoming decades. Our study is based on Gaussian processes (GPs), a well-known hyper-parametric regression method (Rasmussen & Williams, 2005). The structure and implementation of this pipeline are presented in section 5.4.

GPs have been widely used in the literature to reconstruct the shapes of physical functions such as the dark energy equation of state $w(z)$ (Shafieloo, Kim, & Linder, 2012; Gerardi, Martinelli, & Silvestri, 2019), the primordial inflaton's speed of sound (Cañas-Herrera, Torrado, & Achúcarro, 2021) or the mass function of the merging binary black hole systems (Li et al., 2021). GPs are very useful for such functional reconstructions due to their flexibility and simplicity, relying only on a handful of hyper-parameters. In general, binned reconstructions (Crittenden, Pogosian, & Zhao, 2009; Crittenden, Zhao, Pogosian, Samushia, & Zhang, 2012; Zhao, Crittenden, Pogosian, & Zhang, 2012), as well as parametric reconstructions with high-degree polynomials are similar in spirit to GPs while also offering more flexibility in the extrapolation (Colgáin & Sheikh-Jabbari, 2021). The chapter is organized as follows. In section 5.2 we summarize the essential concepts concerning the GW propagation in modified gravity models. In section 5.3 we detail the modelling of the clustering correlations between the GW source population and galaxies. In section 5.4 we explain our reconstruction pipeline for the GW luminosity distance and the bias. Our findings are presented in section 5.5 and further discussed in section 5.6.

Unless stated otherwise, our fiducial cosmology is based on the best fit results from Planck 2018 (Planck Collaboration, Aghanim, Akrami, Arroja, et al., 2020). In our analysis, we use `COLOSSUS` (Diemer, 2018) and `Astropy` (Astropy Collaboration et al., 2013, 2018) for cosmological calculations, `sklearn` (*Scikit-learn 0.19.1 documentation: Gaussian Processes*, 2018) for the GP implementation, `emcee` (Foreman-Mackey et al., 2013) as our posterior sampler and `GetDist` (Lewis, 2019) to plot the final contours. Our analysis pipeline is made publicly available, see (Vardanyan, Cañas-Herrera, & Contigiani, 2021).

5.2 Gravitational wave propagation

In GR, and around a background Friedmann-Lemaitre-Robertson-Walker (FLRW) metric, the amplitude of GWs evolves according to

$$h''_\alpha + 2\mathcal{H}h'_\alpha - \vec{\nabla}^2 h_\alpha = 0, \tag{5.1}$$

where h_α denotes the amplitude of either polarization ($\alpha \in [\times, +]$), primes denote derivatives with respect to the conformal time, and $\mathcal{H}$ is the conformal Hubble function. In this equation, the prefactor of the Laplacian term controls the propagation speed, which we have set to coincide with the speed of light in $c = 1$ units.

The second term is the standard cosmic friction term which causes the strain amplitude to decay as $h_{\alpha(z)} \propto D_{\rm L}^{-1}(z)$, with D_L being the FLRW luminosity distance:

$$D_{\rm L(z)} = (1+z)\int_0^z \frac{{\rm d}\tilde{z}}{H(\tilde{z})}, \tag{5.2}$$

where the Hubble function $H(z)$ is given in terms of the Hubble constant H_0, present-day dark matter abundance $\Omega_{\rm m}$ and dark energy abundance $\Omega_{\rm DE(z)}$ as

$$H(z) = H_0\left[\Omega_m(1+z)^3 + \Omega_{DE}(z)\right]. \tag{5.3}$$

Throughout this chapter we assume a constant equation of state w_0 for dark energy, such that its energy density is given by

$$\Omega_{DE}(z) = (1-\Omega_{\rm m})(1+z)^{3(1+w_0)}. \tag{5.4}$$

The standard ΛCDM cosmology corresponds to $w_0 = -1$.

It is now established that modifications of GR can affect the propagation of GWs. The important effect for us is the modified friction term with respect to the GR expectation in equation (5.1),

$$h''_\alpha + \left[2 + \alpha_M(z)\right]\mathcal{H}h'_\alpha - \vec{\nabla}^2 h_\alpha = 0, \tag{5.5}$$

where we have again imposed the GW speed to be unity as suggested by observations. The modified friction term introduces a new scaling $h_{\alpha(z)} \propto 1/D_{\rm L,GW}(z)$, with

$D_{\rm L,GW}(z) \neq D_{\rm L(}z{\rm)}$ for non-zero $\alpha_M(z)$. The luminosity distance to GW events can be written as:

$$\frac{D_{\rm L,GW}}{D_{\rm L,EM}}(z) = \exp\left\{-\frac{1}{2}\int_0^z {\rm d}\tilde{z}\ \frac{\alpha_M(\tilde{z})}{(1+\tilde{z})}\right\}. \tag{5.6}$$

In this work, we assume that the luminosity distance for EM sources $D_{\rm L,EM}$ is unaffected and is equal to the expression in equation (5.2). The function α_M corresponds to the running of the effective Planck mass, i.e.,

$$\alpha_M = \frac{d\log(M_{\rm eff}/M_{\rm P})^2}{d\log a}, \tag{5.7}$$

where $M_{\rm P}$ is the Planck mass and $M_{\rm eff}$ is its effective value at redshift $z = 1/a - 1$. This function encodes information about extensions of GR such as scalar-tensor theories (Horndeski, 1974; Bellini & Sawicki, 2014) or, more broadly, quantum gravity (Calcagni et al., 2019). The modified friction term is also a natural prediction of non-local modifications of gravity (Dirian, Foffa, Kunz, Maggiore, & Pettorino, 2016)

From an effective field theory point of view $\alpha_M(z)$ is a free function of order unity. In practical studies of modified gravity and dark energy, however, α_M is often assumed to take simple parametric forms. The main guiding principle is the assumption that its effects should be negligible in the early universe, which prompts to choose $\alpha_M(z)$ to be proportional either to the dark energy abundance or simply to some power of the scale factor a.

Such parametrizations make it possible to find a closed form expression for the ratio in equation (5.6) and have inspired a widely used parametrization of the ratio as a monotonic deviation which goes to 1 at present day (Belgacem et al., 2018)

$$\frac{D_{\rm L,GW}}{D_{\rm L,EM}}(z) = \Xi_0 + \frac{1-\Xi_0}{(1+z)^n}. \tag{5.8}$$

In this expression, Ξ_0 and n are two constant parameters, which are typically ~ 1.

5.3 Angular power-spectra

5.3.1 GW sources

We consider GW mergers with a distribution in redshift written as

$$n_{\rm GW}(z) = \frac{n_0}{1+z}, \tag{5.9}$$

where n_0 corresponds to the comoving number density of observed events as a function of redshift, and the term $(1+z)$ takes into account the cosmological time dilation. In our analysis, we use a constant value of $n_0 \approx 3\times10^{-6}\ h^3{\rm Mpc}^{-3}$ (with h denoting the usual normalized Hubble constant), motivated by current LIGO constraints (R. Abbott et al., 2021).

For a given selection of sources along the line of sight, the average number of projected sources can be written using the comoving distance $\chi(z)$:

$$\bar{n}_{\rm gw} = \int_0^\infty dz \frac{\chi^2(z)}{H(z)} S(z) n_{\rm GW}(z). \tag{5.10}$$

The function S encodes the selection and the scatter due to observational errors. In this chapter, simple bins in a range $[D_{\rm L,min}, D_{\rm L,max}]$ are used and we assume a lognormal distribution with fixed scatter $\sigma_{\ln D}$ for the individual sources (Oguri, 2016). In this case, S can be written as:

$$S(z) = \frac{1}{2}\left[x_{\rm min}(z) - x_{\rm max}(z)\right], \tag{5.11}$$

with

$$x_{\rm min}(z) = {\rm erfc}\left[\frac{\ln D_{\rm L,min} - \ln D_{\rm L,GW}(z)}{\sqrt{2}\sigma_{\ln D}}\right], \tag{5.12}$$

and similarly for $x_{\rm max}$. Including this effect makes S resemble a top-hat function with damping tails dictated by $\sigma_{\ln D}$.

The angular power spectrum of these sources can be written using the Limber approximation

$$\begin{aligned} C_{\rm GW}(\ell) = \int_0^\infty dz\ & \frac{H(z)}{\chi^2(z)} W^2_{\rm GW}(z) \\ & b^2_{\rm GW}(z) P\left(\frac{\ell + 1/2}{\chi(z)}, z\right), \end{aligned} \tag{5.13}$$

where $P(k, z)$ is the matter power-spectrum at redshift z and comoving scale k, $b_{\rm GW}$ is the bias of the GW sources, and the window function can be written as

$$W_{\rm GW}(z) = \frac{\chi^2(z)}{H(z)} \frac{n_{\rm GW}(z)}{\bar{n}_{\rm GW}} S(z). \tag{5.14}$$

For the purpose of illustration, we will make use of a few simple parametrization for the GW bias. We will consider either a constant bias $b_{\rm GW}$ with a value of order unity or a more complex form:

$$b_{\rm GW}(z) = b_0\left(1 + \frac{1}{D(z)}\right), \tag{5.15}$$

where $D(z)$ represents the growth factor. The first model, with its low constant value, mimics a PBH origin for the mergers (Bird et al., 2016; Raccanelli et al., 2016), while the second mimics the stellar evolution case by tracking the galaxy linear bias (Oguri, 2016).

5.3.2 Galaxies

Similarly to the GW population, we again assume a constant comoving number density of galaxies. Throughout our analysis we fix

$$n_{\rm gal}(z) = 10^{-3} h^3 {\rm Mpc}^{-3}, \tag{5.16}$$

and we write the autocorrelation signal of galaxies under the Limber approximation as

$$\begin{aligned} C_{\rm gal}(\ell) = \int_0^\infty dz \, \frac{H(z)}{\chi(z)^2} W_{\rm gal}^2(z) \\ b_{\rm gal}^2(z) P\left(\frac{\ell + 1/2}{\chi(z)}, z\right). \end{aligned} \tag{5.17}$$

In this expression the definition of $W_{\rm gal}$ is the same as $W_{\rm GW}$ used in the previous section except for using $n_{\rm gal}(z)$, a different selection function, and $b_{\rm gal}(z)$ is the linear galaxy bias. In our analyses, we assume a known galaxy bias in the form of

$$b_{\rm gal}(z) = 1 + \frac{1}{D(z)}. \tag{5.18}$$

In general, this function is expected to be accurately measured from the galaxy autocorrelation signal alone.

In this chapter, we employ a top-hat selection function for $W_{\rm gal}$, which assumes no uncertainty in galaxy redshift estimates. This choice mimics a spectroscopic galaxy survey or a general redshift survey with negligible uncertainties. As an example, another choice commonly found in the literature is a Gaussian distribution $\mathcal{N}(z, \sigma_{\rm gal})$, where $\sigma_{\rm gal}$ should be much larger than the expected redshift uncertainty for each individual galaxy.

By combining the distribution of GW sources and galaxies one can construct a cross-correlation map. In our formalism, we write the cross-correlation between a GW bin i and a galaxy bin j (fully specified by their respective window functions) as:

$$\begin{aligned} C_\times^{ij}(\ell) = \int_0^\infty {\rm d}z \frac{H(z)}{\chi^2(z)} W_{\rm GW}^i(z) W_{\rm gal}^j(z) \\ \times b_{\rm GW}(z) b_{\rm gal}(z) P\left(\frac{\ell + 1/2}{\chi(z)}, z\right). \end{aligned} \tag{5.19}$$

We conclude this section by pointing out that the power spectra in equation (5.17), (5.13) and (5.19) do not include relativistic terms and do not capture the effects of evolution and lensing bias (see, e.g., (Scelfo et al., 2018; Scelfo, Boco, Lapi, & Viel, 2020) for a detailed treatment). Specifically, while the effects of lensing are expected to be negligible at the redshifts considered here (Oguri, 2016; Contigiani, 2020; S. Mukherjee et al., 2020b), the same is not true for relativistic effects. Therefore, we choose not to consider small values of ℓ in our analysis since the signal at these large angular scales is dominated by them.

5.4 Reconstructing GW physics

The primary goal of the chapter is to demonstrate how to reconstruct the properties of GW propagation and source clustering as a function of redshift. We do so by showing how to recover an assumed fiducial model by using mock angular power spectra with cosmic-variance or shot-noise limited uncertainties.

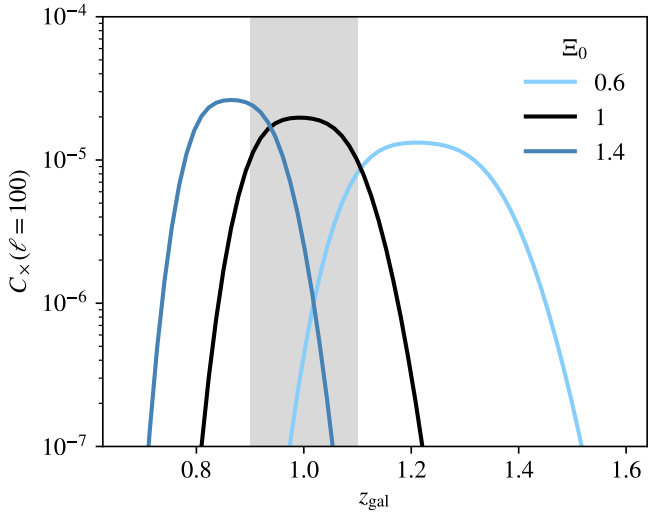

Figure 5.1: The cross-correlation signal between GW sources at $z = [0.9, 1.1]$ (shaded area) and galaxies at different redshifts ($z_{\rm gal}$). If the luminosity distance ratio $D_{\rm L,GW}/D_{\rm L,EM}(z)$ in equation (5.8) is different from its GR assumption ($\Xi_0 \neq 1$), the location of the predicted cross-correlation peak is also affected.

Our methodology hinges on the fact that by cross-correlating a GW luminosity distance bin with multiple galaxy redshift bins we can determine the redshift of the GW sources by matching the clustering properties of the two at the true redshift (Oguri, 2016; Bera, Rana, More, & Bose, 2020).

We demonstrate this idea in Figure 5.1, where we have considered GW sources located at redshift $[0.9, 1.1]$ in a GR cosmology where $D_{\rm L,GW}(z) = D_{\rm L,EM}(z)$. In this figure, we show the expected cross-correlation signal between the angular distribution of these sources and the angular distribution of galaxies located at various redshifts. As expected, in GR ($\Xi_0 = 1$) the signal peaks inside the correct redshift range (shaded area). However, as we depart from the GW luminosity distance relation, the location of this peak is affected.

5.4.1 Mock data and fiducial model

In this section, we describe the recipe used to generate the mock angular power spectra ($C_{\rm gal}, C_{\rm GW}$, and $C_\times$) that are fed into our reconstruction pipeline together with their error covariance matrix. When describing real data, these angular power-spectra are extracted from the autocorrelation and cross-correlation maps representing the

sky distribution of galaxies and GW sources. The recipe has three main ingredients: the details of the fiducial model, a description of the instrumental configuration and a definition of the dominant source of error.

The first ingredient is the fiducial model. Our decision in this case is based on the results of (T. Baker & Harrison, 2021), where present-day constraints on the function α_M appearing in equation (5.6) are presented. As shown in (Belgacem et al., 2019), the results of the $\alpha_M \propto a$ parametrization found in that work can be mapped to the $\Xi(z)$ function in equation (5.8). Using this transformation, we find that the 3σ upper limit roughly corresponds to

$$\Xi_0 \lesssim 1.4, \tag{5.20}$$

with $n = 1$. Thus, we assume a fiducial model with $\Xi_0^{\rm fid} = 1.4$ and $n^{\rm fid} = 1$, representing the limit of our present understanding.

The second ingredient of our forecast is the instrumental configurations. The size of our data vector is given by the number of multipoles ℓ and window functions that we include in our analysis. Since both are largely dictated by observational considerations, in this work we assume an optimistic combination of a network of three Einstein Telescopes (Maggiore et al., 2020; Hall & Evans, 2019) capable of a log-scatter in measured $D_{\rm L,GW}$ of $\sigma_{\ln D} = 0.05$, and a high-z redshift survey with large sky coverage and negligible redshift uncertainties such as SKA (Weltman et al., 2020).

The range of angular scales that we consider is limited by two factors. On small scales, large multipoles ($\ell > 100$) are excluded due to the angular resolution of about 1 degree expected for our GW detector configuration of choice (Hall & Evans, 2019). On large scales, we do not explore values of $\ell < 10$ because our modelling does not take into account the relativistic effects dominating the signal at these scales. Nevertheless, we stress that these multipoles contribute relatively little information compared to larger multipoles since they are dominated by cosmic variance.

Our window functions are distributed in the redshift range $[0.1, 3]$. We assume $N_{\rm gal} = 12$ galaxy bins equally spaced in redshift, and $N_{\rm GW} = 8$ GW luminosity distance bins equally spaced in $D_{\rm L,GW}$. We mention in particular that this choice is not completely arbitrary. The number of GW bins is motivated by forcing well-defined bins such that their width is at least three times the luminosity distance uncertainty $\sigma_{\ln D}$ that we have assumed. Furthermore, we have also verified that the exact number of galaxy bins does not dominate our results as long as $N_{\rm gal} > N_{\rm GW}$.

As for the last ingredient, we assume cosmic-variance or shot-noise limited uncertainties. In this case, we can write the covariance matrix of the auto-correlation and cross-correlation signals defined in equation (5.13), (5.17) and (5.19) as the following:

$$\begin{aligned}\mathrm{Cov}\left[C^{ij}(\ell)C^{mn}(\ell')\right] = \frac{\delta_{\ell\ell'}}{(2\ell+1)f_{\rm sky}}\times \\ \left(\tilde{C}^{im}\tilde{C}^{jn} + \tilde{C}^{in}\tilde{C}^{jm}\right),\end{aligned} \tag{5.21}$$

where the indices i, j, m, n can represent both galaxy or gravitational wave bins. The terms $\tilde{C}^{im}$ contain the shot-noise contribution when they represent the autocorrela-

tions in the same bin:

$$\tilde{C}^{im}(\ell) = C^{im}(\ell) + \frac{\delta_{im}}{\bar{n}}, \tag{5.22}$$

where $\bar{n}$ is the average density of projected objects from equation (5.10). In this work, we assume a survey covering a sky fraction equal to $f_{\rm sky} = 0.5$.

Let us point out that we do not use the cross-bin correlations for the bins of the same type as a signal. However, we properly take into account the $\tilde{C}^{im}$ terms for overlaps between two different GW bins. Similar terms for galaxy bins are completely negligible as there are no overlaps between the spectroscopic redshift bins.

For the setup described in this section, we find a total SNR of the GW-gal and GW-GW angular power-spectra of ~ 37. This value is dominated by the GW-gal cross-correlations since the GW-GW auto-correlations are not well measured (SNR$\lesssim 6$).

To generalize our choices, in subsection 5.5.2 we expand on how different combinations of instrumental specifications can affect the precision of the reconstruction.

5.4.2 Reconstruction Pipeline

In this section, we describe how the mock data presented in the previous section can be used to reconstruct $D_{\rm L,GW}$ and $b_{\rm GW}$ as a function of z. For ease of interpretation and visualization, in our analysis we do not fit these functions directly, but instead focus on the ratios $D_{\rm L,GW}/D_{\rm L,EM}\,(z)$ and $b_{\rm GW}/b_{\rm gal}\,(z)$. We point out that we do not marginalize over different possibilities for $b_{\rm gal}\,(z)$. This is because we are not interested in exploring the properties of the galaxy population, which are expected to be very well constrained by the galaxy-galaxy correlation signal alone.

As emphasized earlier, our approach makes use of a GP regression. This method is often used when the function of interest is directly measured at certain redshifts. These measurements are used as a training sample of the GP model, which then can predict the values of the reconstructed function at redshifts lacking any direct measurements (Belgacem, Foffa, Maggiore, & Yang, 2020; Renzi et al., 2021; P. Mukherjee & Mukherjee, 2021; Levi Said, Mifsud, Sultana, & Zarb Adami, 2021; Perenon et al., 2021).

However, this is not directly applicable for our current problem as neither the $D_{\rm L,GW}/D_{\rm L,EM}\,(z)$ nor the $b_{\rm GW}/b_{\rm gal}\,(z)$ are directly observable. Instead, these functions determine the auto- and cross-angular power spectra, which constitute our direct observables. In order to use GPs for our problem, we consider a certain number of redshift nodes for the two functions, referred to as *training nodes* with a slight abuse of terminology. The amplitudes of the nodes are free and, given a node configuration, we consider GPs which pass through all of these nodes exactly. To render our scenario computationally feasible and not consider many functions for each node configuration, we use the GPs regressor of the `python` package `sklearn` to output the best fit and use this as our function.

Our use of GPs can be thought of as a binning of the functions of interest in redshift space, and imposing a certain prior correlations between the bins. These correlations are specified by the GP kernel function, which in our case is chosen to

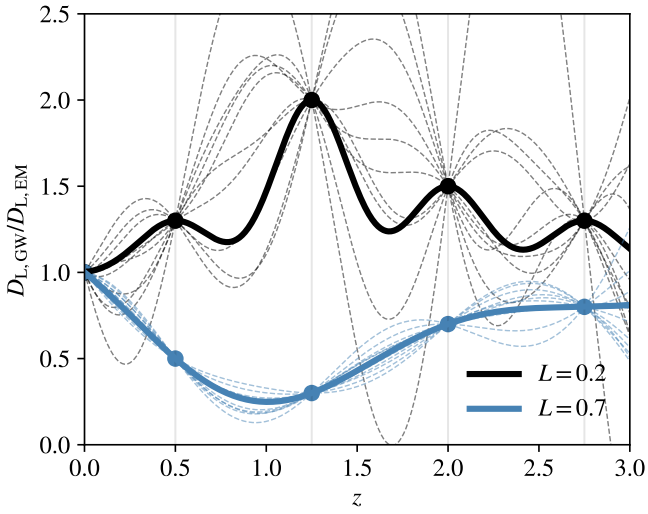

Figure 5.2: Example of GP reconstruction. The function $D_{\mathrm{L,GW}}/D_{\mathrm{L,EM}}\,(z)$ is constructed using 4 nodes at fixed redshifts (filled black dots). By varying the amplitudes of the nodes and the correlation length (L), it is possible to obtain different functional forms. The dashed lines represent a few possible realizations, while the thick lines represents the GP best fit from `sklearn` used in our sampling. Larger correlation lengths produce smoother lines.

be

$$\kappa(z_i, z_j;\, L) \propto \exp\left\{-\frac{1}{2}\left(\frac{|z_i - z_j|}{L}\right)^2\right\}, \tag{5.23}$$

where L is the so-called correlation length. This kernel is flexible enough for our purposes, and we do not expect the detailed choice to have any significant impact on our results. For computational purposes, we generate the GPs using a baseline around $D_{\mathrm{L,GW}}/D_{\mathrm{L,EM}}\,(z) = 1$. This baseline makes the GPs reconstruction to efficiently return to $D_{\mathrm{L,GW}}/D_{\mathrm{L,EM}}\,(z) = 1$ when not pushed towards other values by the training nodes.

This process is described in Figure 5.2 for two values of correlation length L. While pictured in this example, for physical reasons in our analysis we exclude negative-valued functions when exploring $D_{\mathrm{L,GW}}/D_{\mathrm{L,EM}}\,(z)$ and $b_{\mathrm{GW}}/b_{\mathrm{gal}}\,(z)$, and also non-monotonic realizations of $D_{\mathrm{L,GW}}\,(z)$.

The goal of our statistical analysis is to explore the possible constraints on the shape of both $D_{\mathrm{L,GW}}/D_{\mathrm{L,EM}}\,(z)$ and $b_{\mathrm{GW}}/b_{\mathrm{gal}}\,(z)$. For that, we aim to sample the posterior distributions of the amplitudes of the nodes as well as the cosmological parameters. The correlation length L can in principle be fixed based on theoretical priors. Lacking such priors in our case, we only impose a wide uniform prior on L and consider it as a free parameter (see Table 5.1).

Each step of the sampling process consists of producing two GP curves – one for each $D_{\mathrm{L,GW}}/D_{\mathrm{L,EM}}\,(z)$ and $b_{\mathrm{GW}}/b_{\mathrm{gal}}\,(z)$ – given the current set of training node

Parameter	Prior
Node amplitudes	[0, 11] (Uniform)
Correlation length (L)	[1, 10] (Uniform)
Ω_m	1% (Gaussian)
h	1% (Gaussian)
w_0	5% (Gaussian)

Table 5.1: Summary of the priors imposed before reconstructing $b_{\rm GW}/b_{\rm gal}\,(z)$ and $D_{\rm L,GW}/D_{\rm L,EM}\,(z)$ using 4 nodes each. The GP hyper-parameters (i.e., the 2 correlation lengths and the 4×2 amplitudes) are explored independently. The fiducial model is given by $\Xi_0 = 1.4, n = 1, \Omega_{\rm m} = 0.31, h = 0.67, w_0 = -1$.

amplitudes. The curves are used in the calculation of theoretical auto- and cross-correlation power spectra described in section 5.3. The theoretical power spectra enter the Gaussian likelihood together with the generated mock data. The theoretical predictions are computed using our `python` code which is interfaced with the `emcee` sampler. The typical runs with varying cosmology take approximately 10 hours on a modern machine.

After obtaining the posterior distribution of the nodes, we reverse-engineer the problem to obtain confidence contours for each $D_{\rm L,GW}/D_{\rm L,EM}\,(z)$ and $b_{\rm GW}/b_{\rm gal}\,(z)$. For all the sampled node amplitudes we generate the corresponding GP profiles on finite but sufficiently many redshift points and calculate the 68% and 95% confidence intervals at each redshift using the statistical `python` package `GetDist`.

To assess the impact of different cosmological backgrounds and clustering properties, we include in our reconstruction three nuisance parameters: the dark matter abundance $\Omega_{\rm m}$, the Hubble constant H_0 and the dark energy equation of state parameter w_0. A summary of our model parameters and priors used in this work is presented in Table 5.1.

We would like to point out that upcoming galaxy surveys will measure these parameters with very high precision. While GWs alone might be able to provide competitive constraints, the focus of our analysis is not in constraining them. Rather, we would like to quantify how accurately the GW luminosity distance and source properties can be measured. These properties are not accessible to generic redshift surveys and can only be measured with the use of GW-specific observables. This reasoning justifies our tight Gaussian priors on the aforementioned cosmological parameters.

5.5 Results

5.5.1 Reconstructions

In our reconstructions, we always impose the $D_{\rm L,GW}/D_{\rm L,EM}\,(z)$ to become unity at redshift zero by placing a fixed node at $z = 0$ with an amplitude of 1. Besides this fixed node, we have 4 nodes for each of the reconstructed functions. We have arrived at this number by gradually increasing the number of nodes and monitoring

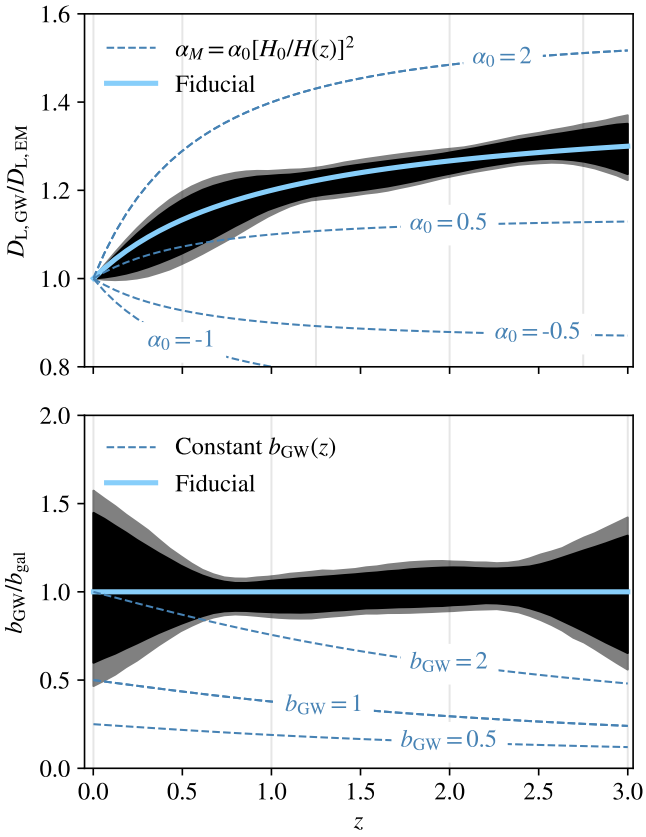

Figure 5.3: Confidence intervals (68%, in black, and 95%, in lighter grey) of the jointly reconstructed functions $D_{\mathrm{L,GW}}/D_{\mathrm{L,EM}}\,(z)$ and $b_{\mathrm{GW}}/b_{\mathrm{gal}}\,(z)$. Together with the assumed fiducial model, we also plot the expectation for different models (see text for more details). The vertical lines mark the fixed location of the nodes used in the GP reconstruction.

the goodness of fit. In practice, we have monitored the AIC information criterion for 1-, 2- and 3- node setups for the GW luminosity distance. Our experiments suggest that, as expected, the 1-node configuration is significantly worse than the presented 4-node setup. The 2- and 3- node setups have similar performance compared to the 4-node case, but the latter still outperforms the former two. We then use the same number of nodes for the bias reconstruction.

In our main analysis, the redshift positions z_i of the nodes are fixed. We have, however, performed an experimental run to assess the impact of letting them free in reduced uniform prior ranges. The result of this experimentation is that the node locations remain unconstrained, and the final posterior of e.g. $D_{\mathrm{L,GW}}/D_{\mathrm{L,EM}}\,(z)$ does not change when the node redshift locations are being sampled as free parameters. This implies that the exact locations of the GP notes are unimportant given they are uniformly distributed in the redshift range of interest.

It is worth emphasizing that when applying our methodology to real data, the number of nodes, as well as their exact redshift placements, should be constrained

by performing similar, but more systematic experiments. Particularly, more accurate measures, such as the Bayesian evidence ratios, should be employed. Also, if enough data is used, some possible constraints could be found by letting the training nodes be completely free in the entire redshift range of interest. As the resulting posteriors are expected to be multimodal, this should be investigated using *nested sampling* algorithms.

As mentioned earlier, we also explore the GP correlation lengths both for the bias and the luminosity distance. On physical grounds, we are interested in smooth GP functions and have imposed the minimum of the uniform prior range for L to be of the order of the inter-node distance so that the smoothness is maintained. We find that, as expected, both of the correlation lengths remain unconstrained within the imposed wide prior ranges.

The results of our joint reconstruction is presented in Figure 5.3. In the same Figure, we also compare these constraints to different theoretical models. In the case of $D_{\rm L,GW}/D_{\rm L,EM}(z)$, we use the parametrization $\alpha_M(z) = \alpha_0 \left[\frac{H_0}{H(z)}\right]^2$ where we use the equation (5.3) with $w_0 = -1$ to obtain the plotted lines (Belgacem et al., 2019). On the other hand, for $b_{\rm GW}/b_{\rm gal}\,(z)$ we plot the lines corresponding to constant values of $b_{\rm GW}\,(z)$, while keeping the galaxy bias fixed to the expression in equation (5.18).

In principle, the output of our sampling can also be used to reconstruct the function $\alpha_M(z)$ by calculating the numerical derivative of $D_{\rm L,GW}/D_{\rm L,EM}\,(z)$. For the purposes of this chapter, however, we chose not to do this. The kernel in equation (5.23) can be interpreted as a smoothness prior and the value of $\alpha_M(z)$ is directly affected by it. Because of this, if one is interested in inferring $\alpha_M(z)$, GPs should be used to sample this function directly.

As expected, we observe how the fiducial models for both $D_{\rm L,GW}/D_{\rm L,EM}\,(z)$ and $b_{\rm GW}/b_{\rm gal}\,(z)$ are well encoded within the reconstructed confidence contours in both panels of Figure 5.3. The constraints at higher redshift ($z \approx 3$) for both reconstructions are broader. This is an effect that could not be seen if a parametric function was used for $D_{\rm L,GW}/D_{\rm L,EM}\,(z)$, for instance, as the parametrization would have fixed the behaviour similarly at low and higher redshifts. Finally, in Figure 5.4 we show the correlation between these functions. We observe weak but non-zero correlations between the GW bias and the luminosity distance. In general, parametric models might induce non-physical correlations. GPs are expected to behave better in this regard, but they can still induce spurious correlations due to finite correlation lengths. For consistency, we have also performed a reconstruction using a redshift binning approach and concluded that we can recover a very similar correlation structure with both of the methods.

5.5.2 Signal-to-Noise scaling

The constraining power of our method crucially depends on a number of observational specifications. The most relevant parameters are 1) the angular sensitivity, specified by the maximum multipole $\ell_{\rm max}$ of the angular power spectra; 2) the number of GW sources, which is specified by the comoving number density $n_{\rm GW}$; and 3) the

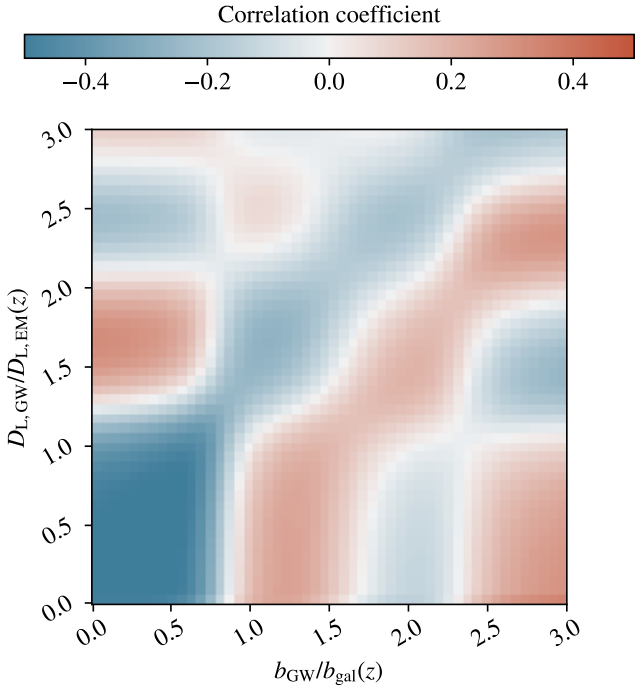

Figure 5.4: Correlation matrix for the reconstructed functions $b_{\mathrm{GW}}/b_{\mathrm{gal}}\,(z)$ and $D_{\mathrm{L,GW}}/D_{\mathrm{L,EM}}\,(z)$. The figure displays how mild, but non trivial correlations can arise from a joint reconstruction.

precision of the GW luminosity distance measurements $\sigma_{\ln D}$. In the case of n_{GW}, we adjust the value of n_0 in equation (5.9) as a way to explore different values of the total number of observed GW events, $N = 4\pi f_{\mathrm{sky}} \bar{n}_{\mathrm{GW}}$. This, in principle, should include selection effects not captured by our formalism. Obviously, for a given experimental configuration the mentioned three variables are not independent, but it is still interesting to find the dependence of our results on each one of them separately. This allows us to reach conclusions without relying on specific experiments, and to suggest potential design guidelines for future GW detectors.

To attain such insights, in this subsection we consider constraints on the parametric expression in equation (5.8), as well as the parametric GW bias given by equation (5.15). For simplicity, we fix $n = 1$ and only constrain the parameter Ξ_0.

When varying ℓ_{max} and N we keep the rest of the configuration (including the luminosity distance binning) fixed. Each case of $\sigma_{\ln D}$, on the other hand, is accompanied by an adjustment in the number of luminosity distance bins. This is done to be consistent with our binning strategy, namely that the luminosity distance width of each bin is at least $\mathcal{O}(3)$ times wider than $\sigma_{\ln D}$.

Our results are summarized in Figure 5.5, where we plot the anticipated uncertainties in Ξ_0 (upper panel) and b_0 (lower panel) as a function of the SNR of the cross-correlation in equation (5.19).

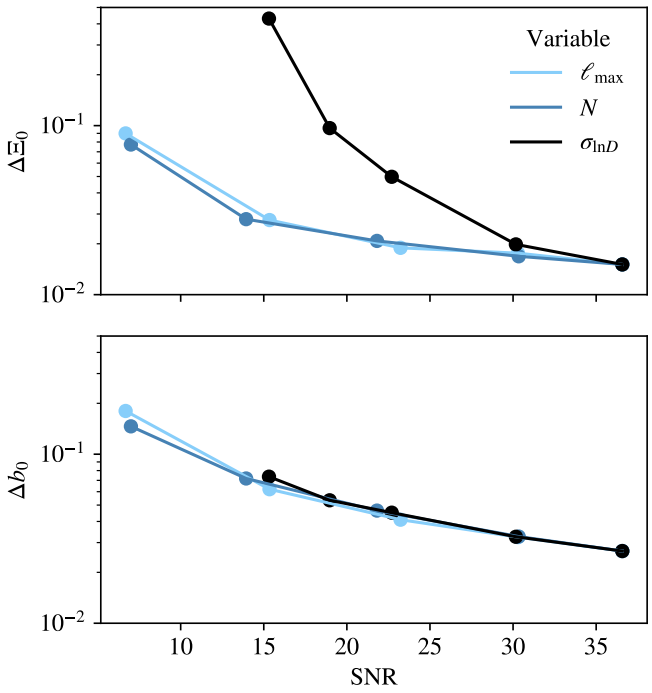

Figure 5.5: Scaling of the observed constraints with the cross-correlation SNR. Using a parametrized model for $b_{\rm GW}/b_{\rm gal}\,(z)$ and $D_{\rm L,GW}/D_{\rm L,EM}\,(z)$ we explore the constraining power of our method as a function of the number of observed GW sources $N = [0.7, 4, 7, 13, 20] \times 10^4$, angular resolution $\ell_{\rm max} = [20, 40, 60, 80, 100]$ and luminosity distance uncertainty $\sigma_{\ln D} = [0.5, 0.3, 0.2, 0.075, 0.05]$. As visible from the figure, the data SNR completely captures the effect on the observed uncertainties Δb_0 and $\Delta\Xi_0$ in the first two cases. In the case of $\sigma_{\ln D}$, we observe that the increase in constraining power for $D_{\rm L,GW}/D_{\rm L,EM}\,(z)$ is steeper due to the larger number of window functions that we can build to sample $D_{\rm L,GW}\,(z)$.

For a fixed $\sigma_{\ln D}$ the constraining power on Ξ_0 and b_0 is almost completely determined by the cross-correlation SNR. This fact suggests that no matter how the given SNR is realized (either by increasing the number of sources or by improving the angular sensitivity), the expected constraints will be the same. This implies that the results presented in this chapter can be easily scaled to different configurations. Unsurprisingly, we find that the constraints scale as 1/SNR.

The situation is somewhat different for the case of varying $\sigma_{\ln D}$ (and the number of luminosity distance bins). The constraints on the bias still follow the same form (see the lower panel), but the scaling of the Ξ_0 constraints, on the other hand, is much steeper than in the cases of varying $\ell_{\rm max}$ and $n_{\rm GW}$, roughly $1/{\rm SNR}^3$. This fact can be qualitatively understood by remembering the importance of the relative positions of GW and galaxy window functions demonstrated in Figure 5.1. Sampling this relation with a larger number of window functions increases the precision of our reconstruction.

The results presented in this section quantify the importance of accurate luminosity distance measurements and demonstrate the benefit that smaller values of $\sigma_{\ln D}$ can bring to a binned approach.

5.6 Discussion and Conclusions

In this chapter, we have shown how the combination of GW observations and redshift surveys can be exploited in the era of GW cosmology. We have identified two essential quantities that characterize this new research field: 1) the luminosity distance relation, a clear imprint of modifications to the propagation of tensor modes (see section 5.2), and 2) the bias of the sources, regulating their spatial distribution and betraying their origin.

Proposed GW detectors such as the Einstein Telescope (Maggiore et al., 2020) or Cosmic Explorer (Reitze et al., 2019) are expected to probe a sizeable fraction of the visible Universe and produce large statistical samples. In this context, we point out that the number of sources assumed for our main analysis, 2×10^5, is particularly conservative and differs from expectations by at least an order of magnitude (Maggiore et al., 2020). This difference is primarily due to our assumption of a constant comoving density of events. While we do not explore other assumptions for the distribution of GW sources over cosmic time, we have investigated similar effects in subsection 5.5.2, where we have shown how our results can be rescaled to other instrument configurations or number of observed sources.

Our formalism, based on binned angular power-spectra and sky maps, is optimal for a large number of sources with no known counterpart. Its main advantages are related to the simple modelling of the theoretical signals and their data covariance matrix. Because no reconstruction of the underlying density field is necessary, the predictions display a clear separation of scales. For example, the angular scales that we have considered hare are all within the linear regime ($k \lesssim 0.1$ Mpc^{-1}). Furthermore, because this formalism is well established, our shot-noise limited covariance matrix can be easily generalized to include additional sources of (co-)variance.

Although a comprehensive comparison between multiple approaches is outside the scope of this work, it is worth discussing how our results compare to others found in the literature. We preface this by saying that one-to-one comparisons, however, are often complicated either by significantly different assumptions or the impossibility of directly translating these assumptions from one prescription to another. Despite this, here we draw a parallel between our method and two other methods.

The first method is the one used in (S. Mukherjee, Wandelt, & Silk, 2021), which has also been shown to be extremely successful in measuring both $b_{\rm GW}(z)$ and $D_{\rm L,GW}(z)$ using parametric models. Similarly to this work, the information is also extracted from the cross-correlation with redshift sources, but no binning of the GW data is performed. In this case, we have verified that such methods perform significantly better than our map-based approach in the case of a low number density of GW sources and large uncertainty in the measured $D_{\rm L,EM}(z)$. These features, in particular, make it especially useful for near-future samples of a few tens of objects.

The second promising method to measure $D_{\mathrm{L,GW}}(z)$ that has been proposed in the literature is offered by GW sources with known counterparts. Such observations give direct access to $D_{\mathrm{L,GW}}$ as a function of redshift and can be combined with similar measurements in the EM spectrum to obtain $D_{\mathrm{L,GW}}/D_{\mathrm{L,EM}}(z)$. The analysis of (Belgacem et al., 2020) is based on this methodology and, similarly to ours, also employs GPs to reconstruct this ratio from an Einstein Telescope sample with $\sim 10^2$ sources.

Ultimately, we expect this counterpart-based formalism and the one described in this work to be complementary: a direct measure of $D_{\mathrm{L,GW}}(z)$ can be used to break the degeneracy between bias and luminosity distance shown in Figure 5.4. However, because the fraction of events with known counterparts that will be observed is heavily dependent on both the GW source distribution and multiple instrumental setups, we do not attempt to combine the two methods here.

In conclusion, the combination of GW resolved events and the clustering of galaxies is expected to improve our current knowledge of the physical properties of the Universe. Our work shows how to reconstruct these properties as a function of redshift in a generic way, and highlights the need for accurate and precise measurements of $D_{\mathrm{L,GW}}$. This will require control over the instrument calibration uncertainties (Cahillane et al., 2017), but also the degeneracy between the inclination of the source and its luminosity distance (Ghosh, Del Pozzo, & Ajith, 2016). In the future, we aim to apply our current analysis pipeline to the next generation of large scale structure surveys and incoming GW observations.

Part III

Data Science for the Euclid mission

Chapter 6

Cosmological Likelihood for Observables in *Euclid*

We can use observations of the large scale structure (LSS) of the universe, as galaxy clustering (GC) and weak lensing (WL), to extract cosmological information (see chapter 1). If we want to exploit all the information encoded in LSS, we need large data sets of galaxies that contain not only their positions in the sky and their shapes but also their redshifts. These large data sets form the so-called galaxy redshift surveys. Some examples of well-known galaxy redshifts surveys are the *Center for Astrophysics Redshift Survey* (CfA) (Tonry & Davis, 1979) and the *Sloan Digital Sky Survey* (SDSS) (Lundgren et al., 2015). These surveys have successfully shown that we can use the information contained in the large scale structure of the universe to improve our cosmological knowledge and constrain not only cosmological parameters but also be able to discern among several cosmological models.

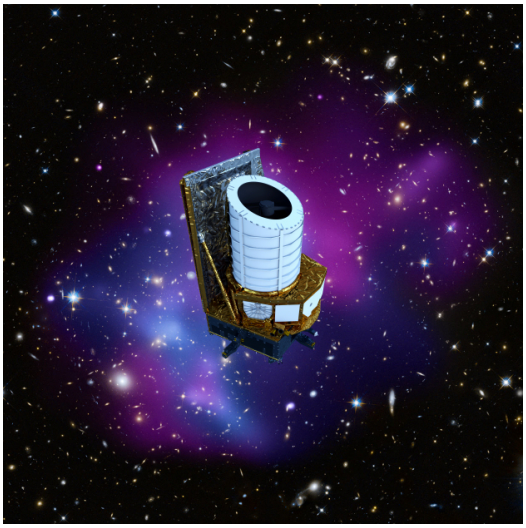

Figure 6.1: Artist's impression of the *Euclid* spacecraft.

In this chapter, we will focus on the future *Euclid* mission, a near-infrared space telescope currently under development by the European Space Agency and the Euclid Consortium, and for which the author of this thesis has dedicated a vast percentage of her doctoral time. In particular, the author works as one of the main *software developers* for the official Bayesian analysis pipeline that will be eventually used to produce the constraints on the cosmological parameters once the data is available: the *Cosmological Likelihood for Observables in Euclid* (`CLOE`).

This chapter is organized as follows. In section 6.1 we present the mission, main goals, structure and technical details of the telescope. In section 6.2 we introduce the features of the future *Euclid* surveys and we show briefly how the photometric and spectroscopic techniques works. In section 6.3 we explain in mathematical detail the theoretical description and modelling of the primary observational probes. In section 6.4 we detail how the fiducial benchmark data and covariance matrices are generated and later used to validate the Bayesian analysis pipeline. In section 6.6 we formally introduce `CLOE`: structure, features and dependencies. In section 6.7 we show the results of the validation of `CLOE` in its earliest version release. Finally, we conclude with the outline of the future plans for `CLOE` in section 6.8.

6.1 The *Euclid* Mission

6.1.1 Telescope, goals and the Euclid Consortium

Euclid[1] is an European Space Agency (ESA) medium-class mission part of the *Cosmic Vision* 2015-2025 program. According to the main scientific objectives listed in the mission's study definition report[2] (Laureijs et al., 2011), the ultimate goals of *Euclid* are to understand the physical origin of the accelerated expansion of the Universe, whose responsible agent is thought to be Dark Energy, and to understand the nature of Dark Matter. Moreover, *Euclid* is meant to study the initial conditions that seeded the early universe and were responsible for the formation of the cosmic structure and ultimately, it will also try to understand the range of validity of General Relativity. To achieve its goals, *Euclid* will map the LSS of the universe by creating one of the largest galaxy catalogues ever, and this is why *Euclid* is considered primarily a cosmological mission.

Euclid satellite's launch is planned for 2024[3] using an Ariane 62 and will travel up to the L2 Sun-Earth Lagrangian point to be on duty for a 6-years mission. It will explore the expansion history of the Universe and the evolution of the LSS by measuring the position, shapes and redshifts of galaxies. For that, it will incorporate a 1.2-meter telescope and two scientific instruments that will observe in the optical and near-infrared bands: a high-quality panoramic optic visible imager (VIS), a near-infrared 3-filter (Y, J and H) photometer (NISP-P) with a slitless spectrograph (NISP-S). The spectrograph consists on one 'blue' grism ($920 - 1250\ nm$) and three 'red' grisms ($1250 - 1850\ nm$ placed on three different orientations). The optical and near-infrared instruments share a common field-of-view of 0.53 deg^2. *Euclid* will cover eventually 15000 deg^2 of the sky up to redshift $z \approx 2$. The primary result product of the mission will be a complete redshift galaxy catalogue up to magnitude 24 in the Y, J, and H bands. It will collect enough information to study the mission's primary observational probes: 30 million spectroscopic redshifts used for spectroscopic galaxy clustering measurements and 2 billion photometric galaxy images, which will be used

[1] Up-to-date information about the mission can be found at `https://www.Euclid-ec.org`

[2] The study definition report receives the popular name of the "Red Book".

[3] The launch of the satellite is yet unknown due to consequences of the war conflict in Ukraine.

for weak lensing and photometric galaxy clustering observations (Amendola, Appleby, et al., 2018; Euclid Collaboration et al., 2020).

Apart from the primary probes, *Euclid* will provide additional cosmological observables such as the cross-correlation with CMB observations, strong lensing, abundance and properties of galaxy clusters and even luminosity distance measurements through supernovae Ia. Moreover, *Euclid* will complement the cosmological observables with astrophysical and astronomical probes of high scientific interest (i.e: cool brown dwarfs, stellar populations in the galaxy and the universe nearby, High-z Lyman Break Galaxies...).

In June 2012, ESA named the *Euclid Consortium* (EC) as the team responsible for the mission, the data production and the scientific instruments (VIS and NISP). The scientific exploitation and interpretation of the massive *Euclid* data collection will be led by the scientists of the EC as well. It comprises approximately 1500 scientists with various backgrounds (i.e: astronomers and astrophysicists, theoretical physicists, engineers, technicians, and managers...). The EC contains researchers from 14 European countries working in their national space agencies or associated research institutes (Austria, Belgium, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Portugal, Romania, Spain, Switzerland and the United Kingdom), scientists from Canada and USA (through NASA and other US laboratories) and few scientists from Japan. The EC is led by the Euclid Consortium Lead (ECL) and the Euclid Consortium Board (ECB), which have also the role of acting as contact points between the EC members and ESA. The activity of the EC is organized in several groups (Tutusaus Lleixa, 2018, September):

- the EC Science Ground Segment (SGS), which is responsible for the design, test, integration and operation of the data processing tools and pipelines and whose activities are sub-arranged in Organizational Units (OUs). Each of the OUs is dedicated to one specific task, for instance, providing the measurements of the positions of galaxies.

- the EC Science Working Groups (SWGs), which have the responsibility of the scientific production and delivery of Euclid data releases and their scientific exploitation. There are three different types of SWGs: the cosmology SWGs (weak lensing, galaxy clustering, galaxy cluster, theory...), the legacy SWGs (exoplanets, Milky Way...) and the cosmological simulations SWG. The work within the SWGs is further structured in Working Packages (WPs).

With *Euclid's* launch fast approaching, it is necessary to improve and update the current cosmological forecasts for the scientific performance of the satellite to assess the impact of the design and technical description of the mission. The first *Euclid* forecasts were shown in the definition study report (Laureijs et al., 2011). Years later, the EC multidisciplinary group "Inter-Science Working Group Taskforce Forecast" (also known as "IST:Forecast") (Euclid Collaboration et al., 2020) unified the implemented definitions of the different *Euclid* primary probes: updating the

information presented in the Red Book by including the latest astrophysical knowledge. Besides, they also validated the data analysis methodology and computational tools to produce Fisher forecasts and verify their robustness. The work carried out by "IST:Forecast", which focused their analysis on the ΛCDM model, remains the current state-of-the-art for the scientific performance of the mission.

Still, as mentioned in chapter 1, section 1.7, it is customary to use Bayesian statistics as the analysis framework for testing cosmological models against data to infer the probability distributions of the parameters of a model. Therefore, the construction of a *software* that allows us to perform a full Bayesian statistical analysis not to only go beyond the "Fisher forecast" scenario but to analyse the incoming data is crucial for the *Euclid* mission. Using a full Bayesian statistics approach will open the possibility of forecasting the scientific performance of the satellite not only for the Standard Cosmological Model but also for extensions of ΛCDM, where, for instance, the posterior distributions of the parameters of interest sometimes cannot be easily approximated by a Gaussian function. To be able to move towards a full Bayesian statistical analysis, it is necessary to construct a software that computes the *likelihood* distribution given the model and its parameters, and some (fiducial) data. Furthermore, we need to model the theoretical observables to make predictions: weak lensing (WL), photometric galaxy clustering (GCph), spectroscopic galaxy clustering (GCsp) and their combinations.

Additionally, to unleash the full power of *Euclid* primary observational probes, it is essential to exploit all the information contained in the non-linear regime of matter density perturbations; that is, at late times and small scales ($k > 0.1$ Mpc). At these scales, the matter distribution is affected by the non-linear evolution of density fluctuations, which induces changes in the shape of the matter power spectrum beyond the predictions of linear perturbation theory (see subsection 1.4.1 for more information). Therefore, the EC is giving special attention to the need of modelling with precision the theoretical predictions of *Euclid* primary observables for the non-linear scales to make the best use of the future *Euclid* data.

6.1.2 The role of IST:L and IST:Non-Linear

The work presented in this chapter, which is a summary of the science contained in a series of upcoming *Euclid* key-project papers, represents the results of the intensive activity of code design, development, review and testing carried out by several scientists from different Science Working Groups within the EC. The work has been conducted by a novel and special multidisciplinary group known as the "Inter-Science Taskforce Likelihood" (IST:L). This group is in charge of developing the official likelihood and *Euclid* software that will produce the theoretical predictions for *Euclid* primary probes and the computation of the *Euclid* likelihood, and will also produce the official constraints on the probability distributions of the cosmological parameters. The results of this work activity, which currently has been taking place for approximately three years, is the software *Cosmological Likelihood for Observables in Euclid* (also known as `CLOE`).

`CLOE`[4], which is currently available only for EC members, can compute the theoretical predictions of the primary probes given the implementation recipes provided by the cosmology Science Working Groups. Furthermore, `CLOE` can be attached to the Bayesian analysis framework `Cobaya` as an *external likelihood* to perform full sampling of the posterior probability distributions of the parameters of interest (see section 6.6). In May 2021 `CLOE` v.1.0 was released, including the first implementation for the theoretical observables' recipe only up to the linear regime. A version 1.1 appeared soon after after fixing a series of bugs and implementing functionalities in the user interface. By the end of 2022, `CLOE` v.2.0 will be available within the EC including a revisited recipe of the observables including corrections for the non-linear scales.

The multidisciplinary group in charge of correcting the primary observables for the non-linear scales is the "Inter-Science Working Group Taskforce Non-Linear" (IST:NL). If these corrections are not implemented, the theoretical models will be incomplete and will eventually bias the the statistical analysis. This is why the work of IST:NL is crucial for `CLOE` to succeed and provide correct constraints of the cosmological parameters. IST:NL is responsible for the non-linear modelling as well as the theory covariances, and they will create the "IST:NL model library" that interfaces with `CLOE`. So far, IST:NL has implemented external codes able to provide non-linear corrections: `bacco` (Aricò, Angulo, & Zennaro, 2021; Angulo et al., 2021; Aricò, Angulo, Contreras, et al., 2021), `euclid emulator 2` (Euclid Collaboration et al., 2021) and `fastPT` (McEwen, Fang, Hirata, & Blazek, 2016; Fang, Blazek, McEwen, & Hirata, 2017). Therefore, IST:L is responsible for the computation of the theoretical predictions for the primary observables and the likelihood, and IST:NL will include the non-linear corrections by implementing the non-linear codes within `CLOE`.

The work in IST:L and IST:NL is organized following an *Agile* inspired philosophy where the working tasks are split into smaller sub-tasks that are assigned to different sub-groups, which use a `SCRUM` framework for developing, delivering, and sustaining software (see (Abrahamsson, Salo, Ronkainen, & Warsta, 2017) for more details about *Agile*). Each sub-group is composed by a *Scrum master, code developer, code reviewer* and several theory *experts*. The software `CLOE` contains unit testing and combines the practices of continuous integration (CI) and continuous delivery (CD), enforcing automation in building, testing and deployment of `CLOE` (see (*SCRUM: practical guide*, 2022) and (Sane, 2021) for more information about good practices in coding). The author of this chapter works as one of the main *code developers* for `CLOE` within IST:L and as a consultant for IST:NL.

6.2 Photometric and Spectroscopic surveys

Galaxy redshift surveys can be classified based on the technique used to extract the redshift measurements. In this sense, we can mainly classify redshift surveys as

[4]For EC members `CLOE` can be downloaded from `https://gitlab.euclid-sgs.uk/pf-ist-likelihood/likelihood-implementation`.

photometric and spectroscopic ones. Spectroscopic redshift surveys use spectrographs to obtain redshifts. These devices can obtain the spectrum of a cosmic object by separating the incoming light into several narrow bins using dispersion. As seen during chapter 1, section 1.1, the light coming from distance objects is redshifted. To extract the redshift measurements, we need to compare the obtained spectra to the one known from an object of the same class at rest. This approach requires a large collection time and previous information about the angular position of the galaxies for targeting. Photometric redshift surveys, on the other hand, use images to extract low-resolution spectra, where the redshift is still inferred by comparing the obtained spectra with other spectra from similar objects at rest. This method does not require target galaxies and it needs less exposition time than the spectroscopic technique, but provides worse redshift estimates.

When *Euclid* arrives at the L2 Lagrangian point, it will start operating for a 6-years long mission, creating two different surveys. The first survey is the *Euclid* Wide Survey covering 15000deg^2 of the sky up to a magnitude of approximately 24 for both VIS and NISP instruments; thus, it will contain both spectroscopic and photometric redshifts. The contamination coming from the Milky Way as well as the contamination coming from the Solar System will be removed. This survey will be used for the study of the *Euclid* primary probes: weak lensing and galaxy clustering. The other survey will consist of three *Euclid* Deep Fields going to 2 magnitudes deeper than the wide survey but covering only, in total, 40 deg^2. The *Euclid* Deep Fields will be used to calibrate the wide survey, apart from providing precious information to study Active Galactic Nuclei, high redshift galaxies and other objects. It is expected that after the 6 operation years, Euclid will have measured the shape and photometric redshifts of 1.5×10^9 galaxies, creating the photometric *Euclid* survey, and about 5×10^7 galaxy spectroscopic redshifts, used for the spectroscopic *Euclid* survey.

The photometric measurements of the *Euclid* wide survey will be used to study weak lensing (WL) and photometric galaxy clustering (GCph). WL is sensitive to both baryonic and dark matter (BM and DM) and the expansion rate of the universe. The spectroscopic measurements of the *Euclid* wide survey will be employed for the spectroscopic galaxy clustering (GCsp), including baryonic acoustic oscillations (BAO) and redshift space distortions (RSD). GCsp is sensitive to the distribution of matter as well as the expansion of the universe and the growth rate of structures. Combining all the probes, as we will see in section 6.7, will allow us to constrain different cosmological parameters and possibly discern among cosmological models once the real data is available.

Moreover, when the redshift of the galaxies is known, we can extract more cosmological information using *tomography*: we can create slices of two-dimensional projected images to recover the three-dimensional distribution of matter in the universe (Hu, 1999). By dividing the galaxies into redshift tomographic bins, $n_i(z)$, we can obtain the time evolution of LSS. For the *Euclid* mission, the distribution of galaxies of both the photometric and spectroscopic surveys will be divided into redshift bins, which will be used to compute the *Euclid* primary probes. In this first approach,

10 redshift bins for the photometric probes and 4 redshift bins for the spectroscopic probe are used[5].

6.3 Theoretical predictions for *Euclid* primary observables

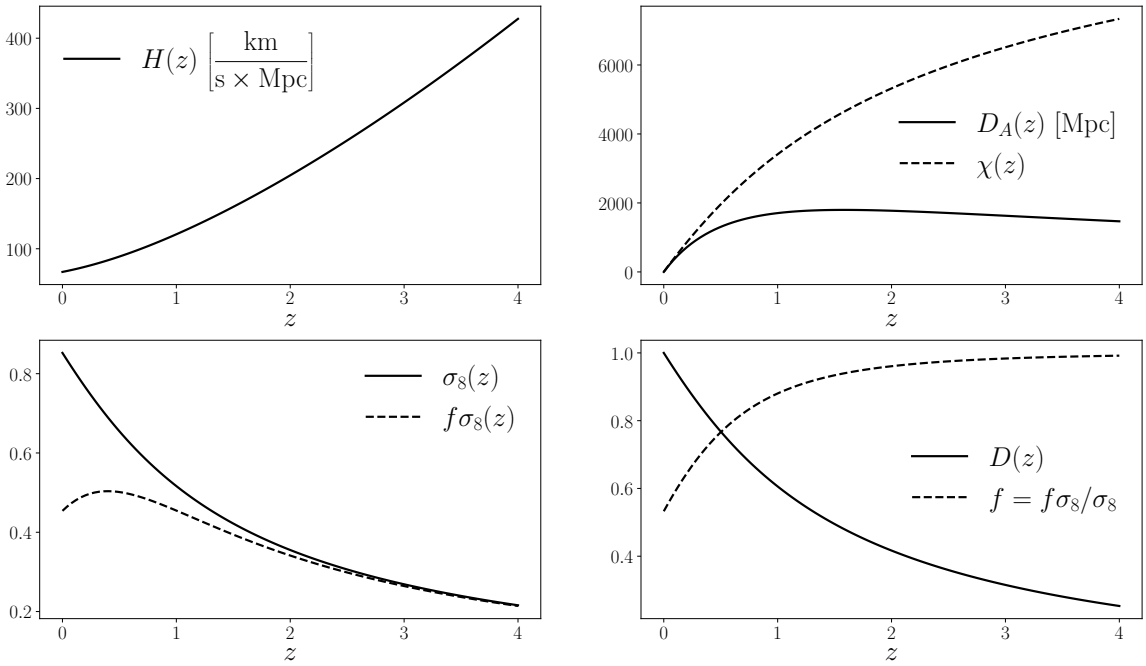

Figure 6.2: Cosmological background quantities requested by `CLOE` to `CAMB` via `Cobaya`, to be used as ingredients of the computation of the *Euclid* primary theoretical observables described in section 6.3. **Top-left** is the Hubble parameter $H(z)$, **top-right** are the comoving $\chi(z)$ and the angular diameter $D_{\mathrm{A}}(z)$ distances, **bottom-left** are the root mean square mass fluctuations amplitude on 8 Mpc/h scales $\sigma_8(z)$ (where h is the reduced Hubble parameter), and the product of σ_8 times the growth rate $f(z)$, and finally, **bottom-right** are the growth factor $D(z)$ and growth rate $f(z)$. All quantities are represented as a function of redshift z. Within `CLOE`, these quantities are obtained in `cobaya_interface.py` and further processed in `cosmo.py`. All plots have been generated taken the fiducial cosmological values presented in section 6.9. All the definitions of these background functions can be found at chapter 1.

In this section, the modelling of the *Euclid* main scientific probes is presented: weak lensing and photometric galaxy clustering for the photometric catalogue with the cross-correlation between the two, and spectroscopic galaxy clustering for the spectroscopic catalogue. For simplicity and completeness, the modelling described

[5]The number of redshift bins is decided by the SWGs and the corresponding OUs, and can be changed at later times

Cosmological probe	Output
Spectroscopic galaxy clustering (GCsp)	
Spectroscopic galaxy bias	$b^{\rm GCsp}(z)$
Redshift-space galaxy-galaxy power spectrum	$P_{\rm gg}^{\rm GCsp}(z,k)$
Perpendicular scaling factor	$q_\perp(z)$
Parallel scaling factor	$q_\parallel(z)$
Cosine of the angle between $\vec{k}$ and the line-of-sight	μ_k
Legendre multipole power spectra ($\ell = m$)	$P_\ell(k,z)$
Weak lensing (WL)	
Galaxy redshift density distribution	$n_i^{\rm WL}$
Intrinsic alignment - intrinsic alignment power spectrum	$P_{\rm IAIA}(z,k)$
Matter density - intrinsic alignment power spectrum	$P_{\delta{\rm IA}}(z,k)$
Intrinsic alignment function	$f_{\rm IA}(z)$
Intrinsic alignment nuisance parameters	$A_{\rm IA}$, $\eta_{\rm IA}$
Shear window function	$W_i^\gamma(z,k)$
Intrinsic alignment window function	$W_i^{\rm IA}(z)$
Intrinsic alignment angular power spectra	$C_{ij}^{\rm IAIA}(\ell)$
Shear angular power spectra	$C_{ij}^{\gamma\gamma}(\ell)$
Weak lensing angular power spectra	$C_{ij}^{\rm WL}(\ell)$
Photometric galaxy clustering (GCph)	
Galaxy redshift density distribution	$n_i^{\rm GCph}$
Galaxy-galaxy power spectrum	$P_{\rm gg}^{\rm GCph}(z,k)$
Photometric galaxy window function	$W_i^{\rm GCph}(z)$
Photometric galaxy bias	$b^{\rm GCph}(z)$
Photometric galaxy clustering angular power spectra	$C_{ij}^{\rm GCph}(\ell)$
Weak Lensing - Photometric Galaxy Clustering Cross-Correlation (XC)	
Galaxy-matter power spectrum	$P_{\delta{\rm g}}^{\rm GCph}(z,k)$
Galaxy-intrinsic alignment power spectrum	$P_{\rm gAI}^{\rm GCph}(z,k)$
Photometric galaxy-matter angular power spectrum	$C_{ij}^{\delta{\rm g}}(\ell)$
Photometric galaxy-intrinsic alignment angular power spectrum	$C_{ij}^{\rm gIA}(\ell)$
Photometric cross-correlation angular power spectrum	$C_{ij}^{\rm XC}(\ell)$

Table 6.1: Summary of the different quantities and symbols used to write the recipe for the *Euclid* primary probes: weak lensing (WL), photometric galaxy clustering (GCph), spectroscopic galaxy clustering (GCsp) and the weak lensing - photometric galaxy clustering cross-correlation (XC). The recipe is written in terms of several different power spectra $P(z,k)$ to simplify the interface with IST:NL, which will provide IST:L with the non-linear corrections for the observables. Note that the notation for the angular power spectra $C_{ij}(\ell)$ is written differently with respect to the notation used in the previous chapters, where here we have decided to write explicitly the dependence on the tomographic redshift bins i,j as subscripts and the symbol of the probe as superscript. The colours used to highlight the different cosmological probes are in concordance with the colours used in the plots of the observables or cosmological constraints.

below corresponds to the recipe v.1.0 implemented in `CLOE` version 1.0 and version 1.1. The primary *Euclid* observables are written in terms of basic cosmological functions described in chapter 1, such as the *comoving distance* $\chi(z)$, the *Hubble parameter* $H(z)$, the *matter power spectrum* $P_{\delta\delta}$, the *growth factor* $D(z)$ and growth rate $f(z)$, the *angular diameter distance* $D_{\rm A}$, and the *Modified Gravity phenomenological function* $\Sigma_{\rm MG}$[6] (see Figure 6.2). In this section, the *speed of light* c factors in the mathematical expressions are kept.

6.3.1 Galaxy Clustering: Spectroscopic (GCsp)

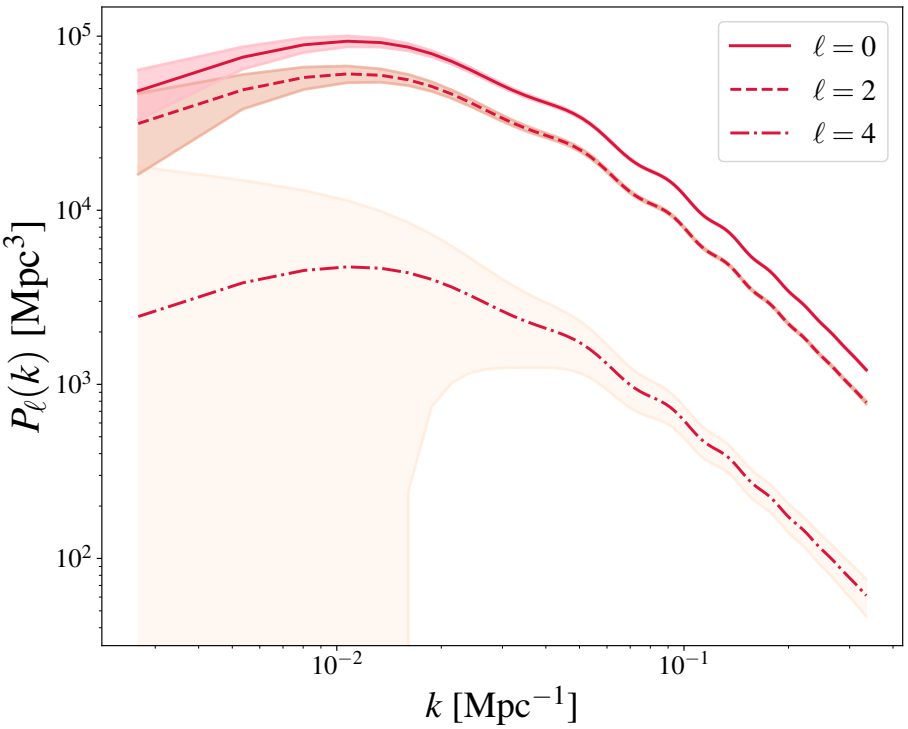

Figure 6.3: Legendre multipoles $P_\ell(k)$ for the spectroscopic galaxy clustering probe (GCps) as a function of scale k for multipoles $\ell = 0$ (solid line), $\ell = 2$ (dashed line) and $\ell = 4$ (dashed-dotted line), corresponding to the second tomographic redshift bin centred at $z = 1.2$. The shades regions correspond to the data uncertainties given the fiducial covariance matrix as explained in subsection 6.4.4. Within `CLOE`, the spectroscopic legendre multipoles are obtained in `spectro.py`. The plot has been generated taking the fiducial cosmological values presented in section 6.9.

As we have seen in subsection 1.5.3, galaxies are biased tracers of the total matter in the Universe. In the linear regime, we can assume a linear deterministic bias

[6]The Modify Gravity function $\Sigma_{\rm MG}$ is implemented in the code, but in this version, it is set equal to 1.

model and write the relation between the galaxy power spectrum $P_{\rm gg}^{\rm GCsp}$ and the matter power spectrum $P_{\delta\delta}(k,z)$ as

$$P_{\rm gg}^{\rm GCsp}(k,\mu_k,z) = \left[b^{\rm GCsp}(z) + f(z)\mu_k^2\right]^2 P_{\delta\delta}(k,z)\,, \tag{6.1}$$

where we know that our observations are affected by redshift space distortions (RSD), whose effect in the power spectrum can be described at the linear level by $f(z)\mu_k$. Also $k \equiv |\vec{k}|$ is the module of the wavevector $\vec{k}$, μ_k is the cosine of the angle between $\vec{k}$ and the line-of-sight direction, and $f(z)$ is the growth rate. The $b^{\rm GCsp}(z)$ is the redshift-dependent linear bias parameter and will take different values for each tomographic redshift bin. The analysis of tomographic galaxy clustering probe using spectroscopic redshifts is expressed in Legendre multipoles. In Fourier space, the Legendre multipoles $P_\ell(k)$ are given by

$$P_\ell(k,z) \equiv \frac{2\ell+1}{2}\int_{-1}^{1} {\rm d}\mu_k\, L_\ell(\mu_k) P_{\rm gg}^{\rm GCsp}(k,\mu_k,z)\,, \tag{6.2}$$

where L_ℓ is the Legendre polynomial. As we have explained already subsection 1.5.3, the true underlying power spectrum is not direct observable. All estimates of these functions based on galaxy surveys require the assumption of a fiducial cosmology to transform the observed redshifts into physical separations. A difference between the fiducial and true cosmologies leads to a re-scaling of the components parallel and perpendicular to the line-of-sight direction, $s_\parallel$ and $s_\perp$, of the total separation vector $\mathbf{s}$ between two galaxies in the survey, defined as

$$s_\perp = q_\perp s_\perp^{\rm fid}, \tag{6.3}$$

$$s_\parallel = q_\parallel s_\parallel^{\rm fid}, \tag{6.4}$$

where 'fid' denotes the quantities in the fiducial cosmology and the scaling factors q are given by the ratios of the comoving angular diameter distance, $D_{\rm A}(z)$, and the Hubble parameter, $H(z)$, respectively, in the true and fiducial cosmologies at the mean redshift of the sample:

$$q_\perp = \frac{D_{\rm A}(z)}{D_{\rm A}^{\rm fid}(z)}, \tag{6.5}$$

$$q_\parallel = \frac{H^{\rm fid}(z)}{H(z)}. \tag{6.6}$$

Equation (6.3) and equation (6.4) can be written in terms of k and μ_k as (Ballinger et al., 1996)

$$k(k^{\rm fid},\mu_k^{\rm fid},z) = k^{\rm fid}\left[q_\parallel^{-2}(z)\,(\mu_k^{\rm fid})^2 + q_\perp^{-2}(z)\left(1-(\mu_k^{\rm fid})^2\right)\right]^{1/2}, \tag{6.7}$$

$$\mu_k(\mu_k^{\rm fid},z) = \mu_k^{\rm fid}\, q_\parallel^{-1}(z)\left[q_\parallel^{-2}(z)\,(\mu_k^{\rm fid})^2 + q_\perp^{-2}(z)\left(1-(\mu_k^{\rm fid})^2\right)\right]^{-1/2}. \tag{6.8}$$

In Fourier space, the fiducial cosmology also leads to a rescaling of the power spectrum amplitude by a factor $\left(q_\perp^2 q_\parallel\right)^{-1}$. These geometric distortions must be applied to

all model predictions before they are compared with real galaxy clustering measurements. Using equations (6.7, 6.8), the prediction for the observed Legendre multipoles $P_{\mathrm{obs},\ell}(k^{\mathrm{fid}})$ including geometric Alcock-Pacyznski distortions can be obtained as

$$P_\ell(k^{\mathrm{fid}}, z) = \frac{1}{q_\perp^2(z) q_\parallel(z)} \frac{2\ell+1}{2} \int_{-1}^{1} L_\ell(\mu_k^{\mathrm{fid}}) P_{\mathrm{gg}}^{\mathrm{GCsp}} \left(k(k^{\mathrm{fid}}, \mu_k^{\mathrm{fid}}), \mu_k(\mu_k^{\mathrm{fid}}); z \right) \mathrm{d}\mu_k^{\mathrm{fid}} \,. \tag{6.9}$$

The plotted expression for equation (6.9) can be found in Figure 6.3.

6.3.2 Weak Lensing (WL)

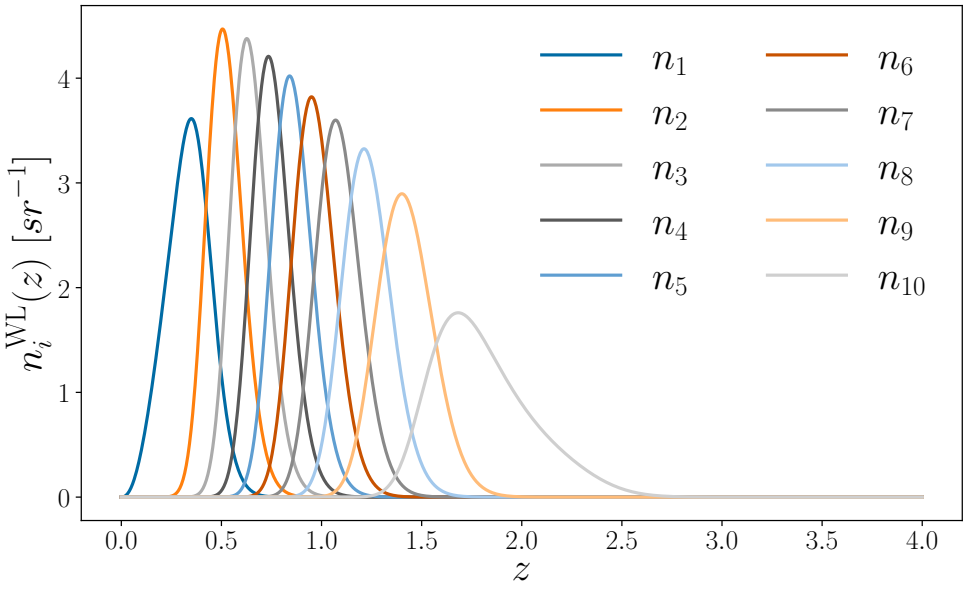

Figure 6.4: Galaxy redshift density distribution $n_i^{\mathrm{WL}}(z)$ as a function of redshift for the weak lensing (WL) probe in the 10 different tomographic redshift bins i. These distributions are read by `CLOE` as external files in `reader.py` and further processed in `redshift_distribution.py`. The benchmark distributions have been generated as explained in subsection 6.4.1. In `CLOE` v.1.1, the density distributions for both weak lensing and photometric galaxy clustering are considered equal: $n_i^{\mathrm{WL}}(z) = n_i^{\mathrm{GCph}}(z)$.

As it was already introduced in subsection 1.5.4, the light coming from distant galaxies bends towards us due to the massive objects that it can find along its trajectory, inducing distortions in their observed shape, allowing us to map the distribution of matter in space. This effect is called gravitational lensing, and in their weakest limit, we can extract cosmological information by studying the tiny distortions in galaxy shapes statistically. The weak lensing probe is written in terms of the tomographic angular power spectra in harmonic space, $C_{ij}^{\mathrm{WL}}(\ell)$, with i, j labelling the corresponding redshift bins, and WL standing for Weak Lensing. The angular power spectrum C_{ij}^{WL} is the Fourier transform of the two-point correlation function, which

can be written as a line-of-sight integral of the matter power spectrum $P_{\delta\delta}$ and a window function $W_i(z)$, which contains information about the probability distribution of observing a source galaxy at a given redshift z and about the physics of gravitational lensing. The weak lensing effect is modelled by taking into account not only the shear power spectrum (γ) but also the intrinsic alignment (IA) contribution, which is an astrophysical systematic error on the shapes of galaxies due to tidal forces applied by the surrounding large scale structure. In total, the weak lensing angular power spectrum $C_{ij}^{\rm WL}(\ell)$ is:

$$C_{ij}^{\rm WL}(\ell) = C_{ij}^{\gamma\gamma}(\ell) + C_{ij}^{\rm IA\gamma}(\ell) + C_{ij}^{\rm IAIA}(\ell). \tag{6.10}$$

We start by introducing the WL galaxy source redshift density distribution convolved with the photo-z uncertainties of the i-th redshift bin $n_i^{\rm WL}(z)$ (see Figure 6.4). The total galaxy density in the i-th redshift bin is given by the equation

$$\bar{n}_i^{\rm WL} = \int_{z_{\rm min}}^{z_{\rm max}} {\rm d}z\, n_i^{\rm WL}(z), \tag{6.11}$$

where $z_{\rm min}$ and $z_{\rm max}$ are the minimum and the maximum redshift value of the redshift bin distribution, respectively. In this recipe, we assumed that the analysis will be done with 10 different tomographic redshift bins. Note that we have explicitly added a superscript 'WL' to remind us that, in general, WL and GCph samples can be different, i.e. $n_i^{\rm WL}(z) \neq n_i^{\rm GCph}(z)$, although, in this recipe so far, both are considered equal. We can now define the window functions $W_i(z)$ in terms of the WL galaxy source redshift distribution $n_i^{\rm WL}(z)$ for the two main contributions included in the modelling of the weak lensing effect (see Figure 6.5): the shear window function W_i^{γ} and the intrinsic alignment $W_i^{\rm IA}$ one, defined as:

$$W_i^{\gamma}(k,z) = \frac{3}{2}\left(\frac{H_0}{c}\right)^2 \Omega_{\rm m}(1+z)\Sigma_{\rm MG}(k,z)\chi(z)\int_z^{z_{\rm max}} {\rm d}z' n_i^{\rm WL}(z')\left[\frac{\chi(z')-\chi(z)}{\chi(z')}\right], \tag{6.12}$$

$$W_i^{\rm IA}(z) = \frac{n_i^{\rm WL}(z)}{\bar{n_i}^{\rm WL}}\frac{H(z)}{c}, \tag{6.13}$$

where $\Sigma_{\rm MG}(k,z)$ a modified gravity function. Moreover, we can relate the window functions W_i with the matter density distribution in the universe encoded in the matter power spectrum $P_{\delta\delta}$. For the case of IA, we need to introduce an intrinsic alignment model. In version 1.0, we model the intrinsic alignment using the redshift-dependent non-linear alignment (zNLA) model explained in (Euclid Collaboration et al., 2020), which is written in terms of the intrinsic alignment function $f_{\rm IA}(z)$:

$$f_{\rm IA}(z) = -A_{\rm IA}\mathcal{C}_{\rm IA}\frac{\Omega_m}{D(z)}(1+z)^{\eta^{\rm IA}}[\langle L\rangle(z)/L_\star(z)]^{\beta^{\rm IA}}, \tag{6.14}$$

where $\langle L\rangle(z)$ is the luminosity redshift-dependent mean, $L_\star(z)$ the characteristic luminosity of source galaxies as computed from the luminosity function, the parameter

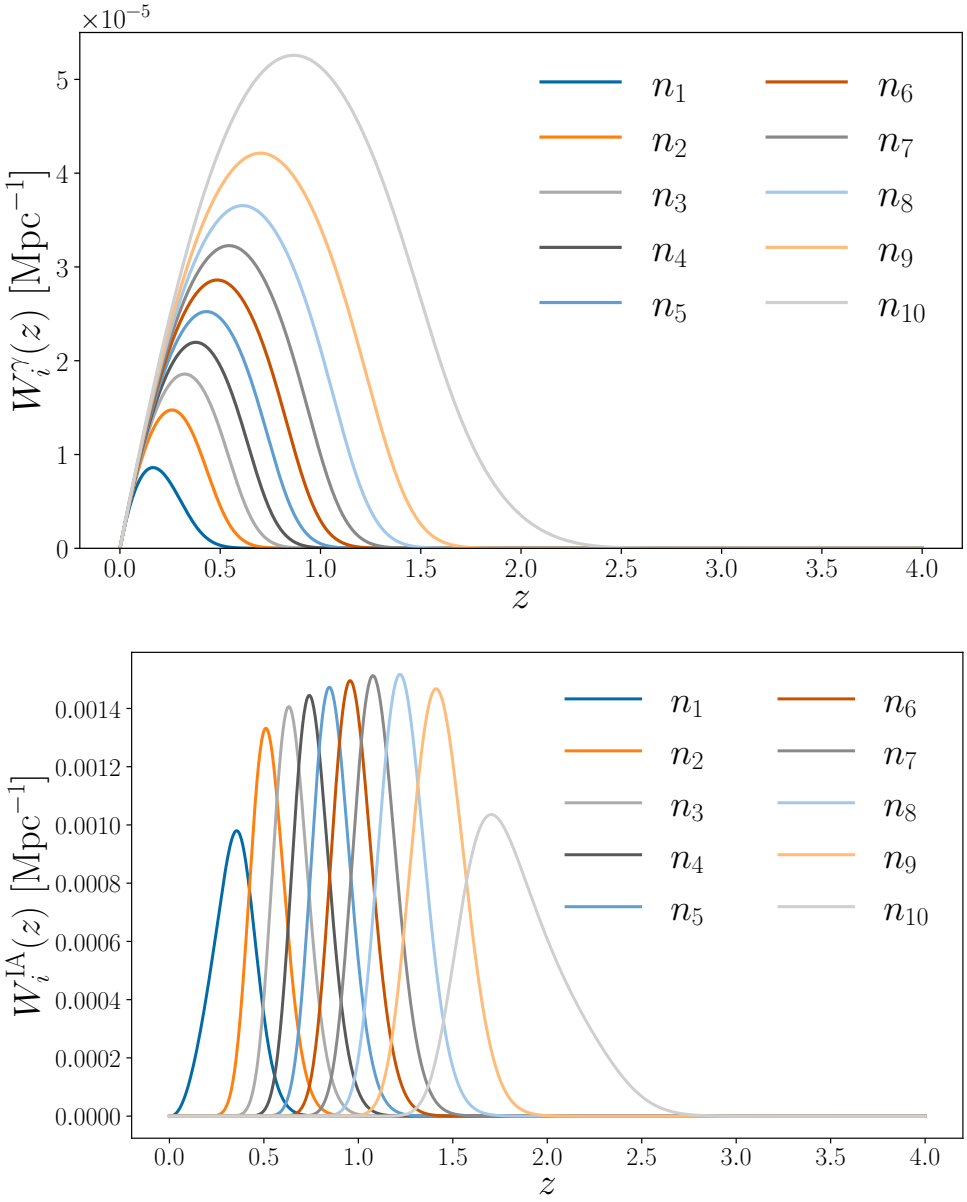

Figure 6.5: Weak lensing window functions: (**top**) the shear window function $W_i^\gamma(z)$ and (**bottom**) the intrinsic alignment $W_i^{\rm IA}(z)$ as a function of redshift z for the 10 photometric tomographic redshift bins i. The plots have been generated taking the fiducial cosmological values presented in section 6.9. Within `CLOE`, these window functions are calculated in `photo.py`.

$\beta^{\rm IA} = 0$, and $\mathcal{C}_{\rm IA} = 0.0134$ is kept fixed as it is degenerated with the rest of the parameters. The parameters $\eta^{\rm IA}$ and $A^{\rm IA}$ are free parameters of the model and should be determined by fitting the data and/or via carefully designed simulations; this is why they are defined as free nuisance parameters in this recipe so that we can marginalize over them. Given the intrinsic alignment model we can define the corresponding power spectra that relates the matter power spectrum $P_{\delta\delta}$ with the intrinsic alignment systematic distortions:

$$P_{\delta{\rm IA}}(k,z) = f_{\rm IA}(z) P_{\delta\delta}(k,z), \tag{6.15}$$

$$P_{\rm IAIA}(k,z) = [f_{\rm IA}(z)]^2 P_{\delta\delta}(k,z). \tag{6.16}$$

Finally, we can write the full expressions for each of the WL angular power spectra components present in equation (6.10) using the Limber approximation:

$$C_{ij}^{\gamma\gamma}(\ell) = c \int {\rm d}z \frac{W_i^{\gamma}\left[k_\ell(z), z\right] W_j^{\gamma}\left[k_\ell(z), z\right]}{H(z)\chi^2(z)} P_{\delta\delta}\left[k_\ell(z), z\right], \tag{6.17}$$

$$C_{ij}^{\gamma{\rm IA}}(\ell) = c \int dz \frac{W_i^{\gamma}\left[k_\ell(z), z\right] W_j^{\rm IA}(z) + W_i^{\rm IA}(z) W_j^{\gamma}\left[k_\ell(z), z\right]}{H(z)\chi^2(z)} P_{\delta{\rm IA}}\left[k_\ell(z), z\right], \tag{6.18}$$

$$C_{ij}^{\rm IAIA}(\ell) = c \int dz \frac{W_i^{\rm IA}(z) W_j^{\rm IA}(z)}{H(z)\chi^2(z)} P_{\rm IAIA}\left[k_\ell(z), z\right]. \tag{6.19}$$

The different components of $C_{ij}^{\rm WL}(\ell)$ are evaluated at $k = k_\ell(z) = \frac{\ell+1/2}{\chi(z)}$. An example plot for the total weak lensing angular power spectrum $C_{ij}^{\rm WL}(\ell)$ can be found in Figure 6.6.

6.3.3 Photometric Galaxy Clustering (GCph)

Similarly to the weak lensing case (section 6.3.2), to analyze the photometric galaxy clustering probe we still focus on the tomographic galaxy clustering angular power spectra $C_{ij}^{\rm GCph}(\ell)$ in the Fourier space using the Limber approximation and remaining up to linear scales. We start by defining the GCph galaxy source redshift density distribution $n_i^{\rm GCph}(z)$ in each tomographic bin i of the photometric survey by $n_i^{\rm CGph}(z)$, being the total angular galaxy density $\bar{n}_i^{\rm GCph}$ in such redshift bin given by

$$\bar{n}_i^{\rm GCph} = \int_{z_{\rm min}}^{z_{\rm max}} {\rm d}z \, n_i^{\rm GCph}(z). \tag{6.20}$$

We can define the window function $W_i^{\rm GCph}$ for a given tomographic bin i as

$$W_i^{\rm GCph}(z) = \frac{n_i^{\rm GCph}(z)}{\bar{n}_i^{\rm GCph}} \frac{H(z)}{c}, \tag{6.21}$$

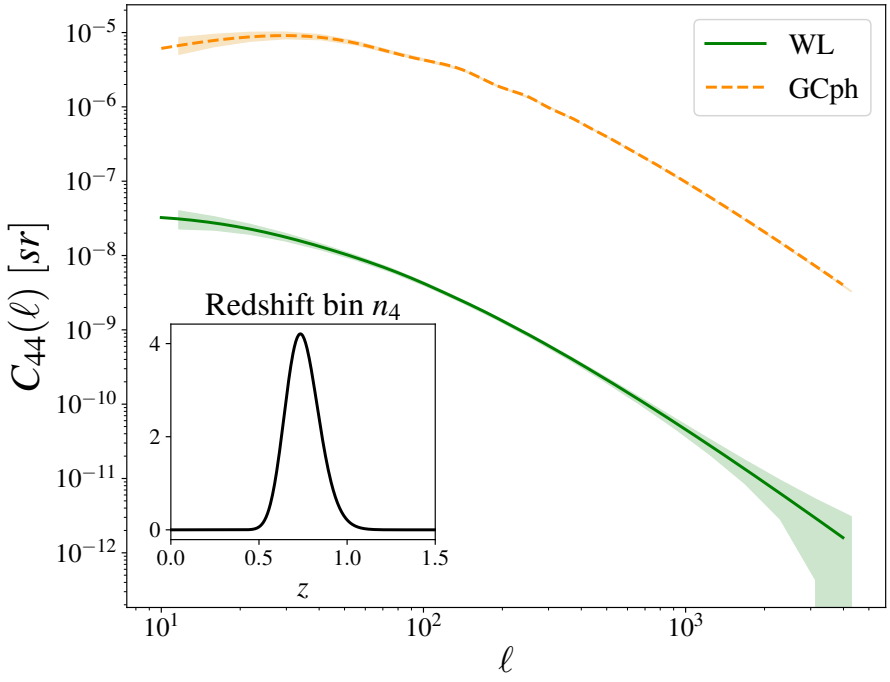

Figure 6.6: Photometric angular power spectra $C(\ell)$ for both weak lensing (WL, in green solid line) and photometric galaxy clustering (GCph, in yellow dashed lines) probes and for only one tomographic redshift bin $n_4(z)$ (that can be seen in the subplot in solid black line). The WL angular power spectrum includes both the contributions from shear and intrinsic alignment. The shades regions (WL, green and GCph, yellow) correspond to the data uncertainties given the fiducial covariance matrix as explained in subsection 6.4.1 and subsection 6.4.2. The plots have been generated taking the fiducial cosmological values presented in section 6.9. Within `CLOE`, these angular power spectra are calculated in `photo.py`.

which contains information about the probability density distribution of source galaxies at a given redshift. Similarly to the weak lensing case, we can relate the matter density power spectrum $P_{\delta\delta}$ with the galaxy-galaxy photometric power spectrum $P_{\rm gg}^{\rm GCph}(k,z)$, defined in this recipe as

$$P_{\rm gg}^{\rm GCph}(k,z) = [b^{\rm GCph}(z)]^2 P_{\delta\delta}(k,z), \tag{6.22}$$

where $b^{\rm GCph}$ is the linear photometric galaxy clustering bias and we expect to have different values for each redshift bin, so that we treat them as nuisance parameters to be sampled and marginalized over. Joining all the information together, the GCph angular power spectrum reads

$$C_{ij}^{\rm GCph}(\ell) = \int {\rm d}z \frac{W_i^{\rm GCph}(z) W_j^{\rm GCph}(z)}{H(z)\chi^2(z)} P_{\rm gg}^{\rm GCph}\left[k_\ell(z), z\right], \tag{6.23}$$

and it is evaluated at $k = k_\ell(z) = \frac{\ell+1/2}{\chi(z)}$. An example plot of equation (6.3.3) can be found at Figure 6.6.

6.3.4 Weak Lensing - Photometric Galaxy Clustering Cross-Correlation (XC)

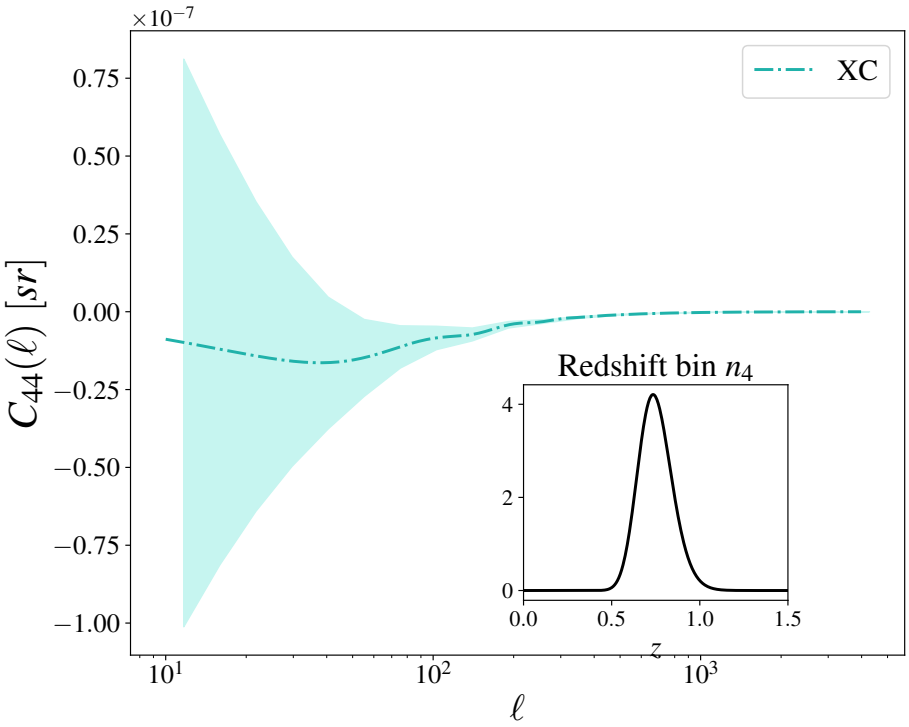

Figure 6.7: Angular power spectrum $C(\ell)$ for the photometric weak lensing - photometric Galaxy Clustering cross-correlation (XC, dashed-dotted blue line) and for only one tomographic redshift bin $n_4(z)$ (that can be seen in the subplot). The shade region (XC, blue) corresponds to the data uncertainties given the fiducial covariance matrix as explained in subsection 6.4.3. The plot has been generated taking the fiducial cosmological values presented in section 6.9. Within `CLOE`, the angular power spectrum is calculated in `photo.py`.

As for the other probes of the photometric catalogue, we focus on the angular cross-correlation between WL and GCph in Fourier space, $C_{ij}^{\rm XC}(\ell)$, with i, j labelling redshift bins, and WL and GCph respectively denoting lensing and (photometric) galaxy clustering. For each single probe that we cross-correlate, we refer to their dedicated sections (see subsection 6.3.2 and subsection 6.3.3). To calculate the cross-correlation, we need to define all the cross power spectra needed to build the corresponding angular power spectra. To take into account the shear-position correlation

we employ the following definition for the galaxy-matter power spectrum:

$$P_{\delta \mathrm{g}}^{\mathrm{GCph}}(k,z) = b^{\mathrm{GCph}}(z) P_{\delta\delta}(k,z), \tag{6.24}$$

where the photometric galaxy bias b^{photo} are those described in subsection 6.3.3. Similarly, to take into account the galaxy-IA correlation we employ the following definition of the density-intrinsic cross-spectrum:

$$P_{\mathrm{gIA}}^{\mathrm{GCph}}(k,z) = [f_{\mathrm{IA}}(z) b^{\mathrm{GCph}}(z)] P_{\delta\delta}(k,z), \tag{6.25}$$

where $f_{\mathrm{IA}}(z)$ is the intrinsic alignment function defined in equation (6.14). The weak lensing - position angular cross-spectrum $C_{\mathrm{ij}}^{\mathrm{XC}}(\ell)$ will be then given by the sum of the galaxy-shear correlation and the galaxy-IA correlation:

$$C_{ij}^{\mathrm{XC}}(\ell) = C_{ij}^{\delta \mathrm{g}}(\ell) + C_{ij}^{\mathrm{gIA}}(\ell), \tag{6.26}$$

which in practice is implemented as

$$\begin{aligned} &C_{ij}^{\mathrm{XC}}(\ell) \\ &\quad = c \int \frac{dz}{H(z) r(z)} \Big[W_i^{\gamma}\left[k_\ell(z), z\right] W_j^{\mathrm{GCph}}(z) P_{\delta \mathrm{g}}^{\mathrm{GCph}}(k_\ell, z) \\ &\qquad + W_i^{\mathrm{IA}}(z) W_j^{\mathrm{GCph}}(z) P_{\mathrm{gIA}}^{\mathrm{GCph}}(k_\ell, z) \Big], \end{aligned} \tag{6.27}$$

resulting in a single integration of the two different integrands, where $W_i^{\mathrm{GCph}}(z)$, $W_i^{\gamma}\left[k_\ell(z), z\right]$ and $W_i^{\mathrm{IA}}(z)$ are respectively defined in equations (6.21), (6.12) and (6.14). An example plot of the full observable in equation (6.27) can be found at Figure 6.7.

6.4 Benchmark data and covariance matrices

After introducing the modelling of the theoretical predictions for the *Euclid* primary observables, in this section, we introduce how the fiducial benchmark data, as well as the covariance matrices, have been generated to validate `CLOE`.

6.4.1 Weak Lensing (WL)

As a test case, we consider 10 equally populated redshift bins over the range $0.001 \leq z \leq 2.5$. We model the photometric redshift uncertainty as the sum of two Gaussian distributions: one for the well determined photometric redshifts and another for the outliers (see (Euclid Collaboration et al., 2020)), parameterized by $1 - f_{out}$ with σ_b uncertainty, and another f_{out} with σ_o uncertainty, respectively:

$$\begin{aligned} p_{\mathrm{ph}}(z_{\mathrm{p}}|z) = &\frac{1 - f_{\mathrm{out}}}{\sqrt{2\pi}\sigma_{\mathrm{b}}(1+z)} \exp\left\{ -\frac{1}{2} \left[\frac{z - c_{\mathrm{b}} z_{\mathrm{p}} - z_{\mathrm{b}}}{\sigma_{\mathrm{b}}(1+z)} \right]^2 \right\} \\ &+ \frac{f_{\mathrm{out}}}{\sqrt{2\pi}\sigma_{\mathrm{o}}(1+z)} \exp\left\{ -\frac{1}{2} \left[\frac{z - c_{\mathrm{o}} z_{\mathrm{p}} - z_{\mathrm{o}}}{\sigma_{\mathrm{o}}(1+z)} \right]^2 \right\}. \end{aligned} \tag{6.28}$$

The number density distribution $n_i(z)$ of the observed galaxies in the ith bin is given by:

$$n_i(z) = \frac{\int_{z_i^-}^{z_i^+} \mathrm{d}z_\mathrm{p}\, n(z) p_\mathrm{ph}(z_\mathrm{p}|z)}{\int_{z_\mathrm{min}}^{z_\mathrm{max}} \mathrm{d}z \int_{z_i^-}^{z_i^+} \mathrm{d}z_\mathrm{p}\, n(z) p_\mathrm{ph}(z_\mathrm{p}|z)}, \tag{6.29}$$

where (z_i^-, z_i^+) are the edges of the ith redshift bin and

$$n(z) \propto \left(\frac{z}{z_0}\right)^2 \exp\left[-\left(\frac{z}{z_0}\right)^{3/2}\right] \tag{6.30}$$

is the galaxy density distribution as defined in the "Red Book" (Laureijs et al., 2011), where $z_0 = z_\mathrm{m}/\sqrt{2}$ with z_m being the median redshift. The values for the different parameters are provided in (Euclid Collaboration et al., 2020). As mentioned in the description of the WL angular power spectra, concerning the Intrinsic Alignment effect, for this recipe we have used a zNLA model (or eNLA with $\beta_\mathrm{IA} = 0$) for the benchmark data array as well as for the covariance matrices. The following fiducial values are used $\{\mathcal{A}_\mathrm{IA}, \eta_\mathrm{IA}, \beta_\mathrm{IA}\} = \{1.72, -0.41, 0.0\}$, while $\mathcal{C}_\mathrm{IA}$ =0.0134 remains fixed in the analysis, as it is degenerated with $\mathcal{A}_\mathrm{IA}$.

Concerning the covariance matrix, we will consider only Gaussian terms and it is given by

$$\begin{aligned}
&\mathrm{Cov}[C_{ij}^\mathrm{WL}(\ell), C_{mn}^\mathrm{WL}(\ell')] \\
&\qquad = \frac{\delta^\mathrm{K}_{\ell\ell'}}{(2\ell+1) f_\mathrm{sky} \Delta\ell} \Big\{ \Big[C_{im}^{WL}(\ell) + N_{im}^\mathrm{WL}(\ell) \Big] \\
&\qquad \Big[C_{jn}^\mathrm{WL}(\ell') + N_{jn}^\mathrm{WL}(\ell') \Big] \\
&\qquad + \Big[C_{in}^\mathrm{WL}(\ell) + N_{im}^\mathrm{WL}(\ell) \Big] \\
&\qquad \Big[C_{jm}^\mathrm{WL}(\ell') + N_{jm}^\mathrm{WL}(\ell') \Big] \Big\} ,
\end{aligned} \tag{6.31}$$

where δ^K is the Kronecker delta symbol, the indexes i, j, m, n run over all tomographic bins, $\Delta\ell$ is the width of the multipoles bins and f_sky is the fraction of the sky covered by the survey (see (Euclid Collaboration et al., 2020) for details). The shot noise term due to the uncorrelated part of the intrinsic alignment ellipticity field is defined as

$$N_{ij}^\mathrm{WL}(\ell) = \frac{\sigma_\epsilon^2}{\bar{n}_i^\mathrm{WL}} \delta_{ij}^\mathrm{K} \tag{6.32}$$

with $\sigma_\epsilon = 0.3$ being the variance of the intrinsic ellipticity. It is worth mentioning that the code implementation reads the observed galaxy distribution (already convolved with the photometric redshift PDF) from an external file since it will be provided by the corresponding SGS OU as well as the covariance matrix as an external file as it will be computed (analytically or from simulations) by an external IST:NL/SWG/OU team.

Any of the validated codes used in the IST:Forecast work can give as outputs the $C_{ij}^\mathrm{WL}(\ell)$ using the galaxy density distribution in equation (6.29), as well as the

Gaussian covariance matrix. We have used the IST:Forecast codes to produce the benchmark data, which is read by `CLOE` as external data. The external benchmark data $C_{ij}^{\rm WL}(\ell)$ is computed by considering 20 logarithmically equispaced values in ℓ over the range $\ell = [10, 4000]$.

6.4.2 Photometric Galaxy Clustering (GCph)

Like in the case for external benchmark for $C_{ij}^{\rm WL}(\ell)$, we consider 10 equally populated redshift bins from $z = 0.001$ up to $z = 2.5$ with 20 multipole bins ℓ logarithmically equally spaced over the range $\ell = [10, 4000]$. The photometric bias $b^{\rm GCph}$ for the benchmark is obtained by linearly interpolating the bias values for the redshift bins $0.001 \leq z \leq 2.5$, and for redshifts above the final bin ($z > 2.5$), we use the bias obtained for the final redshift bin value. Similarly, for redshifts below the first bin ($z < 0.001$), we use the bias obtained for the first redshift bin value. In practice, this is computed as

$$b^{\rm GCph}(z) = \begin{cases} \mathcal{L}(z = 0.001) & \text{for} \quad z < 0.001 \\ \mathcal{L}(z) & \text{for} \quad 0.001 \leq z \leq 2.5 \\ \mathcal{L}(z = 2.5) & \text{for} \quad z > 2.5 \end{cases} \tag{6.33}$$

where $\mathcal{L}$(z) is obtained by linearly interpolating the points

$$[z_1, (1+z_1)^{1/2}, z_2, (1+z_2)^{1/2}, z_3, (1+z_3)^{1/2}, ..., z_{10}, (1+z_{10})^{1/2}], \tag{6.34}$$

with z_i the centre of the i-th redshift bin.

Concerning the data covariance matrix, we consider a theoretical Gaussian covariance matrix as in the Weak Lensing case in equation (6.31), but with a different shot-noise term given by

$$N_{ij}^{\rm GCph}(\ell) = \frac{\delta_{ij}^{\rm K}}{\bar{n}_i^{\rm GCph}}, \tag{6.35}$$

which is scale-independent. Similarly to the weak lensing case, `CLOE` reads both the observed galaxy distribution (already convolved with the photometric redshift PDF, which in this case is assumed to be equal to the one used for the weak lensing observable) and the covariance matrix as external files. The $C_{ij}^{\rm GCph}(\ell)$ as well as the covariance matrix are obtained by using IST:Forecast validated codes.

6.4.3 Weak Lensing - Photometric Galaxy Clustering Cross-Correlation (XC)

The benchmark data is computed as given in sections 6.4.2 and 6.4.1. Concerning the covariance matrix of the XC observable, we consider again a theoretical Gaussian

covariance matrix, which in this case reads

$$\begin{aligned}\mathrm{Cov}[C_{ij}^{\mathrm{XC}}(\ell), C_{kl}^{\mathrm{XC}}(\ell')] \\ &= \frac{\delta^{\mathrm{K}}_{\ell\ell'}}{(2\ell+1) f_{\mathrm{sky}} \Delta\ell} \Big\{ C_{il}^{\mathrm{XC}}(\ell)\, C_{jk}^{\mathrm{XC}}(\ell') + \\ &\Big[C_{ik}^{\mathrm{WL}}(\ell) + N_{ik}^{\mathrm{WL}}(\ell)\Big] \\ &\Big[C_{jl}^{\mathrm{GCph}}(\ell') + N_{jl}^{\mathrm{GCph}}(\ell')\Big]\Big\}\,,\end{aligned} \tag{6.36}$$

where the noise terms are defined in equations (6.32) and (6.35). The covariance matrix as well as the fiducial benchmark data can be read as external files and provided as an input to `CLOE`, and the $C_{ij}^{\mathrm{GCph}}(\ell)$ as well as the covariance matrix is obtained by using IST:Forecast validated codes.

6.4.4 Spectroscopic Galaxy Clustering (GCsp)

As an external benchmark, we consider predictions of the Legendre multipoles in Fourier space computed under the same recipe described in subsection 6.3.1, with covariance matrices computed within the Gaussian approximation as described in (Grieb, Sánchez, Salazar-Albornoz, & Vecchia, 2016). These predictions assume the same underlying cosmology[7] and survey specifications in terms of volume, galaxy number density and bias as of the IST:Forecast paper (Euclid Collaboration et al., 2020) (see tables 1, 2, and 3 of this paper). The analysis of the spectroscopic sample assumes 4 redshift bins from $z = 0.9$ up to $z = 1.8$, whose redshift range is summarised in Table 6.2. All theory predictions are computed at the mean redshift $z = (z_{max} - z_{min})/2$ of each bin. The input linear power spectra used to compute these external benchmarks were obtained with `CAMB`, using the parameter file stored in the `gitlab` repository of the IST:Forecast (which corresponds to the cosmology specified in Table 1 of (Euclid Collaboration et al., 2020), except for the value of σ_8 that here is 0.8156, corresponding to $A_s = 2.12605 \cdot 10^{-9}$).

The Gaussian prediction for the statistical uncertainty on the power spectrum multipoles $P_\ell(k)$ depends on the choice of the k-bin width, Δk. As we are implicitly assuming that measurements of $P_\ell(k)$ are uncorrelated under the Gaussian assumption, the minimal value for Δk should be the effective fundamental frequency defined as $k_{\mathrm{f}}^{\mathrm{eff}} \equiv 2\pi/V^{1/3}$, V being the volume sample (we assume an ideal cubic or otherwise compact volume). A value of Δk smaller than $k_{\mathrm{f}}^{\mathrm{eff}}$ would result in correlated bins even under the Gaussian assumption since $k_{\mathrm{f}}^{\mathrm{eff}}$ is the smallest difference in k we can resolve. The values of $k_{\mathrm{f}}^{\mathrm{eff}}$ for the four redshift bins are reported in Table 6.2. We assume a bin width of $\Delta k = 0.004$, a value larger but close to the largest value for $k_{\mathrm{f}}^{\mathrm{eff}}$ across the four redshift bins. The power spectra are computed in the range $0.002 < k < 0.502$ and are evaluated at the centre of the wavenumber bin. To compute the predictions for the gaussian variance we average over the bins.

[7]Except for the value of σ_8 that is 0.8156, corresponding to $A_s = 2.12605 \cdot 10^{-9}$.

redshifts	$k_{\rm f}^{\rm eff}$
$0.9 < z < 1.1$	0.0031
$1.1 < z < 1.3$	0.0030
$1.3 < z < 1.5$	0.0029
$1.5 < z < 1.8$	0.0025

Table 6.2: Values of $k_{\rm f}^{\rm eff}$ for the four spectroscopic redshift bins.

With respect to the covariance matrix, if we define the multipole expansion of the per-mode covariance in a volume $V_{\rm sur}(z)$ as (Grieb et al., 2016)

$$\sigma^2_{\ell_1\ell_2}(k^{\rm fid};\,z) \equiv \frac{(2\ell_1+1)\,(2\ell_2+1)}{V_{\rm sur}(z)}\times \int_{-1}^{1}\left[\frac{1}{q_\perp^2 q_\parallel}P_{\rm gg}\left(k(k^{\rm fid},\mu_k^{\rm fid}),\mu_k(\mu_k^{\rm fid})\right)+\frac{1}{\bar{n}}\right]^2_{\ell_1}(\mu_k^{\rm fid})\,_{\ell_2}(\mu_k^{\rm fid})\mathrm{d}\mu_k^{\rm fid}\ , \quad (6.37)$$

the Gaussian covariance of the power spectrum multipoles $P_{{\rm obs},\ell}(k^{\rm fid};\,z)$ is given by

$$\mathrm{Cov}\left[P_{{\rm obs},\ell_1}(k_i^{\rm fid};\;z),P_{{\rm obs},\ell_2}\left(k_j^{\rm fid};\,z\right)\right]=\frac{2\,(2\pi)^4}{V^2_{k_i^{\rm fid}}}\delta_{ij}\int_{k_i^{\rm fid}-\Delta k^{\rm fid}/2}^{k_i^{\rm fid}+\Delta k^{\rm fid}/2}\sigma^2_{\ell_1\ell_2}(k^{\rm fid})\,(k^{\rm fid})^2\mathrm{d}k^{\rm fid}\ , \quad (6.38)$$

where the volume of the bin in k-space is $V_{k_i^{\rm fid}} = 4\pi[(k_i^{\rm fid}+\Delta k^{\rm fid}/2)^3-(k_i^{\rm fid}-\Delta k^{\rm fid}/2)^3]/3$. We will neglect any cosmology dependence of the covariance matrix C, which is kept fixed during the analysis. For our linear theory tests, we will compute input covariance matrices using the Gaussian approximation of (Grieb et al., 2016) specifically, the covariance matrices of the Legendre multipoles are given by equations (16) and (18) of (Grieb et al., 2016). All the spectroscopic benchmark data as well as the covariance matrices are pre-computed and given as external input to `CLOE`.

6.5 Standard likelihood analysis

The estimation of constraints on a given set of cosmological parameters, θ, based on the *Euclid* data, $\vec{d}$, will follow a Bayesian framework. According to Bayes' theorem (see chapter 1, section 1.7.1), the key ingredient in the estimation of the posterior distribution of the parameters, $P(\theta|\vec{d},M)$, is the likelihood function $\mathcal{L}(\vec{d}|\theta,M)$, which describes the plausibility of a certain parameter value θ after observing a particular outcome. For all *Euclid* primary probes, in this version of the recipe, we assume that the likelihood function $\mathcal{L}(\vec{d}|\theta,M)$ of these measurements is Gaussian with a cosmology independent covariance matrix C:

$$-2\log\mathcal{L}(\vec{d}|\theta,M)\propto\left(\vec{d}-\vec{T}(\theta)\right)^t C^{-1}\left(\vec{d}-\vec{T}(\theta)\right), \quad (6.39)$$

where C^{-1} is the inverse of the covariance matrix, also known as the precision matrix, $\vec{d}$ is the data vector constructed with the fiducial benchmark data explained in section 6.4 and $\vec{T}(\theta)$ is the theory vector constructed with the predictions for the *Euclid* observables whose recipes are depicted in section 6.3.

6.6 CLOE: specifications and structure

The code *Cosmological Likelihood for Observables in Euclid* (CLOE) is designed to produce theoretical predictions of the primary *Euclid* observables (see section 6.3) as well as the computation of the likelihood (see section 6.5) given some fiducial benchmark data (detailed in section 6.4). The specifications and structure presented in this section are referred to CLOE v.1.1, which contains the theoretical recipe version 1.0. explained in section 6.3. The code CLOE is fully written in python (Van Rossum & Drake Jr, 1995) using a set of common packages described in Table 6.3. It is designed to work as an external likelihood class for the Bayesian Analysis Framework code Cobaya[8]. To build and run CLOE, IST:L provides a dedicated conda environment (*Anaconda Software Distribution*, 2020) with development tools.

Package	Version
astropy	5.0.1
matplotlib	3.5.1
scipy	1.8.0
numpy	1.22.2
pandas	1.4.1
PyYAML	6.0
Cobaya	3.1.1
getdist	1.3.3

Table 6.3: Summary of the python packages that CLOE uses in all its different modules. The versions correspond to those requested by the conda environment provided by IST:L to run CLOE v.1.1.

The structure of CLOE is very modular, as it is designed keeping in mind the flexibility and future expansion of the code. The philosophy behind CLOE's design is to make it as user-friendly as possible so that many scientists within the EC can use it when the first release of data arrives[9]. As all *Euclid* data will be publicly released after a relatively short proprietary period and will constitute for many years the ultimate survey for Cosmology, we also aim CLOE to be very easy to learn and use so that everybody within the Cosmology community can exploit its capacity and features soon.

[8]See the documentation of how to create a customized external likelihood for Cobaya at (*Creating your own cosmological likelihood class*, 2022).

[9]Note that CLOE will eventually be used within different Working Packages of the SWGs. However, IST:L will be the ultimately responsible team for producing the official constraints on the cosmological parameters.

CLOE is the result of a study and analysis of the current state-of-the-art open-science Cosmology codes available in the literature. The first goal of IST:L was to investigate the features, portability, pros and cons of the most widely used cosmological software for Bayesian Statistical analyses: CosmoMC (Lewis & Bridle, 2002; Lewis, 2013), MontePython (Audren, Lesgourgues, Benabed, & Prunet, 2013), CosmoSIS (Zuntz et al., 2015) and Cobaya (Torrado & Lewis, 2021, 2019). In the end, given the collection of requirements requested by different Science Working Groups (SWGs) as well as the goals of the *Euclid* mission, IST:L decided to interface the likelihood code with Cobaya (and in future versions, with CosmoSIS), because:

- both codes allow us to obtain computational theoretical background functions from the most widely used Boltzmann Solvers CAMB (Lewis et al., 2000; Howlett et al., 2012) and CLASS (Essinger-Hileman et al., 2014). Moreover, Cobaya is prepared to interact with modified versions of CAMB and CLASS without further hacking Cobaya source code.

- both codes contain a large number of different samplers (the most important of them explained in chapter 1: nested sampler Polychord (W. J. Handley et al., 2015a, 2015b), Metropolis Hastings, evaluate...)

- both codes accept external self-defined likelihoods with minimal modification of their source codes.

The structure of CLOE v.1.1 can be seen in figure 6.8. The software CLOE consists of a main python class inherited from Cobaya that works as an external likelihood, called cobaya_interface.py (the class being called EuclidLikelihood). This class initializes the class Euclike and collects the main cosmological background functions requested to the Boltzmann Solver (i.e: CAMB or CLASS) through Cobaya. These ingredients are saved into an instance of the class Cosmology (within the cosmo module), where they are subsequently passed on to the nonlinear module. Given the requests selected by the user, the class Euclike invokes the initialization and reading routines of the corresponding selected data and covariance matrices to create the **data vector** (using the classes Reader and Masking in figure 6.8). Given the read data, CLOE computes the theoretical predictions of the observables and produces the **theory vector** calculating the predictions for the *Euclid* primary observables coded in the modules photometric_survey and spectroscopic_survey. Finally, Euclike computes the "log-like" value (see equation 6.39), which is returned to the Cobaya class in cobaya_interface.py. The selection of the observables and further features of the theoretical predictions can be done by the user either following the standard Cobaya input files (i.e: yaml files, python scripts or even using a python interpreter as jupyter) or using the so-called CLOE *overlayer* (a series of scripts provided in CLOE to camouflage the use of Cobaya so that the user is agnostic to the whole CLOE structure). CLOE understands "on-the-fly" the nuisance parameters that are introduced in the theoretical predictions, so that cobaya_interface.py source code does not need to be further modified. The advanced structure of CLOE, seen in detail with

its several python classes, inheritance and dependencies, can be found at the IST:L gitlab repository[10].

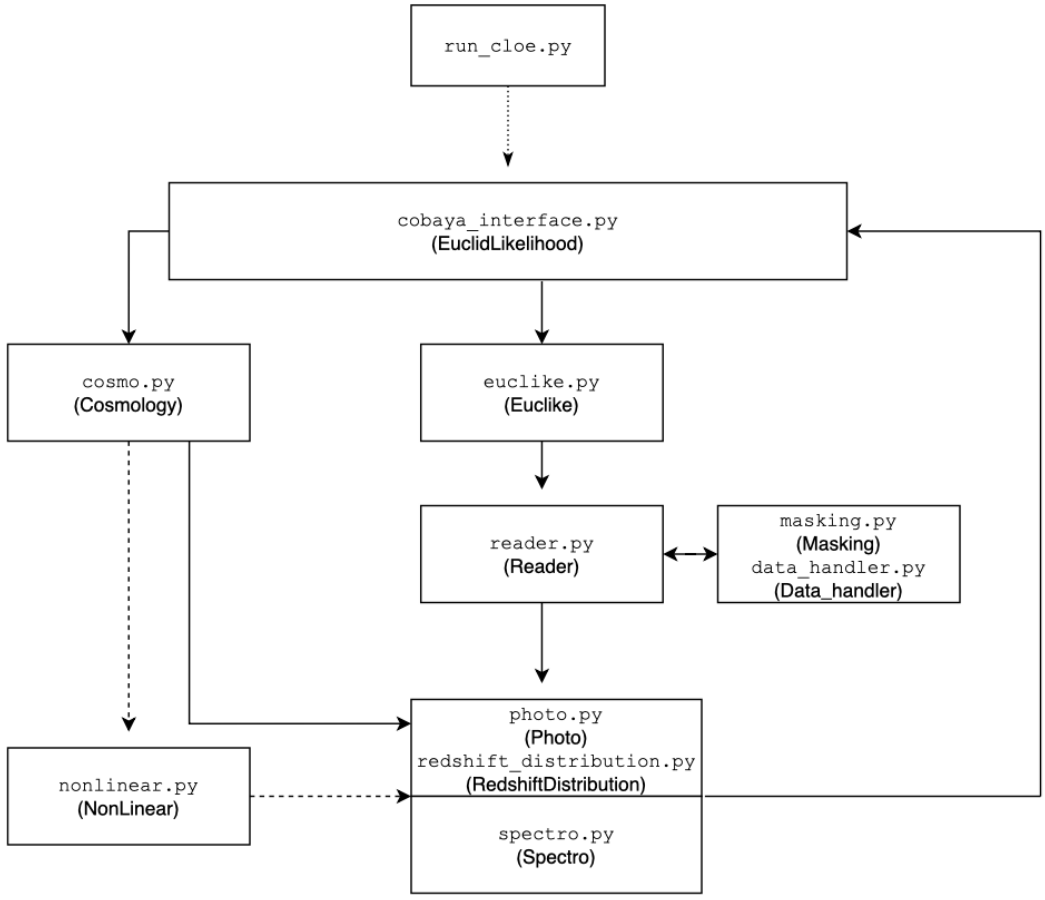

Figure 6.8: Simple diagram of the structure of CLOE in version 1.1, generated using draw.io. In the diagram it can be seen how the different modules within CLOE interact. The module run_cloe.py is informally known as the *overlayer* and the module nonlinear.py is currently implemented within the structure of CLOE but not yet used in this version (IST:NL is the responsible of the structure of this module and in this particular version, the theoretical observables are computed up to linear regime).

The software CLOE contains a README, an automatic generated documentation of the python software[11] as well as a series of scripts and notebooks. In particular, CLOE offers a DEMO jupyter notebook that shows how to compute the *Euclid* primary observables according to the corresponding recipe, internal background functions retrieved by Cobaya, the values of the likelihood, posterior and priors, the redshift

[10]Visit https://gitlab.euclid-sgs.uk/pf-ist-likelihood/likelihood-implementation/-/wikis/uploads/083b51a7591215dffecdbe511446d8b5/classes.png

[11]The automated generated documentation is created using "Read the docs" https://docs.readthedocs.io/en/stable/index.html.

distributions of the galaxies for both the photometric and spectroscopic catalogues and even the photometric window functions and various matter power spectra explained in subsection 6.3.1, subsection 6.3.2, subsection 6.3.3, subsection 6.3.4 and in table 6.3. All the figures of section 6.3 have been produced using this DEMO jupyter notebook.

6.7 Validation of the benchmark data using CLOE

To validate the theoretical recipe implemented in CLOE of the *Euclid* primary probes, we carry out a Monte Carlo Markov Chain (MCMC) sampling technique to retrieve the posterior distributions of the cosmological and nuisance parameters given the fiducial benchmark data explained in section 6.4. We take a flat ΛCDM model as the baseline cosmological model and we use the theoretical prediction of the *Euclid* primary probes up to linear order (non-linear corrections are not included). The sampler used is the Metropolis-Hastings `mcmc` included in `Cobaya`. The Boltzmann Solver is `CAMB`. In all cases, we use the default `Cobaya` convergence criterion (i.e: R-1 (means) < 0.01 and R-1 (standard deviations) < 0.2, see section 1.7.2 for more details about the definition of the convergence criterion R). We show in Table 6.4 the time required for each run to reach convergence according to these conditions while letting the nuisance parameters free. These runs were done in the `xMaris` cluster at the Lorentz Institute. `xMaris` runs on CentOS v7.6 and has as architecture x86_64. To run `CLOE` in this cluster, the `conda` environment it is not used and `python` v.3.7.4, `gcc` v.8.30, `OpenMPI` v.3.1.4 and `mpi4py` v.3.1.1 are used instead. It has been verified that very similar running times were obtained when a different cluster was used (`Alice` cluster at Leiden University).

The list of free sampled parameters included 5 cosmological parameters ($\Omega_{\rm b}$, $\Omega_{\rm m}$, H_0, n_s, A_s, where σ_8 and S_8 are obtained as derived parameters) and up to 16 nuisance parameters (the 10 different photometric bias parameters $b^{\rm GCph}$ for the 10 photometric redshift bins, the 4 different spectroscopic bias parameters $b^{\rm GCsp}$ for the 4 spectroscopic redshift bins, and the 2 intrinsic alignment nuisance parameters $A^{\rm IA}$ and $\eta^{\rm IA}$). The running scripts can be found in the *Euclid* `gitlab` repository of `CLOE`.

All the runs consisted of 8 chains, each of them crossing into 4 cores. The chains achieved a convergence of $R - 1 < 0.1$ in approximately 3 days, where some of them were visually converged, in particular, those runs that included the combinations of different observational probes. However, to reach the standard convergence criterion, the runs needed several days to end. The average computation time for `CAMB` was approximately 4.2 seconds per evaluation and that of `CLOE` around 7.1 seconds per evaluation. Note that this average time for `CLOE` is the sum of both the computation time of the theoretical observables as well as the computation time of the likelihood, and therefore, the average time for the theoretical predictions computed by `CLOE` is around 3 seconds per evaluation. Each chain contains around $\mathcal{O}(5)$ accepted steps.

After reaching convergence, the analysis and post-processing of the chains is performed with `GetDist` (Lewis, 2019). The list of best fit values, together with their 68% confidence intervals, are obtained after marginalizing the posterior distributions.

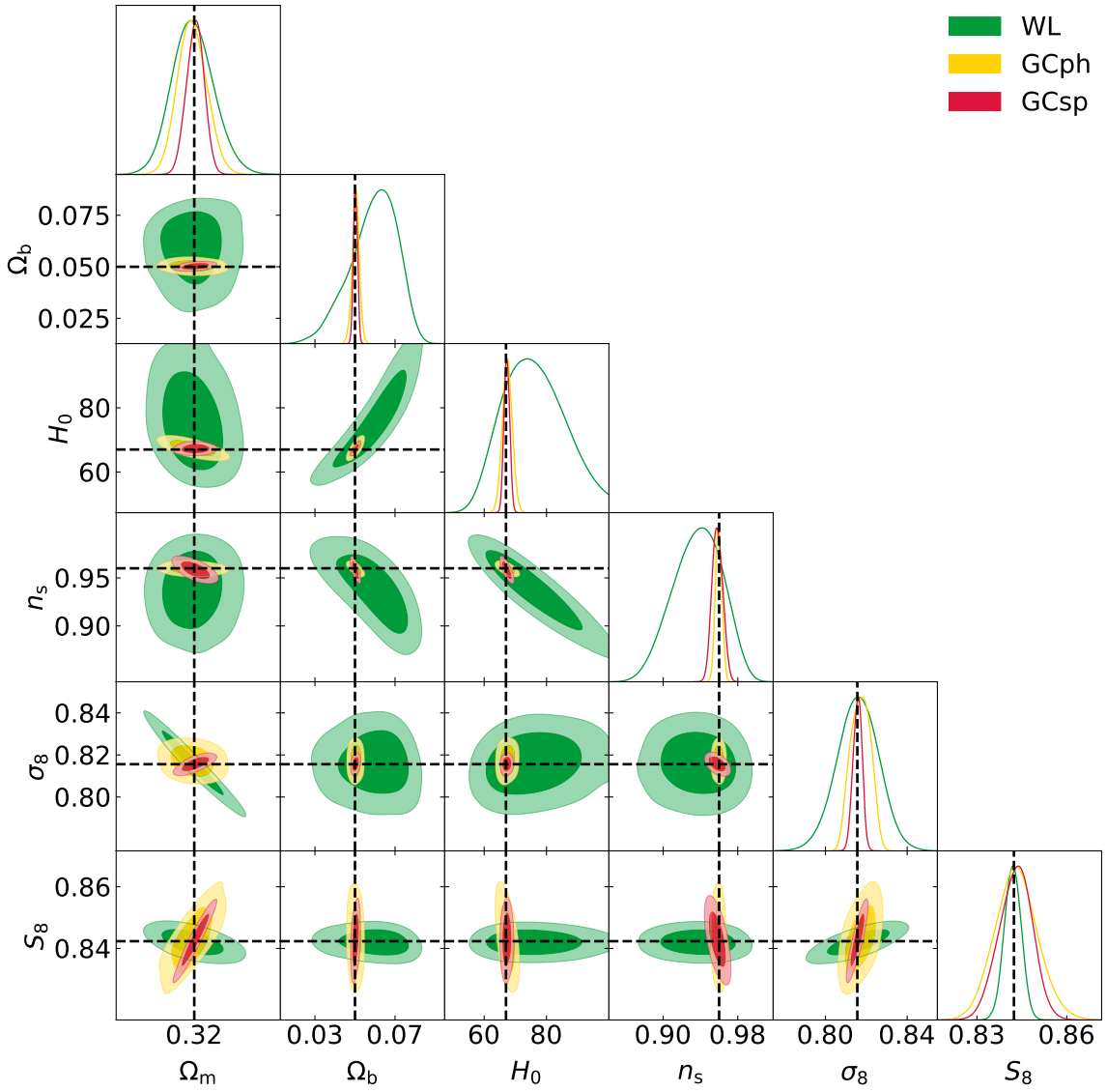

Figure 6.9: Constraints on the cosmological parameters using *Euclid* primary observational probes. This plot shows the different observational primary probes independently, i.e. weak lensing (WL) in green, photometric galaxy clustering (GCph) in yellow and spectroscopic galaxy clustering (GCsp) in red. The red dashed lines show the fiducial values of the parameters.

These best fits and uncertainties are gathered in the appendix section 6.9, in Table 6.5, Table 6.6, Table 6.7, Table 6.8 and Table 6.9. The benchmark data have been generated using a list of fiducial values for the cosmological and nuisance parameters that are also present in those tables. We also use the fiducial values for the cosmological parameters in plots of the posterior distributions, Figure 6.9 and Figure 6.10 for visualisation purposes.

As expected, the individual *Euclid* primary probes alone (WL, GCph and GCsp) do not show a big constraining power on the cosmological parameters. In fact, photo-

Data combination	Node	Elapsed time
GCsp	maris069	14 days, 18 hours, 52 minutes
GCph	maris070	42 days, 5 hours, 32 minutes
WL	maris072	16 days, 3 hours, 24 minutes
XC	maris073	17 days, 19 hours, 34 minutes
GCph+WL	maris070	14 days, 23 hours, 1 minutes
3x2pt	maris073	10 days, 4 hours, 3 minutes
GCph+WL+GCsp	maris073	15 days, 17 hours, 43 minutes
3x2pt+GCsp	maris072	7 days, 20 hours, 14 minutes

Table 6.4: Computation time to reach convergence for each of the considered *Euclid* probe combinations in the case where the nuisance parameters are free to vary. All nodes have as hardware an opteron6376 with 1Gb,R815,highmem. The combination 3x2pt refers to WL+GCph+XC together.

metric galaxy clustering (GCph) shows some degeneracies between the nuisance bias parameters b^{GCph} and the amplitude of the power spectrum A_s, and therefore, also between σ_8 (see Figure 6.9). This is the main reason why the computation time of this probe to reach convergence is the largest with respect to the other observational probes. To achieve convergence, the proposal step of the photometric nuisance bias parameters' priors have been adjusted to the optimal value to avoid the chains to get stuck. Weak lensing (WL) also shows degeneracies in the Hubble parameter H_0 and in the density of baryons Ω_{b}. Moreover, its constraining power on the spectral index n_s is more limited than that of GCph. Contrarily, WL constrains σ_8, due to the fact that WL is very sensitive to the matter density Ω_{m}. The spectroscopic galaxy clustering is overall the primary observational probe that has a good constraining power on all the cosmological parameters.

When combined, we can unleash the full power of *Euclid* primary probes (see Figure 6.10). The combination of both photometric primary probes (WL and GCph) breaks the degeneracies previously shown by the individual photometric probes (in particular, in σ_8, H_0 and Ω_{b}). Including the cross-correlation between weak lensing and photometric galaxy clustering (XC) in the analysis; in order words, the full 3x2pt photometric combination, increases the constraining power in σ_8 and Ω_{m}, although it does not improve the constraining power on the rest of parameters. However, when GCsp is included in the analysis, the confidence intervals of all cosmological parameters reduce significantly, showing how the combination of all *Euclid* primary probes will be the future state-of-the-art of new LSS cosmological analysis, being at the level of constraints obtained by Planck (if not successfully beating it).

For the case of the 3x2pt with GCsp, the analysis included 16 free nuisance parameters. The 12 photometric nuisance parameters, bias and and intrinsic alignment ones, are shown in Figure 6.11. It can be observed how adding the cross-correlation XC between weak lensing and photometric galaxy clustering improves the constraining power on the nuisance parameters, in particular, the intrinsic alignment parameters. In the case of the analysis of the spectroscopic bias parameters (see Figure 6.12),

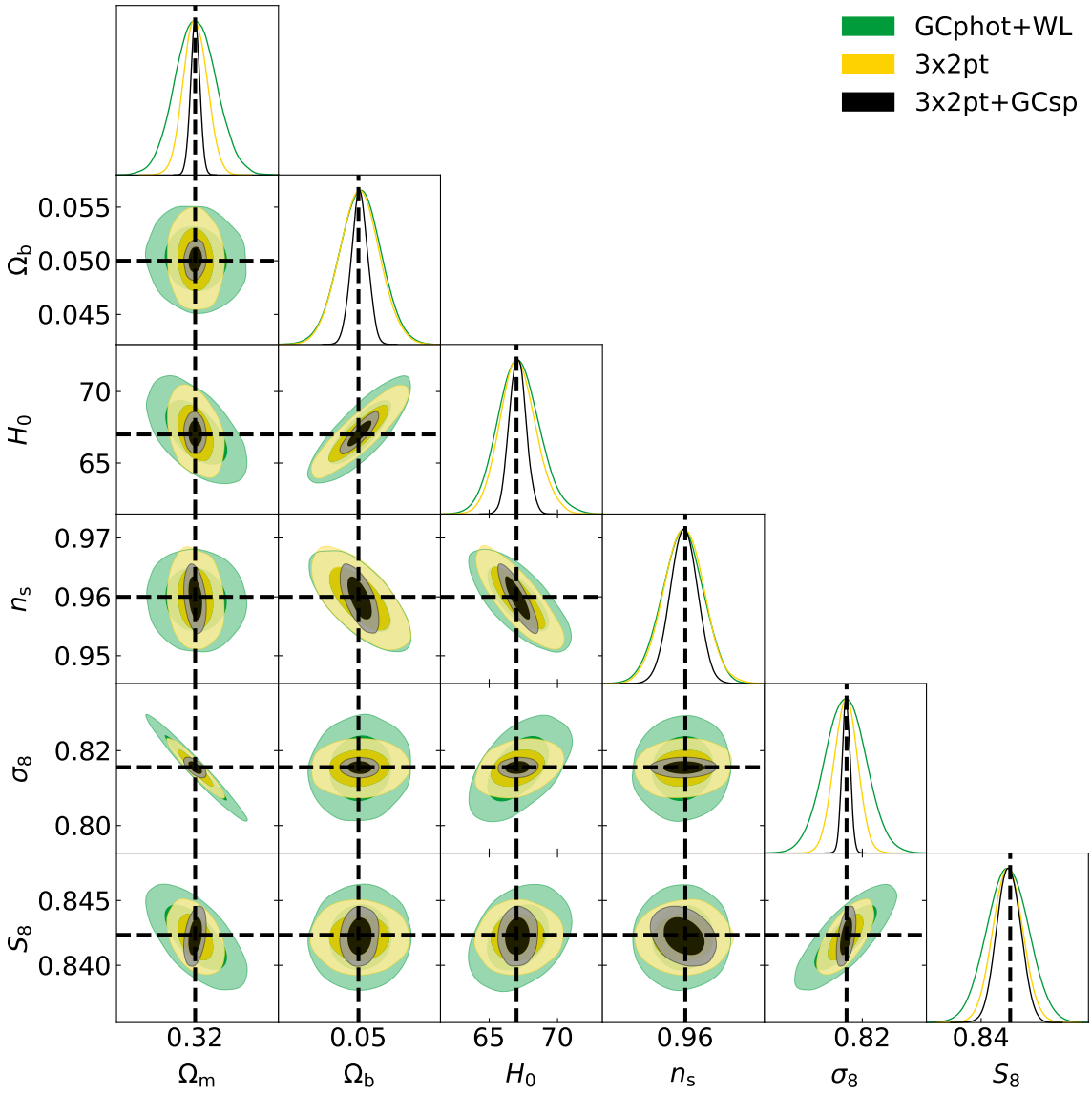

Figure 6.10: Constraints on the cosmological parameters using all possible (available) combinations of *Euclid* primary probes. The green contours show the combination of the two photometric probes, GCph and WL, yellow contours add to this combination the cross-correlation of these two probes (XC), producing the standard 3x2pt combination, and in the black contours, GCsp is added to this last combination as an independent probe (3x2pt+GCsp). The black dashed lines show the fiducial values of the parameters.

the spectroscopic bias parameters get further constrained when the 3x2pt probe is included in the analysis. We conclude that the improvement on all the nuisance parameters is a natural consequence of the improvement in constraining power of the cosmological parameters as more *Euclid* probes are combined.

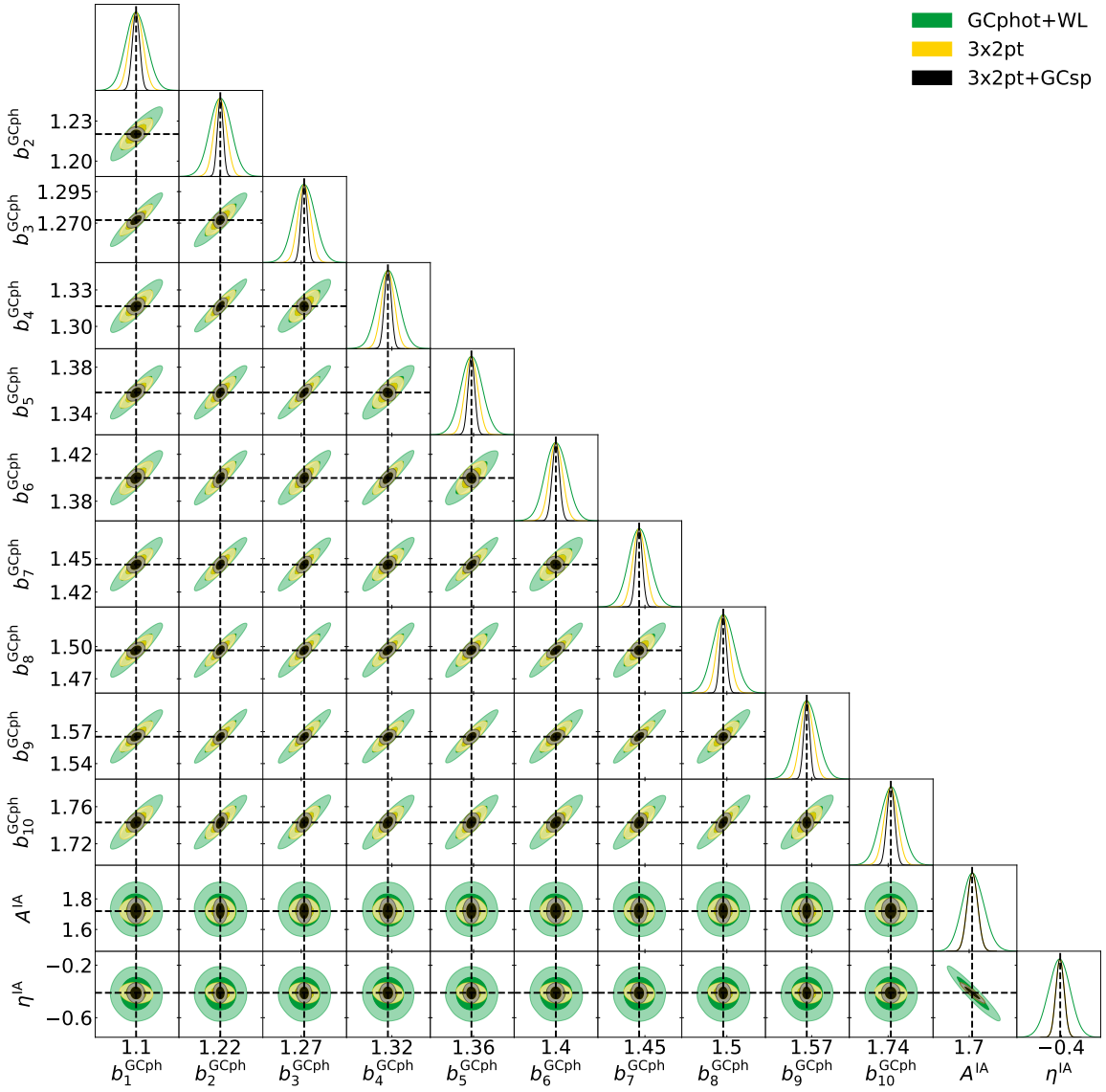

Figure 6.11: Constraints on the nuisance parameters of the photometric observational probes of *Euclid*: WL, GCph and XC. All photometric nuisance parameters of *Euclid* are shown, including the 10 different photometric galaxy clustering bias parameters $b^{\rm GCph}$ and the two weak lensing intrinsic alignment parameters $A^{\rm IA}$ and $\eta^{\rm IA}$. The green contours show the combination of the two photometric probes, GCph and WL, yellow contours add to this combination the cross-correlation of these two probes, producing the standard 3x2pt combination, and in the black contours, GCsp is added to this last combination as an independent probe (3x2pt+GCsp). The black dashed lines show the fiducial values of the nuisance parameters.

6.8 The future of CLOE

In the previous sections, we have seen how we can use the *Cosmological Likelihood for Observables in Euclid* (CLOE) to plot *Euclid* primary observables and other internal

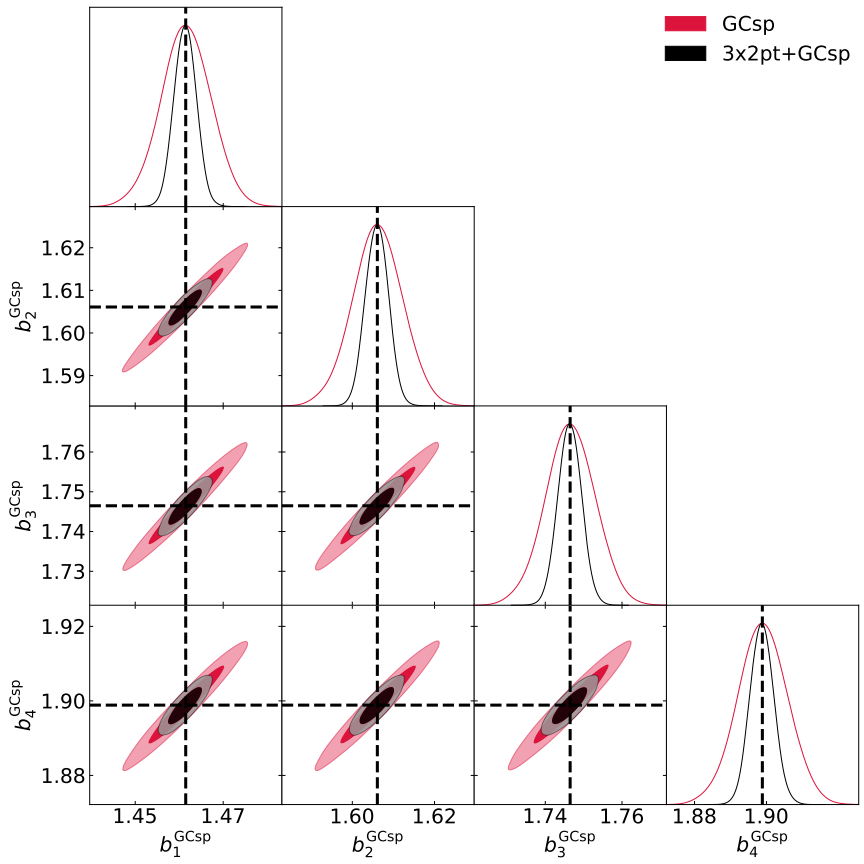

Figure 6.12: Nuisance parameters of the spectroscopic observational probes of *Euclid*. The 4 bias nuisance parameters of the spectroscopic *Euclid* probe are shown. The red contours show the bias parameters when only the Spectroscopic Galaxy Clustering is used, whereas, in the black contours, the photometric 3x2pt probe is added to the analysis. The black dashed lines show the fiducial values of the nuisance parameters.

ingredients needed for the computation of the theoretical recipe. Moreover, we have demonstrated that we can use `CLOE` as an external likelihood for `Cobaya` to perform parameter inference. Nevertheless, to do meaningful science, `CLOE` needs to evolve. In fact, versions 1.0 and 1.1 of `CLOE` can be considered a proof-of-concept where we set up the general structure of code, the interaction among the modules and the interface with `Cobaya` as well as with IST:NL. However, for real scientific performance, several changes are needed.

First, we need to update the recipe for the *Euclid* primary probes to include a more realistic modelling of the observables. For the spectroscopic survey, in addition to the observed Legendre multipoles $P_\ell(k)$, we aim to include the observed two-point correlation $\xi(s)$ function multipoles, where s is the separation between two galaxies in the survey. Furthermore, for the photometric survey, in particular photometric galaxy clustering, we will incorporate new angular power spectra terms in $C_{ij}^{\rm GCph}$ to account for contributions coming from redshift space distortions (RSD) and magnification. For the weak lensing observable, the WL galaxy redshift distribution $n_i^{\rm WL}(z)$ will be

replaced by an effective one to take into account the shear multiplicative bias and the blending terms. Furthermore, due to the sensitivity of the weak lensing probe on the non-linear model used to correct the matter power spectrum, the Bernaredeau - Nishimichi - Taruya (BNT) transform (see (Bernardeau, Nishimichi, & Taruya, 2014)) will be also included in the lensing window functions to minimise the overlap between them so that each tomographic bin only receives contribution from a well defined range in scale. For that, we will follow the recipes from (P. L. Taylor, Bernardeau, & Kitching, 2018; P. L. Taylor et al., 2021). For both cases, WL and GCph, we will allow testing non-flat cosmological models by including the corresponding changes into the equations and requesting new background functions to the Boltzmann solvers through `Cobaya`. Besides, the equations for the photometric observables will be updated to include the Modify Gravity function $\mu_{\rm MG}$ too.

Second, we will need to work towards a complete `CLOE` version that contains the necessary implementations to correct *Euclid* primary probes at non-linear scales. In this sense, a merge of IST:NL software development into `CLOE` is essential for the release of v.2.0 as well as for the validation of the current `nonlinear` module interface. Finally, we need to validate the new theoretical recipe against fiducial benchmark data and we need to check the overall performance of `CLOE`, giving particular attention to the speed of the code, and introduce any software optimisation if required.

`CLOE` is planned to be used for the third *Euclid* Science Performance Verification (SPV3) exercise. This activity aims to assess whether the expected performances of the *Euclid* project are in line with the core science objectives of the nominal mission, and furthermore, it will ensure gaining insights into the performances of the project and identify any critical aspect for the realization of the core goals in terms of systematic errors (either of instrumental, astrophysical of theoretical origin). For this exercise, IST:L will use `CLOE` to produce the official constraints on the cosmological parameters given the w_0w_aCDM model. For this reason, the release of `CLOE` v.2.0, including the above-mentioned modifications on the theoretical recipe of the observables, non-linear corrections and required speed optimisations, is planned before the end of 2022.

6.9 Appendix: fiducial and best fit values for the cosmological and nuisance parameters

	$\Omega_{\rm b}$	$\Omega_{\rm m}$	H_0	$n_{\rm s}$	σ_8
Fiducial	0.05	0.32	67	0.96	0.8156
GCsp	0.0502 ± 0.0020	0.3197 ± 0.0048	$67.3^{+1.5}_{-1.8}$	0.9594 ± 0.0036	$0.700^{+0.031}_{-0.057}$
GCph	0.0504 ± 0.0018	0.3194 ± 0.0047	67.4 ± 1.6	0.9595 ± 0.0035	0.8172 ± 0.0047
WL	$0.0593^{+0.014}_{-0.0089}$	$0.3196^{+0.0062}_{-0.0070}$	76^{+9}_{-10}	$0.937^{+0.029}_{-0.024}$	0.816 ± 0.010
XC	0.0513 ± 0.0053	0.3180 ± 0.0095	$68.7^{+3.9}_{-5.3}$	0.9572 ± 0.0098	0.8172 ± 0.0056
GCph+WL	0.0502 ± 0.0020	0.3201 ± 0.0036	67.2 ± 1.5	0.9595 ± 0.0036	0.8153 ± 0.0058
3x2pt	0.0502 ± 0.0020	0.3200 ± 0.0021	$67.2^{+1.2}_{-1.4}$	0.9596 ± 0.0036	0.8154 ± 0.0033
GCph+WL+GCsp	0.05007 ± 0.00079	0.32000 ± 0.00086	67.06 ± 0.62	0.9598 ± 0.0025	0.8154 ± 0.0012
3x2pt+GCsp	0.05009 ± 0.00079	0.32001 ± 0.00080	67.07 ± 0.60	0.9598 ± 0.0024	0.8155 ± 0.0011

Table 6.5: Recovered best fit values and 68% errors for the cosmological parameters for the considered probe (and probe combinations) when the nuisance parameters were let free.

	$b_1^{\rm GCph}$	$b_2^{\rm GCph}$	$b_3^{\rm GCph}$	$b4^{\rm GCph}$	$b5^{\rm GCph}$
Fiducial	1.0997727037892875	1.220245876862528	1.2723993083933989	1.316624471897739	1.35812370570578
GCph	1.0976 ± 0.0067	1.2177 ± 0.0071	1.2697 ± 0.0076	1.3139 ± 0.0072	1.3549 ± 0.0076
XC	1.101 ± 0.017	1.220 ± 0.011	1.271 ± 0.012	1.3152 ± 0.0083	1.353 ± 0.010
GCph+WL	1.1000 ± 0.0077	1.2205 ± 0.0082	1.2726 ± 0.0087	1.3169 ± 0.0089	1.3583 ± 0.0094
3x2pt	1.0998 ± 0.0050	1.2203 ± 0.0048	1.2724 ± 0.0051	1.3167 ± 0.0054	1.3581 ± 0.0056
GCph+WL+GCsp	1.0998 ± 0.0028	1.2203 ± 0.0024	1.2724 ± 0.0028	1.3167 ± 0.0030	1.3581 ± 0.0032
3x2pt+GCsp	1.0997 ± 0.0027	1.2203 ± 0.0022	1.2724 ± 0.0026	1.3166 ± 0.0029	1.3581 ± 0.0030

Table 6.6: Recovered best fit values and 68% errors for the first five photometric bias nuisance parameters for the considered probe (and probe combinations) when the cosmological parameters were let free.

	$b_6^{\rm GCph}$	$b_7^{\rm GCph}$	$b_8^{\rm GCph}$	$b9^{\rm GCph}$	$b10^{\rm GCph}$
Fiducial	1.3998214171814918	1.4446452851824907	1.4964959071110084	1.5652475842498528	1.7429859437184225
GCph	$1.3971^{+0.0083}_{-0.0095}$	1.4416 ± 0.0085	1.4933 ± 0.0089	1.5621 ± 0.0091	1.739 ± 0.011
XC	1.397 ± 0.013	1.442 ± 0.013	1.496 ± 0.016	1.560 ± 0.016	1.750 ± 0.029
GCph+WL	1.4000 ± 0.0095	1.4449 ± 0.0098	1.497 ± 0.010	1.565 ± 0.010	1.743 ± 0.012
3x2pt	1.3999 ± 0.0058	1.4446 ± 0.0059	1.4965 ± 0.0061	1.5653 ± 0.0060	1.7429 ± 0.0074
GCph+WL+GCsp	1.3999 ± 0.0034	1.4446 ± 0.0033	1.4965 ± 0.0031	1.5653 ± 0.0030	1.7430 ± 0.0044
3x2pt+GCsp	1.3998 ± 0.0033	1.4446 ± 0.0031	1.4965 ± 0.0030	1.5652 ± 0.0029	1.7429 ± 0.0043

Table 6.7: Recovered best fit values and 68% errors for the first five photometric bias nuisance parameters for the considered probe (and probe combinations) when the cosmological parameters were let free.

	$b_1^{\rm GCsp}$	$b_2^{\rm GCsp}$	$b_3^{\rm GCsp}$	$b4^{\rm GCsp}$
Fiducial	1.4614804	1.6060949	1.7464790	1.8988660
GCsp	1.4615 ± 0.0057	1.6061 ± 0.0061	1.7465 ± 0.0065	1.8989 ± 0.0070
3x2pt+GCsp	1.4614 ± 0.0025	1.6060 ± 0.0028	1.7464 ± 0.0031	1.8988 ± 0.0033

Table 6.8: Recovered best fit values and 68% errors for the spectroscopic bias nuisance parameters for the considered probe (and probe combinations) when the cosmological parameters were let free.

	$A^{\rm IA}$	$\eta^{\rm IA}$
Fiducial	1.72	-0.41
WL	1.724 ± 0.076	-0.412 ± 0.094
XC	1.722 ± 0.038	-0.404 ± 0.046
GCph+WL	$1.748^{+0.079}_{-0.064}$	$-0.441^{+0.069}_{-0.093}$
3x2pt	1.722 ± 0.035	-0.411 ± 0.032
GCph+WL+GCsp	1.725 ± 0.073	-0.415 ± 0.086
3x2pt+GCsp	1.722 ± 0.035	-0.411 ± 0.032

Table 6.9: Recovered best fit values and 68% errors for the intrinsic alignment nuisance parameters for the considered probe (and probe combinations) when the cosmological parameters were let free.

References

Abazajian, K., Addison, G. E., Adshead, P., Ahmed, Z., Akerib, D., Ali, A., ... CMB-S4 Collaboration (2022, February). CMB-S4: Forecasting Constraints on Primordial Gravitational Waves. *The Astrophysical Journal*, *926*(1), 54. doi: 10.3847/1538-4357/ac1596

Abazajian, K. N., Adshead, P., Ahmed, Z., Allen, S. W., Alonso, D., Arnold, K. S., ... Kimmy Wu, W. L. (2016, October). CMB-S4 Science Book, First Edition. *arXiv e-prints*, arXiv:1610.02743.

Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., ... Virgo Collaboration (2016, February). Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, *116*(6), 061102. doi: 10.1103/PhysRevLett.116.061102

Abbott, B. P., Abbott, R., Abbott, T. D., Abraham, S., Acernese, F., Ackley, K., ... Virgo Collaboration (2019, July). GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *Physical Review X*, *9*(3), 031040. doi: 10.1103/PhysRevX.9.031040

Abbott, B. P., Abbott, R., Abbott, T. D., Abraham, S., Acernese, F., Ackley, K., ... Virgo Collaboration (2021, March). A Gravitational-wave Measurement of the Hubble Constant Following the Second Observing Run of Advanced LIGO and Virgo. *The Astrophysical Journal*, *909*(2), 218. doi: 10.3847/1538-4357/abdcb7

Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., ... et al. (2017a, October). Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A. *The Astrophysical Journal*, *848*(2), L13. doi: 10.3847/2041-8213/aa920c

Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., ... Serra-Ricart, M. (2017b, November). A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, *551*(7678), 85-88. doi: 10.1038/nature24471

Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., ... Virgo Collaboration (2017c, October). GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Physical Review Letters*, *119*(16), 161101. doi: 10.1103/PhysRevLett.119.161101

Abbott, R., Abbott, T. D., Abraham, S., Acernese, F., Ackley, K., Adams, A., ... Virgo Collaboration (2021, May). Population Properties of Compact Objects

from the Second LIGO-Virgo Gravitational-Wave Transient Catalog. *The Astrophysical Journal*, *913*(1), L7. doi: 10.3847/2041-8213/abe949

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., Aleksić, J., Allam, S., Allen, S., ... Dark Energy Survey Collaboration (2018, August). Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. *Physical Review D*, *98*(4), 043526. doi: 10.1103/PhysRevD.98.043526

Abdalla, E., Abellán, G. F., Aboubrahim, A., Agnello, A., Akarsu, Ö., Akrami, Y., ... Zumalacárregui, M. (2022, June). Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies. *Journal of High Energy Astrophysics*, *34*, 49-211. doi: 10.1016/j.jheap.2022.04.002

Abrahamsson, P., Salo, O., Ronkainen, J., & Warsta, J. (2017, September). Agile Software Development Methods: Review and Analysis. *arXiv e-prints*, arXiv:1709.08439.

Achúcarro, A., Atal, V., Céspedes, S., Gong, J.-O., Palma, G. A., & Patil, S. P. (2012, December). Heavy fields, reduced speeds of sound, and decoupling during inflation. *Physical Review D*, *86*(12), 121301. doi: 10.1103/PhysRevD.86.121301

Achúcarro, A., Atal, V., Hu, B., Ortiz, P., & Torrado, J. (2014, July). Inflation with moderately sharp features in the speed of sound: Generalized slow roll and in-in formalism for power spectrum and bispectrum. *Physical Review D*, *90*(2), 023511. doi: 10.1103/PhysRevD.90.023511

Achúcarro, A., Atal, V., Ortiz, P., & Torrado, J. (2014, May). Localized correlated features in the CMB power spectrum and primordial bispectrum from a transient reduction in the speed of sound. *Physical Review D*, *89*(10), 103006. doi: 10.1103/PhysRevD.89.103006

Achúcarro, A., Gong, J.-O., Hardeman, S., Palma, G. A., & Patil, S. P. (2011a, January). Features of heavy physics in the CMB power spectrum. *Journal of Cosmology and Astroparticle Physics*, *1*, 030. doi: 10.1088/1475-7516/2011/01/030

Achúcarro, A., Gong, J.-O., Hardeman, S., Palma, G. A., & Patil, S. P. (2011b, September). Mass hierarchies and nondecoupling in multi-scalar-field dynamics. *Physical Review D*, *84*(4), 043502. doi: 10.1103/PhysRevD.84.043502

Achúcarro, A., Gong, J.-O., Hardeman, S., Palma, G. A., & Patil, S. P. (2012, May). Effective theories of single field inflation when heavy fields matter. *Journal of High Energy Physics*, *2012*, 66. doi: 10.1007/JHEP05(2012)066

Achúcarro, A., Gong, J.-O., Palma, G. A., & Patil, S. P. (2013, June). Correlating features in the primordial spectra. *Physical Review D*, *87*(12), 121301. doi: 10.1103/PhysRevD.87.121301

Achúcarro, A., Kallosh, R., Linde, A., Wang, D.-G., & Welling, Y. (2018, April). Universality of multi-field α-attractors. *Journal of Cosmology and Astroparticle Physics*, *2018*(4), 028. doi: 10.1088/1475-7516/2018/04/028

Achúcarro, A., & Welling, Y. (2015, February). Multiple Field Inflation and Signatures of Heavy Physics in the CMB. *arXiv e-prints*, arXiv:1502.04369.

Adams, M. R., & Cornish, N. J. (2014, January). Detecting a stochastic gravitational wave background in the presence of a galactic foreground and instrument noise.

Physical Review D, *89*(2), 022001. doi: 10.1103/PhysRevD.89.022001

Ade, P., Aguirre, J., Ahmed, Z., Aiola, S., Ali, A., Alonso, D., ... Simons Observatory Collaboration (2019, February). The Simons Observatory: science goals and forecasts. *Journal of Cosmology and Astroparticle Physics*, *2019*(2), 056. doi: 10.1088/1475-7516/2019/02/056

Ade, P., et al. (2018, November). BICEP2 / Keck Array x: Constraints on Primordial Gravitational Waves using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season. *Physical Review Letters*, *121*, 221301. doi: 10.1103/PhysRevLett.121.221301

Ade, P. A. R., et al. (2016, January). Improved Constraints on Cosmology and Foregrounds from BICEP2 and Keck Array Cosmic Microwave Background Data with Inclusion of 95 GHz Band. *Physical Review Letters*, *116*, 031302. doi: 10.1103/PhysRevLett.116.031302

Adshead, P., & Hu, W. (2014, April). Bounds on nonadiabatic evolution in single-field inflation. *Physical Review D*, *89*(8), 083531. doi: 10.1103/PhysRevD.89.083531

Akaike, H. (1974, December). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723. doi: 10.1109/TAC.1974.1100705

Akrami, Y., Brax, P., Davis, A.-C., & Vardanyan, V. (2018, June). Neutron star merger GW170817 strongly constrains doubly coupled bigravity. *Physical Review D*, *97*(12), 124010. doi: 10.1103/PhysRevD.97.124010

Akrami, Y., Casas, S., Deng, S., & Vardanyan, V. (2021, April). Quintessential α-attractor inflation: forecasts for Stage IV galaxy surveys. *Journal of Cosmology and Astroparticle Physics*, *2021*(4), 006. doi: 10.1088/1475-7516/2021/04/006

Akrami, Y., Kallosh, R., Linde, A., & Vardanyan, V. (2018, June). Dark energy, α-attractors, and large-scale structure surveys. *Journal of Cosmology and Astroparticle Physics*, *06*, 041. doi: 10.1088/1475-7516/2018/06/041

Alonso, D., Cusin, G., Ferreira, P. G., & Pitrou, C. (2020, July). Detecting the anisotropic astrophysical gravitational wave background in the presence of shot noise through cross-correlations. *Physical Review D*, *102*(2), 023002. doi: 10.1103/PhysRevD.102.023002

Amaro-Seoane, P., Aoudia, S., Babak, S., Binétruy, P., Berti, E., Bohé, A., ... Ward, H. (2012, June). Low-frequency gravitational-wave science with eLISA/NGO. *Classical and Quantum Gravity*, *29*(12), 124016. doi: 10.1088/0264-9381/29/12/124016

Amendola, L., Appleby, S., Avgoustidis, A., Bacon, D., Baker, T., Baldi, M., ... Zlosnik, T. (2018, April). Cosmology and fundamental physics with the Euclid satellite. *Living Reviews in Relativity*, *21*(1), 2. doi: 10.1007/s41114-017-0010-3

Amendola, L., Könnig, F., Martinelli, M., Pettorino, V., & Zumalacarregui, M. (2015, May). Surfing gravitational waves: can bigravity survive growing tensor modes? *Journal of Cosmology and Astroparticle Physics*, *2015*(5), 052. doi: 10.1088/1475-7516/2015/05/052

Amendola, L., Sawicki, I., Kunz, M., & Saltas, I. D. (2018, August). Direct de-

tection of gravitational waves can measure the time variation of the Planck mass. *Journal of Cosmology and Astroparticle Physics*, *2018*(8), 030. doi: 10.1088/1475-7516/2018/08/030

Anaconda software distribution. (2020). Anaconda Inc. Retrieved from `https://docs.anaconda.com/`

Ando, S., Benoit-Lévy, A., & Komatsu, E. (2014, July). Mapping dark matter in the gamma-ray sky with galaxy catalogs. *Physical Review D*, *90*(2), 023514. doi: 10.1103/PhysRevD.90.023514

Angulo, R. E., Zennaro, M., Contreras, S., Aricò, G., Pellejero-Ibañez, M., & Stücker, J. (2021, November). The BACCO simulation project: exploiting the full power of large-scale structure for cosmology. *Monthly Notices of the RAS*, *507*(4), 5869-5881. doi: 10.1093/mnras/stab2018

Appleby, S., Gong, J.-O., Hazra, D. K., Shafieloo, A., & Sypsas, S. (2016, September). Direct search for features in the primordial bispectrum. *Physics Letters B*, *760*, 297-301. doi: 10.1016/j.physletb.2016.07.004

Arai, S., & Nishizawa, A. (2018, May). Generalized framework for testing gravity with gravitational-wave propagation. II. Constraints on Horndeski theory. *Physical Review D*, *97*(10), 104038. doi: 10.1103/PhysRevD.97.104038

Aresté Saló, L., Benisty, D., Guendelman, E. I., & Haro, J. d. (2021a, June). α -attractors in quintessential inflation motivated by supergravity. *Physical Review D*, *103*(12), 123535. doi: 10.1103/PhysRevD.103.123535

Aresté Saló, L., Benisty, D., Guendelman, E. I., & Haro, J. d. (2021b, July). Quintessential inflation and cosmological seesaw mechanism: reheating and observational constraints. *Journal of Cosmology and Astroparticle Physics*, *2021*(7), 007. doi: 10.1088/1475-7516/2021/07/007

Aricò, G., Angulo, R. E., Contreras, S., Ondaro-Mallea, L., Pellejero-Ibañez, M., & Zennaro, M. (2021, September). The BACCO simulation project: a baryonification emulator with neural networks. *Monthly Notices of the RAS*, *506*(3), 4070-4082. doi: 10.1093/mnras/stab1911

Aricò, G., Angulo, R. E., & Zennaro, M. (2021, April). Accelerating Large-Scale-Structure data analyses by emulating Boltzmann solvers and Lagrangian Perturbation Theory. *arXiv e-prints*, arXiv:2104.14568.

Arzoumanian, Z., Baker, P. T., Brazier, A., Burke-Spolaor, S., Chamberlin, S. J., Chatterjee, S., ... NANOGrav Collaboration (2018, May). The NANOGrav 11 Year Data Set: Pulsar-timing Constraints on the Stochastic Gravitational-wave Background. *The Astrophysical Journal*, *859*(1), 47. doi: 10.3847/1538-4357/aabd3b

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., ... Astropy Contributors (2018, September). The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *Astronomical Journal*, *156*(3), 123. doi: 10.3847/1538-3881/aabc4f

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., ... Streicher, O. (2013, October). Astropy: A community Python package for astronomy. *Astronomy and Astrophysics*, *558*, A33. doi: 10.1051/0004-6361/201322068

Audren, B., Lesgourgues, J., Benabed, K., & Prunet, S. (2013, February). Conservative constraints on early cosmology with MONTE PYTHON. *Journal of Cosmology and Astroparticle Physics*, *2013*(2), 001. doi: 10.1088/1475-7516/2013/02/001

Audren, B., Lesgourgues, J., Bird, S., Haehnelt, M. G., & Viel, M. (2013, January). Neutrino masses and cosmological parameters from a Euclid-like survey: Markov Chain Monte Carlo forecasts including theoretical errors. *Journal of Cosmology and Astroparticle Physics*, *2013*(1), 026. doi: 10.1088/1475-7516/2013/01/026

Baker, J., Baker, T., Carbone, C., Congedo, G., Contaldi, C., Dvorkin, I., ... Zumalacárregui, M. (2019, August). High angular resolution gravitational wave astronomy. *arXiv e-prints*, arXiv:1908.11410.

Baker, T., Bellini, E., Ferreira, P. G., Lagos, M., Noller, J., & Sawicki, I. (2017, December). Strong Constraints on Cosmological Gravity from GW170817 and GRB 170817A. *Physical Review Letters*, *119*(25), 251301. doi: 10.1103/PhysRevLett.119.251301

Baker, T., & Harrison, I. (2021, January). Constraining scalar-tensor modified gravity with gravitational waves and large scale structure surveys. *Journal of Cosmology and Astroparticle Physics*, *2021*(1), 068. doi: 10.1088/1475-7516/2021/01/068

Ballardini, M. (2019, January). Probing primordial features with the primary CMB. *Physics of the Dark Universe*, *23*, 100245. doi: 10.1016/j.dark.2018.11.006

Ballinger, W. E., Peacock, J. A., & Heavens, A. F. (1996, October). Measuring the cosmological constant with redshift surveys. *Monthly Notices of the RAS*, *282*, 877. doi: 10.1093/mnras/282.3.877

Baumann, D. (2009, July). TASI Lectures on Inflation. *arXiv e-prints*, arXiv:0907.5424.

Belgacem, E., Calcagni, G., Crisostomi, M., Dalang, C., Dirian, Y., Ezquiaga, J. M., ... Sakellariadou, M. (2019, July). Testing modified gravity at cosmological distances with LISA standard sirens. *Journal of Cosmology and Astroparticle Physics*, *2019*(7), 024. doi: 10.1088/1475-7516/2019/07/024

Belgacem, E., Dirian, Y., Foffa, S., & Maggiore, M. (2018, July). Modified gravitational-wave propagation and standard sirens. *Physical Review D*, *98*(2), 023510. doi: 10.1103/PhysRevD.98.023510

Belgacem, E., Foffa, S., Maggiore, M., & Yang, T. (2020, March). Gaussian processes reconstruction of modified gravitational wave propagation. *Physical Review D*, *101*(6), 063505. doi: 10.1103/PhysRevD.101.063505

Bellini, E., & Sawicki, I. (2014, July). Maximal freedom at minimum cost: linear large-scale structure in general modifications of gravity. *Journal of Cosmology and Astroparticle Physics*, *07*, 050. doi: 10.1088/1475-7516/2014/07/050

Bennett, C. L., Larson, D., Weiland, J. L., Jarosik, N., Hinshaw, G., Odegard, N., ... Wright, E. L. (2013, October). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *The Astrophysical Journal, Supplement*, *208*, 20. doi: 10.1088/0067-0049/208/2/20

Benson, B. A., Ade, P. A. R., Ahmed, Z., Allen, S. W., Arnold, K., Austermann, J. E., ... Yoon, K. W. (2014, July). SPT-3G: a next-generation cosmic microwave

background polarization experiment on the South Pole telescope. *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*, *9153*, 91531P. doi: 10.1117/12.2057305

Bera, S., Rana, D., More, S., & Bose, S. (2020, October). Incompleteness Matters Not: Inference of H_0 from Binary Black Hole-Galaxy Cross-correlations. *The Astrophysical Journal*, *902*(1), 79. doi: 10.3847/1538-4357/abb4e0

Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. (2002, September). Large-scale structure of the Universe and cosmological perturbation theory. *Physics Reports*, *367*(1-3), 1-248. doi: 10.1016/S0370-1573(02)00135-7

Bernardeau, F., Nishimichi, T., & Taruya, A. (2014, December). Cosmic shear full nulling: sorting out dynamics, geometry and systematics. *Monthly Notices of the RAS*, *445*(2), 1526-1537. doi: 10.1093/mnras/stu1861

Bertacca, D., Ricciardone, A., Bellomo, N., Jenkins, A. C., Matarrese, S., Raccanelli, A., ... Sakellariadou, M. (2019, September). Projection effects on the observed angular spectrum of the astrophysical stochastic gravitational wave background. *arXiv e-prints*, arXiv:1909.11627.

Bird, S., Cholis, I., Muñoz, J. B., Ali-Haïmoud, Y., Kamionkowski, M., Kovetz, E. D., ... Riess, A. G. (2016, May). Did LIGO Detect Dark Matter? *Physical Review Letters*, *116*(20), 201301. doi: 10.1103/PhysRevLett.116.201301

Bird, S., Viel, M., & Haehnelt, M. G. (2012, March). Massive neutrinos and the non-linear matter power spectrum. *Monthly Notices of the RAS*, *420*(3), 2551-2561. doi: 10.1111/j.1365-2966.2011.20222.x

Blaizot, J., Wadadekar, Y., Guiderdoni, B., Colombi, S. T., Bertin, E., Bouchet, F. R., ... Hatton, S. (2005, June). MoMaF: the Mock Map Facility. *Monthly Notices of the RAS*, *360*(1), 159-175. doi: 10.1111/j.1365-2966.2005.09019.x

Blanchet, L., Damour, T., Iyer, B. R., Will, C. M., & Wiseman, A. G. (1995, May). Gravitational-Radiation Damping of Compact Binary Systems to Second Post-Newtonian Order. *Physical Review Letters*, *74*(18), 3515-3518. doi: 10.1103/PhysRevLett.74.3515

Blanco-Pillado, J. J., Olum, K. D., & Siemens, X. (2018, March). New limits on cosmic strings from gravitational wave observation. *Physics Letters B*, *778*, 392-396. doi: 10.1016/j.physletb.2018.01.050

Bustillo, J. C., Sanchis-Gual, N., Torres-Forné, A., Font, J. A., Vajpeyi, A., Smith, R., ... Leong, S. H. W. (2021, February). GW190521 as a Merger of Proca Stars: A Potential New Vector Boson of 8.7×10^{-13} eV. *Physical Review Letters*, *126*(8), 081101. doi: 10.1103/PhysRevLett.126.081101

Cañas-Herrera, G., Contigiani, O., & Vardanyan, V. (2020, August). Cross-correlation of the astrophysical gravitational-wave background with galaxy clustering. *Physical Review D*, *102*(4), 043513. doi: 10.1103/PhysRevD.102.043513

Cañas-Herrera, G., Contigiani, O., & Vardanyan, V. (2021, September). Learning How to Surf: Reconstructing the Propagation and Origin of Gravitational Waves with Gaussian Processes. *The Astrophysical Journal*, *918*(1), 20. doi: 10.3847/1538-4357/ac09e3

Cañas-Herrera, G., & Renzi, F. (2021, November). Current and future constraints on single-field α-attractor models. *Physical Review D*, *104*(10), 103512. doi:

10.1103/PhysRevD.104.103512

Cañas-Herrera, G., Torrado, J., & Achúcarro, A. (2021, June). Bayesian reconstruction of the inflaton's speed of sound using CMB data. *Physical Review D*, *103*(12), 123531. doi: 10.1103/PhysRevD.103.123531

Cabass, G., Pagano, L., Salvati, L., Gerbino, M., Giusarma, E., & Melchiorri, A. (2016, March). Updated Constraints and Forecasts on Primordial Tensor Modes. *Physical Review D*, *93*(6), 063508. doi: 10.1103/PhysRevD.93.063508

Cahillane, C., Betzwieser, J., Brown, D. A., Goetz, E., Hall, E. D., Izumi, K., ... Weinstein, A. J. (2017, November). Calibration uncertainty for Advanced LIGO's first and second observing runs. *Physical Review D*, *96*(10), 102001. doi: 10.1103/PhysRevD.96.102001

Calcagni, G., Kuroyanagi, S., Marsat, S., Sakellariadou, M., Tamanini, N., & Tasinato, G. (2019, December). Quantum gravity and gravitational-wave astronomy. *Journal of Cosmology and Astroparticle Physics*, *10*, 012. doi: 10.1088/1475-7516/2019/10/012

Calculadora Cosmologica: code that solves Friedman Equations numerically given values for components of universe. (2015, December). `https://github.com/gcanasherrera/CalculadoraCosmologica`. (Accessed: 27-06-2018)

Calore, F., Cuoco, A., Regimbau, T., Sachdev, S., & Serpico, P. D. (2020, February). Cross-correlating galaxy catalogs and gravitational waves: a tomographic approach. *arXiv e-prints*, arXiv:2002.02466.

Cannone, D., Bartolo, N., & Matarrese, S. (2014, June). Perturbative unitarity of inflationary models with features. *Physical Review D*, *89*(12), 127301. doi: 10.1103/PhysRevD.89.127301

Carrasco, J. J. M., Kallosh, R., Linde, A., & Roest, D. (2015, August). Hyperbolic geometry of cosmological attractors. *Physical Review D*, *92*(4), 041301. doi: 10.1103/PhysRevD.92.041301

Carroll, B. W., & Ostlie, D. A. (2014). *An introduction to modern astrophysics* (second ed.). Pearson.

Cartis, C., Fiala, J., Marteau, B., & Roberts, L. (2018, March). Improving the Flexibility and Robustness of Model-Based Derivative-Free Optimization Solvers. *arXiv e-prints*, arXiv:1804.00154.

Cartis, C., Roberts, L., & Sheridan-Methven, O. (2018, December). Escaping local minima with derivative-free methods: a numerical investigation. *arXiv e-prints*, arXiv:1812.11343.

Céspedes, S., Atal, V., & Palma, G. A. (2012, May). On the importance of heavy fields during inflation. *Journal of Cosmology and Astroparticle Physics*, *5*, 008. doi: 10.1088/1475-7516/2012/05/008

Chabanier, S., Millea, M., & Palanque-Delabrouille, N. (2019, October). Matter power spectrum: from Ly α forest to CMB scales. *Monthly Notices of the RAS*, *489*(2), 2247-2253. doi: 10.1093/mnras/stz2310

Chen, H.-Y., Fishbach, M., & Holz, D. E. (2018, October). A two per cent Hubble constant measurement from standard sirens within five years. *Nature*, *562*(7728), 545-547. doi: 10.1038/s41586-018-0606-0

Chen, X. (2010, January). Primordial Non-Gaussianities from Inflation Models. *Advances in Astronomy*, *2010*, 638979. doi: 10.1155/2010/638979

Chen, Z.-C., Huang, F., & Huang, Q.-G. (2019, January). Stochastic Gravitational-wave Background from Binary Black Holes and Binary Neutron Stars and Implications for LISA. *The Astrophysical Journal*, *871*(1), 97. doi: 10.3847/1538-4357/aaf581

Cheung, C., Fitzpatrick, A. L., Kaplan, J., Senatore, L., & Creminelli, P. (2008, March). The effective field theory of inflation. *Journal of High Energy Physics*, *3*, 014-014. doi: 10.1088/1126-6708/2008/03/014

Chevallier, M., & Polarski, D. (2001, January). Accelerating Universes with Scaling Dark Matter. *International Journal of Modern Physics D*, *10*(2), 213-223. doi: 10.1142/S0218271801000822

Chiang, H. C., et al. (2010, March). Measurement of CMB Polarization Power Spectra from Two Years of BICEP Data. *The Astrophysical Journal*, *711*, 1123-1140. doi: 10.1088/0004-637X/711/2/1123

Chluba, J., Hamann, J., & Patil, S. P. (2015, June). Features and new physical scales in primordial observables: Theory and observation. *International Journal of Modern Physics D*, *24*, 1530023. doi: 10.1142/S0218271815300232

Clarke, T. J., Copeland, E. J., & Moss, A. (2020, October). Constraints on primordial gravitational waves from the Cosmic Microwave Background. *Journal of Cosmology and Astroparticle Physics*, *10*, 002. doi: 10.1088/1475-7516/2020/10/002

Clesse, S., & García-Bellido, J. (2017, March). The clustering of massive Primordial Black Holes as Dark Matter: measuring their mass distribution with Advanced LIGO. *Physics of the Dark Universe*, *15*, 142–147. doi: 10.1016/j.dark.2016.10.002

Colgáin, E. Ó., & Sheikh-Jabbari, M. M. (2021, January). Elucidating cosmological model dependence with H_0. *arXiv e-prints*, arXiv:2101.08565.

Contigiani, O. (2020, March). Lensing efficiency for gravitational wave mergers. *Monthly Notices of the RAS*, *492*(3), 3359-3363. doi: 10.1093/mnras/staa026

Courbin, F., Saha, P., & Schechter, P. L. (2002). Quasar Lensing. *Gravitational Lensing: An Astrophysical Tool*, *608*, 1.

Creating your own cosmological likelihood class. (2022, January). Retrieved from `https://cobaya.readthedocs.io/en/latest/cosmo_external_likelihood_class.html` (Accessed: 04-04-2022)

Creminelli, P., & Vernizzi, F. (2017, December). Dark Energy after GW170817 and GRB170817A. *Physical Review Letters*, *119*(25), 251302. doi: 10.1103/PhysRevLett.119.251302

Crittenden, R. G., Pogosian, L., & Zhao, G.-B. (2009, December). Investigating dark energy experiments with principal components. *Journal of Cosmology and Astroparticle Physics*, *2009*(12), 025–025. Retrieved from `http://dx.doi.org/10.1088/1475-7516/2009/12/025` doi: 10.1088/1475-7516/2009/12/025

Crittenden, R. G., Zhao, G.-B., Pogosian, L., Samushia, L., & Zhang, X. (2012, February). Fables of reconstruction: controlling bias in the dark energy equation of state. *Journal of Cosmology and Astroparticle Physics*, *2012*(02), 048–048.

Retrieved from http://dx.doi.org/10.1088/1475-7516/2012/02/048 doi: 10.1088/1475-7516/2012/02/048

Cusin, G., Durrer, R., & Ferreira, P. G. (2019, January). Polarization of a stochastic gravitational wave background through diffusion by massive structures. *Physical Review D*, *99*(2), 023534. doi: 10.1103/PhysRevD.99.023534

Cusin, G., Dvorkin, I., Pitrou, C., & Uzan, J.-P. (2018a, November). Comment on the article "Anisotropies in the astrophysical gravitational-wave background: The impact of black hole distributions" by A.C. Jenkins et al. [arXiv:1810.13435]. *arXiv e-prints*, arXiv:1811.03582.

Cusin, G., Dvorkin, I., Pitrou, C., & Uzan, J.-P. (2018b, June). First Predictions of the Angular Power Spectrum of the Astrophysical Gravitational Wave Background. *Physical Review Letters*, *120*(23), 231101. doi: 10.1103/PhysRevLett.120.231101

Cusin, G., Dvorkin, I., Pitrou, C., & Uzan, J.-P. (2019, September). Properties of the stochastic astrophysical gravitational wave background: Astrophysical sources dependencies. *Physical Review D*, *100*(6), 063004. doi: 10.1103/PhysRevD.100.063004

Cusin, G., Dvorkin, I., Pitrou, C., & Uzan, J.-P. (2020, March). Stochastic gravitational wave background anisotropies in the mHz band: astrophysical dependencies. *Monthly Notices of the RAS*, *493*(1), L1-L5. doi: 10.1093/mnrasl/slz182

Cusin, G., Pitrou, C., & Uzan, J.-P. (2017, November). Anisotropy of the astrophysical gravitational wave background: Analytic expression of the angular power spectrum and correlation with cosmological observations. *Physical Review D*, *96*(10), 103019. doi: 10.1103/PhysRevD.96.103019

Cusin, G., Pitrou, C., & Uzan, J.-P. (2018, June). The signal of the gravitational wave background and the angular correlation of its energy density. *Physical Review D*, *97*(12), 123527. doi: 10.1103/PhysRevD.97.123527

Cyburt, R. H., Fields, B. D., & Olive, K. A. (2003, August). Primordial nucleosynthesis in light of WMAP. *Physics Letters B*, *567*, 227-234. doi: 10.1016/j.physletb.2003.06.026

Debackere, S. N. B., Schaye, J., & Hoekstra, H. (2020, February). The impact of the observed baryon distribution in haloes on the total matter power spectrum. *Monthly Notices of the RAS*, *492*(2), 2285-2307. doi: 10.1093/mnras/stz3446

Deffayet, C., & Menou, K. (2007, October). Probing Gravity with Spacetime Sirens. *Astrophysical Journal, Letters*, *668*(2), L143-L146. doi: 10.1086/522931

Diemer, B. (2018). COLOSSUS: A python toolkit for cosmology, large-scale structure, and dark matter halos. *Astrophys. J. Suppl.*, *239*(2), 35. doi: 10.3847/1538-4365/aaee8c

Dirian, Y., Foffa, S., Kunz, M., Maggiore, M., & Pettorino, V. (2016, May). Non-local gravity and comparison with observational datasets. ii. updated results and bayesian model comparison with cdm. *Journal of Cosmology and Astroparticle Physics*, *2016*(05), 068–068. Retrieved from http://dx.doi.org/10.1088/1475-7516/2016/05/068 doi: 10.1088/1475-7516/2016/05/068

Di Valentino, E., Holz, D. E., Melchiorri, A., & Renzi, F. (2018, October). The cosmological impact of future constraints on H_0 from gravitational-wave standard

sirens. *Physical Review D*, *98*(8), 083523. doi: 10.1103/PhysRevD.98.083523

Dodelson, S. (2003). *Modern cosmology*. San Diego, CA: Academic Press. Retrieved from `https://cds.cern.ch/record/1282338`

Durakovic, A., Hunt, P., Mukherjee, S., Sarkar, S., & Souradeep, T. (2018, February). Reconstruction of a direction-dependent primordial power spectrum from Planck CMB data. *Journal of Cosmology and Astroparticle Physics*, *2018*(2), 012. doi: 10.1088/1475-7516/2018/02/012

Durakovic, A., Hunt, P., Patil, S., & Sarkar, S. (2019, October). Reconstructing the EFT of inflation from cosmological data. *SciPost Physics*, *7*(4), 049. doi: 10.21468/SciPostPhys.7.4.049

Dyson, F. W., Eddington, A. S., & Davidson, C. (1920, January). A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London Series A*, *220*, 291-333. doi: 10.1098/rsta.1920.0009

Eckhardt, R. (1987). Stan Ulam, John von Neumann, and the Monte Carlo Method. *Los Alamos Science Special Issue*. Retrieved from `http://www-star.st-and.ac.uk/~kw25/teaching/mcrt/MC_history_3.pdf`

Einstein, A. (1916, January). Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin*, 688-696.

Einstein, A. (1918). Über Gravitationswellen. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 154-167.*.

Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., ... York, D. G. (2005, November). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *The Astrophysical Journal*, *633*(2), 560-574. doi: 10.1086/466512

Essinger-Hileman, T., Ali, A., Amiri, M., Appel, J. W., Araujo, D., Bennett, C. L., ... Zeng, L. (2014, July). CLASS: the cosmology large angular scale surveyor. *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII*, *9153*, 91531I. doi: 10.1117/12.2056701

Euclid Collaboration, Blanchard, A., Camera, S., Carbone, C., Cardone, V. F., Casas, S., ... Zucca, E. (2020, October). Euclid preparation. VII. Forecast validation for Euclid cosmological probes. *Astronomy and Astrophysics*, *642*, A191. doi: 10.1051/0004-6361/202038071

Euclid Collaboration, Knabenhans, M., Stadel, J., Potter, D., Dakin, J., Hannestad, S., ... Zucca, E. (2021, August). Euclid preparation: IX. EuclidEmulator2 - power spectrum emulation with massive neutrinos and self-consistent dark energy perturbations. *Monthly Notices of the RAS*, *505*(2), 2840-2869. doi: 10.1093/mnras/stab1366

Ezquiaga, J. M., & Zumalacárregui, M. (2017, December). Dark Energy After GW170817: Dead Ends and the Road Ahead. *Physical Review Letters*, *119*(25), 251304. doi: 10.1103/PhysRevLett.119.251304

Ezquiaga, J. M., & Zumalacárregui, M. (2018, December). Dark Energy in light of Multi-Messenger Gravitational-Wave astronomy. *Frontiers in Astronomy and Space Sciences*, *5*, 44. doi: 10.3389/fspas.2018.00044

Ezquiaga, J. M., & Zumalacárregui, M. (2020, December). Gravitational wave lensing beyond general relativity: birefringence, echoes and shadows. *Physical Review D*, *102*(12), 124048. doi: 10.1103/PhysRevD.102.124048

Fabian, A. C., & Barcons, X. (1992). The origin of the x-ray background. *Annual Review of Astronomy and Astrophysics*, *30*(1), 429–456. doi: 10.1146/annurev.aa.30.090192.002241

Fang, X., Blazek, J. A., McEwen, J. E., & Hirata, C. M. (2017, February). FAST-PT II: an algorithm to calculate convolution integrals of general tensor quantities in cosmological perturbation theory. *Journal of Cosmology and Astroparticle Physics*, *2017*(02), 030–030. doi: 10.1088/1475-7516/2017/02/030

Farr, W. M., Fishbach, M., Ye, J., & Holz, D. E. (2019, October). A Future Percent-level Measurement of the Hubble Expansion at Redshift 0.8 with Advanced LIGO. *The Astrophysical Journal*, *883*(2), L42. doi: 10.3847/2041-8213/ab4284

Feroz, F., & Hobson, M. P. (2008, February). Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *Monthly Notices of the RAS*, *384*(2), 449-463. doi: 10.1111/j.1365-2966.2007.12353.x

Feroz, F., Hobson, M. P., & Bridges, M. (2009, October). MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the RAS*, *398*(4), 1601-1614. doi: 10.1111/j.1365-2966.2009.14548.x

Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. (2019, November). Importance Nested Sampling and the MultiNest Algorithm. *The Open Journal of Astrophysics*, *2*(1), 10. doi: 10.21105/astro.1306.2144

fftlog - python wrapper FFTLog. (2016). `https://github.com/prisae/fftlog`. (Accessed: 01-09-2019)

Finke, A., Foffa, S., Iacovelli, F., Maggiore, M., & Mancarella, M. (2021, August). Cosmology with LIGO/Virgo dark sirens: Hubble parameter and modified gravitational wave propagation. *Journal of Cosmology and Astroparticle Physics*, *2021*(8), 026. doi: 10.1088/1475-7516/2021/08/026

Font-Ribera, A., McDonald, P., Mostek, N., Reid, B. A., Seo, H.-J., & Slosar, A. (2014, May). DESI and other dark energy experiments in the era of neutrino mass measurements. *Journal of Cosmology and Astroparticle Physics*, *05*, 023. doi: 10.1088/1475-7516/2014/05/023

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013, March). emcee: The MCMC Hammer. *Publications of the ASP*, *125*(925), 306. doi: 10.1086/670067

Friedmann, A. (1924, December). Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift fur Physik*, *21*(1), 326-332. doi: 10.1007/BF01328280

Frusciante, N., & Perenon, L. (2020, May). Effective field theory of dark energy: A review. *Physics Reports*, *857*, 1–63. doi: 10.1016/j.physrep.2020.02.004

Gariazzo, S., & Mena, O. (2019, January). Cosmology-marginalized approaches in Bayesian model comparison: The neutrino mass as a case study. *Physical Review D*, *99*(2), 021301. doi: 10.1103/PhysRevD.99.021301

Garoffolo, A., Tasinato, G., Carbone, C., Bertacca, D., & Matarrese, S. (2020, November). Gravitational waves and geometrical optics in scalar-tensor theories. *Journal of Cosmology and Astroparticle Physics*, *11*, 040. doi: 10.1088/1475-7516/2020/11/040

Gelman, A., Gilks, W. R., & Roberts, G. O. (1997, February). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, *7*(1), 110 – 120. doi: 10.1214/aoap/1034625254

Gelman, A., & Rubin, D. B. (1992, November). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.

Gerardi, F., Martinelli, M., & Silvestri, A. (2019, July). Reconstruction of the Dark Energy equation of state from latest data: the impact of theoretical priors. *Journal of Cosmology and Astroparticle Physics*, *07*, 042. doi: 10.1088/1475-7516/2019/07/042

Ghosh, A., Del Pozzo, W., & Ajith, P. (2016, November). Estimating parameters of binary black holes from gravitational-wave observations of their inspiral, merger and ringdown. *Physical Review D*, *94*(10), 104070. doi: 10.1103/PhysRevD.94.104070

Giarè, W., Di Valentino, E., & Melchiorri, A. (2019, June). Testing the inflationary slow-roll condition with tensor modes. *Physical Review D*, *99*, 123522. doi: 10.1103/PhysRevD.99.123522

Giarè, W., & Melchiorri, A. (2021, April). Probing the inflationary background of gravitational waves from large to small scales. *Physics Letters B*, *815*, 136137. doi: 10.1016/j.physletb.2021.136137

Giarè, W., & Renzi, F. (2020, October). Propagating speed of primordial gravitational waves. *Physical Review D*, *102*(8), 083530. doi: 10.1103/PhysRevD.102.083530

Grayson, J. A., et al. (2016, July). BICEP3 performance overview and planned Keck Array upgrade. *Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VIII*, *9914*, 99140S. doi: 10.1117/12.2233894

Grieb, J. N., Sánchez, A. G., Salazar-Albornoz, S., & Vecchia, C. D. (2016, February). Gaussian covariance matrices for anisotropic galaxy clustering measurements. *Monthly Notices of the RAS*, *457*(2), 1577–1592. doi: 10.1093/mnras/stw065

Hall, E. D., & Evans, M. (2019, November). Metrics for next-generation gravitational-wave detectors. *Classical and Quantum Gravity*, *36*(22), 225002. doi: 10.1088/1361-6382/ab41d6

Hamilton, A. J. S. (1998, January). Linear Redshift Distortions: a Review. In D. Hamilton (Ed.), *The evolving universe* (Vol. 231, p. 185). doi: 10.1007/978-94-011-4960-0_17

Hamilton, A. J. S. (2000, February). Uncorrelated modes of the non-linear power spectrum. *Monthly Notices of the RAS*, *312*(2), 257-284. doi: 10.1046/j.1365-8711.2000.03071.x

Handley, W., & Millea, M. (2019, March). Maximum-Entropy Priors with Derived Parameters in a Specified Distribution. *Entropy*, *21*(3), 272. doi: 10.3390/e21030272

Handley, W. J., Hobson, M. P., & Lasenby, A. N. (2015a, June). POLYCHORD: nested sampling for cosmology. *Monthly Notices of the RAS*, *450*, L61-L65.

doi: 10.1093/mnrasl/slv047

Handley, W. J., Hobson, M. P., & Lasenby, A. N. (2015b, November). POLYCHORD: next-generation nested sampling. *Monthly Notices of the RAS*, *453*(4), 4384-4398. doi: 10.1093/mnras/stv1911

Handley, W. J., Lasenby, A. N., Peiris, H. V., & Hobson, M. P. (2019, November). Bayesian inflationary reconstructions from Planck 2018 data. *Physical Review D*, *100*(10), 103511. doi: 10.1103/PhysRevD.100.103511

Hastings, W. K. (1970, April). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97-109. doi: 10.1093/biomet/57.1.97

Hazra, D. K., Shafieloo, A., & Souradeep, T. (2013, July). Primordial power spectrum: a complete analysis with the WMAP nine-year data. *Journal of Cosmology and Astroparticle Physics*, *2013*(7), 031. doi: 10.1088/1475-7516/2013/07/031

Hazumi, M., Ade, P. A. R., Adler, A., Allys, E., Arnold, K., Auguste, D., ... Zonca, A. (2020, December). LiteBIRD satellite: JAXA's new strategic L-class mission for all-sky surveys of cosmic microwave background polarization. *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, *11443*, 114432F. doi: 10.1117/12.2563050

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015, March). The extent and consequences of p-hacking in science. *PLOS Biology*, *13*(3), 1-15. doi: 10.1371/journal.pbio.1002106

Hebecker, A., Jaeckel, J., Rompineve, F., & Witkowski, L. T. (2016, November). Gravitational waves from axion monodromy. *Journal of Cosmology and Astroparticle Physics*, *2016*(11), 003. doi: 10.1088/1475-7516/2016/11/003

Heitmann, K., Lawrence, E., Kwan, J., Habib, S., & Higdon, D. (2014, January). The Coyote Universe Extended: Precision Emulation of the Matter Power Spectrum. *The Astrophysical Journal*, *780*(1), 111. doi: 10.1088/0004-637X/780/1/111

Henderson, S. W., Allison, R., Austermann, J., Baildon, T., Battaglia, N., Beall, J. A., ... Wollack, E. J. (2016, August). Advanced ACTPol Cryogenic Detector Arrays and Readout. *Journal of Low Temperature Physics*, *184*(3-4), 772-779. doi: 10.1007/s10909-016-1575-z

Heymans, C., Tröster, T., Asgari, M., Blake, C., Hildebrandt, H., Joachimi, B., ... Wolf, C. (2020, July). KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints. *arXiv e-prints*, arXiv:2007.15632.

Holz, D. E., & Hughes, S. A. (2005, August). Using Gravitational-Wave Standard Sirens. *The Astrophysical Journal*, *629*(1), 15-22. doi: 10.1086/431341

Horndeski, G. W. (1974, September). Second-Order Scalar-Tensor Field Equations in a Four-Dimensional Space. *International Journal of Theoretical Physics*, *10*(6), 363-384. doi: 10.1007/BF01807638

Howlett, C., Lewis, A., Hall, A., & Challinor, A. (2012, April). CMB power spectrum parameter degeneracies in the era of precision cosmology. *Journal of Cosmology and Astroparticle Physics*, *2012*(4), 027. doi: 10.1088/1475-7516/2012/04/027

Hu, B., Raveri, M., Frusciante, N., & Silvestri, A. (2014, May). Effective field theory of cosmic acceleration: An implementation in CAMB. *Physical Review*

D, *89*(10), 103530. doi: 10.1103/PhysRevD.89.103530

Hu, B., & Torrado, J. (2015, March). Searching for primordial localized features with CMB and LSS spectra. *Physical Review D*, *91*(6), 064039. doi: 10.1103/PhysRevD.91.064039

Hu, W. (1999, September). Power spectrum tomography with weak lensing. *The Astrophysical Journal*, *522*(1), L21–L24. doi: 10.1086/312210

Hubble, E. (1929, March). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, *15*(3), 168-173. doi: 10.1073/pnas.15.3.168

Hulse, R. A., & Taylor, J. H. (1975, January). Discovery of a pulsar in a binary system. *Astrophysical Journal, Letters*, *195*, L51-L53. doi: 10.1086/181708

Hunt, P., & Sarkar, S. (2014, January). Reconstruction of the primordial power spectrum of curvature perturbations using multiple data sets. *Journal of Cosmology and Astroparticle Physics*, *2014*(1), 025. doi: 10.1088/1475-7516/2014/01/025

Iarygina, O., Sfakianakis, E. I., Wang, D.-G., & Achucarro, A. (2019, June). Universality and scaling in multi-field α-attractor preheating. *Journal of Cosmology and Astroparticle Physics*, *06*, 027. doi: 10.1088/1475-7516/2019/06/027

Iarygina, O., Sfakianakis, E. I., Wang, D.-G., & Achúcarro, A. (2020, May). Multi-field inflation and preheating in asymmetric α-attractors. *arXiv e-prints*, arXiv:2005.00528.

Ivezic, Z., Connolly, A. J., VanderPlas, J. T., & Gray, A. (2014). *Statistics, data mining, and machine learning in astronomy: A practical python guide for the analysis of survey data*. Princeton, NJ, USA: Princeton University Press.

Jenkins, A. C., Romano, J. D., & Sakellariadou, M. (2019, October). Estimating the angular power spectrum of the gravitational-wave background in the presence of shot noise. *Physical Review D*, *100*(8), 083501. doi: 10.1103/PhysRevD.100.083501

Jenkins, A. C., & Sakellariadou, M. (2018, September). Anisotropies in the stochastic gravitational-wave background: Formalism and the cosmic string case. *Physical Review D*, *98*(6), 063509. doi: 10.1103/PhysRevD.98.063509

Jenkins, A. C., & Sakellariadou, M. (2019, September). Shot noise in the astrophysical gravitational-wave background. *Physical Review D*, *100*(6), 063508. doi: 10.1103/PhysRevD.100.063508

Jenkins, A. C., Sakellariadou, M., Regimbau, T., & Slezak, E. (2018, September). Anisotropies in the astrophysical gravitational-wave background: Predictions for the detection of compact binaries by LIGO and Virgo. *Physical Review D*, *98*(6), 063501. doi: 10.1103/PhysRevD.98.063501

Jenkins, A. C., Sakellariadou, M., Regimbau, T., Slezak, E., O'Shaughnessy, R., & Wysocki, D. (2019, January). Response to Cusin et al's comment on arXiv:1810.13435. *arXiv e-prints*, arXiv:1901.01078.

Kaiser, N. (1987, July). Clustering in real space and in redshift space. *Monthly Notices of the RAS*, *227*, 1-21. doi: 10.1093/mnras/227.1.1

Kaiser, N. (1998, May). Weak Lensing and Cosmology. *The Astrophysical Journal*, *498*(1), 26-42. doi: 10.1086/305515

Kallosh, R., & Linde, A. (2019, December). CMB targets after the latest Planck data release. *Physical Review D*, *100*(12), 123523. doi: 10.1103/PhysRevD.100.123523

Kallosh, R., Linde, A., & Roest, D. (2013, November). Superconformal Inflationary α-Attractors. *Journal Of High Energy Physics*, *11*, 198. doi: 10.1007/JHEP11(2013)198

Kallosh, R., Linde, A., Roest, D., & Yamada, Y. (2017, July). $\overline{D3}$ induced geometric inflation. *Journal Of High Energy Physics*, *07*, 057. doi: 10.1007/JHEP07(2017)057

Kojima, K., Kajino, T., & Mathews, G. J. (2010). Generation of Curvature Perturbations with Extra Anisotropic Stress. *Journal of Cosmology and Astroparticle Physics*, *02*, 018. doi: 10.1088/1475-7516/2010/02/018

Lagos, M., Fishbach, M., Landry, P., & Holz, D. E. (2019, April). Standard sirens with a running Planck mass. *Physical Review D*, *99*(8), 083504. doi: 10.1103/PhysRevD.99.083504

Laureijs, R., Amiaux, J., Arduini, S., Auguères, J. L., Brinchmann, J., Cole, R., ... Zucca, E. (2011, October). Euclid Definition Study Report. *arXiv e-prints*, arXiv:1110.3193.

Leclercq, F., Pisani, A., & Wandelt, B. D. (2014, March). Cosmology: from theory to data, from data to theory. *arXiv e-prints*, arXiv:1403.1260.

Lee, A., Abitbol, M. H., Adachi, S., Ade, P., Aguirre, J., Ahmed, Z., ... Zonca, A. (2019, September). The Simons Observatory. , *51*, 147.

Lemaître, G. (1931, March). Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *Monthly Notices of the RAS*, *91*, 483-490. doi: 10.1093/mnras/91.5.483

Lesgourgues, J. (2018). *Cosmo Tools 2018: A school designed to help young researchers understand the black box of cosmology codes.* `https://indico.cern.ch/event/688110/timetable/?print=1&view=standard`. (Accessed: 2018-05-18)

Levi Said, J., Mifsud, J., Sultana, J., & Zarb Adami, K. (2021, June). Reconstructing teleparallel gravity with cosmic structure growth and expansion rate data. *Journal of Cosmology and Astroparticle Physics*, *2021*(6), 015. doi: 10.1088/1475-7516/2021/06/015

Lewis, A. (2013, May). Efficient sampling of fast and slow cosmological parameters. *Physical Review D*, *87*(10), 103529. doi: 10.1103/PhysRevD.87.103529

Lewis, A. (2019, October). GetDist: a Python package for analysing Monte Carlo samples. *arXiv e-prints*, arXiv:1910.13970.

Lewis, A., & Bridle, S. (2002, November). Cosmological parameters from CMB and other data: A Monte Carlo approach. *Physical Review D*, *66*(10), 103511. doi: 10.1103/PhysRevD.66.103511

Lewis, A., Challinor, A., & Lasenby, A. (2000, August). Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. *The Astrophysical Journal*, *538*(2), 473-476. doi: 10.1086/309179

Li, Y.-J., Wang, Y.-Z., Han, M.-Z., Tang, S.-P., Yuan, Q., Fan, Y.-Z., & Wei, D.-M. (2021, August). A Flexible Gaussian Process Reconstruction Method and the Mass Function of the Coalescing Binary Black Hole Systems. *The Astrophysical Journal*, *917*(1), 33. doi: 10.3847/1538-4357/ac0971

LIGO Scientific collaboration and VIRGO collaboration. (2016a, June). GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence. *Physical Review Letters*, *116*(24), 241103. doi: 10.1103/PhysRevLett.116.241103

LIGO Scientific collaboration and VIRGO collaboration. (2016b, February). Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, *116*(6), 061102. doi: 10.1103/PhysRevLett.116.061102

LIGO Scientific collaboration and VIRGO collaboration. (2017a, June). GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Physical Review Letters*, *118*(22), 221101. doi: 10.1103/PhysRevLett.118.221101

LIGO Scientific collaboration and VIRGO collaboration. (2017b, December). GW170608: Observation of a 19 Solar-mass Binary Black Hole Coalescence. *The Astrophysical Journal*, *851*(2), L35. doi: 10.3847/2041-8213/aa9f0c

LIGO Scientific collaboration and VIRGO collaboration. (2017c, October). Gw170814: A three-detector observation of gravitational waves from a binary black hole coalescence. *Physical Review Letters*, *119*, 141101. doi: 10.1103/PhysRevLett.119.141101

LIGO Scientific collaboration and VIRGO collaboration. (2017d, October). GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Physical Review Letters*, *119*(16), 161101. doi: 10.1103/PhysRevLett.119.161101

LIGO Scientific collaboration and VIRGO collaboration. (2019, September). Directional limits on persistent gravitational waves using data from Advanced LIGO's first two observing runs. *Physical Review D*, *100*(6), 062001. doi: 10.1103/PhysRevD.100.062001

Linder, E. V. (2003, March). Exploring the expansion history of the universe. *Physical Review Letters*, *90*(9). doi: 10.1103/physrevlett.90.091301

Lombriser, L., & Taylor, A. (2016, March). Breaking a dark degeneracy with gravitational waves. *Journal of Cosmology and Astroparticle Physics*, *2016*(03), 031–031. doi: 10.1088/1475-7516/2016/03/031

LSST Science Collaboration, Abell, P. A., et al. (2009, December). LSST Science Book, Version 2.0. *arXiv e-prints*, arXiv:0912.0201.

Lundgren, B., Kinemuchi, K., Zasowski, G., Lucatello, S., Diamond-Stanic, A. M., Tremonti, C. A., ... Gallagher, J. S. (2015, August). The SDSS-IV in 2014: A Demographic Snapshot. *Publications of the ASP*, *127*(954), 776. doi: 10.1086/682387

Maggiore, M., Broeck, C. V. D., Bartolo, N., Belgacem, E., Bertacca, D., Bizouard, M. A., ... Sakellariadou, M. (2020, March). Science case for the Einstein telescope. *Journal of Cosmology and Astroparticle Physics*, *2020*(03), 050–050. doi: 10.1088/1475-7516/2020/03/050

Mannerkoski, M., Johansson, P. H., Pihajoki, P., Rantala, A., & Naab, T. (2019, December). Gravitational Waves from the Inspiral of Supermassive Black Holes in Galactic-scale Simulations. *The Astrophysical Journal*, *887*(1), 35. doi: 10.3847/1538-4357/ab52f9

María Ezquiaga, J. (2021, April). Hearing gravity from the cosmos: GWTC-2 probes general relativity at cosmological scales. *arXiv e-prints*, arXiv:2104.05139.

Markov, A. A. (2006, January). An Example of Statistical Investigation of the Text Eugene Onegin Concerning the Connection of Samples in Chains. *Science in Context*, *19*(4), 591–600. doi: 10.1017/S0269889706001074

Max, K., Platscher, M., & Smirnov, J. (2017, September). Gravitational Wave Oscillations in Bigravity. *Physical Review Letters*, *119*(11), 111101. doi: 10.1103/PhysRevLett.119.111101

Max, K., Platscher, M., & Smirnov, J. (2018, March). Decoherence of gravitational wave oscillations in bigravity. *Physical Review D*, *97*(6), 064009. doi: 10.1103/PhysRevD.97.064009

McEwen, J. E., Fang, X., Hirata, C. M., & Blazek, J. A. (2016, September). FAST-PT: a novel algorithm to calculate convolution integrals in cosmological perturbation theory. *Journal of Cosmology and Astroparticle Physics*, *2016*(09), 015–015. doi: 10.1088/1475-7516/2016/09/015

Mead, A. J., Heymans, C., Lombriser, L., Peacock, J. A., Steele, O. I., & Winther, H. A. (2016, March). Accurate halo-model matter power spectra with dark energy, massive neutrinos and modified gravitational forces. *Monthly Notices of the RAS*, *459*(2), 1468–1488. doi: 10.1093/mnras/stw681

Messenger, C., & Read, J. (2012, March). Measuring a Cosmological Distance-Redshift Relationship Using Only Gravitational Wave Observations of Binary Neutron Star Coalescences. *Physical Review Letters*, *108*(9), 091101. doi: 10.1103/PhysRevLett.108.091101

Miranda, T., Fabris, J. C., & Piattella, O. F. (2017, September). Reconstructing a $f(R)$ theory from the α-Attractors. *Journal of Cosmology and Astroparticle Physics*, *09*, 041. doi: 10.1088/1475-7516/2017/09/041

Mukhanov, V. (2005). *Physical Foundations of Cosmology*. Cambridge: Cambridge Univ. Press. doi: 10.1017/CBO9780511790553

Mukhanov, V. F. (1988, July). The quantum theory of gauge-invariant cosmological perturbations. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, *94*, 1-11.

Mukherjee, P., & Mukherjee, A. (2021, June). Assessment of the cosmic distance duality relation using Gaussian process. *Monthly Notices of the RAS*, *504*(3), 3938-3946. doi: 10.1093/mnras/stab1054

Mukherjee, S., & Wandelt, B. D. (2018, August). Beyond the classical distance-redshift test: cross-correlating redshift-free standard candles and sirens with redshift surveys. *arXiv e-prints*, arXiv:1808.06615.

Mukherjee, S., Wandelt, B. D., & Silk, J. (2020a, May). Multimessenger tests of gravity with weakly lensed gravitational waves. *Physical Review D*, *101*(10), 103509. doi: 10.1103/PhysRevD.101.103509

Mukherjee, S., Wandelt, B. D., & Silk, J. (2020b, May). Probing the theory of gravity with gravitational lensing of gravitational waves and galaxy surveys. *Monthly*

Notices of the RAS, *494*(2), 1956-1970. doi: 10.1093/mnras/staa827

Mukherjee, S., Wandelt, B. D., & Silk, J. (2021, March). Testing the general theory of relativity using gravitational wave propagation from dark standard sirens. *Monthly Notices of the RAS*, *502*(1), 1136-1144. doi: 10.1093/mnras/stab001

Mukohyama, S., Namba, R., Peloso, M., & Shiu, G. (2014, August). Blue Tensor Spectrum from Particle Production during Inflation. *Journal of Cosmology and Astroparticle Physics*, *08*, 036. doi: 10.1088/1475-7516/2014/08/036

Namba, R., Peloso, M., Shiraishi, M., Sorbo, L., & Unal, C. (2016, January). Scale-dependent gravitational waves from a rolling axion. *Journal of Cosmology and Astroparticle Physics*, *01*, 041. doi: 10.1088/1475-7516/2016/01/041

Namikawa, T., Nishizawa, A., & Taruya, A. (2016, July). Detecting Black-Hole Binary Clustering via the Second-Generation Gravitational-Wave Detectors. *Physical Review D*, *94*(2), 024013. doi: 10.1103/PhysRevD.94.024013

Oguri, M. (2016, April). Measuring the distance-redshift relation with the cross-correlation of gravitational wave standard sirens and galaxies. *Physical Review D*, *93*(8), 083511. doi: 10.1103/PhysRevD.93.083511

Palma, G. A. (2015, April). Untangling features in the primordial spectra. *Journal of Cosmology and Astroparticle Physics*, *04*, 035. doi: 10.1088/1475-7516/2015/04/035

Peacock, J. A., & Dodds, S. J. (1994, April). Reconstructing the Linear Power Spectrum of Cosmological Mass Fluctuations. *Monthly Notices of the RAS*, *267*, 1020. doi: 10.1093/mnras/267.4.1020

Peirone, S. (2020). Probing gravity at cosmic scales. *Casimir PhD Series*. Retrieved from `https://hdl.handle.net/1887/137440`

Peloso, M., Sorbo, L., & Unal, C. (2016, September). Rolling axions during inflation: perturbativity and signatures. *Journal of Cosmology and Astroparticle Physics*, *09*, 001. doi: 10.1088/1475-7516/2016/09/001

Penzias, A. A., & Wilson, R. W. (1965, July). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *The Astrophysical Journal*, *142*, 419-421. doi: 10.1086/148307

Perenon, L., Martinelli, M., Ilić, S., Maartens, R., Lochner, M., & Clarkson, C. (2021, December). Multi-tasking the growth of cosmological structures. *Physics of the Dark Universe*, *34*, 100898. doi: 10.1016/j.dark.2021.100898

Perotto, L., Lesgourgues, J., Hannestad, S., Tu, H., & Y Y Wong, Y. (2006, October). Probing cosmological parameters with the CMB: forecasts from Monte Carlo simulations. *Journal of Cosmology and Astroparticle Physics*, *2006*(10), 013. doi: 10.1088/1475-7516/2006/10/013

Pettorino, V., & Amendola, L. (2015, March). Friction in gravitational waves: A test for early-time modified gravity. *Physics Letters B*, *742*, 353-357. doi: 10.1016/j.physletb.2015.02.007

Pitrou, C., Cusin, G., & Uzan, J.-P. (2020, April). Unified view of anisotropies in the astrophysical gravitational-wave background. *Physical Review D*, *101*(8), 081301. doi: 10.1103/PhysRevD.101.081301

Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., ... Zonca, A. (2014a, November). Planck 2013 results.

XVI. Cosmological parameters. *Astronomy and Astrophysics*, *571*, A16. doi: 10.1051/0004-6361/201321591

Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., ... Zonca, A. (2014b, November). Planck 2013 results. XXII. Constraints on inflation. *Astronomy and Astrophysics*, *571*, A22. doi: 10.1051/0004-6361/201321569

Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Arroja, F., Ashdown, M., ... et al. (2016, September). Planck 2015 results. XX. Constraints on inflation. *Astronomy and Astrophysics*, *594*, A20. doi: 10.1051/0004-6361/201525898

Planck Collaboration, Aghanim, N., Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., ... Zonca, A. (2020, September). Planck 2018 results. I. Overview and the cosmological legacy of Planck. *Astronomy and Astrophysics*, *641*, A1. doi: 10.1051/0004-6361/201833880

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., ... Zonca, A. (2020a, September). Planck 2018 results. V. CMB power spectra and likelihoods. *Astronomy and Astrophysics*, *641*, A5. doi: 10.1051/0004-6361/201936386

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., ... Zonca, A. (2020b, September). Planck 2018 results. VI. Cosmological parameters. *Astronomy and Astrophysics*, *641*, A6. doi: 10.1051/0004-6361/201833910

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., ... Zonca, A. (2020c, September). Planck 2018 results. VIII. Gravitational lensing. *Astronomy and Astrophysics*, *641*, A8. doi: 10.1051/0004-6361/201833886

Planck Collaboration, Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C., ... Zonca, A. (2020, September). Planck 2018 results. X. Constraints on inflation. *Astronomy and Astrophysics*, *641*, A10. doi: 10.1051/0004-6361/201833887

Powell, M. J. D. (2009, January). The BOBYQA algorithm for bound constrained optimization without derivatives. *Technical Report, Department of Applied Mathematics and Theoretical Physics*.

Pozdeeva, E. O. (2020, July). Generalization of cosmological attractor approach to Einstein-Gauss-Bonnet gravity. *European Physical Journal C*, *80*(7), 612. doi: 10.1140/epjc/s10052-020-8176-3

Pozdeeva, E. O., & Vernov, S. Y. (2021, July). Construction of inflationary scenarios with the Gauss-Bonnet term and nonminimal coupling. *European Physical Journal C*, *81*(7), 633. doi: 10.1140/epjc/s10052-021-09435-8

Raccanelli, A., Kovetz, E. D., Bird, S., Cholis, I., & Munoz, J. B. (2016, July). Determining the progenitors of merging black-hole binaries. *Physical Review D*, *94*(2), 023516. doi: 10.1103/PhysRevD.94.023516

Raccanelli, A., Vidotto, F., & Verde, L. (2018, August). Effects of primordial black holes quantum gravity decay on galaxy clustering. *Journal of Cosmology and Astroparticle Physics*, *08*, 003. doi: 10.1088/1475-7516/2018/08/003

Rasmussen, C. E., & Williams, C. K. I. (2005). *Gaussian processes for machine learning (adaptive computation and machine learning)*. The MIT Press.

Ravenni, A., Verde, L., & Cuesta, A. J. (2016, August). Red, Straight, no bends: primordial power spectrum reconstruction from CMB and large-scale structure. *Journal of Cosmology and Astroparticle Physics*, *2016*(8), 028. doi: 10.1088/1475-7516/2016/08/028

Raveri, M., Hu, B., Frusciante, N., & Silvestri, A. (2014, August). Effective Field Theory of Cosmic Acceleration: constraining dark energy with CMB data. *Phys. Rev.*, *D90*(4), 043513. doi: 10.1103/PhysRevD.90.043513

Reitze, D., Adhikari, R. X., Ballmer, S., Barish, B., Barsotti, L., Billingsley, G., ... Zucker, M. (2019, September). Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. In *Bulletin of the american astronomical society* (Vol. 51, p. 35).

Renzi, F., Cabass, G., Di Valentino, E., Melchiorri, A., & Pagano, L. (2018, August). The Impact of Primordial Magnetic Fields on Future CMB Bounds on Inflationary Gravitational Waves. *Journal of Cosmology and Astroparticle Physics*, *08*, 038. doi: 10.1088/1475-7516/2018/08/038

Renzi, F., Hogg, N. B., Martinelli, M., & Nesseris, S. (2021, May). Strongly lensed supernovae as a self-sufficient probe of the distance duality relation. *Physics of the Dark Universe*, *32*, 100824. doi: 10.1016/j.dark.2021.100824

Renzi, F., Shokri, M., & Melchiorri, A. (2020, January). What is the amplitude of the gravitational waves background expected in the Starobinski model? *Physics of the Dark Universe*, *27*, 100450. doi: 10.1016/j.dark.2019.100450

Reyes, R. P., Dieste, O., C., E. R. F., & Juristo, N. (2020, April). Publication bias. *Proceedings of the Evaluation and Assessment in Software Engineering*. doi: 10.1145/3383219.3383233

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., ... Tonry, J. (1998, September). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *Astronomical Journal*, *116*, 1009-1038. doi: 10.1086/300499

Robert, C. P. (2011, November). Error and Inference: an outsider stand on a frequentist philosophy. *arXiv e-prints*, arXiv:1111.5827.

Robertson, H. P. (1935, November). Kinematics and World-Structure. *The Astrophysical Journal*, *82*, 284. doi: 10.1086/143681

Rodrigues, J. G., Santos da Costa, S., & Alcaniz, J. S. (2021, April). Observational constraints on α-attractor inflationary models with a Higgs-like potential. *Physics Letters B*, *815*, 136156. doi: 10.1016/j.physletb.2021.136156

Rubin, V. C. (1983, June). The Rotation of Spiral Galaxies. *Science*, *220*(4604), 1339-1344. doi: 10.1126/science.220.4604.1339

Sakstein, J., & Jain, B. (2017, December). Implications of the Neutron Star Merger GW170817 for Cosmological Scalar-Tensor Theories. *Physical Review Letters*, *119*(25), 251303. doi: 10.1103/PhysRevLett.119.251303

Sane, P. (2021, February). A Brief Survey of Current Software Engineering Practices in Continuous Integration and Automated Accessibility Testing. *arXiv e-prints*, arXiv:2103.00097.

Sasaki, M. (1983, August). Gauge-Invariant Scalar Perturbations in the New Inflationary Universe. *Progress of Theoretical Physics*, *70*(2), 394-411. doi: 10.1143/PTP.70.394

Sasaki, M., Suyama, T., Tanaka, T., & Yokoyama, S. (2016, August). Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914. *Physical Review Letters*, *117*(6), 061101. doi: 10.1103/PhysRevLett.117.061101

Sasaki, M., Suyama, T., Tanaka, T., & Yokoyama, S. (2018, March). Primordial black holes—perspectives in gravitational wave astronomy. *Classical and Quantum Gravity*, *35*(6), 063001. doi: 10.1088/1361-6382/aaa7b4

Scelfo, G., Bellomo, N., Raccanelli, A., Matarrese, S., & Verde, L. (2018, September). GW×LSS: chasing the progenitors of merging binary black holes. *Journal of Cosmology and Astroparticle Physics*, *2018*(09), 039–039. doi: 10.1088/1475-7516/2018/09/039

Scelfo, G., Boco, L., Lapi, A., & Viel, M. (2020, October). Exploring galaxies-gravitational waves cross-correlations as an astrophysical probe. *Journal of Cosmology and Astroparticle Physics*, *10*, 045. doi: 10.1088/1475-7516/2020/10/045

Schafer, C. M. (2015, April). A Framework for Statistical Inference in Astrophysics. *Annual Review of Statistics and Its Application*, *2*(1), 141-162. doi: 10.1146/annurev-statistics-022513-115538

Schafer, C. M., & Stark, P. B. (2003). Inference in microwave cosmology: A frequentist perspective. *Statistical Challenges in Astronomy*, 215–219.

Schaye, J., Crain, R. A., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., ... Trayford, J. (2015, January). The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *Monthly Notices of the RAS*, *446*(1), 521-554. doi: 10.1093/mnras/stu2058

Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., Schommer, R. A., Clocchiatti, A., Kirshner, R. P., ... Ciardullo, R. (1998, November). The high-z supernova search: Measuring cosmic deceleration and global curvature of the universe using type ia supernovae. *The Astrophysical Journal*, *507*(1), 46–63. doi: 10.1086/306308

Schneider, A., Teyssier, R., Stadel, J., Chisari, N. E., Le Brun, A. M. C., Amara, A., & Refregier, A. (2019, March). Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation. *Journal of Cosmology and Astroparticle Physics*, *2019*(3), 020. doi: 10.1088/1475-7516/2019/03/020

Schutz, B. F. (1986, September). Determining the Hubble constant from gravitational wave observations. *Nature*, *323*(6086), 310-311. doi: 10.1038/323310a0

Scikit-learn 0.19.1 documentation: Gaussian Processes. (2018). `http://scikit-learn.org/stable/modules/gaussian_process.html`. (Accessed: 10-06-2019)

Scikit-learn 1.1.1 documentation: Gaussian Mixture Model Selection. (2022). `https://scikit-learn.org/stable/auto_examples/mixture/plot_gmm_selection.html#sphx-glr-auto-examples-mixture-plot-gmm-selection-py`. (Accessed: 20-05-2022)

Scrum: practical guide. (2022). Retrieved from `https://www.scrum.org/about`

Shafieloo, A., Kim, A. G., & Linder, E. V. (2012, June). Gaussian process cosmography. *Physical Review D*, *85*(12), 123530. doi: 10.1103/PhysRevD.85.123530

Shokri, M., Renzi, F., & Melchiorri, A. (2019, March). Cosmic Microwave Background constraints on non-minimal couplings in inflationary models with power law potentials. *Physics of the Dark Universe*, *24*, 100297. doi: 10.1016/j.dark.2019.100297

Skilling, J. (2006, December). Nested sampling for general bayesian computation. *Bayesian Analysis*, *1*(4), 833–859. doi: 10.1214/06-BA127

Slosar, A., Chen, X., Dvorkin, C., Meerburg, D., Wallisch, B., Green, D., & Silverstein, E. (2019, May). Scratches from the past: Inflationary archaeology through features in the power spectrum of primordial fluctuations. *Bulletin of the American Astronomical Society*, *51*(3). (https://baas.aas.org/pub/2020n3i098)

Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., ... Couchman, H. M. P. (2003, June). Stable clustering, the halo model and non-linear cosmological power spectra. *Monthly Notices of the RAS*, *341*(4), 1311-1332. doi: 10.1046/j.1365-8711.2003.06503.x

Smith, T. L., Pierpaoli, E., & Kamionkowski, M. (2006, July). New Cosmic Microwave Background Constraint to Primordial Gravitational Waves. *Physical Review Letters*, *97*(2), 021301. doi: 10.1103/PhysRevLett.97.021301

Soares-Santos, M., Palmese, A., Hartley, W., Annis, J., Garcia-Bellido, J., Lahav, O., ... Virgo Collaboration (2019, May). First Measurement of the Hubble Constant from a Dark Standard Siren using the Dark Energy Survey Galaxies and the LIGO/Virgo Binary-Black-hole Merger GW170814. *The Astrophysical Journal*, *876*(1), L7. doi: 10.3847/2041-8213/ab14f1

Sousa, L., Avelino, P. P., & Guedes, G. S. F. (2020, February). Full analytical approximation to the stochastic gravitational wave background generated by cosmic string networks. *arXiv e-prints*, arXiv:2002.01079.

Stewart, A., & Brandenberger, R. (2008, August). Observational Constraints on Theories with a Blue Spectrum of Tensor Modes. *Journal of Cosmology and Astroparticle Physics*, *08*, 012. doi: 10.1088/1475-7516/2008/08/012

Stoica, P., & Selen, Y. (2004, July). Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, *21*(4), 36-47. doi: 10.1109/MSP.2004.1311138

Suzuki, A., et al. (2018, December). The litebird satellite mission: Sub-kelvin instrument. *Journal of Low Temperature Physics*, *193*(5), 1048–1056.

Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. (2012, December). Revising the Halofit Model for the Nonlinear Matter Power Spectrum. *The Astrophysical Journal*, *761*(2), 152. doi: 10.1088/0004-637X/761/2/152

Taylor, P. L., Bernardeau, F., & Kitching, T. D. (2018, October). k -cut cosmic shear: Tunable power spectrum sensitivity to test gravity. *Physical Review D*, *98*(8), 083514. doi: 10.1103/PhysRevD.98.083514

Taylor, P. L., Kitching, T., Cardone, V. F., Ferté, A., Huff, E. M., Bernardeau, F., ... Zoubian, J. (2021, July). Euclid: Forecasts for k-cut 3×2 Point Statistics. *The Open Journal of Astrophysics*, *4*(1), 6. doi: 10.21105/astro.2012.04672

Taylor, S. R., Vallisneri, M., Ellis, J. A., Mingarelli, C. M. F., Lazio, T. J. W., & van Haasteren, R. (2016, March). Are We There Yet? Time to Detection of Nanohertz Gravitational Waves Based on Pulsar-timing Array Limits. *"Astrophysical Journal, Letters"*, *819*(1), L6. doi: 10.3847/2041-8205/819/1/L6

Thrane, E., Ballmer, S., Romano, J. D., Mitra, S., Talukder, D., Bose, S., & Mand ic, V. (2009, December). Probing the anisotropies of a stochastic gravitational-wave background using a network of ground-based laser interferometers. *Physical Review D*, *80*(12), 122002. doi: 10.1103/PhysRevD.80.122002

To, C., Krause, E., Rozo, E., Wu, H., Gruen, D., Wechsler, R. H., ... DES Collaboration (2021, April). Dark Energy Survey Year 1 Results: Cosmological Constraints from Cluster Abundances, Weak Lensing, and Galaxy Correlations. *Physical Review Letters*, *126*(14), 141301. doi: 10.1103/PhysRevLett.126.141301

Tonry, J., & Davis, M. (1979, October). A survey of galaxy redshifts. I. Data reduction techniques. *Astronomical Journal*, *84*, 1511-1525. doi: 10.1086/112569

Torrado, J., Hu, B., & Achúcarro, A. (2017, October). Robust predictions for an oscillatory bispectrum in Planck 2015 data from transient reductions in the speed of sound of the inflaton. *Physical Review D*, *96*(8), 083515. doi: 10.1103/PhysRevD.96.083515

Torrado, J., & Lewis, A. (2019, October). *Cobaya: Bayesian analysis in cosmology.* Astrophysics Source Code Library, record ascl:1910.019.

Torrado, J., & Lewis, A. (2021, May). Cobaya: code for Bayesian analysis of hierarchical physical models. *Journal of Cosmology and Astroparticle Physics*, *2021*(5), 057. doi: 10.1088/1475-7516/2021/05/057

Trotta, R. (2017, January). Bayesian Methods in Cosmology. *arXiv e-prints*, arXiv:1701.01467.

Tutusaus Lleixa, I. (2018, September). *Study of the dark components of the Universe with the Euclid mission* (Theses, Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)). Retrieved from `https://tel.archives-ouvertes.fr/tel-01909574`

Ungarelli, C., & Vecchio, A. (2001, November). Studying the anisotropy of the gravitational wave stochastic background with LISA. *Physical Review D*, *64*, 121501. doi: 10.1103/PhysRevD.64.121501

Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Vardanyan, V., Cañas-Herrera, G., & Contigiani, O. (2021). *GW-cosmo.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.4905445` doi: 10.5281/zenodo.4905445

Verlinde, E. (2017, May). Emergent Gravity and the Dark Universe. *SciPost Physics*, *2*(3), 016. doi: 10.21468/SciPostPhys.2.3.016

Viechtbauer, W. (2007, November). Publication bias in meta-analysis: Prevention, assessment and adjustments. *Psychometrika*, *72*(2), 269-271.

Walker, A. G. (1937, January). On Milne's Theory of World-Structure. *Proceedings of the London Mathematical Society*, *42*, 90-127. doi: 10.1112/plms/s2-42.1.90

Wang, Y.-F., Huang, Q.-G., Li, T. G. F., & Liao, S. (2020, March). Searching for primordial black holes with stochastic gravitational-wave background in the space-based detector frequency band. *Physical Review D*, *101*(6), 063019. doi: 10.1103/PhysRevD.101.063019

Weinberg, S. (2005, August). Quantum contributions to cosmological correlations. *Physical Review D*, *72*(4), 043514. doi: 10.1103/PhysRevD.72.043514

Weltman, A., Bull, P., Camera, S., Kelley, K., Padmanabhan, H., Pritchard, J., ... Gaensler, B. M. (2020, January). Fundamental physics with the Square Kilometre Array. *Publications of the Astron. Soc. of Australia*, *37*, e002. doi: 10.1017/pasa.2019.42

WMAP. (2003, September). First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results. *The Astrophysical Journal, Supplement*, *148*(1), 1-27. doi: 10.1086/377253

Wyithe, J. S. B., & Loeb, A. (2003, June). Low-Frequency Gravitational Waves from Massive Black Hole Binaries: Predictions for LISA and Pulsar Timing Arrays. *The Astrophysical Journal*, *590*(2), 691-706. doi: 10.1086/375187

Ye, C., & Fishbach, M. (2021, August). Cosmology with standard sirens at cosmic noon. *Physical Review D*, *104*(4), 043507. doi: 10.1103/PhysRevD.104.043507

Zhao, G.-B., Crittenden, R. G., Pogosian, L., & Zhang, X. (2012, October). Examining the Evidence for Dynamical Dark Energy. *Physical Review Letters*, *109*(17), 171301. doi: 10.1103/PhysRevLett.109.171301

Zuntz, J., Paterno, M., Jennings, E., Rudd, D., Manzotti, A., Dodelson, S., ... Kowalkowski, J. (2015, September). CosmoSIS: Modular cosmological parameter estimation. *Astronomy and Computing*, *12*, 45-59. doi: 10.1016/j.ascom.2015.05.005

Summary

The fascination with the night sky. A simple gaze at the wonderful landscape above us opens a window toward imagination. This fascination together with our innate human curiosity paved the passion for understanding the world we live in. From ancient cultures to nowadays, we were all concerned about the same questions: "*Where did we come from?*" and "*What is our place in the Universe?*".

Understanding and learning more about the cosmos is not only appealing to those who dedicate their lives to science but also to the lay population in general. A quick search within the most popular forums on the internet demonstrates that all have threads up-to-date dedicated to conversations about the universe[12] Moreover, every time there is an exciting astronomical discovery, it spreads within a few hours, as was the case for the first-ever image captured by the Event Horizon Telescope (EHT) of the supermassive black hole Sgr A*.

The composition of the universe, as well as its believed origin, and the main force dominating, gravity, are part of the popular culture of our civilization. But, how is (and was) the Universe studied? We rely on astronomical observations of, for instance, stars and other massive compact objects present in the Universe, to build our scientific knowledge based on the comparison of those observations with some physical assumptions, which form a cosmological model. For example, an average lay citizen is familiar with the concept of Big Bang.

Indeed, cosmologists believe that the universe was formed approximately 13.77 billion years ago from a very dense and hot state. The young Universe cooled down while it expanded, and it is still expanding today at a faster rate. Our Universe is composed of two main ingredients: an unknown matter substance that interacts gravitationally denominated dark matter and the agent responsible of the accelerated expansion of the universe called dark energy. The matter we are all made of, baryonic matter, accounts only for approximately 5% of the content of the universe. Therefore, it is not surprising that concepts such as dark energy and dark matter are surrounded by a halo of mystery, which attracts the interest of the public. Talking in scientific terms, most of the scientific knowledge about the composition of the universe was obtained from observations of the cosmic microwave background: the relic radiation

[12]See for instance `https://www.quora.com/Why-do-you-love-to-gaze-at-the-stars-of-the-night-sky`.

emitted in the early universe when the first neutral atoms were formed and photons could finally scatter away and move freely throughout the Universe. From these observations, we concluded that our Universe is flat.

As the Universe cooled down and expanded, gravity started to play a more significant role and the current structures observed in our Universe started to form: from the first stars to the first galaxies, and from the first galaxies to the first clusters of galaxies. These objects formed the so-called large scale structure of the universe, which looks like a web. However, gravity is not the only player in the formation of the large scale structure of the universe. We need the existence of an underlying Gaussian distribution of primordial density perturbations that can explain the origin of the current structure. So far, inflation, an exponential expansion phase in the primordial universe, is believed to be the mechanism of production of these primordial seeds. The simplest model of inflation is compatible with the assumption of a Gaussian distribution of the primordial density perturbations, apart from explaining why the cosmological observations indicate that the universe is causally connected and flat. The inflationary paradigm together with the assumption that our Universe is mostly composed of dark matter and dark energy are the basis of the standard cosmological model.

In the coming years, the cosmological observations available to study the Universe are expected to increase. In particular, we will exploit the information encoded in the large scale structure of the universe. For that, new missions and experiments are focusing on creating large galaxy surveys that will contain information about their positions in the sky, their shapes and also their redshifts. With the catalogues, cosmologists will study the composition of the universe using observables such as galaxy clustering (GC) and weak lensing (WL). Among these experiments, the European Space Agency medium-size *Euclid* mission, whose satellite is expected to be launched in the near future, aims to understand the physical origin of the accelerated expansion of the universe, as well as understand the nature of dark matter. *Euclid* will also study the initial conditions that seeded the early universe and were responsible for the formation of the cosmic-web structure. For that, *Euclid* will map the large scale structure of the universe by creating one of the largest galaxy catalogues ever known.

What is the methodology used to compare astrophysical observations with physical models to extract conclusions about the universe we live in? Cosmologists have used statistics, in particular, a Bayesian statistical approach, during the last 30 years to test different cosmological models against data. Bayesian statistics are based on Bayes' theorem. This theorem tells us that our degree of belief can be encoded in a probability distribution. Bayesian statistics assume that the probability distribution of some parameters of a model, given some observed data, is directly proportional to the product of two probability distributions: the distribution that gives the probability of observing the data given the model M and the parameter values, times the probability distribution that encodes some prior knowledge information known about the theory or the experiment. In the last decades, the success of Bayesian statis-

tics applied to cosmology is mostly due to the exponential increase in computational power that made massive numerical inference feasible for the first time.
Bayesian statistics can be used for a large set of scientific questions. For instance, it allows us to constrain, given a model and some data, the best fit values of the parameters of that model. On top of that, with this statistical approach we can also discern whether an alternative cosmological model is statistically favoured with respect to the standard cosmological model. Moreover, Bayesian statistics are also useful in forecasting analyses, where we aim to test how sensitive a future experiment will be to the possible detection of a new observable or even, given the experimental set-up of a future mission, if there is any chance that some extensions of the standard cosmological model can be ruled out statistically.

This thesis is dedicated to the Bayesian statistical analyses of extensions of the standard cosmological model using several astronomical data sets, and to the forecast of new observables or experiments. The use of this methodology is the common motto in all the chapters of the thesis. After **Chapter 1**, where we introduce all the main concepts of cosmology as well as the basics of Bayesian statistics, the thesis is divided into three different parts depending on which goal the Bayesian statistical methodology was used for:

- The first part focuses on data science and inflation, and it aims to constrain inflationary models using advanced inference techniques and forecasting tools. **Chapter 2** shows the first-ever results of the reconstruction of the speed of sound of the field responsible for driving inflation, using the latest cosmic microwave background (CMB) data from Planck 2018 and modern algorithms (Gaussian processes). **Chapter 3** is dedicated to the forecast of a particular class of single-field inflation models, known as α-attractors, for a future CMB stage-IV experiment using a model-dependent alternative approach for the sampling of the inflationary parameters based on current constraints obtained by cosmic microwave background and large scale structure data.

- The second part of the thesis is dedicated to the novel concept in cross-correlations of gravitational-wave (GW) physics and large scale structure observables; in particular, galaxy clustering. In two projects we study how we can exploit the information contained in these new observables by forecasting their behaviour and possible detection using future experimental set-ups and Bayesian statistics. **Chapter 4** studies unresolved GW events that form the astrophysical gravitational-wave background, and how we can use the cross-correlation of the anisotropies of that background with galaxy clustering to extract both astrophysical and cosmological information. On the other hand, in **Chapter 5** we investigate how machine learning techniques can be used to reconstruct the propagation of tensor perturbations by combining the spatial correlation between resolved GW mergers and galaxies.

- The third part of this thesis is dedicated to the *Euclid* mission and the tasks of the Euclid Consortium: goals, main survey features and the primary *Euclid* observational probes. In particular, **Chapter 6** focuses on a crucial data science analysis software for the mission: the code *Cosmological Likelihood for Observables in Euclid*, also known as `CLOE`. The Bayesian analysis tool `CLOE` is designed to be able to constrain the values of the parameters of a model given the future *Euclid* data through performing Bayesian inference. In this chapter, we describe the implemented cosmological recipe of the primary probes, as well as the description of the *Euclid* likelihood. The results concerning the structure of `CLOE` and its performance to constrain cosmological parameters are a preview of the current state of the *Euclid* mission, for which I have dedicated a vast percentage of my time.

Samenvatting

De fascinatie voor de nachtelijke hemel. Een eenvoudige blik op het prachtige landschap boven ons opent een venster naar de verbeelding. Uit deze fascinatie samen met onze aangeboren menselijke nieuwsgierigheid, is de passie voor het begrijpen van de wereld waarin wij leven geboren. Van oude culturen tot nu, dezelfde vragen hielden ons bezig: "Waar komen we vandaan?" en "Wat is onze plaats in het heelal?".

Het begrijpen en bestuderen van de kosmos is een onderwerp dat niet alleen wetenschappers, maar ook de bevolking in het algemeen. Een snelle zoektocht op de populairste fora op het internet toont aan dat ze allemaal actieve berichten hebben die gewijd zijn aan gesprekken over het heelal[13] Bovendien wordt elke spannende astronomische ontdekking binnen een paar uur verspreid, zoals het geval was bij de allereerste foto van het superzware zwarte gat Sgr A*, gemaakt door de Event Horizon Telescope (EHT).

De samenstelling van het heelal, de vermoedelijke oorsprong ervan en de belangrijkste dominante kracht, de zwaartekracht, maken deel uit van de popcultuur van onze beschaving. Maar hoe wordt (en werd) het heelal bestudeerd? We vertrouwen op astronomische waarnemingen, bijvoorbeeld van sterren en andere massieve, compacte objecten in het heelal, om onze wetenschappelijke kennis op te bouwen, gebaseerd op de vergelijking van die waarnemingen met voorspellingen gebaseerd op een aantal fysische veronderstellingen, die een kosmologisch model vormen. De gemiddelde leek is bijvoorbeeld bekend met het begrip "Big Bang".

Kosmologen hebben gevonden dat het heelal ongeveer 13,77 miljard jaar geleden is ontstaan uit een zeer dichte en hete toestand. Het jonge heelal koelde af terwijl het uitdijde, en het dijt nu nog steeds sneller uit. Ons heelal is opgebouwd uit twee hoofdbestanddelen: een onbekende materie die gravitationeel op elkaar inwerkt, genaamd donkere materie, en de stof die verantwoordelijk is voor de versnelde uitdijing van het heelal, donkere energie. De materie waar wij allemaal van gemaakt zijn, de baryonische materie, maakt slechts ongeveer 5% uit van de inhoud van het heelal. Het is dan ook niet verwonderlijk dat begrippen als donkere energie en donkere materie omgeven zijn door een waas van mysterie die de belangstelling van het publiek wekt. Wetenschappelijk gezien is het grootste deel van de kennis over de samenstelling van

[13]Zie bijvoorbeeld `https://www.quora.com/Why-do-you-love-to-gaze-at-the-stars-of-the-night-sky`.

het heelal verkregen door waarnemingen van de kosmische achtergrondstraling (Cosmic Microwave Background): de reststraling die in het vroege heelal is uitgezonden toen de eerste neutrale atomen werden gevormd en fotonen eindelijk konden verstrooien en zich vrij door het heelal konden bewegen. Uit deze waarnemingen hebben we ook geconcludeerd dat ons heelal geometrisch vlak is.

Naarmate het heelal afkoelde en uitdijde, begon de zwaartekracht een grotere rol te spelen en ontstonden de huidige structuren die in ons heelal waar te nemen zijn: van de eerste sterren tot de eerste sterrenstelsels, en van de eerste sterrenstelsels tot de eerste clusters van sterrenstelsels. Deze objecten vormden de zogenaamde grootschalige structuur van het heelal, die eruitziet als een web. De zwaartekracht is echter niet de enige speler bij de vorming van de grootschalige structuur van het heelal. Er moet een onderliggende Gaussische verdeling van primordiale dichtheidsverstoringen bestaan die de oorsprong van de huidige structuur kan verklaren. Tot nu toe wordt aangenomen dat inflatie, een exponentiële expansiefase in het oerheelal, het mechanisme is voor de productie van deze primordiale kiemen. Het eenvoudigste model van inflatie is verenigbaar met de veronderstelling van een Gaussische verdeling van de primordiale dichtheidsverstoringen, en verklaart tevens waarom de kosmologische waarnemingen erop wijzen dat het heelal causaal verbonden en vlak is. Het inflatieparadigma vormt samen met de aanname dat ons heelal voor het grootste deel uit donkere materie en donkere energie bestaat, de basis van het standaard kosmologisch model.

In de komende jaren zullen naar verwachting steeds meer kosmologische waarnemingen beschikbaar komen om het heelal te bestuderen. In het bijzonder zal gebruik worden gemaakt van de informatie die is gecodeerd in de grootschalige structuur van het heelal. Daartoe zijn nieuwe missies en experimenten gericht op het maken van grote verzamelingen van gegevens van van sterrenstelsels, die informatie zullen bevatten over hun posities aan de hemel, hun vormen en ook hun roodverschuivingen. Met de catalogi zullen kosmologen de samenstelling van het heelal bestuderen aan de hand van waarnemingen zoals clustering van sterrenstelsels (GC) en zwakke lenseffecten (WL). Een van deze experimenten, de middelgrote *Euclid* missie van de Europese ruimtevaartorganisatie (European Space Agency), waarvan de satelliet naar verwachting binnenkort zal worden gelanceerd, is erop gericht de fysische oorsprong van de versnelde uitdijing van het heelal te begrijpen en inzicht te krijgen in de aard van donkere materie. De *Euclid* missie zal ook de begincondities bestuderen die aan de basis lagen van het ontstaan van het heelal en die verantwoordelijk waren voor de vorming van de structuur van het kosmische web. Daartoe zal *Euclid* de grootschalige structuur van het heelal in kaart brengen door een van de grootste catalogi van sterrenstelsels ooit te maken.

Wat is de methodologie die wordt gebruikt om astrofysische waarnemingen te vergelijken met fysische modellen om conclusies te trekken over het heelal waarin we leven? Kosmologen hebben in de laatste 30 jaar gebruik gemaakt van statistiek, in het bijzonder van een Bayesiaanse statistische benadering, om verschillende kos-

mologische modellen te toetsen aan data. Bayesiaanse statistiek is gebaseerd op het theorema van Bayes. Dit theorema zegt ons dat onze graad van geloof kan worden gecodeerd in een waarschijnlijkheidsverdeling. De Bayesiaanse statistiek gaat ervan uit dat de kansverdeling van een aantal parameters van een model, gegeven een aantal geobserveerde gegevens, recht evenredig is met het product van de kansverdeling die de waarschijnlijkheid geeft van het observeren van de gegevens gegeven het model M en de parameterwaarden maal de kansverdeling die een aantal *a priori* bekende informatie over de theorie of het experiment codeert. In de laatste decennia is het succes van de Bayesiaanse statistiek toegepast op de kosmologie vooral te danken aan de exponentiële toename van de rekenkracht van computers die massale numerieke inferentie voor het eerst haalbaar maakte.

Bayesiaanse statistiek kan worden gebruikt voor een groot aantal wetenschappelijke vraagstukken. Het stelt ons bijvoorbeeld in staat om, gegeven een model en enkele gegevens, de best passende waarden van de parameters van dat model te bepalen. Daarnaast kunnen wij met deze statistische benadering ook nagaan of een alternatief kosmologisch model volgens de statistiek de voorkeur geniet ten opzichte van het standaard kosmologisch model. Bovendien zijn Bayesiaanse statistieken ook nuttig in voorspellingsanalyses, waarin we willen testen hoe gevoelig een toekomstig experiment zal zijn voor de mogelijke detectie van een nieuw waarneembaar object of zelfs, gegeven de experimentele opzet van een toekomstige missie, of er een kans is dat sommige uitbreidingen van het standaard kosmologisch model statistisch kunnen worden uitgesloten.

Dit proefschrift is gewijd aan de Bayesiaanse statistische analyses van uitbreidingen van het standaard kosmologisch model met behulp van verschillende astronomische datasets, en aan de voorspelling van nieuwe waarneembare gegevens of experimenten. Het gebruik van deze methodologie is de rode draad van het proefschrift. Na hoofdstuk 1, waarin we de belangrijkste concepten van de kosmologie en de basisprincipes van de Bayesiaanse statistiek introduceren, is het proefschrift verdeeld in drie verschillende delen, afhankelijk van het doel waarvoor de Bayesiaanse statistische methodologie werd gebruikt:

- Het eerste deel richt zich op datawetenschap en inflatie, en het heeft als doel inflatiemodellen te begrenzen met behulp van geavanceerde inferentietechnieken en voorspellingsinstrumenten. **Hoofdstuk 2** toont de allereerste resultaten van de reconstructie van de geluidssnelheid van het veld dat verantwoordelijk is voor de inflatie, met behulp van de nieuwste gegevens over de kosmische achtergrondstraling (CMB) van Planck 2018 en moderne algoritmes (Gaussische processen). **Hoofdstuk 3** is gewijd aan de voorspelling van een bepaalde klasse van inflatiemodellen met één veld, bekend als -attractoren, voor een toekomstig CMB fase-IV-experiment met behulp van een model-afhankelijke, alternatieve benadering voor de verzameling van de inflatieparameters op basis van de huidige beperkingen die verkregen zijn met gegevens over de kosmische achtegrondstraling en de grote schaal structuur.

- Het tweede deel van het proefschrift is gewijd aan de nieuwe interesse in de cross-correlaties van gravitatie-golf (GW) fysica en grote schaal structuur waarnemingen; en in het bijzonder, clustering van sterrenstelsels. In deze twee projecten bestuderen we hoe we de informatie in deze nieuwe waarneemgegevens kunnen exploiteren door hun gedrag en mogelijke detectie te voorspellen met behulp van toekomstige experimenten en Bayesiaanse statistiek. In **hoofdstuk 4** bestuderen we de ruis van overlappende GW signalen die de astrofysische zwaartekrachtgolfachtergrond vormen, en hoe we de correlatie tussen de structuur van die achtergrond en de clustering van melkwegstelsels kunnen gebruiken om zowel astrofysische als kosmologische informatie te verkrijgen. Anderzijds onderzoeken we in **hoofdstuk 5** hoe machine learning technieken kunnen worden gebruikt om de propagatie van tensorverstoringen te reconstrueren door de ruimtelijke correlatie tussen individueel herkenbare GW bronnen en melkwegstelsels te combineren.

- Het derde deel van dit proefschrift is gewijd aan de *Euclid* missie en de taken van het Euclid Consortium: doelstellingen, hoofdkenmerken van de survey en de primaire *Euclid*-observatiesondes. In het bijzonder wordt in **hoofdstuk 6** aandacht besteed aan cruciale software voor de data-analyse voor de missie: de code *Cosmological Likelihood for Observables in Euclid*, ook bekend als `CLOE`. Dit Bayesiaanse analyse-instrument is ontworpen om de parameterwaarden van een model te kunnen afleiden uit de toekomstige metingen van de *Euclid* missie. In dit hoofdstuk beschrijven we het geïmplementeerde kosmologische recept van de primaire sondes, evenals de beschrijving van de statistische waarschijnlijkheid van *Euclid* data. De resultaten met betrekking tot de structuur van `CLOE` en hoe goed de code presteert bij het beperken van kosmologische parameters zijn een voorproefje van wat we hopen te leren van de *Euclid* missie, ik een groot deel van mijn tijd heb besteed.

List of Publications

- *Bayesian reconstruction of the inflaton's speed of sound using CMB data*
  Guadalupe Cañas-Herrera, Jesús Torrado Cacho and Ana Achúcarro
  2020, Physical Review D 103, 12, 123531

- *Current and future constraints on single-field α-attractor models*
  Guadalupe Cañas-Herrera, Fabrizio Renzi
  2021, Physical Review D 104, 10, 103512

- *Cross-correlation of the astrophysical gravitational-wave background with galaxy clustering*
  Guadalupe Cañas-Herrera, Omar Contigiani and Valeri Vardanyan.
  2020, Physical Review D 102, 4, 043513

- *Learning how to surf: Reconstructing the propagation and origin of gravitational waves with Gaussian Processes*
  Guadalupe Cañas-Herrera, Omar Contigiani and Valeri Vardanyan.
  2021, The Astrophysical Journal, 918, 20.

Curriculum Vitae

I was born on the 8th of April 1993 in Griñón (Madrid), Spain, to Miguel Cañas Muñoz and Guadalupe Herrera Herrera. When I was three years old we moved to Cantabria. I pursued my primary education at the public school *Miguel Primo de Rivera* in Ampuero, Cantabria, and I graduated with honours in 2011 from the secondary school *IES Fuente Fresnedo*, within a French bilingual program, in Laredo, Cantabria. Since I was a child I knew I would become a scientist one day, and when I was only fourteen years old, I was determined to study physics. I combined my compulsory studies with my interest in classical music studying piano at the Regional Conservatory. I became a professional pianist graduating at Conservatory *Jesús de Monasterio* in 2013.

In 2011 I began my Bachelor's studies in Physics at *University of Cantabria*, pursuing an Erasmus exchange year at *Utrecht University*, the Netherlands, between 2013 and 2014. I completed my Bachelor's research project during a summer internship in 2015 at *Brown University*, USA, in the field of "Weak Lensing measurements". In spring 2016, after working as a research intern and a teaching assistant for the *Physics Institute of Cantabria* during the fall semester, I obtained my Bachelor's degree in Physics with a minor in "Fundamental Physics".

After graduating, I spent the summer of 2016 working as a research intern at DESY, Germany, in astroparticle physics. In September 2016, thanks to a LION scholarship, I joined *Leiden University* as a master's student in Physics to pursue the specialization in Cosmology at the *De Sitter* programme. I graduated *cum laude* in July 2018. I combined my studies with organizational positions at the educational board and intensive Dutch language courses.

Soon after, I started my PhD as a *De Sitter* fellow in both the *Lorentz Institute for Theoretical Physics* and *Leiden Observatory* under the supervision of Prof. dr. Ana Achúcarro and Dr. Alessandra Silvestri. I studied various topics in Cosmology such as inflation, gravitational waves and data analysis. During my PhD time, I was teaching assistant of the courses *Theory of General Relativity* and *Quantum Field Theory*, visited several research institutions, participated in many conferences and schools, co-organized the bi-weekly *Cosmology seminars* and the Leiden sessions of the *Theoretical Cosmology Meetings*. I organized my first ever workshop at the

Lorentz Center taking place in September 2022.

Besides, I had the opportunity to join the Euclid Consortium, which is the responsible scientific team of the European Space Agency *Euclid* mission. In this collaboration, I work as an active member of the Theory Science Working Group and as a core developer of the Inter-Science Taskforce Likelihood coding the software *Cosmological Likelihood for Observables in Euclid* (`CLOE`), which is a crucial analysis computational tool for the *Euclid* mission. I participate weekly as a consultant in the Inter-Science Taskforce Non-Linear, I co-organized the 2021 Theory Science Working Group meeting, and I chaired a plenary session of the 2022 Annual Euclid Consortium Meeting. My work within the consortium was acknowledged in 2021 with the STAR prize.

In fall 2022 I will join the European Space Agency as a postdoctoral Research Fellow at the facility location "European Space and Technology Centre" (ESTEC) in Noordwijk (Zuid-Holland, the Netherlands), where I will continue working for the *Euclid* mission.

Acknowledgments

Many thanks to my supervisor, Ana Achúcarro, for being an outstanding mentor during these years and for all the trust she has put in me. Thank you, Alessandra Silvestri, for adopting me in your group meetings and for your ultimate guidance in the *Euclid* Theory Science Working group.

Special thanks go to my collaborators Fabrizio, Jesús, Omar and Valeri, with whom I have written the papers presented in this thesis. I enjoyed our scientific discussions and all the input you gave me at the beginning of my career.

To my *Euclid* IST:Likelihood and IST:Non-Linear colleagues, thanks a lot for teaching me new things about Large Scale Structure, project management, software development and efficient communication. I have always believed that names are important so here they go: Santiago, Chiara, Pedro, Virginia, Marco, Stefano D, Stefano C, Davide, Isaac, Linda, Stéphane, Sergio, Amandine, Andrea, Ziad, Sam, Konstantinos.

Further, my dearest thanks go to my mentor Matteo Martinelli, who was the person that brought me into the *Euclid* world.

I would like to thank also the *Euclid* working group leaders: Valeria, Vincenzo, Shahab, Ariel, Alkistis, Martín. I appreciate beyond words the opportunities you gave me within the consortium. Special thanks go to Henk Hoekstra for supporting my work within *Euclid* and for reading the manuscript of my thesis.

I thank my friend Michelle for helping me with the translation of the summary and for useful Dutch tips and to Koen Kuijkens for checking the Dutch summary with an astronomer's eye.

My experience at the Lorentz Institute would have not been the same without my group colleagues (Gang, Oksana, Alice, Anna, Yashar, Simone). Further, I thank Fran Ouwerkerk for every minute she has spent taking care of my bureaucratic things.

Finally, I would like to thank my friends and family outside the cosmology world for all the support during the PhD time. My friends Susana, Marie, Ángela, Linda and Julia, Ed, Sarah, Sanne and Systse. A mis queridos Marco, Carlos y Zaida. A

mi familia política: Laura, Íñigo, mi ahijado Marco, Ana y Javi. A mi primo Jaime por hacerme la portada más maravillosa del mundo.

A mi madre, por dedicarme los mejores años de su vida. To my father, I want to thank the fantastic job he has done with the painting used in the cover of my thesis. Gracias por apoyarme todos estos años.

To my husband, Juan Claramunt González, thanks for checking the accuracy of the statistical terms and methodology introduced in the first chapter of this thesis. On a more personal note, thank you, Juan, for your patience during these four years but foremost, for your unconditional love and support.