Doctoral course in Physics

Cycle XXXVIII

Coordinator: Prof. Paolo Lenisa

GIORGIA ZAGATTI

From Foreground To Foreshadower: Modelling Extragalactic Emissions *in the* cosmic microwave background *Era to Learn About Astrophysics and Cosmology*

Scientific/Disciplinary Sector (SDS) FIS/02

A thesis submitted in satisfaction of the requirements for the degree of

Doctor of Philosophy

Department of Physics and Earth Sciences
Faculty of Physics
Univeristy of Ferrara

Supervisor
Dr. Martina Gerbino

Co-advisor
Prof. Luca Pagano

November 2025

SUPERVISOR:
Dr. Martina Gerbino

CO-SUPERVISOR:
Prof. Luca Pagano

LOCATION:
Ferrara (IT)
Dipartimento di Fisica e Scienze della Terra,
Università degli Studi di Ferrara, Via G. Saragat 1, I-44122

WEBSITE:
https://www.fe.infn.it/cosmologia/

E-MAIL:
giorgia.zagatti@unife.it (academic)
giorgia.zagatti@gmail.com (personal)

A Gloria e Cristiano,
i miei genitori,
la mia Casa.

ABSTRACT

We are facing an era where future Cosmic Microwave Background (CMB) experiments, both satellite and ground-based, together with Large–scale structure (LSS) surveys, will reach unprecedented precision. These future achievements, while exciting in terms of the advanced physics that can be extracted, will pose new challenges [Alonso et al., 2021] to extract the maximum amount of information from upcoming data and fulfil the synergy between different experiments. In this thesis, I study a physically motivated model of extragalactic emissions, which are, on the one hand, contaminants of the primary CMB signal, on the other hand, contain a great deal of information, both about the formation and the evolution of the LSS of the Universe, and about astrophysics. My objective is to find a model capable of describing these emissions at all frequencies and angular scales, so as to fully exploit the synergy between different experiments and to gain a better understanding of the small angular scales of the CMB power spectrum, which are the primary target of upcoming ground-based experiments.
After an introductory part that sets out the cosmological basis of my work, the second part of this thesis is devoted to the introduction of the main elements behind the modelling I have developed. In the third part I present the details of the modelling of the extragalactic emissions, starting with a focus on the Cosmic Infrared Background (CIB) only, where the novelty is a two–galaxy–population description of the CIB within the Halo Model (HM) framework [Zagatti et al., 2024]. I constrain the free parameters of the model using existing datasets from *Planck* [Planck Collaboration, 2014b] and *Herschel*–SPIRE [Viero et al., 2019]. However, this first analysis highlights a tension between the two, which I am not yet able to resolve; specifically, the open question is whether the issue lies in the modelling, in the datasets, or in both. With current data, it is not possible to answer this question. In the near future, high-precision data from Simons Obervatory (SO) Large Aperture Telescope (LAT) [Abitbol et al., 2025; Ade et al., 2019] will help answer this question, but in order to exploit them, other extragalactic emissions must be included in the modelling. For this reason, I extend the CIB modelling to other extragalactic emissions, such as the Sunyaev-Zel'dovich (SZ) effect, both thermal and kinetic, the cross-correlation between CIB and thermal Sunyaev-Zel'dovich (tSZ), and the contributions coming from radio sources and Galactic cirrus. With this comprehensive modelling in hand, I create a dataset which mimics the upcoming SO LAT data – theoretical C_ℓs generated from my model, together with a power-spectrum-based covariance matrix built from the SO LAT noise curves (including beams and sky fraction). Starting from this dataset, I first go through a Fisher analysis that aims

to evaluate the correlations among the model parameters, to forecast the constraining power of the upcoming data, and to estimate, from a qualitative point of view, the impact that this extragalactic-emission modelling will have on cosmological inference. After this part, I then perform an Markov Chain Monte Carlo (MCMC) analysis, finding promising results in terms of the constraining power that future SO LAT data will have on the different models of the extragalactic emissions.
This study is accompanied by the code I developed to generate the theoretical predictions of my modelling. This code is public, available at `https://github.com/giorgiazagatti/CIB_halomodel.git` and will be used by the SO collaboration, of which I have been a member since 2024.

SOMMARIO

Stiamo entrando in un'epoca in cui i futuri esperimenti di CMB – sia da satellite sia da terra –insieme alle survey di LSS raggiungeranno una precisione senza precedenti. Questi progressi, per quanto entusiasmanti dal punto di vista della fisica avanzata che potremo estrarre, comporteranno nuove sfide [Alonso et al., 2021] per riuscire a trarre la massima informazione dai dati in arrivo e realizzare la piena sinergia tra esperimenti diversi. In questa tesi studio una modellizzazione delle emissioni extragalattiche che, da un lato, contaminano il segnale primario di CMB e, dall'altro, racchiudono un'enorme quantità di informazione sia sulla formazione ed evoluzione della LSS dell'Universo, sia sull'astrofisica delle sorgenti. Il mio obiettivo è costruire un modello capace di descrivere tali emissioni a tutte le frequenze e scale angolari, così da sfruttare appieno la sinergia fra esperimenti e comprendere meglio le piccole scale dello spettro di potenza di CMB, che sono il principale target dei prossimi esperimenti da terra.
Dopo una parte introduttiva che pone le basi cosmologiche del lavoro, la seconda parte di questa tesi è dedicata alla presentazione degli elementi principali alla base della modellizzazione che ho sviluppato. Nella terza parte presento i dettagli del modello delle emissioni extragalattiche, iniziando dal solo CIB, il quale modellizzo a partire dal HM e in cui la novità consiste in una descrizione a due popolazioni di galassie [Zagatti et al., 2024]. Dopo aver implementato il modello, il mio obiettivo è quello di vincolare i parametri liberi utilizzando i dataset esistenti di *Planck* [Planck Collaboration, 2014b] e *Herschel*–SPIRE [Viero et al., 2019]. Questa prima analisi evidenzia tuttavia una tensione tra i due dataset che al momento non sono in grado di risolvere; in particolare, resta aperta la questione se l'origine del problema risieda nella modellizzazione, nei dati, oppure in entrambi. Con i dati attuali non è possibile rispondere in modo definitivo. Nel prossimo futuro, saranno disponibili dati ad alta precisione di SO LAT [Abitbol et al., 2025; Ade et al., 2019], i quali potranno aiutare a chiarire questo punto; per sfruttarli

appieno, però, è necessario includere nella modellizzazione anche le altre emissioni extragalattiche. Per questo motivo, ho esteso il modello del CIB ad altre componenti quali l'effetto SZ (termico e cinetico), la cross-correlazione CIB–tSZ, e i contributi da radio–sorgenti e cirri Galattici. Con questa modellizzazione completa ho costruito quindi un dataset che riproduce i dati di SO LAT – C_ℓ teorici generati dal mio modello, insieme a una matrice di covarianza basata sullo spettro di potenza e costruita a partire dalle curve di rumore di SO LAT (includendo beam e frazione di cielo). A partire da questo dataset, svolgo innanzitutto un'analisi di Fisher per valutare le correlazioni tra i parametri del modello, prevedere il potere vincolante dei dati in arrivo e stimare, in termini qualitativi, l'impatto che questa modellizzazione delle emissioni extragalattiche avrà sull'inferenza cosmologica. Successivamente eseguo un'analisi MCMC, ottenendo risultati promettenti in termini di potere vincolante che i futuri dati SO LAT potranno fornire sui diversi modelli di emissione extragalattica.
Questo studio è accompagnato dal codice che ho sviluppato per generare le predizioni teoriche del mio modello. Il codice è pubblico, disponibile all'indirizzo `https://github.com/giorgiazagatti/CIB_halomodel.git`, e verrà utilizzato dalla collaborazione SO, di cui faccio parte dal 2024.

PUBLICATIONS

The content of this thesis is based on the following publications:

Zagatti, G., M. Gerbino, and L. Pagano (2025). "Joint Modelling of CIB, SZ Effects, Radio Sources, and Cirrus for Small-Scale CMB Analyses: Forecasts for the Simons Observatory LAT." In preparation.

Zagatti, G., E. Calabrese, C. Chiocchetta, M. Gerbino, M. Negrello, and L. Pagano (2024). "A halo model approach to describe clustering and emission of the two main star-forming galaxy populations for cosmic infrared background studies." In: *Astron. Astrophys.* **692**, A190. DOI: 10.1051/0004-6361/202451424. arXiv: 2406.19069 [astro-ph.CO].

The following papers were written during my PhD but are not included in the main body of this thesis, in order to maintain a coherent overview and preserve the flow of the manuscript.

Arcari, S., N. Bartolo, G. Fabbian, A. Greco, A. Gruppuso, M. Lattanzi, P. Natoli, L. Pagano, and G. Zagatti (Sept. 2025). "Stairway to Axions: the cross-correlation of birefringence and galaxies from NPIPE and Quaia." In: arXiv: 2509.22273 [astro-ph.CO].

Zagatti, G., M. Bortolami, A. Gruppuso, P. Natoli, L. Pagano, and G. Fabbian (2024). "Planck constraints on cosmic birefringence and its cross-correlation with the CMB." In: *JCAP* **05**, p. 034. DOI: 10.1088/1475-7516/2024/05/034. arXiv: 2401.11973 [astro-ph.CO].

ACKNOWLEDGMENTS

This thesis reflects the work and support of many people. I am grateful to all those who, in different ways and at different times, shaped these years and helped me grow as a physicist.

First and foremost, I thank my supervisors, **Martina Gerbino** and **Luca Pagano**. Your steady guidance and timely support accompanied me from my first steps to working independently by the end of this journey. You opened many opportunities, for which I am sincerely thankful.

A special thank you goes to Luca, who has been by my side since my Master's. In connection with that period, I also wish to thank **Paolo Natoli**, **Alessandro Gruppuso**, **Giulio Fabbian**, and **Marco Bortolami**. Although my research direction has since shifted, our work together played an important role in shaping the physicist I am today.

I am deeply grateful to **Erminia Calabrese** for welcoming me to Cardiff and for *"bringing me back to my orbit*" at a time when I felt a bit lost. My thanks go also to **Serena Giardiello** for the shared "Italian time" in Wales, and in general, to everyone in the cosmology group in Cardiff.

I warmly thank **Laura Salvati** for hosting me in Paris and giving me the chance to experience the French research environment, which – though brief – left a lasting impression.

My gratitude extends to the entire cosmology group – seniors, staff, post-docs, and PhD students. I spent three years in a stimulating, yet always friendly and familiar, environment.

Finally, to my PhD friends, especially those who have been around since my Master's days (Nicola, Nicolò, Stefano, Margherita, Luca, and Lorenzo). This last year has felt a bit emptier without you around.

Alla mia famiglia, mia costante fonte di supporto. A metà del mio percorso sono andata a vivere da sola, Giulia ormai a casa non c'è da una vita, e pure la Fre ora ha la sua casetta. È cambiato cosí tanto da come eravamo noi cinque, tre anni fa, eppure nulla è cambiato. Il *'buongiorno'* e la *'buonanotte'* a te, mamma, non saltano un giorno. Due chiacchiere davanti a uno spritz o a del buon vino con te, papi, di certo non mancano. Francesca continua a bersi cappuccini con un sacco di schiuma, non importa in che casa se li faccia, e Giulia a trattare Plutino come fosse suo figlio ogni volta che torna a casa. Non sarei nulla di tutto quello che sono oggi senza di voi.

CONTENTS

LIST OF FIGURES xix
LIST OF TABLES xxi
PREFACE xxiii
CONVENTIONS xxv

I FUNDAMENTALS

1 STANDARD COSMOLOGICAL MODEL 3
1.1 Kinematics: cosmological principle, redshift and the Hubble law 4
1.1.1 Cosmological principle . . . 4
1.1.2 Redshift, scale factor and the Hubble law . . . 4
1.2 Geometry: the FLRW metric . . . 6
1.3 Dynamics: Einstein equations, Friedmann equations and distances . . . 9
1.4 Linear perturbations: Newtonian gauge and the Boltzmann equation . . . 13
2 COSMIC MICROWAVE BACKGROUND 23
2.1 Photon perturbations . . . 24
2.1.1 Large-scale anisotropies . . . 24
2.1.2 Acoustic peak region . . . 25
2.1.3 Free stream . . . 27
2.2 CMB power spectrum . . . 29
2.2.1 Large angular scales . . . 29
2.2.2 Acoustic peak region . . . 30
2.3 Polarized CMB . . . 31
2.3.1 Compton scattering . . . 33
2.3.2 Polarisation modes . . . 34
3 STRUCTURE GROWTH IN LINEAR THEORY 39
3.1 Matter perturbations . . . 39
3.1.1 Transfer function . . . 40
3.1.2 Growth function . . . 43
3.2 Linear matter power spectrum . . . 44

II EXTRAGALACTIC EMISSIONS

4 THE HALO MODEL FRAMEWORK 49
4.1 Spherical Collapse Model . . . 49
4.2 Main Ingredients . . . 51
4.2.1 Halo Mass Function . . . 52
4.2.2 Halo Bias . . . 55
4.2.3 Halo Density Profile . . . 58
4.3 Spatial Power Spectra Computation . . . 60

4.3.1 Matter Power Spectrum . . . 64
4.3.2 Galaxy Power Spectrum . . . 65
5 EXTRAGALACTIC EMISSIONS 73
5.1 Cosmic Infrared Background . . . 75
5.2 Sunyaev-Zel'dovich Effect . . . 79
5.2.1 Thermal Sunyaev-Zel'dovich Effect . . . 79
5.2.2 Kinetic Sunyaev-Zel'dovich Effect . . . 83
5.3 CIB and tSZ Cross Correlation . . . 86
5.4 Radio Sources . . . 89
5.5 Galactic Cirrus . . . 90

III A PHYSICALLY MOTIVATED MODELLING OF THE EXTRAGALACTIC EMISSIONS

6 THE COSMIC INFRARED BACKGROUND MODEL 95
6.1 The Model . . . 95
6.1.1 Modeling of the two galaxy populations . . . 96
6.1.2 Modelling of the shot noise term . . . 104
6.2 Available CIB Data . . . 105
6.2.1 Planck data . . . 105
6.2.2 Lenz data . . . 107
6.2.3 Hershel-SPIRE data . . . 108
6.3 Model fit to data . . . 109
6.3.1 P14 data analysis . . . 110
6.3.2 L19 data analysis . . . 115
6.3.3 V19 data analysis . . . 123
6.3.4 P14 and V19 comparison . . . 127
6.4 Conclusions about the CIB modelling . . . 128
7 COMPREHENSIVE MODELLING OF THE EXTRAGALACTIC EMISSIONS 131
7.1 Modelling of the extragalactic components . . . 132
7.2 Creation of a smooth dataset . . . 135
7.2.1 Smooth power spectra . . . 135
7.2.2 Smooth covariance matrix . . . 137
7.3 Results . . . 139
7.3.1 Fisher Forecasts . . . 139
7.3.2 MCMC analysis . . . 144
7.4 Conclusions and future prospects . . . 147
CONCLUSIONS 151

IV APPENDICES

A HALO MODEL CALCULATIONS 157
A.1 Different halo mass function parametrisation . . . 157
A.2 CIB-tSZ power spectrum: full calculations . . . 158
B WEBSKY SIMULATIONS FOR CIB MODELLING 161
C ANISOTROPIC COSMIC BIREFRINGENCE 165

C.1 Introduction . . . 165
C.2 Harmonic estimator . . . 165
C.2.1 EB-only implementation . . . 166
C.3 Power spectrum and de-biasing . . . 167
C.4 Data and simulations . . . 167
C.5 Results . . . 168
C.5.1 Full-mission NPIPE . . . 168
C.5.2 Component-separation . . . 169
C.5.3 Consistency checks . . . 170
C.5.4 Cross-correlations . . . 171
C.6 Birefringence–Galaxy Cross-correlation and Constraints on Axion-like Fields . . . 171
C.7 Conclusions . . . 173
ACRONYMS 175
BIBLIOGRAPHY 177

LIST OF FIGURES

Figure 1.1 Alexander Friedmann. 4
Figure 1.2 Hubble diagram. 5
Figure 1.3 Parallel transport of a vector along geodesics of a sphere. 7
Figure 1.4 Interactions among the different components. 18
Figure 2.1 Evolution of different k–modes up to recombination . . 26
Figure 2.2 CMB power spectrum: low-multipole region 29
Figure 2.3 CMB power spectrum: acoustic-peak region 30
Figure 2.4 Parity even E-modes. 31
Figure 2.5 Parity odd B-modes. 31
Figure 2.6 Compton scattering: single wave. 33
Figure 2.7 Compton scattering: monopole. 33
Figure 2.8 Compton scattering: dipole. 34
Figure 2.9 Compton scattering: quadrupole. 34
Figure 2.10 CMB power spectra. 36
Figure 2.11 CMB temperature map. 37
Figure 2.12 CMB polarisation map. 37
Figure 3.1 Evolution of gravitational potential and matter perturbations. 44
Figure 3.2 Linear matter power spectrum. 46
Figure 4.1 HMF for different redshift values. 56
Figure 4.2 HMF for different halo mass values. 56
Figure 4.3 Halo bias for different redshift values. 58
Figure 4.4 Halo bias for different halo mass values. 58
Figure 4.5 FT of the NFW profile for different redshift values. . . 61
Figure 4.6 FT of the NFW profile for different halo mass values. . 61
Figure 4.7 Scheme of the 1h and 2h contributions. 64
Figure 4.8 Matter power spectrum in a HM framework. 65
Figure 4.9 Galaxy power spectrum in a HM framework. 68
Figure 4.10 HOD predictions for the number of central and satellite galaxies. 70
Figure 4.11 Scheme of the HM. 71
Figure 5.1 Emissivity functions computed at 217, 353, 545 GHz. . 77
Figure 5.2 Prediction of the CIB power spectrum computed in a HM framework. 78
Figure 5.3 Scheme of the inverse Compton scattering. 79
Figure 5.4 Spectral distortion of the CMB due to SZ effect. . . . 80
Figure 5.5 SZ mechanism. 80
Figure 5.6 HMF predictions for different density contrasts at different redshifts. 82

Figure 5.7 Theoretical prediction of the tSZ power spectrum evaluated in a HM framework. 84
Figure 5.8 Predictions of the kSZ power spectrum evaluated with a template-based approach. 85
Figure 5.9 Theoretical prediction of the CIBxtSZ power spectrum evaluated in a HM framework. 88
Figure 5.10 Theoretical prediction of the radio sources power spectrum evaluated using a template-based approach. . . . 90
Figure 5.11 Theoretical prediction of the Galactic cirrus power spectrum evaluated using a template-based approach. 91
Figure 6.1 Galaxy distribution in a colour-magnitude diagram. . . 96
Figure 6.2 Two-galaxy-populations power spectrum. 100
Figure 6.3 Templates SEDs. 103
Figure 6.4 Two-galaxy-populations emissivity functions. 104
Figure 6.5 Planck satellite. 105
Figure 6.6 Hershel spacecraft. 108
Figure 6.7 P14: model predictions vs. observed data. 112
Figure 6.8 P14: posterior distributions of the free model parameters. 113
Figure 6.9 P14: posterior distributions of the clustering parameters with and without the 217 GHz frequency channel. 114
Figure 6.10 P14: posterior distributions when the correlation coefficients are fixed to unity and when they are free to vary. 115
Figure 6.11 P14: unconstrained α_{ET}. 116
Figure 6.12 L19: posterior distributions of the free model parameters. 117
Figure 6.13 L19: posterior distributions of the clustering parameters only. 117
Figure 6.14 L19–P14 comparison of the clustering parameters. . . . 118
Figure 6.15 L19: model predictions vs. observed data. 119
Figure 6.16 L19: covariance matrix. 120
Figure 6.17 L19: correlation matrix. 120
Figure 6.18 L19 CIB datasets for different masking thresholds. . . . 121
Figure 6.19 L19: posterior distributions for single frequency analysis. 123
Figure 6.20 V19: model predictions vs. observed data. 124
Figure 6.21 V19: posterior distributions of the free model parameters. 126
Figure 6.22 V19: posterior distributions of the clustering parameters only. 126
Figure 6.23 V19–P14 comparison of the clustering parameters. . . . 128
Figure 7.1 Simons Observatory site in the Atacama Desert, *Chile*. . 131
Figure 7.2 SO LAT emissivity functions 136
Figure 7.3 SO LAT auto-frequency spectra with error bars. 138
Figure 7.4 SO noise curves. 138
Figure 7.5 Fisher correlation matrix for foregrounds parameters only. 140

Figure 7.6 Fisher SNR for foregrounds parameters only. 141
Figure 7.7 Fisher forecasts results for varying cosmology (Fixed cosmology in foregrounds modelling). 142
Figure 7.8 Fisher forecasts results for varying cosmology (Varying cosmology in foregrounds modelling). 143
Figure 7.9 Comparison of the Fisher SNR for the cosmological parameters. 144
Figure 7.10 Posterior distributions of the free model parameters. . 148
Figure b.1 Map and mask of the CIB from Websky. 163
Figure b.2 `Websky` power spectra; comparison with P14 and L19. . 163
Figure c.1 PR4: de-biased CB power spectrum. 168
Figure c.2 CB power spectrum: consistency checks for different component separation methods. 169
Figure c.3 CB power spectrum: consistency checks for different data splits. 169
Figure c.4 CB power spectrum: consistency checks for different multipoles included in the analysis. 170
Figure c.5 CB power spectrum: consistency checks for different masks. 170
Figure c.6 Cross-spectra between CB and CMB temperature and polarisation fields. 171
Figure c.7 CB-galaxy number counts cross-correlation. 172
Figure c.8 Upper bounds on the axion-photon coupling. 173

LIST OF TABLES

Table 5.1 Conversion factors (CF) for the most used frequency channels throughout my work. 75
Table 6.1 Clustering parameters from previous analyses. 99
Table 6.2 Best-fit parameters from the model fit to P14 data. . . . 111
Table 6.3 P14 correlation coefficients. 115
Table 6.4 Best-fit parameters from the model fit to L19 data. . . . 116
Table 6.5 Best-fit parameters from the model fit to V19 data. . . . 125
Table 7.1 Values of the model parameters employed for the creation of the smooth power spectra. 137
Table 7.2 Best-fit parameters from the model fit to the smooth data set for the baseline and case A presented in this study. 146
Table 7.3 Best-fit parameters from the model fit to the smooth data set for cases B and C presented in this study. . . . 146
Table b.1 Temperature sensitivity for *Planck* HFI frequency channels. 162

PREFACE

Cosmology has become a precision science, and one of the pillars is the study of the Cosmic Microwave Background (CMB). The CMB represents a unique scientific target since its primary anisotropies encode the physics of the early Universe, while secondary anisotropies probe the matter distribution and physical processes at late times.

Beyond these, a variety of extragalactic emissions (*e.g.* the Cosmic Infrared Background (CIB), radio sources) contribute additional microwave and infrared sky signals. These are neither primary nor secondary CMB anisotropies, but they trace the Large–scale structure (LSS), and therefore act both as astrophysical foregrounds in CMB analyses and as cosmological probes in clustering studies.

A deep investigation into the modelling of these extragalactic emissions is the central focus of this thesis.

The work is organised as follows:

FIRST CHAPTER introduces the Λ–Cold–Dark–Matter cosmological model, outlining the composition and background evolution of the homogeneous Universe. The chapter then relaxes homogeneity to present linear perturbations, establishing the theoretical foundations used in the subsequent chapters for modelling CMB anisotropies and LSS.

SECOND CHAPTER presents the CMB. It follows the evolution of photon perturbations up to recombination, derives the temperature anisotropy power spectrum – separating the large–scale and acoustic–peak regimes – and then introduces CMB polarisation, explaining its generation by Compton scattering and its description in harmonic space.

THIRD CHAPTER develops the linear theory to explain structure formation and serves as the bridge between primordial perturbations and present–day matter clustering. It examines the growth of cold dark matter perturbations across horizon entry in radiation– and matter–dominated epochs. This framework leads to the prediction of the linear matter power spectrum.

FOURTH CHAPTER introduces the Halo Model (HM) as the tool used to describe non–linear matter and galaxy clustering. Combining the ingredients of the HM, it derives the prescription for the evaluation of the power spectrum between two generic fields, and applies it to matter and galaxy power spectra.

FIFTH CHAPTER defines all non–CMB signals relevant at small scales (Cosmic Infrared Background (CIB), thermal Sunyaev-Zel'dovich (tSZ), kinetic

Sunyaev-Zel'dovich (kSZ), CIB×tSZ, radio sources, Galactic cirrus) and describes how they project into angular power spectra.

SIXTH CHAPTER builds a HM two–galaxy–population description of the CIB, then fits it to current observations of the CIB power spectra. A persistent tension between the datasets prevents a joint fit, hinting at either data systematics or model incompleteness.

SEVENTH CHAPTER builds a unified, HM–based and template–based framework for all major extragalactic emissions, and generates a multi–frequency dataset mimicking upcoming observations to forecast the expected constraining power of next–generation experiments.

CONVENTIONS

Even though sometimes I will keep the fundamental constants explicit in the equations, in this work I will largely adopt the so-called natural units: $c = \hbar = k_B = 1$.

Useful conversions are

- 1 kg $= 5.610 \times 10^{26}\ GeV\, c^{-2}$
- 1 m $= 5.068 \times 10^{15}\ GeV^{-1} (\hbar c)$
- 1 s $= 1.519 \times 10^{24}\ GeV^{-1} \hbar$
- 1 K $= 8.6 \times 10^{-14}\ GeV$
- 1 Mpc $= 3.08 \times 10^{22}$ m $= 1.56 \times 10^{38}\ GeV^{-1}$

while useful physical constants are

- speed of light in vacuum$\equiv c = 2.9979 \times 10^{8}\ ms^{-1}$
- Planck's constant$\equiv \hbar = 1.05457 \times 10^{-34}\ Js$
- Electron volt$\equiv 1eV = 1.6022 \times 10^{-19}\ J$
- Boltzmann's constant$\equiv k_B = 1.380 \times 10^{-23}\ \ JK^{-1}$
- Newton's constant$\equiv G = 6.674 \times 10^{-11}\ m^3 kg^{-1} s^{-2}$
- Planck mass$\equiv m_{pl} \equiv \sqrt{\hbar c / G} = 1.220 \times 10^{19}\ GeV$
- Reduced Planck mass$\equiv M_{pl} \equiv \sqrt{\hbar c / 8\pi G} = 2.435 \times 10^{18}\ GeV$
- Planck length$\equiv l_{pl} \equiv \sqrt{\hbar G / c^3} = 1.616 \times 10^{-35}\ m$
- Planck time$\equiv t_{pl} \equiv \sqrt{\hbar G / c^5} = 5.391 \times 10^{-44}\ s$

and in particle physics

- Electron mass$\equiv m_e = 0.510\ MeV$
- Proton mass$\equiv m_p = 938\ MeV$
- Neutron mass$\equiv m_n = 939\ MeV$
- Proton–neutron mass difference$\equiv Q = 1.293\ MeV$
- Neutron lifetime$\equiv \tau_n = 879\ s$
- Deuteron mass$\equiv m_D = 1875\ MeV$
- Thomson cross section$\equiv \sigma_T = 0.665 \times 10^{-28}\ m^2$
- Fermi's constant$\equiv G_F = 1.116 \times 10^{-5}\ GeV^{-2}$.

Some useful cosmological parameters, using the values from the best fit of Planck 2018 for the 6-parameter ΛCDM model, are

- CMB temperature today$\equiv T_0 = 2.7255\ K$
- Number density of photons$\equiv n_\gamma = 410.7\ cm^{-3}$
- CMB density$\equiv \Omega_\gamma = 5.38 \times 10^{-5}$

- Number density of baryons$\equiv n_b = 2.515 \times 10^{-7}\ cm^{-3}$
- Baryon density$\equiv \Omega_b = 0.0493$
- Dark matter density$\equiv \Omega_c = 0.265$
- Matter density$\equiv \Omega_m = 0.315$
- Dark energy density$\equiv \Omega_\Lambda = 0.685$
- Hubble expansion rate$\equiv H_0 = 100h\, kms^{-1}Mpc^{-1}$
- Hubble parameter$\equiv h = 0.674$
- Critical density$\equiv \rho_{crit} = 8.532 \times 10^{-30}\ gcm^{-3}$

Some important epochs in the universe are

- BBN$\rightarrow z_{BBN} = 4 \times 10^8\ (3min)$
- Electron–positron annihilation$\rightarrow z_{e^+e^-} = 2 \times 10^9\ (6s)$
- Neutrino decoupling$\rightarrow z_{\nu\, dec} = 6 \times 10^9\ (1s)$
- Matter–radiation equality$\rightarrow z_{eq} = 3402$
- Recombination$\rightarrow z_{rec} = 1270$
- Photon decoupling$\rightarrow z_\star = 1089$
- Last scattering$\rightarrow z_{LSS} = 1089$
- Baryon decoupling$\rightarrow z_{bdec} = 1060$
- Half reionisation$\rightarrow z_{re} = 7.7$

Throughout this thesis, I use the following derivative conventions:

- the overdot, $\dot{x}$, indicates the derivative with respect to cosmic time, t: $\dot{x} = dx/dt$;
- the derivative with respect to the conformal time η is indicated with a *prime*: $x' = dx/d\eta$;
- the partial derivatives with respect to a specific component, x^i, are indicated with a comma: $x_{,i} = \partial x/\partial x^i$.

Part I

FUNDAMENTALS

This part of the thesis reviews the key elements of the standard cosmological model. I introduce the fundamental ingredients describing our Universe and then discuss the linear perturbation theory of its components. This framework sets the starting point for understanding the physics of the Cosmic Microwave Background (CMB) and the formation of Large–scale structure (LSS) in the linear regime.

1 STANDARD COSMOLOGICAL MODEL

In this chapter, I introduce the standard cosmological model, ΛCold Dark Matter (CDM), which assumes a Universe that is homogeneous and isotropic on sufficiently large scales and whose energy budget today is dominated by a cosmological constant Λ and CDM. The model has proven remarkably successful at explaining a broad range of observations, from the anisotropies of the Cosmic Microwave Background (CMB) to Large–scale structure (LSS), Baryon Acoustic Oscillations (BAO), to the Hubble diagram of Type Ia supernovae.

Throughout this chapter, in section 1.1, I first establish the kinematic and geometric framework for a homogeneous and isotropic spacetime, and in section 1.3, I derive the background dynamics governed by the Einstein equations and the Friedmann equations, and introduce the main cosmic components and their equations of state. I then relax homogeneity in section 1.4 to outline linear perturbation theory, the Newtonian–gauge description, and the Boltzmann equation for photons.

The background ΛCDM model must be supplemented by a specification of the initial perturbations. Observations favour nearly Gaussian and adiabatic fluctuations. The most quoted mechanism that explains these properties is cosmic inflation: a brief phase of accelerated expansion in the early Universe that not only sets the initial conditions to study the evolution of perturbations, but also solves causal problems, as for example the horizon problem. In practice, inflation provides the initial conditions for the gravitational potentials Φ, Ψ and the species perturbations at early times; these are then evolved with the coupled Einstein–Boltzmann system developed in this chapter and used in chapter 2 and chapter 3. The initial conditions for linear perturbations are commonly specified by the statistically homogeneous, isotropic curvature perturbation $\mathcal{R}(\mathbf{k})$ generated during inflation. For nearly Gaussian fluctuations, all the information is encoded in the two–point function, $\langle \mathcal{R}(\mathbf{k})\, \mathcal{R}^*(\mathbf{k}')\rangle = (2\pi)^3 \delta_D(\mathbf{k}-\mathbf{k}')\, P_{\mathcal{R}}(k)$. It is convenient to use the dimensionless spectrum $\Delta^2_{\mathcal{R}}(k) \equiv \frac{k^3}{2\pi^2} P_{\mathcal{R}}(k)$, which observations show is well described by a power law,

$$\Delta^2_{\mathcal{R}}(k) = A_s \left(\frac{k}{k_p}\right)^{n_s - 1},$$

with amplitude A_s and spectral index n_s defined at a pivot scale k_p (commonly 0.05 Mpc^{-1} or 0.002 Mpc^{-1}).

1.1 KINEMATICS: COSMOLOGICAL PRINCIPLE, REDSHIFT AND THE HUBBLE LAW

The minimal ΛCDM model rests on the following main elements:

1. The cosmological principle, asserting that on large scales ($\gtrsim$ 100 Mpc) the Universe is statistically homogeneous and isotropic [Maartens, 2011].
2. The second important element is General Relativity (GR), since gravity is described by Einstein's field equations with a cosmological constant Λ.
3. Lastly, the constituents of the Universe: cold (pressureless) dark matter, baryons, photons, (nearly) massless neutrinos, and dark energy modelled as Λ (with $w = -1$).

A convenient parameter set for a proper description of our Universe is $\{\Omega_b, \Omega_c, \Omega_\Lambda, \Omega_K, H_0, A_s, n_s, \tau_{reio}\}$, with Ω_K the spatial curvature (vanishing in a flat Universe), A_s and n_s the amplitude and tilt of primordial scalar perturbations, and τ_{reio} the optical depth to reionisation. Details concerning these three central points are provided in the following sections.

1.1.1 Cosmological principle

The cosmological principle asserts statistical homogeneity and isotropy of the matter distribution on sufficiently large scales (see *e.g.* Maartens, 2011). Homogeneity means spatial averages are independent of position; isotropy means no preferred directions for those averages. Under these assumptions, the most general spacetime metric is the Friedmann–Lemaitre–Robertson–Walker (FLRW) metric (see section 1.2).

1.1.2 Redshift, scale factor and the Hubble law

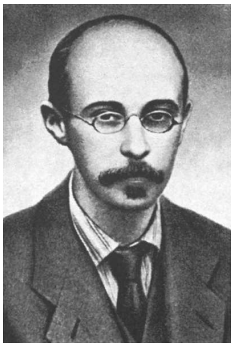

Figure 1.1: Alexander Friedmann.

Another important aspect when dealing with the description of our Universe is the Hubble law, an empirical law that relates the velocities and distances of observed galaxies and provides evidence that the Universe is expanding.

The theoretical idea that we live in an expanding Universe arose in 1922 when Alexander Friedmann derived the so-called Friedmann equations, starting from Einstein's field equations. I analyse them in the following sections. However, the experimental proof arrived in 1929 from Hubble and, independently, from Lemaître. The relation between velocity and distance

can be visualised in the Hubble diagram (see figure 1.2). Since galaxies are characterised by spectra with absorption and emission lines, it is possible to constrain the velocity of a galaxy by measuring the redshift of the spectral lines. For this reason, the Hubble diagram sometimes presents the relation between redshift and distance, rather than velocity and distance.

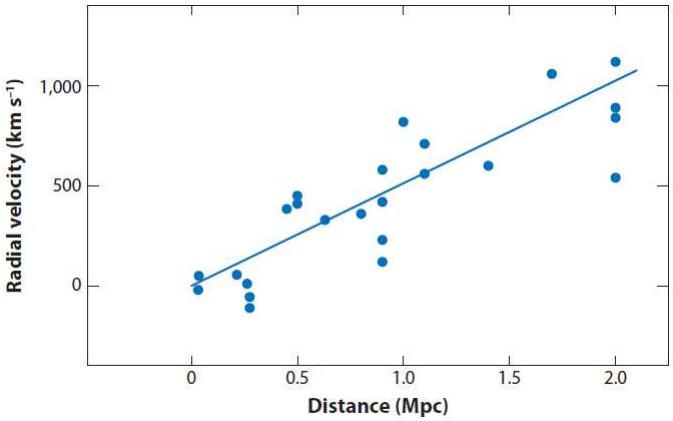

Figure 1.2: Hubble diagram for small distances [Freedman and Madore, 2010].

If the Universe were not expanding, we would expect to observe as many redshifted (receding) galaxies as blueshifted (approaching). The Hubble observations show that most galaxies are redshifted, leading to the conclusion that we live in an expanding Universe.

The relation between the velocity of galaxies and their distance from us (as well as the redshift–distance relation) is provided by the so-called Hubble parameter, $H(t)$. To better understand the Hubble parameter, let us start by introducing the scale factor, $a(t)$, needed when dealing with an expanding Universe since it quantifies how fast the Universe is expanding at a given cosmic time, t. Define the scale factor $a(t)$ such that $a(t_0) = 1$ today. The redshift of light emitted at t_{em} and observed at t_0 is

$$1 + z \equiv \frac{a(t_0)}{a(t_{em})} = \frac{1}{a(t_{em})}.$$

The Hubble parameter,

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)}, \tag{1.1}$$

sets the fractional expansion rate; its present value is the Hubble constant $H_0 = H(t_0)$. For nearby galaxies ($z \ll 1$), the recessional velocity satisfies $v \simeq H_0 d$, *i.e.*, the linear Hubble law (see figure 1.2; [Freedman and Madore, 2010]). The empirical dominance of redshifts (over blueshifts) is clear evidence for cosmic expansion.

Thanks to these two fundamental background elements, I can now move to the following section and explain why the FLRW metric can indeed describe the separation between two events in spacetime for a homogeneous, isotropic, and expanding Universe.

1.2 GEOMETRY: THE FLRW METRIC

The most general metric consistent with homogeneity and isotropy (and signature $-,+,+,+$) is

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-Kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right], \tag{1.2}$$

where $K \in \{-1, 0, +1\}$ encodes constant spatial curvature (open, flat, closed). In conformal time, $d\eta \equiv dt/a$, this becomes

$$ds^2 = a^2(\eta)\left[-d\eta^2 + \gamma_{ij}dx^i dx^j\right] = g_{\mu\nu}dx^\mu dx^\nu, \tag{1.3}$$

with γ_{ij} the metric of a maximally symmetric 3–space of curvature K, and $g_{\mu\nu}$ the rank-2 metric tensor describing the metric of a curved spacetime, with gravity already incorporated in the metric itself, and defined as

$$g_{\mu\nu} = \begin{pmatrix} -a^2(\eta) & 0 & 0 & 0 \\ 0 & a^2(\eta) & 0 & 0 \\ 0 & 0 & a^2(\eta) & 0 \\ 0 & 0 & 0 & a^2(\eta) \end{pmatrix}. \tag{1.4}$$

The metric determines the Christoffel symbols $\Gamma^\alpha_{\mu\nu}$ (via equation 1.5 below), geodesics, and all curvature tensors. In local inertial patches the metric reduces to Minkowski, but globally the curvature induced by matter–energy and expansion cannot be neglected.

The metric is the starting point for understanding how particles move in spacetime. Specifically, a free particle moving in a given spacetime always follows the geodesic, defined as the straightest possible curve between two points. For a flat spacetime, as we are going to see later in this section, the geodesic is a straight line and the orientation of a vector parallel transported along it does not change. On the contrary, when dealing with a curved spacetime, the geodesic is not a straight line anymore and this has implications for the orientation of a parallel transported vector. This can be visualised in figure 1.3. A key quantity for describing the motion of particles within a given spacetime are the Christoffel symbols, detailed in the following.

CHRISTOFFEL SYMBOLS AND GEODESICS. The parallel transport of a vector in a curved spacetime has implications for the computation of derivatives. The different orientation of the vector due to its parallel transport translates into a change of its coordinates, encoded through an additional term when computing derivatives, related to the Christoffel symbols. Furthermore, since this additional term results from a curved spacetime only, the Christoffel symbols have to be related to the metric tensor, $g_{\mu\nu}$. Specifically, the Christoffel symbols are defined through

$$\Gamma^\alpha_{\mu\nu} = \frac{1}{2}g^{\alpha\beta}\left(g_{\beta\mu,\nu} + g_{\nu\beta,\mu} - g_{\mu\nu,\beta}\right), \tag{1.5}$$

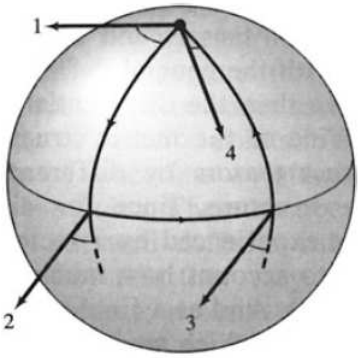

Figure 1.3: Parallel transport of a vector along geodesics of a sphere; after its parallel transport along a closed path, the final orientation of the vector is different from the initial one [Ohanian and Ruffini, 2013].

where $g^{\alpha\beta}$ is the inverse of the metric tensor, $g_{\alpha\beta}$, defined in equation 1.4, and the comma indicates the partial derivative, *i.e.*, $g_{\beta\mu,\nu} = \dfrac{\partial g_{\beta\mu}}{\partial x^{\nu}}$. Thanks to the above equation, it is possible to obtain a direct way to compute the Christoffel symbols for an expanding, homogeneous, and isotropic Universe with flat spatial sections ($K = 0$):

$$\Gamma^0_{00} = 0,$$

$$\Gamma^0_{0i} = \Gamma^0_{i0} = 0,$$

$$\Gamma^0_{ij} = \dot{a}\, a\, \delta_{ij},$$

$$\Gamma^i_{0j} = \Gamma^i_{j0} = \frac{\dot{a}}{a}\, \delta_{ij},$$

$$\Gamma^i_{jk} = 0.$$

It is important to notice that for the specific case of a flat Minkowski metric all Christoffel symbols are identically zero.

Having defined the Christoffel symbols, the geodesic equation, which describes the motion of particles within spacetime, can be obtained starting from the variational principle, which asserts that the action is stationary along the geodesic of the spacetime, $\delta \int ds = 0$. Starting from the variational principle and by making use of the Euler–Lagrange equation resulting from the variational principle itself, it is possible to obtain the generic formulation of the geodesic equation

$$\frac{d^2 x^{\mu}}{d\lambda^2} + \Gamma^{\mu}_{\alpha\beta} \frac{dx^{\alpha}}{d\lambda} \frac{dx^{\beta}}{d\lambda} = 0, \tag{1.6}$$

where λ is an affine parameter which has no physical meaning, but is used to label the trajectory in the four-dimensional spacetime. Instead of λ it is also possible to use the proper time, $d\tau$, which is a natural parametrisation

with a specific physical meaning. Furthermore, it is also possible to define λ in terms of the comoving momentum, P^α, as

$$P^\alpha = \frac{dx^\alpha}{d\lambda}. \tag{1.7}$$

As for all the comoving quantities, the comoving momentum refers to the momentum that the observer measures and is fixed; these quantities have to be distinguished from the physical quantities, which grow with the expansion. Focusing on the momentum, the physical and the comoving momenta are related through

$$p^\mu = a(t)\, P^\mu. \tag{1.8}$$

Since P^μ is the comoving four-momentum, its relation to the physical one is:

$$P^\mu = (P^0, P^i) = (E, \frac{p^i}{a}), \tag{1.9}$$

where E is the energy, p^i is the physical three-momentum, and a is the scale factor.
Starting from the geodesic equation, we can clearly see why a free particle moving within a flat spacetime follows a straight line. As already said, the Christoffel symbols for a flat metric are all zero, therefore the only remaining term of the geodesic equation is the first one

$$\frac{d^2x^\mu}{d\lambda^2} = 0. \tag{1.10}$$

The above equation represents the equation of motion of a free particle in a flat spacetime and indeed such a particle follows a straight trajectory.

Before going on with the final expression of the geodesic equation for the expanding, homogeneous, and isotropic Universe, a rewrite of the geodesic equation in a different formulation, where the comoving momentum explicitly appears, is needed.
To end up with the required formulation I make use of the relation between the affine parameter λ and the comoving momentum (see equation 1.7), to obtain

$$E\frac{dP^\mu}{dt} + \Gamma^\mu_{\alpha\beta} P^\alpha P^\beta = 0. \tag{1.11}$$

Written in these terms, we can write the two general geodesic equations for the temporal and the spatial components, separately. For the $\mu = 0$ component, the only non-zero Christoffel symbol with upper index o is $\Gamma^0_{ij} = \dot{a}(t)a(t)\delta_{ij}$; while for the $\mu = i$ component, the only non-zero Christoffel symbol with upper index i is $\Gamma^i_{j0} = \frac{\dot{a}(t)}{a(t)}\delta_{ij}$. Adding all pieces together we obtain

$$\frac{dE}{dt} + \frac{\dot{a}}{a}E = 0, \tag{1.12}$$

$$\frac{dp^i}{dt} + \frac{\dot{a}}{a}p^i = 0. \tag{1.13}$$

It is important to underline that both geodesic equations are based on the Christoffel symbols, which are computed from the metric of the given spacetime. As I am going to show at the end of this chapter, the introduction of linear perturbations in the spacetime requires a different metric, therefore different Christoffel symbols associated with the perturbed metric have to be computed, and so we expect to end up with a different geodesic equation, or, in other words, a different motion of particles within the perturbed spacetime.

1.3 DYNAMICS: EINSTEIN EQUATIONS, FRIEDMANN EQUATIONS AND DISTANCES

By looking at the geodesic equation, it is possible to appreciate the connection between the physical description and the geometry of spacetime. This is because it has been derived starting from the Christoffel symbols, which are strictly related to the metric. The Einstein equations make this connection even more explicit, since they relate the metric to the constituents of the Universe. The general expression is

$$G_{\mu\nu} = 8\pi G\, T_{\mu\nu}, \tag{1.14}$$

with $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ being the Einstein tensor. In its definition $R_{\mu\nu}$ and R are the Ricci tensor and the Ricci scalar, respectively. Both of these objects are defined in terms of Christoffel symbols and therefore, they depend on the metric, as

$$R_{\mu\nu} = \Gamma^{\alpha}_{\mu\nu,\alpha} - \Gamma^{\alpha}_{\mu\alpha,\nu} + \Gamma^{\gamma}_{\mu\nu}\Gamma^{\alpha}_{\gamma\alpha} - \Gamma^{\alpha}_{\nu\gamma}\Gamma^{\gamma}_{\mu\alpha}, \tag{1.15}$$

$$R = g^{\mu\nu}R_{\mu\nu}. \tag{1.16}$$

Both the Ricci tensor and the Ricci scalar are completely determined once the Christoffel symbols are known and, since they have already been computed, their evaluation is straightforward and leads to

$$R_{00} = -3\frac{\ddot{a}(t)}{a(t)},$$

$$R_{i0} = R_{0i} = 0,$$

$$R_{ij} = (\ddot{a}(t)a(t) + 2\dot{a}(t)^2)\delta_{ij},$$

$$R = 6\left[\frac{\ddot{a}(t)}{a(t)} + \left(\frac{\dot{a}(t)}{a(t)}\right)^2\right].$$

For a perfect fluid in an FLRW spacetime, the energy–momentum tensor is

$$T_{\mu\nu} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix}, \tag{1.17}$$

with ρ the energy density, and P the pressure of all the species in the Universe, including dark energy.

In a Universe described by the FLRW metric, the computation of the time and space components of the Einstein equations leads to the following two relations:

- 00 component of the Einstein equations

$$G_{00} = 8\pi G T_{00}, \tag{1.18}$$

where $G_{00} = 3\left(\frac{\dot{a}(t)}{a(t)}\right)^2$ and $T_{00} = -\rho$. Making this explicit gives

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3}\rho. \tag{1.19}$$

- ij component of the Einstein equations

$$G_{ij} = 8\pi G T_{ij}, \tag{1.20}$$

where $G_{ij} = -(2\ddot{a}(t)a(t) + \dot{a}(t)^2)\delta_{ij}$ and $T_{ij} = P\,\delta_{ij}$, so that the second relation becomes

$$\frac{1}{2}\left(\frac{\dot{a}(t)}{a(t)}\right)^2 + \frac{\ddot{a}(t)}{a(t)} = -4\pi G P. \tag{1.21}$$

The $0i$ component is zero.

The time component of the Einstein equations represents the first Friedmann equation, while in order to obtain the usual form of the second Friedmann equation we have to combine the two relations. In their final formulation, the Friedmann equations for the scale factor are

$$H^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2} + \frac{\Lambda}{3}, \tag{1.22}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3}, \tag{1.23}$$

where H is the Hubble parameter defined in equation 1.1. These two equations describe the expansion of the Universe and its evolution with time.

Specifically, the first equation gives the expansion rate (or in other words, the "speed" of expansion), while the second provides information about the acceleration or deceleration of that expansion.
To conclude, by imposing energy–momentum conservation, we end up with another important law regulating the Universe, the continuity equation:

$$\dot{\rho} + 3H(\rho + P) = 0. \tag{1.24}$$

CRITICAL DENSITY AND Ω PARAMETERS. The critical density is the reference energy density that yields a spatially flat Universe in the Friedmann equation. At any epoch it is

$$\rho_c(z) \equiv \frac{3H^2(z)}{8\pi G}, \tag{1.25}$$

and in particular today $\rho_{c,0} \equiv \rho_c(z{=}0) = 3H_0^2/(8\pi G)$. It is customary to work with dimensionless density parameters (or density fractions) defined with respect to $\rho_{c,0}$:

$$\Omega_i \equiv \frac{\rho_{i,0}}{\rho_{c,0}}, \qquad \Omega_\Lambda \equiv \frac{\Lambda}{3H_0^2}, \qquad \Omega_K \equiv -\frac{K}{a_0^2 H_0^2}, \tag{1.26}$$

where $i \in \{\mathrm{b}, \mathrm{c}, \mathrm{m}, \mathrm{r}, \nu, \ldots\}$ refers to baryons, cold dark matter, total matter, radiation, neutrinos, etc., $K \in \{-1, 0, +1\}$ is the constant spatial curvature parameter, and $a_0 \equiv a(z{=}0) = 1$.

The first Friedmann equation evaluated today gives the closure relation

$$\Omega_\mathrm{r} + \Omega_\mathrm{m} + \Omega_\Lambda + \Omega_K = 1, \tag{1.27}$$

and, more generally, at any redshift z one may define time-dependent density parameters

$$\Omega_i(z) \equiv \frac{\rho_i(z)}{\rho_c(z)} = \frac{\Omega_i\,(1+z)^{3(1+w_i)}}{E^2(z)}, \qquad E(z) \equiv \frac{H(z)}{H_0}, \tag{1.28}$$

where $w_i \equiv P_i/\rho_i$ is the (assumed constant) Equation of State (EoS) parameter of component i ($w = 0$ for non-relativistic matter; $w = 1/3$ for radiation; $w = -1$ for a cosmological constant). In the minimal ΛCDM model the Hubble rate can be written as

$$H(z) = H_0 \sqrt{\Omega_\mathrm{r}(1+z)^4 + \Omega_\mathrm{m}(1+z)^3 + \Omega_K(1+z)^2 + \Omega_\Lambda}, \tag{1.29}$$

with $\Omega_m = \Omega_b + \Omega_c$ and Ω_r including photons and (relativistic) neutrinos.[1]

It is often convenient to express curvature in terms of an effective density:

$$\Omega_K(z) = -\frac{K}{a^2 H^2} = \frac{\Omega_K(1+z)^2}{E^2(z)}, \tag{1.30}$$

1 For practical work it is useful to split $\Omega_r = \Omega_\gamma + \Omega_\nu$. The photon density today is fixed by the CMB temperature $T_{\gamma,0} = 2.7255\,K$ [Fixsen, 2009], giving $\Omega_\gamma h^2 \simeq 2.47 \times 10^{-5} (T_{\gamma,0}/2.7255\,K)^4$. For standard-model neutrinos that are still relativistic today, $\Omega_\nu \simeq 0.2271\, N_{eff}\, \Omega_\gamma$, with $N_{eff} \simeq 3.044$. Here $h \equiv H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$.

and to associate a curvature radius $R_{curv} \equiv \frac{1}{H_0\sqrt{|\Omega_K|}}$ (in units with $c = 1$), which characterises the comoving scale at which spatial curvature becomes manifest if $\Omega_K \neq 0$.

> **Numerical scales.** For reference, the present-day critical density is $\rho_{c,0} \simeq 1.879 \times 10^{-26}\, h^2\ \mathrm{kg\,m^{-3}} \simeq 1.054 \times 10^{4}\, h^2\ \mathrm{eV\,cm^{-3}}$, so that $\Omega_i h^2$ directly sets the physical density of component i.

COSMOLOGICAL DISTANCE MEASURES In a Robertson–Walker spacetime the line element can be written as $ds^2 = -dt^2 + a^2(t)\{d\chi^2 + S_K^2(\chi)\, d\Omega^2\}$, where χ is the comoving radial coordinate and $S_K(\chi)$ encodes the constant spatial curvature.[2] The line-of-sight comoving distance to redshift z is

$$\chi(z) \;=\; \int_0^z \frac{dz'}{H(z')}\,, \tag{1.31}$$

and the transverse comoving distance (often denoted D_M) is

$$D_M(z) \;=\; S_K\!\big(\chi(z)\big), \qquad S_K(\chi) = \begin{cases} \dfrac{\sin\!\left(\sqrt{K}\,\chi\right)}{\sqrt{K}}, & K > 0 \ \ \text{(closed)},\\[2mm] \chi, & K = 0 \ \ \text{(flat)},\\[2mm] \dfrac{\sinh\!\left(\sqrt{-K}\,\chi\right)}{\sqrt{-K}}, & K < 0 \ \ \text{(open)}, \end{cases} \tag{1.32}$$

where $K \in \{+1, 0, -1\}$ in the usual normalisation. It is often convenient to re-express equation 1.32 in terms of the present-day curvature density $\Omega_K = -K/(a_0^2 H_0^2)$:

$$D_M(z) \;=\; \begin{cases} \dfrac{1}{H_0\sqrt{\Omega_K}}\,\sinh\!\left(H_0\sqrt{\Omega_K}\,\chi\right), & \Omega_K > 0,\\[2mm] \chi, & \Omega_K = 0,\\[2mm] \dfrac{1}{H_0\sqrt{|\Omega_K|}}\,\sin\!\left(H_0\sqrt{|\Omega_K|}\,\chi\right), & \Omega_K < 0, \end{cases} \tag{1.33}$$

which makes the dependence on curvature explicit.

The two most commonly used observables are the angular–diameter distance and the luminosity distance, defined by

$$D_A(z) \;=\; \frac{D_M(z)}{1+z}, \qquad D_L(z) \;=\; (1+z)\, D_M(z), \tag{1.34}$$

and related by Etherington's reciprocity relation,

$$D_L(z) \;=\; (1+z)^2\, D_A(z), \tag{1.35}$$

which holds in any metric theory of gravity with photon number conservation and photons travelling on null geodesics [Hogg, 1999; Peacock, 1999].

2 Throughout this section I set $c = 1$.

Useful derivatives follow directly from the definitions:

$$\frac{d\chi}{dz} = \frac{1}{H(z)}, \qquad \frac{dD_M}{dz} = \frac{1}{H(z)} \times \begin{cases} \cos\left(\sqrt{K}\chi\right), & K > 0, \\ 1, & K = 0, \\ \cosh\left(\sqrt{-K}\chi\right), & K < 0, \end{cases} \tag{1.36}$$

and hence $\frac{dD_A}{dz} = \frac{1}{1+z}\frac{dD_M}{dz} - \frac{D_M}{(1+z)^2}$. At low redshift, expanding $H(z)$ around $z = 0$ yields the series

$$D_L(z) \simeq \frac{z}{H_0}\left[1 + \frac{1}{2}(1 - q_0)z - \frac{1}{6}(1 - q_0 - 3q_0^2)z^2 + \cdots\right], \tag{1.37}$$

where q_0 is the present-day deceleration parameter.
The **comoving volume element** per redshift and solid angle is

$$\frac{dV_c}{dz\, d\Omega} = \frac{D_M^2(z)}{H(z)}, \tag{1.38}$$

which is simply the volume measure encoding for Universe expansion.

1.4 LINEAR PERTURBATIONS: NEWTONIAN GAUGE AND THE BOLTZMANN EQUATION

The aim of this section is to appreciate the implications of the presence of perturbations, without going into detail with calculations.
At linear order, perturbations to the metric and to the energy–momentum tensor can be decomposed into scalar, vector, and tensor parts. In Fourier space these three classes evolve independently. Scalar modes are sourced by density and pressure perturbations and by velocity potentials; in Newtonian gauge they are described by the two gravitational potentials Φ and Ψ that appear in the perturbed metric. Vector modes represent vorticity; in standard cosmology they decay with the expansion and are usually negligible. Tensor modes are transverse–traceless perturbations of the metric, *i.e.* gravitational waves. The introduction of perturbations implies a spatial dependence of the metric. The following description is done within the so-called Newtonian gauge framework, where the spatial dependence of the metric is provided by

the presence of two gravitational potentials, $\Psi(\vec{x},t)$ and $\Phi(\vec{x},t)$. The metric tensor for the perturbed Universe becomes

$$g_{\mu\nu} = \begin{pmatrix} -1-2\Psi & 0 & 0 & 0 \\ 0 & a^2(1+2\Phi) & 0 & 0 \\ 0 & 0 & a^2(1+2\Phi) & 0 \\ 0 & 0 & 0 & a^2(1+2\Phi) \end{pmatrix}, \tag{1.39}$$

where the time and space dependence of the potentials is implicit. These gravitational potentials are of the order of $10^{-4/-6}$[3], so in order to describe the Universe through this deviation from the FLRW metric it is possible to work in the framework of first-order perturbation theory. In the absence of anisotropic stress from matter and radiation, GR implies $\Phi = \Psi$. At linear order, spatial gradients and time derivatives of Φ, Ψ lens and redshift photons, and source structure growth.

GEODESICS AND MOMENTA Perturbations modify the timelike and space-like components of geodesics. Having the new expression of the metric, the Christoffel symbols associated with a perturbed Universe can be computed by making use of equation 1.5. As already stated in the previous section, different Christoffel symbols also imply a different geodesic equation. For a particle with comoving momentum $P^\mu = (E, \mathbf{p}/a)$

$$P^\mu = \left((1-\Psi)E; \frac{1-\Phi}{a}p^i\right). \tag{1.40}$$

The spatial dependence of the metric and the presence of the two gravitational potentials complicate the calculations to end up with the Einstein equations, even if their general structure is left unchanged.
Left-hand side of the Einstein equation (time-time component)

$$G^0_0 = -6H\Phi_{,0} + 6H^2\Psi - 2\frac{k^2}{a^2}\Phi. \tag{1.41}$$

Left-hand side of the Einstein equation (space component)

$$P^i_j G^i_j = \frac{2}{3}\frac{k^2}{a^2}(\Phi+\Psi), \tag{1.42}$$

where:

- k appears when moving from real space to Fourier space, since $\frac{\partial}{\partial x^i} \to ik^i$, and represents the Fourier modes.

3 During recombination the relation

$$\frac{\Delta T}{T} \simeq \frac{1}{3}\Phi,$$

holds. The observed temperature anisotropies are of order $\Delta T/T \sim 10^{-5}$, which leads to $\Phi \sim 3 \times 10^{-5}$.

- P^i_j is the projection operator, defined as $P_{ij} = \hat{k}_i\hat{k}_j - \frac{1}{3}\delta_{ij}$, and it is introduced in order to absorb all terms multiplied by δ_{ij} and simplify a lot the calculations.

I only compute the left-hand side of the Einstein equation since the evaluation of the right-hand side requires the Boltzmann equation, detailed in the following.

At first order

$$\frac{dx^i}{d\eta} = \frac{p}{E}\,\hat{p}^i\,(1-\Phi+\Psi), \qquad \frac{dp}{d\eta} = -\left(\mathcal{H}+\Phi'\right)p - E\,\hat{p}^i\,\partial_i\Psi, \tag{1.43}$$

where $\mathcal{H} \equiv a'/a$ and a prime denotes $d/d\eta$. The first correction encodes lensing; the second shows cosmological and gravitational redshifts.

EINSTEIN CONSTRAINTS In Fourier space ($\partial_i \to ik_i$), the linearised 00 and traceless–ij Einstein equations read

$$k^2\Phi + 3\mathcal{H}\left(\Phi' + \mathcal{H}\Psi\right) = 4\pi G a^2\,\delta\rho, \tag{1.44}$$

$$k^2\left(\Phi - \Psi\right) = 12\pi G a^2\left(\rho + P\right)\sigma, \tag{1.45}$$

with $\delta\rho$ the total density perturbation and σ the total anisotropic stress (zero if only perfect fluids are present).

BOLTZMANN EQUATION AND COLLISIONS In order to describe the components of our Universe, we are interested in their macroscopic behaviour, so a statistical description is required. The mathematical tool that allows for this macroscopic description is the distribution function, $f(\vec{x},\vec{p},t)$, implicitly defined through the following relation:

$$N(\vec{x},\vec{p},t) = f(\vec{x},\vec{p},t)\frac{(\Delta x)^3(\Delta p)^3}{(2\pi)^3}, \tag{1.46}$$

where $N(\vec{x},\vec{p},t)$ is the number of particles within the volume $(\Delta x)^3(\Delta p)^3$ in the 6-dimensional phase space.

Therefore, the distribution function allows one to infer the macroscopic behaviour of each species, and the equation that evaluates the distribution function itself is the Boltzmann equation, indeed. In an exactly homogeneous and isotropic FLRW background the phase–space distribution is angle- and space-independent: $f = f(p,t)$. Hence $\partial f/\partial x^i = 0$ and $\partial f/\partial\hat{\mathbf{p}} = 0$. We will decompose $f = f + \delta f$, where the perturbation acquires spatial and angular dependence, $\delta f = \delta f(\mathbf{x}, p, \hat{\mathbf{p}}, t)$.

The most general formulation of the Boltzmann equation encodes for interactions among particles. In order to account for these interactions we have to introduce the collision term, $C[f]$, which is a functional of the distribution function itself. The Boltzmann equation in the case of collisions reads

$$\frac{df}{dt} = C[f], \tag{1.47}$$

so the total derivative with respect to time is equal to the collision term, where $\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{dx^i}{dt}\frac{\partial f}{\partial x^i} + \frac{d\vec{p}}{dt}\frac{\partial f}{\partial \vec{p}}$. In order to end up with the final expression of the Boltzmann equation for a homogeneous, isotropic, and expanding Universe the following two elements are needed:

- the expression of the total derivative of the distribution function;
- the expression of the collision term.

The total time derivative of the physical momentum $\vec{p}$ splits in two parts; the derivative of the magnitude of the momentum, and the derivative of the direction of the momentum. Since we do not expect any dependence on the direction, the latter is zero and the total derivative reduces to

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{dx^i}{dt}\frac{\partial f}{\partial x^i} + \frac{dp}{dt}\frac{\partial f}{\partial p}. \tag{1.48}$$

In the background limit, $f = f(p,t)$ so $\partial f/\partial x^i = 0$ and the operator reduces to

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{dp}{dt}\frac{\partial f}{\partial p}.$$

The missing terms in order to end up with the final expression of the left-hand side of the Boltzmann equation are $\frac{dx^i}{dt}$ –which can be inferred starting from equation 1.7– and $\frac{dp}{dt}$ –which can be evaluated starting from the geodesic equation–

$$\frac{dx^i}{dt} = \frac{p}{E}\frac{d\hat{p}^i}{a}, \tag{1.49}$$

$$\frac{dp}{dt} = -H(t)p. \tag{1.50}$$

In order to define the collision term, a proper characterisation of the scattering process is needed. Specifically, I work assuming a two-body elastic scattering, where both the energy and the momentum are conserved. Furthermore, to end up with the required characterisation, one has to define also the reaction rate as the difference between the backward and forward reactions. For a general reaction of the type

$$(1)_p + (2)_q \leftrightarrow (3)_{p'} + (4)_{q'},$$

for which we are interested in the distribution function of particle 1, $C[f_1(\vec{p_1})]$, the reaction rate, neglecting quantum-mechanical effects, is proportional to $(f_3 f_4 - f_1 f_2)$.
Enforcing the energy and the momentum conservation, the general formula for the collision term is

$$C[f_1(\vec{p_1})] = \frac{1}{2E_1}\int \frac{d\vec{q}}{(2\pi)^3 2E_2}\int \frac{d\vec{p'}}{(2\pi)^3 2E_3}\int \frac{d\vec{q'}}{(2\pi)^3 2E_4}|M|(2\pi)^4 \\ \delta(E_1+E_2-E_3-E_4)\delta^3(\vec{p}+\vec{q}-\vec{p'}-\vec{q'})(f_3 f_4 - f_1 f_2), \tag{1.51}$$

and we end up with the Boltzmann equation for a homogeneous, isotropic, and expanding Universe as

$$\frac{\partial f}{\partial t} + \frac{p}{E}\frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} - H(t)p\frac{\partial f}{\partial p} = C[f]. \tag{1.52}$$

Now it is possible to write the Boltzmann equation for a perturbed Universe, recalling that we are working at first order in perturbation theory, so all second-order terms are already neglected in the following equation:

$$\frac{\partial f}{\partial t} + \frac{p}{E}\frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} - \frac{\partial f}{\partial p}[(H+\dot{\Phi})p + \frac{E}{a}\hat{p}^i\Psi_{,i}] = C[f]. \tag{1.53}$$

The key point in order to describe the observed CMB anisotropies is to encode for perturbations of the distribution function associated to each species.
For the unperturbed Universe case, the distribution function is only a function of the time and the magnitude of the momentum; when perturbations are present, $\delta f = \delta f(\mathbf{x}, p, \hat{\mathbf{p}}, t)$ acquires positional and angular dependence.

In order to describe the CMB anisotropies we have to solve the Boltzmann equation for photons with a non-zero collision term. However, there also are many interactions among all the components of the Universe – as shown in figure 1.4 –, including the metric itself. Specifically,

- photons are affected by gravity – the metric – and they are coupled to electrons via Compton scattering;
- electrons are affected by gravity and they are coupled to protons via Coulomb scattering;
- protons are affected by gravity;
- the metric is influenced by all the components, including neutrinos and dark matter.

From this general picture it is clear that, if the final goal is the evaluation of the distribution function of a single species, it is required to solve the Boltzmann and Einstein equations associated to all the components.

Focusing on the Boltzmann equation for photons, in order to allow for perturbations in their distribution function, I introduce anisotropies to the background temperature. In other words, the overall photon distribution can be seen as the summation of a zeroth-order term, representing the background and homogeneous temperature, and a first-order term, representing the linear perturbation around the background temperature itself, in a way such that we move from T to $T + \delta T$.
The unperturbed distribution function of photons, being bosons, at equilibrium follows the Bose-Einstein distribution function

$$f(\vec{x}, \vec{p}, t) = \frac{1}{e^{\frac{p}{T}} - 1}, \tag{1.54}$$

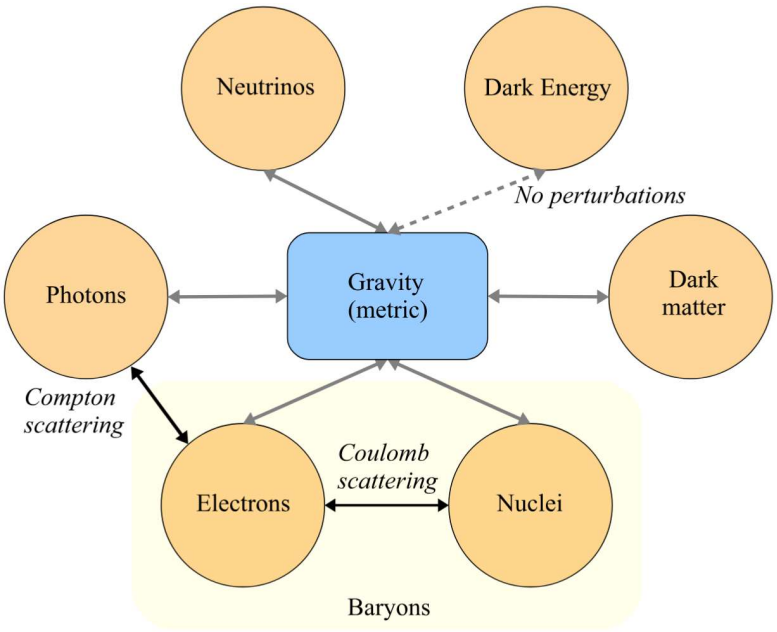

Figure 1.4: Interactions among different components of the Universe [Dodelson and Schmidt, 2020].

and introducing a perturbation in the temperature,

$$T + \delta T = T\left(1 + \frac{\delta T}{T}\right),$$

$$\frac{\delta T}{T} = \theta(\vec{x}, \hat{p}, t), \tag{1.55}$$

where $\theta(\vec{x}, \hat{p}, t)$ is the fractional temperature perturbation and represents the inhomogeneous part of the photon distribution. So, once we introduce this fractional temperature variation, the overall distribution function of photons changes maintaining the same shape, but with a temperature that depends on the direction. We can indeed expand the distribution function up to first order distinguishing the unperturbed distribution function and the perturbed one.

The starting point in order to appreciate the couplings between photons and all the other species is the Boltzmann equation for photons in a perturbed Universe. Recalling that photons are massless particles, so we can replace $E = p$ in equation 1.53, obtaining

$$\frac{\partial f}{\partial t} + \frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} - p\frac{\partial f}{\partial p}\left[(H + \dot{\Phi}) + \frac{\hat{p}^i}{a}\Psi_{,i}\right] = C[f]. \tag{1.56}$$

The Boltzmann equation can be split as:

- a zeroth-order Boltzmann equation, which encodes for the homogeneous and isotropic part of the photon distribution and so presents no collision term.

- a first-order Boltzmann equation, where now we are looking for deviations of the photon distribution from its zeroth-order temperature. At first-order perturbation we expect a non-zero scattering term, and such a scattering term is the source of the couplings between the different components of the Universe.

With regard to photons, they are affected by Compton scattering, that is the scattering process between a photon and an electron which can be regarded as a nearly elastic scattering process.
In the evaluation of the collision term needed in order to solve the first-order Boltzmann equation for photons, I apply the general definition provided by equation 1.51, ending up with the following expression

$$C[f(\vec{p})] = -n_e \sigma_T p \frac{\partial f^{(0)}}{\partial p}[\theta_0 - \theta(\hat{p}) + \hat{p}\vec{v}_b], \tag{1.57}$$

where:

- n_e is the electron number density;
- σ_T is the Thomson scattering[4] cross section;
- $f^{(0)}$ is the zeroth-order distribution function of photons –so, the unperturbed one–;
- θ_0 is the radiation monopole, defined as an average over all the possible directions

$$\theta_0(\vec{x}, t) = \frac{1}{4\pi} \int d\Omega' \theta(\hat{p}', \vec{x}, t); \tag{1.58}$$

- $\hat{p}\vec{v}_b$ is the radiation dipole term and v_b represents the velocity of electrons, which is assumed to be equal to the baryonic one due to the tight coupling between these two components via Coulomb scattering.

In order to study perturbations to the photon distribution we need to solve the Boltzmann equation also for baryons, which are influenced and influence gravity; while gravity is affected by all the other species.
We are indeed supposed to solve the following set of Boltzmann equations for both matter and radiation (assuming a negligible contribution coming from dark energy since at this level we are mainly interested in the early Universe and non-interacting neutrinos, but still relativistic).
Photons:

$$\theta' + ik\mu\theta = -\Phi' - ik\mu\Psi - \tau'[\theta_0 - \theta + \mu v_b].$$

4 Thomson scattering is a low-energy elastic scattering process between electromagnetic radiation and a free electron, in the regime where $h\nu \ll m_e c^2$, such that the photon energy remains unchanged in the electron rest frame.

Cold dark matter

$$\delta' + iku_c = -3\Phi',$$

$$u_c' + \frac{a'}{a}u_c = -ik\Psi.$$

Baryons

$$\delta' + iku_b = -3\Phi',$$

$$u_b' + \frac{a'}{a}u_b = -ik\Psi + \frac{\tau}{R}[3i\theta_1 + u_b].$$

Neutrinos

$$\mathcal{N}' + ik\mu\mathcal{N} = -\Phi' - ik\mu\Psi.$$

This set of Boltzmann equations contains quantities defined in the Fourier space, where:

- the primed quantities are derived with respect to the conformal time. The conformal time is a time variable defined as $d\eta = \frac{dt}{a}$;
- θ is the fractional temperature variation;
- k is the photon wave vector, pointing in the direction along which the temperature is changing;
- μ is a quantity defined as $\mu = \frac{k^i p^i}{k}$ and it describes the direction of propagation of photons. Working under the assumption of irrotational velocity of matter, during the calculations it is possible to make use of the fact that $\hat{p}\vec{v_b} = \mu v_b$;
- τ' is defined through the relation: $\tau' = -n_e\sigma_T a$ and is the derivative of the optical depth, τ;
- δ' is the derivative of $\delta = \frac{\delta n_b}{n_b}$, so the ratio between the perturbation in the baryonic distribution and the unperturbed baryonic distribution;
- u_c is the velocity of cold dark matter;
- u_b is the baryon velocity;
- R is the ratio between baryon and radiation densities;
- θ_1 is the dipole of radiation. The radiation dipole appears in the Boltzmann equation for baryons, making the coupling between photons and baryons explicit;
- $\mathcal{N}$ is the linear order expansion of the neutrinos temperature.

Since all the constituents move within the spacetime, their motions follow the geodesic equation –governed by the metric– and the metric is determined by the constituents themselves. To solve the full set of the Boltzmann equations we also need two equations for the gravitational potentials. These two equations can be derived by evaluating the time and the space components of the Einstein equation.

Gravitational potentials

$$k^2\Phi + 3\frac{a'}{a}\left(\Phi' - \frac{a'}{a}\Psi\right) = 4\pi G a^2(\rho_c\delta_c + \rho_b\delta_b + 4\rho_\gamma\theta_0 + 4\rho_\nu\mathcal{N}_0), \quad (1.59)$$

$$k^2(\Phi + \Psi) = -32\pi G(\theta_2\rho_\gamma + \mathcal{N}_2\rho_\nu), \quad (1.60)$$

where θ_2 and $\mathcal{N}_2$ are the quadrupoles of the photon and the neutrino backgrounds.

We now have the eight equations needed to properly describe an expanding and perturbed Universe. Having specified the metric perturbations and their constraints, we now turn to the dynamics of species perturbations via their phase–space evolution. Let $f(\mathbf{x}, \mathbf{p}, \eta)$ be the photon phase–space distribution. At linear order,

$$\frac{\partial f}{\partial \eta} + \frac{p}{E}\hat{p}^i\frac{\partial f}{\partial x^i} - \frac{\partial f}{\partial p}\left[(\mathcal{H} + \Phi')p + E\,\hat{p}^i ik_i\Psi\right] = C[f], \quad (1.61)$$

with $C[f]$ the collision term. Perturbing the blackbody as $T \to T(1+\theta)$ and expanding in multipoles yields the familiar photon hierarchy [Hu and Dodelson, 2002; Ma and Bertschinger, 1995]. Thomson scattering off free electrons of density n_e gives (schematically)

$$C[f] \propto -\,a\,n_e\sigma_T\,p\,\frac{\partial f^{(0)}}{\partial p}\left[\theta(\hat{\mathbf{p}}) - \theta_0 - \hat{\mathbf{p}}\cdot\mathbf{v}_b\right], \quad (1.62)$$

where σ_T is the Thomson cross–section, θ_0 the monopole (directional average), and $\mathbf{v}_b$ the baryon velocity. The tight–coupling era (pre–recombination) tightly links photons and baryons, producing acoustic oscillations in the photon–baryon fluid that seed the CMB acoustic peaks.

FLUID EQUATIONS FOR MATTER. Although our immediate target is the photon distribution, the gravitational potentials (Φ, Ψ) that lens and redshift photons are sourced by the total stress–energy. CDM dominates the clustering matter density and thus drives the potentials, while baryons are dynamically coupled to photons before recombination. For this reason the photon hierarchy must be evolved together with the CDM and baryon continuity/Euler equations to obtain a closed, self–consistent system. For CDM and baryons, linear continuity and Euler equations (in Fourier space) read

$$\delta_c' + ik\,v_c = -3\Phi', \qquad v_c' + \mathcal{H}v_c = -ik\,\Psi, \quad (1.63)$$

$$\delta_b' + ik\,v_b = -3\Phi', \qquad v_b' + \mathcal{H}v_b = -ik\,\Psi + \dot{\tau}\,(3i\theta_1 + v_b), \quad (1.64)$$

where $\dot{\tau} \equiv -a\, n_e \sigma_T$, and θ_1 is the photon dipole. Together with equations (1.44)–(1.45) and the photon hierarchy, these form a closed linear system solved by Einstein–Boltzmann codes (*e.g.* CLASS, CAMB) [Blas et al., 2011; Lewis et al., 2000].

COSMIC MICROWAVE BACKGROUND

In this chapter I describe the Cosmic Microwave Background (CMB), starting from the temperature anisotropy power spectrum, and then moving to the polarisation field of the CMB.

Regarding the anisotropy power spectrum, we have to distinguish between what happens to the perturbations at the last scattering surface, and how these perturbations propagate from the last scattering surface to us, since what we actually observe are photons that have propagated through the entire expanding Universe. This is detailed in section 2.1.
In evaluating these two parts, I also split the description of photon behaviour into the large–scale anisotropy region (section 2.1.1) and the acoustic–peak region (section 2.1.2).

Once I obtain the theoretical predictions for the CMB power spectrum in section 2.2, and note a very good agreement with observations, in section 2.3, I move on to describing the polarised CMB, to explain the Compton–scattering mechanism that explains the origin of polarisation and its formal description.

Review: Key epochs of pre-recombination cosmology

Tightly coupled photon–baryon plasma Photons, electrons and protons behave as a single fluid due to rapid Thomson and Coulomb scattering ($\tau \gg 1$). Acoustic waves propagate with sound speed $c_s \simeq c/\sqrt{3(1+R)}$ (with $R \equiv 3\rho_b/4\rho_\gamma$), while diffusion damping erases small-scale anisotropies. Temperatures are $\gtrsim$eV ($\gtrsim 10^4$ K); redshifts $z \gtrsim 1100$.

Matter–radiation equality At $z_{eq} \simeq 3400$ ($T_\gamma \sim 0.75\,\mathrm{eV} \approx 8700\,\mathrm{K}$), the Universe transitions from radiation to matter domination. Gravitational potentials stop decaying, setting the overall envelope of the acoustic spectrum and sourcing the early Integrated Sachs–Wolfe contribution near the first peaks.

Recombination and last scattering Hydrogen recombines and the free–electron fraction drops rapidly; the photon mean free path jumps, defining the (finite-width) last-scattering surface. This occurs at $z_* \simeq 1080-1100$ with $T_\gamma \simeq 0.26\,\mathrm{eV} \approx 3000\,\mathrm{K}$.

Before detailing the ingredients of the power spectra, it is useful to summarise the thermal history of the photon field relevant for the CMB. At early times, frequent Thomson scattering off free electrons keeps photons, electrons and protons in a tightly coupled plasma that supports acoustic oscillations; on sufficiently small scales, repeated scatterings lead to diffusion

damping. As the Universe cools to $T_\gamma \simeq 3000\,\mathrm{K}$ ($z_* \approx 1080-1100$), recombination reduces the free–electron fraction and the photon mean free path rises: photons last scatter in a narrow epoch (the last–scattering surface), after which they free–stream to us. Much later, the first luminous sources reionise the intergalactic medium ($6 \lesssim z_{re} \lesssim 10$), introducing an optical depth τ_{reio} that suppresses primary small–scale anisotropies by $e^{-2\tau_{reio}}$ and generates a characteristic large–angle EE "reionisation bump". Between last scattering and today, additional late–time effects (such as gravitational lensing) further modify the primary anisotropies. This qualitative picture underlies the calculations in sections 2.1 to 2.3.

2.1 PHOTON PERTURBATIONS

As already mentioned, in this chapter I describe how perturbations to the photon distribution evolve up to recombination. We obtain a general idea of what we expect to see at the last scattering surface, as photons thermally decouple from baryons.

2.1.1 Large-scale anisotropies

Large–scale anisotropies are the anisotropies predicted by studying the evolution of perturbations of super-horizon modes. For a given photon wavenumber k, super-horizon modes are those with $k\eta \ll 1$ (or equivalently $k < \mathcal{H} = aH$), and therefore they show very little evolution.

The starting point for evaluating the large–scale anisotropies is the relation between the radiation monopole and the gravitational potentials, provided by the initial conditions set by inflation:

$$\theta_{r,0}(k, \eta_i) = \frac{1}{2}\Phi(k, \eta_i), \tag{2.1}$$

where η_i refers to the conformal time at the end of inflation.
By substituting this relation in the time–time Einstein equation for the gravitational potentials (equation 1.59), and using the fact that recombination takes place during the matter–dominated era, one obtains

$$\theta_{r,0}(k, \eta_{rec}) = \frac{2}{3}\Phi(k, \eta_{rec}). \tag{2.2}$$

However, we also have to account for the Sachs-Wolfe (SW) effect, which describes the loss of energy of photons when travelling out of the potential wells. Thus, the observed anisotropy must encode the gravitational redshift as

$$(\theta_0 + \Psi)(k, \eta_{rec}) = \frac{1}{3}\Psi(k, \eta_{rec}). \tag{2.3}$$

2.1.2 Acoustic peak region

Now I move to the evaluation of sub-horizon modes, *i.e.* those with $k\eta \gtrsim 1$. Sub-horizon modes are expected to evolve, depending on the wavenumber, before recombination.
Since they are sub-horizon, they are coupled to electrons via Thomson scattering, and electrons are coupled to protons via Coulomb scattering. Before recombination occurs, it is possible to work under the assumption of a high optical depth ($\tau \gg 1$).

Photons are massless particles, so in principle all the moments of the Boltzmann equation are relevant in order to describe them. However, as they are tightly coupled to electrons – and so to baryons – it can be shown that they can be described as a fluid, the so–called baryon–photon fluid, where only the monopole and the dipole are relevant.
In order to describe the behaviour of photons with a sub-horizon wavenumber k, only two expressions, one for the monopole and the other for the dipole, are needed. However, as already said, the dipole term contains the coupling with baryons. In order to get the overall behaviour of sub-horizon modes it is therefore required to solve a set of three equations; one for the monopole, another for the dipole and the third for the baryon velocity.
The solution of this system leads to the tight–coupling equation for the radiation monopole:

$$\theta_0'' + \frac{a'}{a}\frac{R}{1+R}\theta_0' + k^2 c_s^2 \theta_0 = F(k,\eta), \tag{2.4}$$

where:

- $R = \frac{3\rho_b}{4\rho_\gamma}$, with ρ_b the baryon density and ρ_γ the radiation density;
- $c_s = \sqrt{\frac{1}{3(1+R)}}$ is the sound speed within the baryon–photon fluid and depends on the baryon density. This dependence indicates that baryons have a non–negligible effect, since they lower the sound speed of the baryon–photon oscillations;
- $F(k,\eta) = -\frac{k^2}{3}\Psi - \frac{a'}{a}\frac{R}{1+R}\Phi' - \Phi''$ is the driving term, which contains the gravitational potentials sourced by dark matter. Since it is only related to the gravitational potentials, it enhances compressions (overdensities) and opposes rarefactions.

The overall behaviour of the baryon–photon fluid is an oscillatory behaviour of the radiation monopole both in k–space and conformal time.
Up to recombination the baryon–photon fluid oscillates, and the observed peaks and troughs correspond to those modes that have reached a maximum (compression) or minimum (rarefaction) at recombination. This mechanism can be better appreciated in figure 2.1. To get the final equation for the

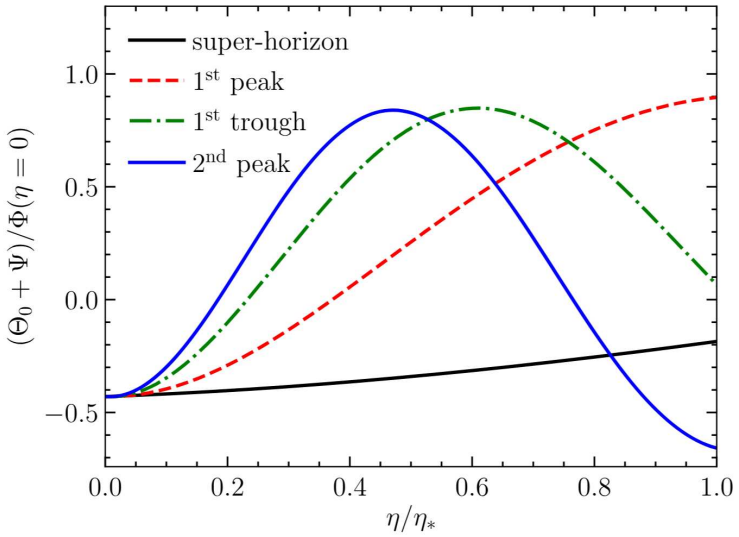

Figure 2.1: Evolution of different k–modes up to recombination. Super-horizon modes (black solid line) show very little evolution. The red dotted line represents the mode which has reached its maximum at recombination (1st peak), so a compression phase of the baryon–photon fluid. The green dotted line represents the mode with an amplitude close to zero at recombination (1st trough). The blue line represents a k–mode with enough time to be compressed and then rarefied (2nd peak) [Dodelson and Schmidt, 2020].

monopole as a function of k and η, initial conditions provided by inflation are required.

Inflation only excites cosine modes, which are the only modes that survive, so it is possible to evaluate the positions of peaks, which are well predicted from the solution to the tight–coupling equation. By contrast, the amplitude of the expected oscillation cannot be properly described assuming that photons behave as a fluid at all scales.

The optical depth is indeed high, but not infinitely high; in other words, the mean free path – that is, the path travelled by a photon between two different scattering events – is small, but not zero. Therefore, on scales smaller than the mean free path photons begin to free–stream, higher moments become important, and diffusion damping suppresses power.

At small scales we have the so–called diffusion damping process, which can be described starting from the same system of three equations for the monopole, the dipole and the baryon velocity, with the assumption that on small scales the contributions coming from the gravitational potentials are negligible. What changes is that it is no longer possible to neglect higher–order moments, so in the equation for the dipole there is the additional contribution coming from the quadrupole moment of the radiation and, in principle, also from even higher orders.

The full k–dependence of the monopole encodes a real part associated with

the oscillation of the baryon–photon fluid and an imaginary one related to the damping process, going as

$$\theta_0 \propto e^{ik\int c_s d\eta}\, e^{-\frac{k^2}{k_D^2}}, \tag{2.5}$$

where k_D is the damping wavenumber, and it depends on the baryon density and the expansion rate of the Universe.

2.1.3 Free stream

In the previous section I provided the basic elements to describe the behaviour of photons up to recombination. This allowed us to arrive at a picture of the expected perturbations on the last scattering surface. In order to understand the actual observations one has to propagate these inhomogeneities from the last scattering surface to us.

From a general point of view, photon perturbations do not grow after decoupling since the gravitational potential wells are too weak to trap them, so they preserve the perturbation at the level of decoupling.
However, as they decouple from baryons, they move freely, so all the higher–order moments must be included, $\theta_\ell(k, \eta_0)$. In order to understand what we are expected to observe, we have to go through the following two main steps:

- express the generic moments $\theta_\ell(k, \eta_0)$ as functions of the monopole and the dipole, since we know their theoretical expressions at the last scattering surface (via the line–of–sight integral and the visibility function);
- relate the θ_ℓ moments, which are abstract quantities, to the observable anisotropy pattern, $T(\vec{x}, \hat{p}, \eta_0)$.

In order to end up with the expression of a generic ℓ–th moment, θ_ℓ, I start from the Boltzmann equation for photons and, assuming that the surface of last scattering is indeed a thin surface – that is, neglecting its finite thickness – I obtain

$$\begin{aligned}\theta_\ell(k, \eta_0) =& [\theta_0(k, \eta_{rec}) + \Psi(k, \eta_{rec})] J_\ell[k(\eta_0 - \eta_{rec})] + \\ & 3\theta_1(k, \eta_{rec}) J_{\ell-1}[k(\eta_0 - \eta_{rec})] - (\ell+1)\frac{J_\ell[k(\eta_0 - \eta_{rec})]}{k(\eta_0 - \eta_{rec})} + \\ & \int_0^{\eta_0} d\eta\, e^{-\tau}(\Psi' - \Phi') J_\ell[k(\eta_0 - \eta)],\end{aligned} \tag{2.6}$$

where:

- the term $[\theta_0(k, \eta_{rec}) + \Psi(k, \eta_{rec})]$ represents the observed monopole anisotropy, so both the temperature anisotropy and the gravitational redshift;
- θ_1 is the dipole;

- the last term contains time–dependent potentials and it becomes relevant after recombination, providing the Integrated Sach–Wolfe (ISW)–briefly analysed in the following section;
- the J_ℓ and $J_{\ell-1}$ are the spherical Bessel functions, and they play a crucial physical role since they project plane–wave perturbations of wavenumber k onto angular multipoles ℓ, with a contribution peaking around $\ell \sim k(\eta_0 - \eta)$.

It is possible to conclude that the anisotropies today are a convolution of the monopole, the dipole and the spherical Bessel functions, and in order to solve them the knowledge of the monopole, the dipole, and the potentials at recombination is required.
All three required quantities are known from the previous section, where I evaluated the behaviour of the monopole – and of the dipole as a consequence – up to recombination.

At this point, it is possible to relate the multipoles to the observed anisotropy pattern:

$$T(\vec{x}, \hat{p}, \eta_0) = T(\eta)\,[1 + \theta(\vec{x}, \hat{p}, \eta_0)], \tag{2.7}$$

where $T(\eta)$ is the background temperature, while $\theta(\vec{x}, \hat{p}, \eta_0)$ is the fractional temperature perturbation defined in equation 1.55 as $\theta = \frac{\delta T}{T}$.

Moving to harmonic space, where the spherical harmonics $Y_{\ell m}$ form the complete set of eigenfunctions on the 2D celestial sphere, we expand the fractional temperature perturbation in terms of scalar spherical harmonics

$$\theta(\vec{x}, \hat{p}, \eta_0) = \sum_{\ell=0}^{+\infty} \sum_{m=-\ell}^{+\ell} T_{\ell m} Y_{\ell m}(\hat{p}). \tag{2.8}$$

The $T_{\ell m}$ coefficients provide all the information about the temperature field. One basic assumption regarding these coefficients is that they are Gaussian random variables with a zero ensemble average, $\langle T_{\ell m} \rangle = 0$, but a non-zero variance. The variance is the quantity that corresponds to the anisotropy power spectrum in harmonic space.
Assuming that the cosmological principle holds for the anisotropy pattern, we expect rotational invariance, so no dependence on the azimuthal number m. The variance tells us the correlation between two different points in the sky, separated by a given angle α; due to rotational invariance, the variance only depends on the angular separation and not on the position of the points under scrutiny. The variance $C(\alpha)$ can be defined both in real and harmonic spaces as

$$C(\alpha) = \langle \theta(\hat{p}_1)\theta(\hat{p}_2) \rangle,$$

$$\langle T_{\ell m} T^*_{\ell' m'} \rangle = C_\ell \delta_{\ell\ell'} \delta_{mm'},$$

where in the last formulation I make use of the fact that no m–dependence is expected, so the variance only depends on ℓ, and tells us how much

anisotropy there is at a given angular scale.
The variance is the actual observable, thus the final target is to relate the variance – the observable quantity – and the multipoles, θ_ℓ, via

$$C_{\ell,\text{theo}} = \frac{2}{\pi}\int_0^{+\infty} dk\, k^2\, P(k) \left|\frac{\theta_\ell(k)}{\theta(k)}\right|^2, \tag{2.9}$$

where $P(k)$ is the primordial power spectrum (introduced at the beginning of chapter 1) and $\left|\frac{\theta_\ell(k)}{\theta(k)}\right|$ is the transfer function.
This sets the needed relation between the multipoles and the observable quantities, so I now move to the description of the CMB power spectrum.

2.2 CMB POWER SPECTRUM

As for the photon perturbations, the expected CMB power spectrum naturally separates into the large–angular–scale regime and the acoustic–peak region.

2.2.1 Large angular scales

Large angular scales refer to modes that entered the horizon only recently, meaning those modes that remain close to their primordial form. In the small–ℓ region, applying the definition of the C_ℓ coefficients as functions of θ_ℓ leads to an approximately constant variance of temperature anisotropies per logarithmic interval in ℓ: $\ell(\ell+1)C_\ell = D_\ell = \text{const.}$
As I will show shortly, this region is also affected by the late–ISW effect, which breaks the expected constant behaviour.
The overall behaviour, including deviations from the flat trend, can be appreciated in figure 2.2.

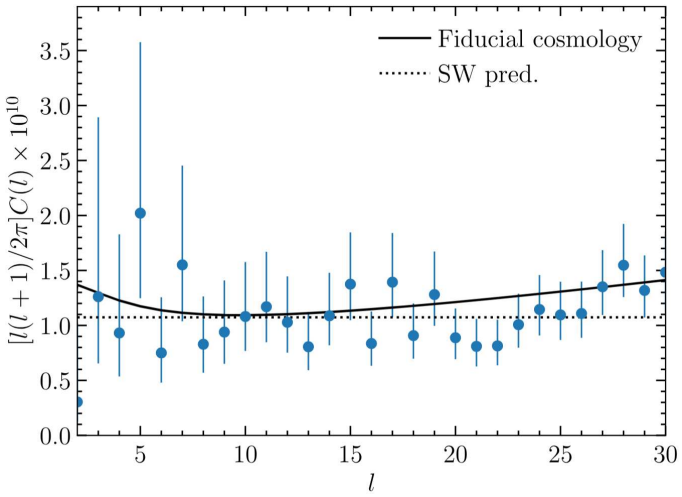

Figure 2.2: Low–ℓ region of the CMB power spectrum [Dodelson and Schmidt, 2020].

2.2.2 Acoustic peak region

In the acoustic–peak region the θ_ℓ are provided by the monopole, the dipole, and the evolution of the gravitational potentials. As already touched upon, the time evolution of the gravitational potentials leads to the ISW effect.
Each time the Universe changes its EoS – meaning when the Universe became matter dominated ($t_{RM} \simeq 4.7 \times 10^4$ yr) and, more recently, when it became dominated by the cosmological constant ($t_{M\Lambda} \simeq 9.8$ Gyr) – there is a time evolution of the gravitational potentials. In general, the larger the conformal-time interval over which the potentials evolve, the lower the affected multipoles; the net effect on those scales is a boost in power.
It is possible to distinguish two different types of ISW effect:

1. The early–ISW effect, associated with the transition from radiation to matter domination, affecting the acoustic–peak region.

2. The late–ISW effect, associated with the recent transition to a Universe dominated by the cosmological constant, which affects the large angular scale region.

In the acoustic–peak region (see figure 2.3), the leading contribution comes from the oscillating monopole, damped at small angular scales; then there is the contribution from the boost of power of the early–ISW, which adds coherently to the monopole, and the contribution coming from the dipole, which does not add coherently. The contribution coming from the dipole is mostly appreciable in troughs, since it pushes them away from zero.

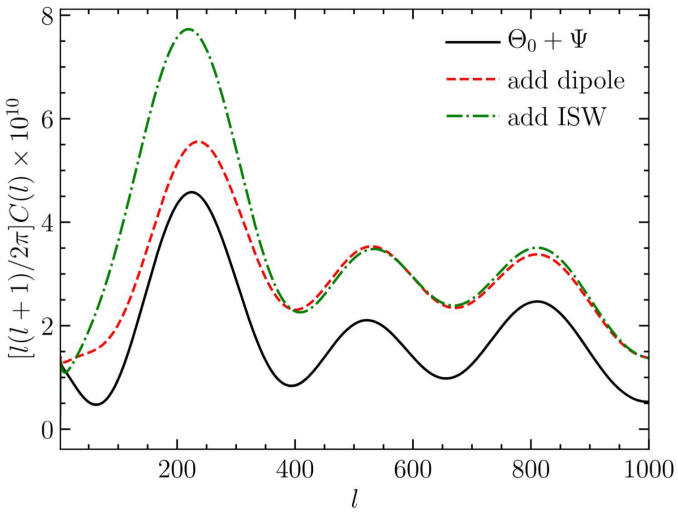

Figure 2.3: Acoustic–peak region of the CMB power spectrum [Dodelson and Schmidt, 2020].

On angular scales $\ell \gtrsim 10^3$ the finite photon mean free path before recombination erases anisotropy, producing the damping tail of the spectrum: C_ℓ is suppressed roughly as $e^{-\ell^2/\ell_D^2}$, where $\ell_D \simeq k_D(\eta_*)(\eta_0 - \eta_*)$ projects the

comoving diffusion scale at last scattering. The damping increases when the photon mean free path is larger (smaller n_e) and when the expansion rate is higher, both of which increase the diffusion damping length and reduce high-ℓ power.

The temperature anisotropy power spectrum can be visualised in figure 2.10.

2.3 POLARIZED CMB

Now we have all the elements in order to understand why the CMB shows a polarisation pattern, what is the mechanism responsible for such a polarisation and what are the expectations in terms of observations.

The polarisation field of the CMB can be described on the celestial sphere through its tensorial formulation as

$$P(\hat{n}) = \frac{1}{\sqrt{2}} \begin{pmatrix} Q(\hat{n}) & -U(\hat{n})\sin(\theta) \\ -U(\hat{n})\sin(\theta) & -Q(\hat{n})\sin^2(\theta) \end{pmatrix}. \tag{2.10}$$

Therefore, the polarisation field can be written as a real, symmetric and trace–free tensor on the celestial sphere. Instead, its decomposition by making use of tensor harmonics is

$$P_{ab}(\hat{n}) = \sum_{\ell=2}^{+\infty} \sum_{m=-\ell}^{m=+\ell} [E_{\ell m} Y^E_{\ell m}(\hat{n})_{ab} + B_{\ell m} Y^B_{\ell m}(\hat{n})_{ab}], \tag{2.11}$$

where:

- $E_{\ell m}$ and $B_{\ell m}$ are the tensor spherical–harmonic coefficients and are the so–called E–modes and B–modes, respectively. The E–modes are parity even, while B–modes are parity odd.
- $Y^E_{\ell m}$ and $Y^B_{\ell m}$ are the gradient and the curl tensor harmonics and they are related to spin–weighted harmonics through the relations

$$Y^{B,*,ab}_{\ell m} Y^B_{\ell' m' ab} = \frac{1}{2}\left({}_{-2}Y^*_{\ell m}\, {}_{-2}Y_{\ell' m'} + {}_{+2}Y^*_{\ell m}\, {}_{+2}Y_{\ell' m'}\right), \tag{2.12}$$

$$Y^{E,*,ab}_{\ell m} Y^E_{\ell' m' ab} = \frac{i}{2}\left({}_{+2}Y^*_{\ell m}\, {}_{+2}Y_{\ell' m'} - {}_{-2}Y^*_{\ell m}\, {}_{-2}Y_{\ell' m'}\right). \tag{2.13}$$

Now I move to the description of the mechanism responsible for the CMB polarisation pattern, Compton scattering.

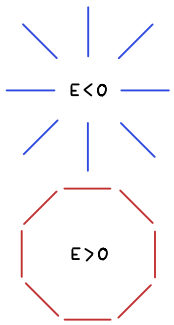

Figure 2.4: Parity even E-modes.

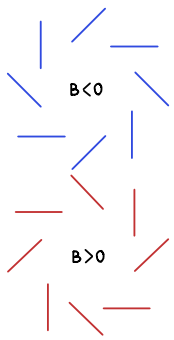

Figure 2.5: Parity odd B-modes.

Review: basic concepts about polarisation

From a general point of view, given an electromagnetic wave, the electric field can be decomposed into its x and y components. If the two components are correlated in phase and amplitude, then the radiation is polarised.

x component:

$$E_x = a_x(t)\cos[\omega_0 t - \theta_x(t)],$$

y component:

$$E_y = a_y(t)\cos[\omega_0 t - \theta_y(t)].$$

The Stokes parameters are the mathematical tools to quantify the polarisation of an electromagnetic wave and they are related to the intensity of the radiation itself as follows:

$$I = \langle a_x^2 \rangle + \langle a_y^2 \rangle,$$

$$Q = \langle a_x^2 \rangle - \langle a_y^2 \rangle,$$

$$U = \langle a_x a_y \cos(\theta_x - \theta_y) \rangle,$$

$$V = \langle a_x a_y \sin(\theta_x - \theta_y) \rangle,$$

where the $\langle \cdot \rangle$ brackets denote time averages. The first Stokes parameter, I, provides information about the intensity of the radiation. The two parameters that describe the state of the linear polarisation are Q and U; if both $Q = U = 0$ the radiation is unpolarised. The last parameter V refers to circular polarisation, which is expected to be zero since in the framework of the standard model there is no mechanism capable of providing circular polarisation in the early Universe.

The I and V Stokes parameters are scalar quantities, whereas Q and U are not scalars; they are the two components of a spin-2 polarisation field. In particular, the combinations

$$(Q \pm iU),$$

transform as spin-2 quantities. Intuitively, Q and U describe a headless vector at each sky position. A headless vector is a bar, so the polarisation field can be seen as a collection of bars and, since they are not scalar quantities, when moving to harmonic space their decomposition requires the use of spin–weighted spherical harmonics instead of scalar ones (as for the case of the temperature field).

2.3.1 Compton scattering

From a general point of view, Compton scattering is a mechanism capable of generating polarised radiation. In order to capture the fundamental mechanism I focus on a very simple and unrealistic case of an electron which scatters, along the $\hat{z}$ direction, an incoming unpolarised electromagnetic wave. From an intuitive point of view, unpolarised radiation can be visualised with equally long components, while if the two components have different lengths, then we can conclude that we are dealing with polarised radiation.
Since electromagnetic waves are transverse, they can only oscillate in the plane perpendicular to the propagation direction. Figure 2.6 shows the basic mechanism capable of generating a polarised outgoing radiation, representing the scattering of a single plane wave along the $\hat{z}$ direction. Since the oscillation is forced to be in the plane perpendicular to the direction of propagation, only one component survives after the scattering process and the outgoing radiation is clearly polarised. However, the case of a single incoming plane

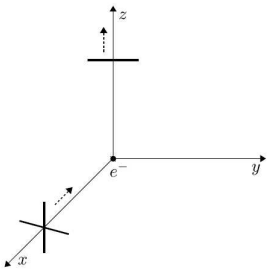

Figure 2.6: Compton scattering of a single plane wave [Dodelson and Schmidt, 2020].

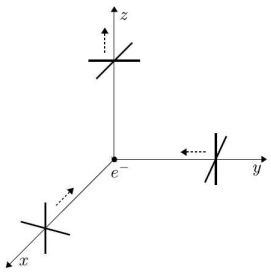

Figure 2.7: Compton scattering of a radiation monopole [Dodelson and Schmidt, 2020].

wave is not realistic; in order to understand how CMB polarisation can be generated we shall examine the cases of a radiation monopole, dipole and quadrupole.
For the monopole, considering two incoming unpolarised electromagnetic waves, they have the same amplitude. Firstly, there is the selection of the $\hat{y}$–component for the plane wave coming from the $\hat{x}$ direction; secondly, the selection of the $\hat{x}$ component for the plane wave coming from the $\hat{y}$ direction (as shown in figure 2.7), but the length of the two selected components is the same, so the outgoing radiation is unpolarised. The monopole cannot generate the observed polarisation pattern.
Moving to the dipole, the electron scatters two incoming unpolarised electromagnetic waves with different amplitudes. However, as shown in figure 2.8, the outgoing wave is still unpolarised. From the positive $\hat{x}$ direction the incoming wave is hotter than the average, while from the negative $\hat{x}$ direction the wave is colder. From the $\hat{x}$ direction there is the selection of the y component, so the net result is an average to the mean background temperature. From the y direction the electron receives radiation with the average back-

ground temperature and then there is the selection of the x component. Both the selected components have the same amplitude, that is the one provided by the average temperature, so the outgoing radiation is not polarised.

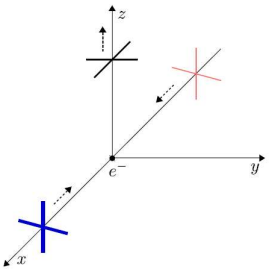

Figure 2.8: Compton scattering of a radiation dipole [Dodelson and Schmidt, 2020].

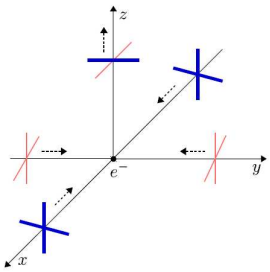

Figure 2.9: Compton scattering of a radiation quadrupole [Dodelson and Schmidt, 2020].

Since the dipole also cannot explain the polarisation pattern of the CMB, I now move to the quadrupole, even if it is mostly negligible before decoupling. Figure 2.9 shows how the radiation quadrupole looks like: along the x axis there are two incoming unpolarised waves with a temperature colder than the average, while along the y axis the two incoming unpolarised waves show a temperature hotter than the average. In the case of a quadrupole, the selected components from the x axis and the y axis are different, providing an outgoing polarised radiation.
To conclude, in order to obtain a polarisation pattern of the CMB, the photon distribution has to show a quadrupole anisotropy at last scattering.
Because photons and baryons are tightly coupled before recombination, the radiation quadrupole θ_2–the piece that Thomson scattering converts into linear polarisation–is strongly suppressed. As recombination proceeds, the photon mean free path increases rapidly while a residual free–electron fraction still exists. Immediately afterwards the free–electron density drops as electrons bind into neutral hydrogen, Thomson scattering effectively switches off, and no further primary polarisation is produced. This thin temporal shell both limits the conversion efficiency and explains why the polarisation signal is much weaker than temperature anisotropies.

2.3.2 Polarisation modes

The description of the polarisation pattern of the CMB requires the introduction of the Boltzmann equation for polarisation in the set of Boltzmann equations introduced in the previous chapter.
From a general point of view, we could both have E–modes and B–modes on the last scattering surface. However, B–modes are parity odd and, since the underlying physics of the observed polarisation pattern is electromagnetism – so, parity conserving – B–modes can only couple to tensor perturbations.

While E–modes, being parity even, can couple to both scalar and tensor perturbations.
The relation between scalar and tensor perturbations can be parametrised through the tensor–to–scalar ratio parameter, r, defining the ratio of the amplitude of the primordial gravitational waves (meaning tensor perturbations for what stated in section 1.4) and the amplitude of primordial density fluctuations (scalar perturbations) in the early Universe. Recent study from Tristram et al., 2022 have constrained the tensor-to-scalar ratio to be $r < 0.032$. Tensor perturbations are much smaller than scalar ones and, since pure B–modes have not been observed yet, from this point I work under the assumption of dealing only with scalar–generated E–modes on the last scattering surface. This assumption allows for the employment of only one additional Boltzmann equation to describe CMB polarisation,

$$\theta'_E + ik\mu\theta_E = -\tau' \left[b(1-\mu^2)\theta_2 - \theta_E\right], \tag{2.14}$$

where the coupling with the radiation quadrupole θ_2 is clear, since it explicitly appears in the Boltzmann equation. Inside the square brackets there are two different contributions: the source term, where the quadrupole appears and b is simply a numerical factor encoding the Compton scattering kernel, and the loss term, encoding the loss of polarisation information after a certain number of scatterings.

The polarisation field can be decomposed in harmonic space by making use of spherical harmonics, where the coefficients of the expansion are the scalar quantities $E_{\ell m}$ and $B_{\ell m}$[1] (see equation 2.11), and they can be seen as random Gaussian variables with zero average, but non–zero variance, as for the case of the expansion coefficients of the temperature field.
As already said for temperature anisotropies, the actual observed quantity is the variance, so the correlation between different points in the sky. With the introduction of polarisation in the overall scenario, we expect to be able to observe different power spectra from the possible cross–correlations between $T_{\ell m}$, $E_{\ell m}$ and $B_{\ell m}$.
The possible cross–correlations are:

$$\begin{aligned}
\left\langle T^*_{\ell' m'} T_{\ell m} \right\rangle &= C_\ell^{TT} \delta_{\ell\ell'} \delta_{mm'}, \\
\left\langle E^*_{\ell' m'} E_{\ell m} \right\rangle &= C_\ell^{EE} \delta_{\ell\ell'} \delta_{mm'}, \\
\left\langle T^*_{\ell' m'} E_{\ell m} \right\rangle &= C_\ell^{TE} \delta_{\ell\ell'} \delta_{mm'}, \\
\left\langle T^*_{\ell' m'} B_{\ell m} \right\rangle &= 0, \\
\left\langle E^*_{\ell' m'} B_{\ell m} \right\rangle &= 0, \\
\left\langle B^*_{\ell' m'} B_{\ell m} \right\rangle &= C_\ell^{BB} \delta_{\ell\ell'} \delta_{mm'},
\end{aligned}$$

1 Since the $E_{\ell m}$ and $B_{\ell m}$ coefficients describe the polarisation field, $\ell = 0$ and $\ell = 1$ are not defined, so $\ell \geq 2$.

where the spectra that are expected to be zero are zero because we are working in a scenario where parity is not violated. Recalling also the discussion about tensor perturbations and B–modes, the BB–spectrum is assumed to be zero too during this analysis.
The non–zero fiducial power spectra can be visualised in figure 2.10. The presented power spectra are in band power, that is: $D_\ell = \frac{\ell(\ell+1)}{2\pi} C_\ell$.
In the analytic discussion above I set the primordial BB spectrum to zero (*i.e.* $r = 0$) and work in a parity-conserving scenario, which implies $\langle TB \rangle = \langle EB \rangle = 0$. However, even with $r = 0$ the late-time gravitational lensing of E-modes by large-scale structure inevitably converts part of E into B, generating a lensing $C_\ell^{BB} \neq 0$ peaking at $\ell \sim 10^3$. The BB curve shown in figure 2.10 is this lensing-induced component; it is not an inflationary (tensor) signal and does not contradict the $r = 0$ assumption used for the equations.

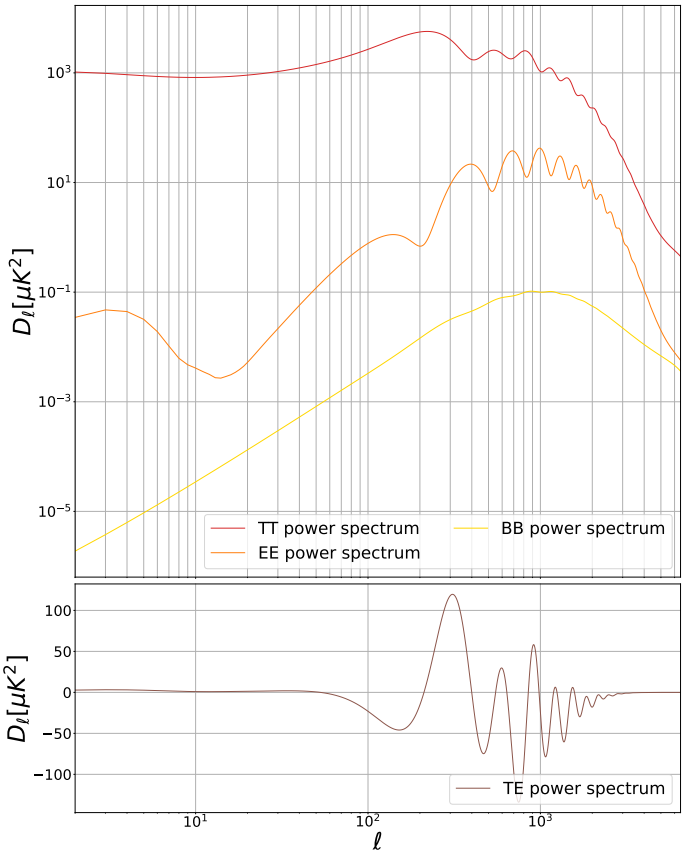

Figure 2.10: CMB power spectra from fiducial maps: C_ℓ^{TT}, C_ℓ^{EE}, C_ℓ^{BB}, C_ℓ^{TE}.

To conclude this chapter entirely devoted to the CMB, it is possible to visualise the CMB maps measured by the Planck satellite [Planck Collaboration, 2020a] for the temperature and polarisation fields in figures 2.11 and 2.12, respectively.

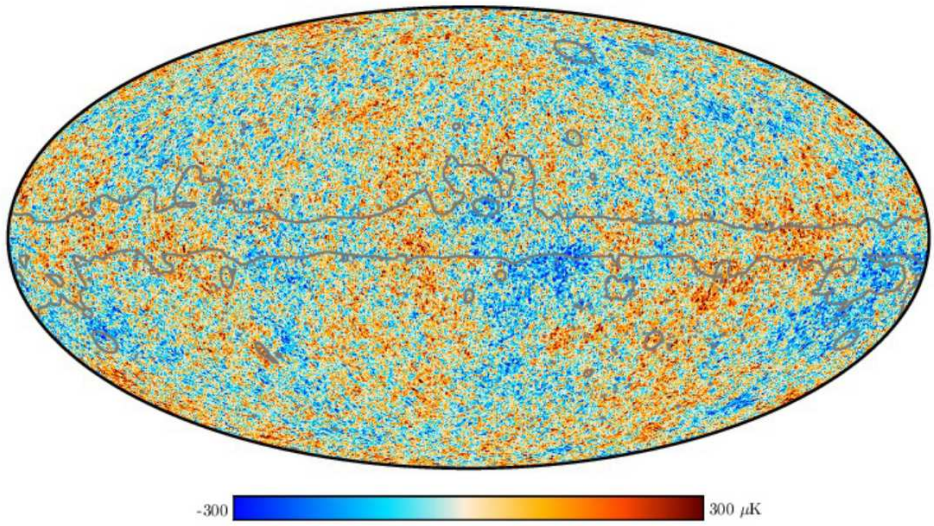

Figure 2.11: CMB temperature map from Planck Collaboration, 2020a.

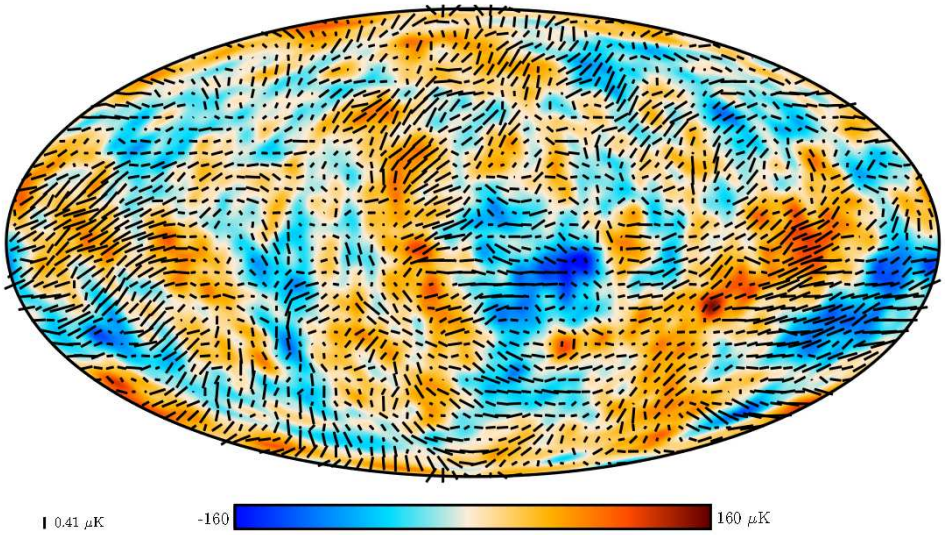

Figure 2.12: CMB polarisation field from Planck Collaboration, 2020a.

3 STRUCTURE GROWTH IN LINEAR THEORY

This chapter provides a first step into the wider description of the Large–scale structure (LSS) of the Universe, which is responsible for distortions and contamination of the CMB power spectra.

This entire part of the thesis is focused on the linear regime; therefore, the very first step in approaching LSS is the study of the evolution of dark matter perturbations, detailed in section 3.1. I conclude this chapter by making use of concepts introduced in chapter 1, and arrive at a prediction of the linear matter power spectrum in section 3.2.

3.1 MATTER PERTURBATIONS

As already noted, the very starting point is the understanding of the growth of dark matter perturbations, whose seeds reside in primordial gravitational instability. At this point it is important to keep in mind how the various components of the Universe are closely intertwined with one another (see figure 1.4). Specifically, dark matter is coupled via gravity to all the other components; while being self-gravitating, and therefore not influenced by photons, it influences baryons. Gravity amplifies matter inhomogeneities, whereas inflation provides the initial curvature perturbations (*i.e.* the primordial gravitational potential) from which growth proceeds; these define our initial conditions.

From a general point of view, matter tends to accumulate in initially overdense regions. This growth process is regulated by a competing force represented by the thermal pressure. There are two main factors to take into account in order to properly describe the growth of perturbations:

- first, the balancing of these two forces happens within the horizon, therefore, modes lying inside or outside the horizon are expected to show different behaviour in the growth of perturbations;
- the competing force, *i.e.* the thermal pressure, depends on the dominant component contributing to the energy budget of the Universe at horizon crossing. This means that we also expect different behaviour depending on whether the Universe is radiation- or matter-dominated.

Having all this in mind, since we observe the matter distribution at late epochs, a crucial point is to relate the potential today to the primordial one. Specifically, we expect a distortion of the behaviour of gravitational perturbations because of both the expansion of the Universe and the pressure

of a given dominant component in the overall energy budget. To account for these aspects, it is possible to write

$$\Phi(k,a) = \Phi_P(k) \times \mathcal{T}(k) \times \mathcal{G}(a), \tag{3.1}$$

where $\Phi_P(k)$ is the primordial potential, and $\mathcal{T}(k)$, and $\mathcal{G}(a)$ are the transfer and growth functions, respectively, which are the subjects of the two following subsections.

In order to predict the evolution of dark matter perturbations, we need the two Boltzmann equations for the radiation monopole and dipole, the linear continuity and Euler equations for Cold Dark Matter (CDM), and the first Einstein equation for the gravitational potential (see chapter 1 for the detailed derivation of these equations). Given the full system of five equations

$$\begin{cases} \theta'_{r,0} + k\theta_{r,1} = -\Phi' \\ \theta'_{r,1} - \dfrac{k}{3}\theta_{r,0} = -\dfrac{k}{3}\Phi \\ \delta'_c + iku_c = -3\Phi' \\ u'_c + \dfrac{a'}{a}u_c = -ik\Phi \\ k^2\Phi + 3\dfrac{a'}{a}\left(\Phi' + \dfrac{a'}{a}\Phi\right) = 4\pi G a^2(\rho_m\delta_m + 4\rho_r\theta_{r,0}) \end{cases} \tag{3.2}$$

it is possible to numerically find solutions for every mode. However, in the following section, I focus on the behaviour of δ and Φ for four specific cases, namely

1. super-horizon waves;
2. waves crossing the horizon during matter domination;
3. waves crossing the horizon before equality;
4. waves crossing the horizon soon after equality.

3.1.1 Transfer function

The transfer function describes the evolution of perturbations through the epoch of horizon crossing if that happens before or soon after matter-radiation equality.
By convention, it is normalised at the value of the gravitational potential at late times, so that today it is equal to unity and the evolution of perturbations is solely driven by the growth factor (see section 3.1.2). Therefore, the transfer function is defined as

$$\mathcal{T}(k) = \frac{\Phi(k, a_{late})}{\Phi_{Large-scale}(k, a_{late})},\ a_{late} \gg a_{eq}. \tag{3.3}$$

As stated at the end of the previous section, in the following I describe the behaviour of δ and Φ in four different regimes, to better understand the effects of the transfer function.

1. The case of super-horizon modes refers to large modes lying outside the horizon. On these scales $k/(aH) \ll 1$, so gradient terms ($\propto k$, k^2) and the velocity sourced by $-ik\Phi$ are suppressed ($ikv = \mathcal{O}[(k/\mathcal{H})^2]$) and can be neglected. Therefore, by looking at the set of equations presented in 3.2, it is possible to neglect all terms multiplied by the wavenumber k, the dipole, and the fluid velocity, so as to obtain the following reduced system of equations

$$\begin{cases} \theta'_{r,0} = -\Phi' \\ \delta'_c = -3\Phi' \\ 3\dfrac{a'}{a}\left(\Phi' + \dfrac{a'}{a}\Phi\right) = 4\pi G a^2(\rho_m\delta_m + 4\rho_r\theta_{r,0}). \end{cases} \tag{3.4}$$

   By introducing the parameter y defined as

$$y \equiv \frac{a}{a_{eq}} = \frac{\rho_m}{\rho_r}, \tag{3.5}$$

   the Einstein equation can be written as

$$3\frac{a'}{a}\left(\Phi' + \frac{a'}{a}\Phi\right) = 4\pi G a^2 \rho_m \delta_m \left(1 + \frac{4}{3}\frac{1}{y}\right). \tag{3.6}$$

   The solution of the system can be computed in terms of the parameter y and is

$$\Phi = \frac{\Phi(0)}{10}\frac{1}{y^3}[16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16]. \tag{3.7}$$

   In a matter-dominated Universe, the parameter y is large and dominates the brackets, leading to $\Phi \to \dfrac{9}{10}\Phi(0)$. This tells us that the potential drops by a factor $1/10$ as the Universe becomes matter-dominated, even in the case of modes outside the horizon. The factor $9/10$ is the normalisation factor used to normalise the transfer function to large-scale values.

2. When considering horizon-crossing in a matter-dominated Universe, it is safe to neglect the first two equations of 3.2 since they refer to radiation. The solution of the system

$$\begin{cases} \delta'_c + iku_c = -3\Phi' \\ u'_c + \dfrac{a'}{a}u_c = -ik\Phi \\ k^2\Phi + 3\dfrac{a'}{a}\left(\Phi' + \dfrac{a'}{a}\Phi\right) = 4\pi G a^2 \rho_m \delta_m, \end{cases} \tag{3.8}$$

leads to a constant potential. The transfer function for all modes with $k \ll k_{eq} = 0.073\,\mathrm{Mpc}^{-1}\,\Omega_m h^2$ is equal to unity, with k_{eq} being the comoving wavenumber of the mode that enters the horizon at matter–radiation equality.

3. If the horizon-crossing happens in the radiation-dominated era, it is now possible to neglect the equations describing matter perturbations ending up with

$$\begin{cases} \theta'_{r,0} + k\theta_{r,1} = -\Phi' \\ \theta'_{r,1} - \dfrac{k}{3}\theta_{r,0} = -\dfrac{k}{3}\Phi \\ k^2\Phi + 3\dfrac{a'}{a}\left(\Phi' + \dfrac{a'}{a}\Phi\right) = 16\pi G a^2 \rho_r \theta_{r,0}. \end{cases} \tag{3.9}$$

After some calculations, the solution for the potential is

$$\Phi = 3\Phi_P\left(\frac{\sin\left(k\eta/\sqrt{3}\right) - (k\eta/\sqrt{3})\cos\left(k\eta/\sqrt{3}\right)}{(k\eta/\sqrt{3})^3}\right). \tag{3.10}$$

This solution tells us that if the horizon-crossing happens in the radiation dominated era, the potential starts a decaying phase followed by an oscillating behaviour. The amplitude of the oscillation decreases with the square of the conformal time (*i.e.*, $\Phi \sim \theta_0/\eta^2$).
During this time, the matter perturbations, using the equations in 3.2, can be written as

$$\delta_c(k,\eta) = \mathcal{A}\Phi_P \ln(\mathcal{B}) + \mathcal{A}\Phi_P \ln(k\eta), \tag{3.11}$$

where $\mathcal{A}$ and $\mathcal{B}$ are two constants. It is possible to conclude that, despite the decaying potential, the matter perturbations are still experiencing a growth, even if suppressed with respect to the matter-dominated case.

4. Soon after equality, the Universe is not fully matter-dominated, and there is still a non-negligible amount of radiation. However, within this picture, because of radiation pressure, radiation perturbations do not grow. On the contrary, matter perturbations do and, at some point, they become larger than radiation perturbations and start evolving with the potential, unaffected by radiation. The evolution of matter perturbations in this regime is regulated by the Meszaros equation [Meszaros, 1974]

$$\frac{\partial^2\delta_c}{\partial y^2} + \frac{2+3y}{2y(y+1)}\frac{\partial\delta_c}{\partial y} - \frac{3}{2y(y+1)}\delta_c = 0, \tag{3.12}$$

where y is the same parameter defined in equation 3.5. The solution to the Meszaros equation is

$$\delta_c(k,y) = C_1 D_1(y) + C_2 D_2(y),\ y \gg y_H. \tag{3.13}$$

In the above equation y_H is obtained by dividing the scale factor at the horizon entry by the scale factor at equality. The C_1 and C_2 constants are determined so that they match the solution of the previous case. The matter perturbations now evolve as the sum of a linearly evolving mode, $D_1(y) = 1 + \frac{3}{2}y$, and a decaying mode, $D_2(y) \propto y^{-3/2}$.

3.1.2 Growth function

The growth function describes the evolution of the potential at late times. Therefore, it does not depend on the wavelength, since it simply provides the time evolution because of the expansion of the Universe.
As in the case of the transfer function, the growth function is normalised as

$$\mathcal{G}(a) = \frac{\Phi(k,a)}{\Phi(k,a_{late})} \equiv \frac{D_+(a)}{a},\ a > a_{late}. \tag{3.14}$$

In the equation above, $D_+(a)$ is the so-called growth factor, defined as the potential evaluated in an epoch after the regime where the transfer function dominates, divided by the potential in a flat matter-dominated scenario.

At redshifts $z \lesssim 10$ it is possible to observe that all scales within the horizon evolve in the same way. This happens because the smoothing of modes due to radiation pressure is now negligible. In this regime, the transfer function tends to unity and the growth of perturbations is driven by the growth function and, more specifically, the growth factor. Specifically, after modifying the Meszaros equation so as to encode the contribution from dark energy too, it is possible to find the expression for the growth factor at a specific epoch as

$$D_+(a) = \frac{5\Omega_m}{2}\frac{H(a)}{H_0}\int_0^a \frac{da'}{(a'H(a')/H_0)^3} \tag{3.15}$$

In the specific case of a flat, matter-dominated Universe, the growth factor is simply the scale factor. However, because of the presence of dark energy, the evolving modes experience a suppression at late times.

To conclude this section, it is possible to explicitly write the equation that relates the gravitational potential at a given epoch to the primordial one as

$$\Phi(k,a) = \frac{9}{10}\Phi_P(k)T(k)\frac{D_+(a)}{a},\ a > a_{late}. \tag{3.16}$$

Figure 3.1 summarises the evolution of the gravitational potential, in the left panel, and the matter perturbations, in the right panel, for different values of wavenumber k. Specifically, from the left panel of figure 3.1 it is possible to appreciate different behaviours depending on the specific time and the value of the wavenumber:

- at very early times, the gravitational potential tends to be constant for any wavenumber. This is because all modes are outside the horizon.

- For the higher k-modes, we see the oscillatory behaviour describing the way in which radiation influences gravity.
- During matter domination, each mode follows a constant behaviour.
- At matter-radiation equality we observe a drop-off of the potential. This drop-off can also be observed at late times, due to the change in the EoS when entering the epoch dominated by the cosmological constant.

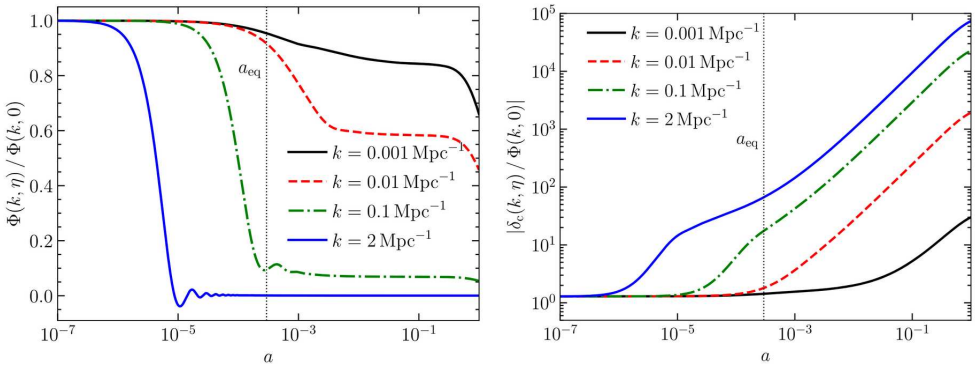

Figure 3.1: Left panel: evolution of the gravitational potential as a function of the scale factor for different k-modes. Right panel: evolution of the matter perturbations as a function of the scale factor for different k-modes. [Dodelson and Schmidt, 2020]

Analogous observations can be made for the right panel of figure 3.1, describing the evolution of matter perturbations:

- at early times all modes are normalised at the same amplitude.
- During the matter-dominated epoch the modes evolve linearly.
- At very late times we appreciate a slight suppression of the linear growth.

3.2 LINEAR MATTER POWER SPECTRUM

In the previous section I provided a description of how to relate matter perturbations to both the primordial power spectrum and the late-time observations. At this point, it is possible to end up with a prediction of the linear matter power spectrum. To do that, the starting point is the first Einstein equation which, evaluated in a matter-dominated Universe, reads

$$k^2\Phi(k,a) = 4\pi G a^2 \rho_m \delta_m. \tag{3.17}$$

From equation 1.25, we can write the matter density, ρ_m, in terms of the critical density, $\rho_{c,0}$, as

$$\rho_{\mathbf{m}} = \frac{\Omega_m}{a^3}\rho_{c,0} = \frac{3H_0^2\Omega_m}{8\pi G\, a^3}. \tag{3.18}$$

By substituting the above expression for the matter density in the first Einstein equation, we end up with

$$k^2\Phi(k,a) = \frac{3}{2}\frac{H_0^2}{a}\Omega_m\delta_m(k,a), \tag{3.19}$$

or, making the matter perturbation explicit,

$$\delta_m(k,a) = \frac{2}{3}\frac{k^2 a}{H_0^2\Omega_m}\Phi(k,a). \tag{3.20}$$

From equation 3.16, we can rewrite the expression for the matter perturbations in terms of the primordial gravitational potential

$$\delta_m(k,a) = \frac{3}{5}\Phi_P(k)\frac{k^2}{\Omega_m H_0^2}T(k)D_+(a). \tag{3.21}$$

The above expression is the starting point for the evaluation of the matter power spectrum.

$$\begin{aligned}\langle\delta_m(k,a)\delta_m^*(k,a)\rangle &= \left\langle\frac{3}{5}\Phi_P(k)\frac{k^2}{\Omega_m H_0^2}T(k)D_+(a)\times\right.\\ &\qquad\left.\frac{3}{5}\Phi_P^*(k)\frac{k^2}{\Omega_m H_0^2}T(k)D_+(a)\right\rangle,\\ \langle\delta_m(k,a)\delta_m^*(k,a)\rangle &= \frac{9}{25}\frac{k^4}{\Omega_m^2 H_0^4}T^2(k)D_+^2(a)\langle\Phi_P(k)\Phi_P^*(k)\rangle.\end{aligned}$$

The angular brackets in the right-hand side of the last equation represent the power spectrum of the gravitational potential, which goes as

$$\langle\Phi_P(k)\Phi_P^*(k)\rangle \propto \frac{\mathcal{A}_s}{k^3}\left(\frac{k}{k_p}\right)^{n_s-1}, \tag{3.22}$$

where the above power-law form is the standard inflationary prediction, with $\mathcal{A}_s$ and n_s defining the amplitude and scale dependence of the primordial scalar spectrum at a chosen pivot scale, k_p (*e.g.* $k_p = 0.05Mpc^{-1}$).

Employing the formulation above for the power spectrum of the primordial gravitational potential allows us to end up with the expression of the matter power spectrum

$$P_{mm}(k) \propto \frac{\mathcal{A}_s}{\Omega_m^2 H_0^4}T^2(k)D_+^2(a)\frac{k^{n_s}}{k_p^{n_s-1}}. \tag{3.23}$$

Figure 3.2 shows the predictions for the matter power spectrum at different redshift values. The peak represents the modes that entered the horizon at matter–radiation equality. For small-k modes we observe the expected linear growth, while for larger-k modes the spectrum decreases.

It is important to stress that this expression describes the linear matter power spectrum, meaning the matter power spectrum in a regime where

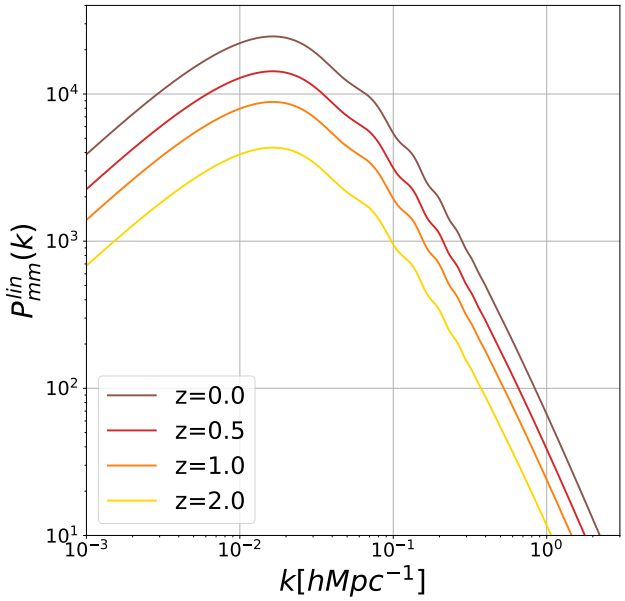

Figure 3.2: Linear matter power spectrum, computed using the `CAMB` Boltzmann solver. The linear matter power spectrum is shown for four different redshifts. The peak in the power spectrum represents the modes that enter the horizon at equality.

matter perturbations are $\ll 1$. As $\delta_m \sim 1$, the density fluctuations become significant and we enter the non-linear regime. The next chapter provides a prescription to add non-linearities to the overall picture.

Throughout this chapter we have implicitly followed the total matter field, but it is important to recall that CDM and baryons do not evolve identically. Before recombination, baryons are tightly coupled to photons and undergo acoustic oscillations with diffusion damping, while CDM grows essentially pressureless. After decoupling, baryons fall into dark-matter potential wells, yet a small residual offset imprints oscillatory features–the Baryon Acoustic Oscillations (BAO)–in the late-time transfer functions. These appear as the wiggles in $P_{mm}(k)$ right after the peak and can be visualised in figure 3.2.

From an observational perspective, the total matter power spectrum is not directly accessible. In practice we measure tracers of the matter field. This motivates the transition in the next chapter: we will relate the dark-matter distribution to its tracers within a physically motivated framework to connect theoretical predictions to data.

Part II

EXTRAGALACTIC EMISSIONS

This part of the thesis moves beyond linear theory to describe structure formation in the non-linear regime using the Halo Model (HM) framework. Within this setting, I model the main extragalactic emissions that contaminate the primary Cosmic Microwave Background (CMB) signal: the Cosmic Infrared Background (CIB), the thermal and kinetic Sunyaev-Zel'dovich (SZ) effects, and contributions from radio sources and Galactic cirrus.

4 THE HALO MODEL FRAMEWORK

In this chapter I provide a detailed description of the Halo Model (HM), a tool capable of predicting the matter – and its tracers' – distribution, capturing their intrinsic non-linear behaviour. Section 4.1 provides a description of the spherical collapse model, which explains how dark matter haloes are distributed in our Universe starting from the primordial perturbations in the matter density field. In section 4.2, I delve into the details of the three main ingredients of the HM, while section 4.3 is devoted to the computation of non-linear power spectra in the HM framework.

4.1 SPHERICAL COLLAPSE MODEL

I use the spherical collapse model [Cooray and Sheth, 2002; Gunn and Gott, 1972] as a first step to relate the abundance of non-linear objects to the initial perturbations of the matter density field.
Let us start at an initial time t_0 with a region of comoving Lagrangian size R_0 and initial overdensity within that radius, δ_i. Within the spherical collapse model, the root–mean–square (rms) of the initial Gaussian fluctuations on scale R_0 is small (meaning $\sigma(R_0) \ll 1$), ensuring $\delta_i \ll 1$ for typical fluctuations[1]. Given these initial assumptions, the mass M_0 within the Lagrangian region is

$$M_0 = \frac{4}{3}\pi R_0^3 \bar{\rho}(1+\delta_i) \approx \frac{4}{3}\pi R_0^3 \bar{\rho}, \tag{4.1}$$

where $\bar{\rho}$ is the comoving background density.
As the Universe evolves, the size of the region changes accordingly. Specifically, there is a deterministic relation between the size $R(t)$ at a generic time t, the initial size R_0, and the overdensity of the perturbation δ. This relation can be obtained by equating the mass inside the region at the initial time t_0 from equation 4.1 with the one at a generic time t:

$$M = \frac{4}{3}\pi R(t)^3 \bar{\rho}(1+\delta), \tag{4.2}$$

so that

$$\left(\frac{R_0}{R(t)}\right)^3 = 1+\delta. \tag{4.3}$$

1 The fact that $\delta_i \ll 1$ comes from its definition evaluated at the initial time, *i.e.*

$$\delta_i = \frac{\rho(\vec{x}) - \bar{\rho}}{\bar{\rho}} \ll 1.$$

Assuming a flat Universe described by a FLRW metric, it is possible to obtain a parametric solution for $R(z)$ as a function of θ:

$$\frac{R(z)}{R_0} = \frac{1+z}{5/3|\delta_0|}\frac{1-\cos\theta}{2}, \tag{4.4}$$

where δ_0 is the initial overdensity δ_i extrapolated at the present time using the linear perturbation theory. Using the relation

$$\frac{1}{1+z} = \left(\frac{3}{4}\right)^{2/3}\frac{(\theta-\sin\theta)^{2/3}}{5/3|\delta_0|}, \tag{4.5}$$

equation 4.4 can be written as

$$\frac{R_0}{R(z)} = \frac{6^{2/3}}{2}\frac{(\theta-\sin\theta)^{2/3}}{(1-\cos\theta)}. \tag{4.6}$$

θ has no physical meaning, rather it is a parameter describing the stages of the collapse in the spherical collapse model which appears when evaluating its parametric solutions. Specifically, there are five different phases in the spherical collapse model identified by the different values assumed by this θ parameter.

1. $\theta = 0$, beginning of expansion;
2. $0 < \theta < \pi$, expansion phase;
3. $\theta = \pi$, turnaround, characterised by zero velocity and maximum radius;
4. $\pi < \theta < 2\pi$, collapse phase;
5. $\theta = 2\pi$, collapse to a point, but replaced in reality with virialisation.

Equation 4.4 tells us how the size of the overdense region evolves. Specifically, in the spherical collapse model framework, the collapse starts with a phase of expansion at $\theta = 0$, then at $\theta = \pi$ there is a "turnaround", and the initial perturbation completely collapses at $\theta = 2\pi$. From a theoretical viewpoint, at collapse $R(z_{col}) = 0$ and the density inside the perturbation is infinite. However, this does not happen in practice because, instead of vanishing, it virialises at some non-zero size, forming a bound object.
In practice, the formal collapse to zero radius is prevented by virialisation: non-linear gravitational interactions lead the system to a stable, bound equilibrium configuration satisfying the virial theorem. Assume now that the initial perturbation virialises at a radius that is half of the turnaround radius in order to provide an estimate of the average density of the final virialised object. If this is the case, since the density scales as R^{-3}, the final object is 8 times denser than the perturbation at the turnaround, $\rho_{vir} = 8\rho_{ta}$. In the

meantime, the background is expanding. Specifically, evaluating equation 4.5 both at the turnaround ($\theta = \pi$) and at collapse ($\theta = 2\pi$), it is possible to infer that the background Universe expanded by a factor $\frac{1+z_{ta}}{1+z_{col}} = 2^{2/3}$ and, consequently, the density at the turnaround was 4 times that at z_{vir}, meaning $\rho^b_{ta} = 4\rho^b_{vir}$, where the superscript b stands for *'background'*. This allows us to determine the so-called density contrast, meaning the density of the virialised object with respect to the background density, as

$$\Delta_{vir} = \frac{\rho_{vir}}{\rho^b_{vir}} = \frac{8\rho_{ta}}{1/4\rho^b_{ta}} = \frac{\rho_{ta}}{\rho^b_{ta}} \times 4 \times 8 = \frac{R_0^3}{R(z_{ta})^3} \times 4 \times 8, \tag{4.7}$$

and evaluating equation 4.4 at the turnaround, we obtain

$$\Delta_{vir} = \left(\frac{9}{16}\pi^2\right) \times 8 \times 4 = 18\pi^2. \tag{4.8}$$

This means that the density of the bound object after it virialised is $18\pi^2$ times the background density.
To conclude, it is also possible to infer the initial overdensity at the time of collapse by imposing $\theta = 2\pi$ in Equation 4.5:

$$\frac{\delta_c(z)}{1+z} = \frac{3}{5}\left(\frac{3}{2}\pi\right)^{2/3}. \tag{4.9}$$

If evaluated today – *i.e.*, at $z = 0$ – the critical overdensity at collapse is $\delta_c = 1.686$.

In the spherical collapse model framework, because of equation 4.3, there exists a relation between the actual overdensity δ, the one predicted by linear theory, δ_0, and this relation does not depend on R_0. In other words, the ratio R/R_0 is determined by δ_i, rather than R itself. Given that $M \propto R^3$, the critical density for collapse δ_c is the same for all objects and does not depend on their mass.

Recap of the different overdensity notation used within this section.

- δ_i is the initial overdensity at the initial time;
- δ_0 is the initial overdensity extrapolated at present time using linear theory;
- δ_c is the critical overdensity for collapse;
- δ is the overdensity evaluated at any time $t > t_0$.

4.2 MAIN INGREDIENTS

In the previous section I described a model providing a way to relate the initial perturbations in the matter density field to the actual virialised object.

The HM framework identifies those bound objects as dark matter haloes and relies on three main ingredients: the Halo Mass Function (HMF), the halo bias, and the halo density profile.
The HMF describes the comoving number density of bound objects of mass m at redshift z; the halo bias relates the dark matter halo overdensity to the matter overdensity, and the halo density profile describes the dark matter distribution around the halo centre.
In the following, I provide a detailed description of these ingredients as well as of the main models present in the literature, explaining their limitations and possibilities.

4.2.1 Halo Mass Function

The HMF, representing the comoving number density of bound objects – *i.e.* dark matter haloes – per unit mass at a given redshift, is related to the number of regions in the initial matter density field which were dense enough to start the collapse. Therefore, in order to estimate the number of bound objects, we first have to estimate the number density of regions in the initial matter density field that were dense enough to collapse. This can be done through a function describing the mass fraction of matter in peaks of a given height defined as:

$$\nu = \frac{\delta_{c,0}^2}{\sigma^2(m)}, \tag{4.10}$$

where $\delta_{c,0}$ is the critical overdensity at collapse evaluated at present time (so $\delta_{c,0} = 1.686$ from the previous section), and $\sigma^2(m)$ is the variance of the initial density field when smoothed with a top-hat filter of scale $R = \left(\frac{3}{4}\frac{m}{\pi\bar{\rho}}\right)^{1/3}$ extrapolated to present time through linear theory:

$$\sigma_{lin}^2(m) = \int \frac{dk}{k} \frac{k^3 P_{lin}(k)}{2\pi^2} |W(kR)|^2, \tag{4.11}$$

where $P_{lin}(k)$ is the linear matter power spectrum (described in section 3.2 of the previous chapter) and $W(kR)$ is the spatial filter defined as

$$W(x) = \left(\frac{3}{x^3}\right)[\sin(x) - x\cos(x)]. \tag{4.12}$$

A first model to relate the HMF, $dn/dm(m,z)$, to the function describing the number density of peaks, $\nu f(\nu)$, is provided in Press and Schechter, 1974 as:

$$\frac{m^2 dn/dm(m,z)}{\bar{\rho}} \frac{dm}{m} = \nu f(\nu) \frac{d\nu}{\nu}. \tag{4.13}$$

The function describing the number density of peaks in equation 4.13 is what actually predicts the shape of the specific HMF under study. There are several

analytical or semi-analytical models for the HMF present in the literature. Existing HMF formulations can be broadly classified into two categories: models motivated by excursion-set theory, which provide an analytic framework for halo formation but rely on simplifying assumptions, and empirical or semi-empirical fitting functions calibrated on N-body simulations, which offer higher accuracy over specific mass and redshift ranges at the expense of reduced theoretical transparency. In practice, the choice of a given HMF depends on the target mass and redshift ranges, as well as on the required accuracy for a specific application. In the following, I list and briefly discuss the main HMF models proposed in the literature, focusing mainly on the halo mass and redshift ranges of validity of the various HMF formulations.

- An excursion-set–motivated extension of Sheth and Tormen, 1999 propose an HMF modelling effective for a halo mass range of $\sim 10^{11} M_\odot h^{-1}$ to $\sim 10^{15} M_\odot h^{-1}$, and for $0 \leq z \leq 4$. Outside these mass and redshift ranges, the validity of their formulation degrades quickly.

- Empirical fits to N-body simulations Jenkins et al., 2001 present a universal HMF. The halo mass range at $z = 0$ goes from $3 \times 10^{11} M_\odot h^{-1}$ to $5 \times 10^{15} M_\odot h^{-1}$ and their universal fit has been tested in the redshift range $0 - 5$.

- Warren et al., 2006 provide a high-precision calibration of an HMF at $z = 0$ only, without giving a redshift-dependent calibration. At $z = 0$ the halo mass over which their parametrisation is effective is $\sim 10^{10} - 10^{15} M_\odot h^{-1}$. Other works use the results of Warren et al., 2006; specifically, Peacock, 2007 present an HMF based on a fit of their HMF, while Crocce et al., 2010 uses a functional form of Warren et al., 2006.

- Reed et al., 2007 provide a high-redshift, low-mass parameterisation of the HMF. Specifically, their formulation covers mass and redshift ranges of $10^5 - 10^{12} M_\odot h^{-1}$ and $10 - 30$, respectively.

- The HMF model proposed in T08 and T10 – Tinker et al., 2008 and Tinker et al., 2010a respectively – cover a halo mass range from $10^{11} M_\odot h^{-1}$ to $10^{15} M_\odot h^{-1}$ and redshift from 0 to 2.5. Outside this redshift range, the universality of the HMF does not hold and an extrapolation is needed. The difference between T08 and T10 only lies in the parametrisation of the HMF: T08 uses σ as the mass parameter, while T10 presents the HMF in terms of ν.

- Watson et al., 2013 explores a wide redshift range – from 0 up to 30 – and halo masses ranging from $7.3 \times 10^{14} M_\odot h^{-1}$ to $\sim 10^{15} M_\odot h^{-1}$.

- Despali et al., 2016 introduces a cosmology-dependent overdensity threshold in their HMF evaluation. The mass and redshift ranges of validity are $10^9 - 10^{15} M_\odot h^{-1}$ and $0 - 1.25$, respectively.

- Similarly, the HMF presented in Del Popolo et al., 2017 is a semi-analytical model based on a mass-dependent threshold.
- In Bocquet et al., 2020 the principal components of the HMF are directly emulated; the halo masses range from $10^{13} M_{\odot} h^{-1}$ to $10^{15} M_{\odot} h^{-1}$, and the redshift range is $0 - 2$.

In this work, I adopt the To8 HMF because it provides a robust and widely used calibration based on large-volume N-body simulations, with explicit dependence on the halo overdensity definition. This feature is particularly important in the present analysis, where different overdensity thresholds (e.g. $\Delta = 200$ and $\Delta = 500$) are required for consistency between CIB and thermal Sunyaev-Zel'dovich (tSZ) modelling. While more recent HMF formulations and emulator-based approaches exist, they are typically optimised for specific cosmological parameter spaces or restricted mass and redshift ranges, and do not offer a clear advantage for the mass scales and redshifts relevant to this work. In To8, the authors evaluate the HMF in a set of cosmological simulations of flat Λ-CDM cosmology, deriving fitting functions for a wide range of overdensities Δ. To obtain the To8 expression, one has to express equation 4.13 in terms of σ instead of ν[2], so that:

$$\frac{dn}{d\ln m} = -\frac{1}{2} f(\sigma, z) \frac{\bar{\rho}}{m} \frac{d\ln \sigma^2}{d\ln m}. \tag{4.14}$$

The expression of the function $f(\sigma, z)$[3] – representing the shape of the HMF – is expected to be universal with respect to redshift and cosmology and is parametrised as (see Tinker et al., 2008)

$$f(\sigma, z) = A(z) \left[\left(\frac{\sigma}{b(z)} \right)^{-a(z)} + 1 \right] e^{-c(z)/\sigma^2}, \tag{4.15}$$

where σ has been defined in equation 4.11. The other parameters that appear in the above equation – *i.e.* A, b, a, and c – are constants which define the HMF shape and which have to be calibrated using simulations. In particular, the A parameter determines the overall amplitude of the HMF; the parameters a and b set the slope and the amplitude of the low-mass power law, respectively; and c represents the cut-off scale at which the dark matter halo abundance exponentially decreases.
From simulations, it is evident that the HMF shows a redshift evolution,

2 In the formulation of the HMF, it is more convenient to use either σ or ν as mass variables, instead of the mass itself. The reason is that the first two variables are more stable and exhibit close-to-universal behaviour with respect to cosmology and redshift (see Asgari et al., 2023; Jenkins et al., 2001 as reference).

3 Details on how to move from the function describing the number density of peaks, $\nu f(\nu)$, to the functional form in terms of σ are provided in chapter a.

which is encoded in the redshift dependence of the parameters entering the definition of the $f(\sigma, z)$ function:

$$\begin{aligned} A(z) &= A_0(1+z)^{-0.14} \\ a(z) &= a_0(1+z)^{-0.06} \\ b(z) &= b_0(1+z)^{-\alpha}, \end{aligned} \tag{4.16}$$

where the power-law coefficient α depends on the density contrast Δ as

$$\log_{10}\alpha(\Delta) = -\left(\frac{0.75}{\log_{10}(\Delta/75)}\right)^{1.2}. \tag{4.17}$$

The values of A_0, a_0, b_0 and c_0 at redshift $z = 0$ for $\Delta = 200$ dark matter haloes, calibrated through simulations, are

$$A_0 = 0.186,\ a_0 = 1.47,\ b_0 = 2.57,\ c_0 = 1.19.$$

However, there is a slight dependence of the density contrast on cosmology, since $\Delta = 200/\Omega_m(z)$. For this reason, To8 considered a wide range of density contrasts and, in order to obtain the values of A_0, a_0, b_0 and c_0 at different Δ, an interpolation over the different density contrast values is required[4].

In the main work of this thesis, I implemented the To8 HMF. In figures 4.1 and 4.2, I show the To8 HMF predictions for different redshift values as a function of the mass, and for different halo mass values as a function of the redshift, respectively. It is possible to notice how the cut-off scale changes depending on both redshift and mass. Specifically, recalling that the cut-off scale determines the exponential decrease of dark matter halo abundance, from figure 4.1 we notice that the higher the redshift, the lower the cut-off mass is. If, instead, we look at figure 4.2, showing the HMF for fixed values of the mass as a function of redshift, we notice that the higher the mass, the earlier in redshift the cut-off is. Intuitively, going to higher redshifts implies that we see less massive dark matter haloes.

4.2.2 Halo Bias

From the previous section we obtain a tool telling us the comoving number density of dark matter haloes of mass m at a given z. I now move to the description of how to relate dark matter halo overdensities with matter overdensity, with dark matter being a biased tracer of the underlying matter density field.

Let us start by dividing the space into cells of comoving volume V. From the most general point of view, these cells have different masses and so different densities, so that $M/V \equiv \bar{\rho}(1+\delta)$. Let $N(m, z_1|M, V, z_0)$ be a function describing the average number of haloes of mass m that collapsed at z_1

4 To8 considered the following values of density contrast: $\Delta = 200, 400, 800, 1600$ and 3200.

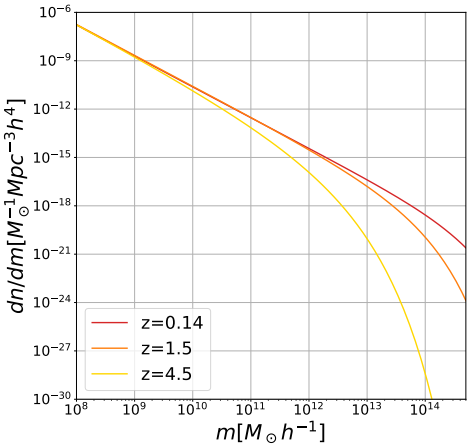

Figure 4.1: HMF as a function of the halo masses for different redshift values. $z = 0.14$ in red, $z = 1.5$ in orange, and $z = 4.5$ in yellow.

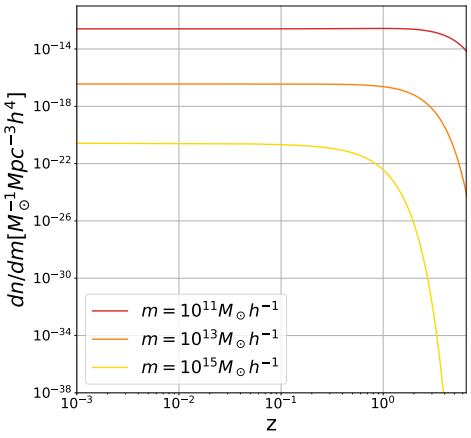

Figure 4.2: HMF as a function of the redshift for different halo masses. $m = 10^{11}M_\odot h^{-1}$ in red, $m = 10^{13}M_\odot h^{-1}$ in orange, and $m = 10^{15}M_\odot h^{-1}$ in yellow.

and occupy cells of size V and mass M today, *i.e.* z_0. The dark matter halo overdensity in such cells can therefore be written as

$$\delta_h(m, z_1|M, V, z_0) = \frac{N(m, z_1|M, V, z_0)}{dn/dm(m, z_1)} \frac{1}{V} - 1. \tag{4.18}$$

The denominator of equation 4.18 is known from the previous section, so one has to find an expression for $N(m, z_1|M, V, z_0)$.
Dark matter haloes are regions that were dense enough to collapse, so the number of haloes in a cell of size V equals the volume of the cell times the number density of regions within the cell that were sufficiently dense to collapse and form haloes, meaning

$$N(m, z_1|M, V, z_0) = dn/dm(m, z_1|M, V, z_0)V(1 + \delta). \tag{4.19}$$

In the above equation, $dn/dm(m, z_1|M, V, z_0)$ is the HMF conditioned on the fact that the haloes under scrutiny have to occupy the cell we are considering. In this case, the HMF is a function of the critical overdensity required for a collapse at that time, so its computation comes after replacing $\delta_c(z)$ with $\delta_c(z_1) - \delta_0(\delta, z_0)$ (see Mo and White, 1996). Here $\delta_0(\delta, z_0)$ represents the initial overdensity extrapolated with linear theory that a region must have in order to have overdensity δ at z_0.
An estimate of the conditioned HMF which appears in equation 4.19 can be obtained by replacing $\delta_c(z)$ in equation 4.13 by

$$\frac{m^2 dn/dm(m, z_1|M, V, z_0)}{\bar{\rho}} \frac{dm}{m} = \nu_{1,0} f(\nu_{1,0}) \frac{d\nu_{1,0}}{\nu_{1,0}}, \tag{4.20}$$

where

$$\nu_{1,0} = \frac{[\delta_c(z_1) - \delta_0(\delta, z_0)]^2}{\sigma^2(m) - \sigma^2(M)}. \tag{4.21}$$

At this point, it is interesting to consider two limits.

- The small cell limit, meaning $V \to 0$, $\delta \to \infty$, and $\delta_0 \to \delta_c(z_0)$; if we consider a region of mass M, that region is what we defined to be a halo of the same mass M. Therefore, $N(m, z_1|M, V = 0, z_0)$ represents the average number of the halo progenitors of mass m which eventually merged to form the dark matter halo of mass M. In other words, the small cell limit provides information related to the halo evolution.
- The large cell limit, satisfying the conditions that $V \to \infty$ and $M \to \infty$, leads to $\sigma^2(M) \to 0$ and $|\delta| \to 0$. In this regime, equation 4.20 reduces to $dn/dm(m, z_1)$.

Focusing on the large cell limit, and working under the assumptions that $\sigma^2(M) \to 0$ and $|\delta| \to 0$, we can expand equation 4.20 as

$$\frac{dn}{dm}(m, z_1|M, V, z_0) \approx \frac{dn}{dm}(m, z_1) - \delta_0(\delta, z_0)\left(\frac{\partial(dn/dm(m, z_1))}{\partial\delta_c}\right)\Bigg|_{\delta_c(z_1)}, \quad (4.22)$$

which, in terms of halo overdensity, is

$$\delta_h(m, z_1|M, V, z_0) \approx \delta - (1+\delta)\delta_0(\delta, z_0)\left(\frac{\partial \ln(dn/dm(m, z_1))}{\partial\delta_c}\right)\Bigg|_{\delta_c(z_1)}. \quad (4.23)$$

If we now focus on the lowest order in the expansion and explicitly replace the expression of the HMF in the above equation, we obtain a relation of the type

$$\delta_h(m, z_1|M, V, z_0) \approx b(m, z_1)\delta, \quad (4.24)$$

where $b(m, z_1)$ is the halo bias, representing the relation between dark matter haloes and matter overdensities.

Similarly to the HMF, there are various formulations of the halo bias. For the sake of coherence, I now focus on the T10 formulation, which is also the one employed in my work. Specifically, the halo bias reads

$$b(m, z_1) = 1 - A\frac{\nu^a}{\nu^a + \delta_c^a} + B\nu^b + C\nu^c, \quad (4.25)$$

where the constants in equation 4.25 depend on the density contrast under consideration. Taking the specific case of a halo with density contrast $\Delta = 200$, the constants in equation 4.25, calibrated with simulations, are:

$$A = 1.04,\ a = 0.132,\ B = 0.183,\ b = 1.5, C = 0.262,\ c = 2.4.$$

In figures 4.3 and 4.4, I show the predictions for the halo bias for the T10 model. Figures 4.3 and 4.4 show the halo bias for different redshift values as a function of the halo masses, and for different halo masses as a function of the redshifts, respectively.
It is clear from both figures that the higher the halo mass, the larger the

halo bias. In particular, since $b(m) \gg 1$ for the most massive haloes, as a consequence, it is possible to state that the most massive haloes also occupy the densest cells. Furthermore, in linear theory the halo bias quantifies how the spatial fluctuations of haloes trace those of the underlying matter field. In particular, the two-point correlation function of haloes is amplified with respect to that of matter by a factor b^2, implying that haloes residing in denser regions exhibit stronger spatial clustering. As a consequence, the most massive haloes, which are characterised by large bias values, are also the most strongly clustered objects.

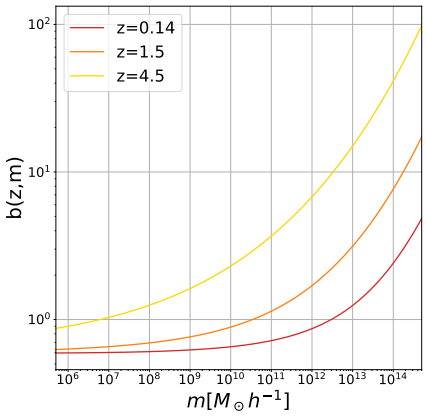

Figure 4.3: Halo bias as a function of the halo masses for different redshift values. $z = 0.14$ in red, $z = 1.5$ in orange, and $z = 4.5$ in yellow.

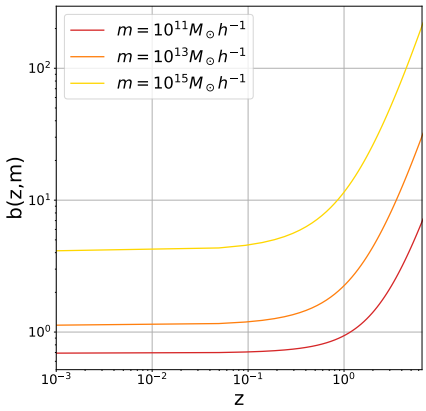

Figure 4.4: Halo bias as a function of the redshift for different halo masses. $m = 10^{11} M_\odot h^{-1}$ in red, $m = 10^{13} M_\odot h^{-1}$ in orange, and $m = 10^{15} M_\odot h^{-1}$ in yellow.

4.2.3 Halo Density Profile

The halo density profile describes the radial distribution of dark matter around the halo centre.

Previous works, such as those by Bertschinger, 1985; Fillmore and Goldreich, 1984, suggest that the radial density profile of a dark matter halo, which describes its inner structure, depends on its formation time and on the initial density distribution of the region that collapsed. Since in this framework dark matter haloes are identified as peaks in the initial matter density field, the more massive the halo, the higher the peak in the initial fluctuation field. Furthermore, the density around a high peak is shallower than the one around a smaller peak (see Bardeen et al., 1986). This explains the prediction, confirmed by Navarro et al., 1996, that more massive haloes are less concentrated than low-mass haloes.

The last aspect which is worth considering in order to better capture the inner structure of dark matter haloes is that the halo density at turnaround is proportional to the density of the Universe at its formation. As a consequence,

early-forming structures are more concentrated, while late-time structures tend to be larger and less concentrated.

To describe the density profile of a dark matter halo as a function of its radius, the most commonly used functional form, able to capture the expected behaviour of the density around the halo centre, is

$$\rho(r|m) = \frac{\rho_s}{(r/r_s)^{\alpha}(1+r/r_s)^{\beta}}, \tag{4.26}$$

where r_s and ρ_s define the scale radius and the density evaluated at that scale radius, respectively.
Depending on the values of (α, β) we have different halo profiles. Specifically, $(\alpha, \beta) = (1, 3)$ corresponds to the prescription of Hernquist, 1990, $(\alpha, \beta) = (3/2, 3/2)$ is the formulation presented in Moore et al., 1999, while $(\alpha, \beta) = (1, 2)$ corresponds to the so-called Navarro–Frenk–White (NFW) profile, which is the functional form presented in Navarro et al., 1996, and employed in this work.

Therefore, for a NFW profile, equation 4.26 becomes

$$\rho(r|m) = \frac{\rho_s}{(r/r_s)(1+r/r_s)^2}. \tag{4.27}$$

The NFW profile scales as $\approx r^{-3}$, leading to a logarithmic divergence of the mass, computed as $m = \int_0^{\infty} 4\pi r^2 \rho(r|m) dr$, at large radii. In order to prevent the logarithmic divergence, the edge of a dark matter halo is its virial radius, and the computation of its mass is therefore truncated at r_{vir} so that

$$m = \int_0^{r_{vir}} 4\pi r^2 \rho(r|m) dr. \tag{4.28}$$

By substituting equation 4.26 in the above expression, we obtain

$$m = 4\pi \rho_s r_s^3 \left[\ln(1+c) - \frac{c}{1+c} \right], \tag{4.29}$$

where $c = r_{vir}/r_s$ is known as the concentration parameter.
Simulations (see Bullock et al., 2001a; Jing, 1999) show that for haloes of the same mass, their concentration is well described by a log-normal distribution with mean $\bar{c}(m, z)$ and width σ_c^2:

$$p(c|m, z) d\ln c = \frac{d\ln c}{\sqrt{2\pi\sigma_c^2}} \exp\left\{ -\frac{\ln^2[c/\bar{c}(m,z)]}{2\sigma_c^2} \right\}. \tag{4.30}$$

The mean concentration $\bar{c}$ depends on both the halo mass and redshift, while the width of the distribution does not.
For a NFW profile, the mean concentration parameter is parametrised following Bullock et al., 2001b as

$$\bar{c}(m, z) = \frac{9}{1+z} \left[\frac{m}{m_*(z)} \right]^{-0.13}, \tag{4.31}$$

where $m_*(z)$ is a characteristic scale so that $\nu(m,z) = 1$. In the parametrisation of Bullock et al., 2001b the distribution width is taken as $\sigma_c^2 \sim 0.25$. Equation 4.31 confirms what was stated at the beginning of this section, *i.e.* that the higher the halo mass, the smaller the concentration.

For reasons that will be clear in the next section, it is useful to have the expression of the normalised Fourier Transform (FT) of the density profile

$$\tilde{u}(\vec{k}|m) = \frac{\int d^3\vec{x}\rho(\vec{x}|m)e^{-i\vec{k}\cdot\vec{x}}}{\int d^3\vec{x}\rho(\vec{x}|m)}, \tag{4.32}$$

where the numerator represents the unnormalised FT, $u(\vec{k}|m)$. Assuming spherical symmetry, the above equation simplifies to

$$\tilde{u}(k,m) = \int_0^{r_{vir}} dr 4\pi r^2 \frac{\sin kr}{kr}\frac{\rho(r|m)}{m}, \tag{4.33}$$

and, computing the integral for the specific case of a NFW profile, we obtain

$$\begin{aligned}\tilde{u}(k,m) = \mathcal{A}\Big\{&\sin(kr_s)[Si(kr_s[1+c]) - Si(kr_s)] - \frac{\sin(ckr_s)}{kr_s(1+c)} \\ &+ \cos(kr_s)[Ci(kr_s[1+c]) - Ci(kr_s)]\Big\},\end{aligned} \tag{4.34}$$

where $\mathcal{A}$ is the NFW profile amplitude evaluated as $\dfrac{1}{\ln(1+c) - \dfrac{c}{1+c}}$, and $Si(x), Ci(x)$ are the sine and cosine integrals defined as

$$Si(x) = \int_0^x \frac{\sin(t)}{t}dt,\ Ci(x) = -\int_0^x \frac{\cos(t)}{t}dt.$$

Figures 4.5 and 4.6 show the predictions for the FT of the NFW profile evaluated at a fixed halo mass of $m = 10^{12}M_\odot h^{-1}$ for three different redshift values, and at a fixed redshift $z = 1.5$ for three different halo masses, respectively.

Figures 4.5 and 4.6 present the FT of the normalised NFW profile $\tilde{\rho}(r|m)$ – meaning $\tilde{\rho}(r|m) = \rho(r|m)/m$ –, such that

$$\int 4\pi r^2 dr\tilde{\rho}(r|m) \equiv 1. \tag{4.35}$$

As a consequence of equation 4.35, $\tilde{u}(k \to 0|m) = 1$, and this explains the behaviour we observe at $k \to 0$. In other words, at large scales a halo acts as a point mass, *i.e.* assuming a constant behaviour in Fourier space.

4.3 SPATIAL POWER SPECTRA COMPUTATION

In this section, I present how it is possible to obtain predictions for non-linear power spectra within an HM framework. In the first part, I follow the

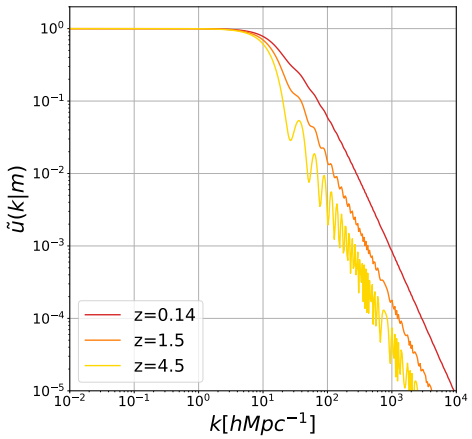

Figure 4.5: Fourier transform of the NFW profile at a fixed halo mass of $m = 10^{12} M_\odot h^{-1}$ for different redshift values. $z = 0.14$ in red, $z = 1.5$ in orange, and $z = 4.5$ in yellow.

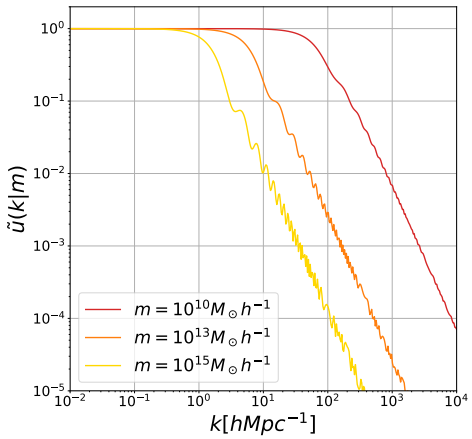

Figure 4.6: Fourier transform of the NFW profile at a fixed redshift $z = 1.5$ for different halo masses. $m = 10^{10} M_\odot h^{-1}$ in red, $m = 10^{13} M_\odot h^{-1}$ in orange, and $m = 10^{15} M_\odot h^{-1}$ in yellow.

approach of Asgari et al., 2023 and present the HM predictions for the power spectrum between two generic fields, u and v.

Let us start from a generic field in real space u, identified with $\theta_u(\vec{x})$. Working under the HM assumption that everything in the field is contained within haloes distributed in the Universe with a density profile $\rho(\vec{x}|m)$ then

$$\theta_u(\vec{x}) = \sum_i N_i \rho_{u,i}(|\vec{x} - \vec{x}_i||m), \tag{4.36}$$

where the sum runs over all the volume elements and $N_i = \{0, 1\}$ depending on whether there is a halo in the volume element under scrutiny.
The correlation between two fields in real space, u and v, is

$$\langle \theta_u(\vec{x}) \theta_v(\vec{x}') \rangle = \xi_{uv}(|\vec{x} - \vec{x}'|), \tag{4.37}$$

with $\xi_{uv}(|\vec{x} - \vec{x}'|)$ being the two-point correlation function between the fields u and v, which depends only on the separation $|\vec{x} - \vec{x}'|$.

Moving the discussion to Fourier space, equation 4.36 becomes

$$\hat{\theta}_u(\vec{k}) \equiv \int \theta_u(\vec{x}) e^{-i\vec{k}\cdot\vec{x}} d^3\vec{x} = \sum_i N_i \int \rho_{u,i}(|\vec{x} - \vec{x}_i||m) e^{-i\vec{k}\cdot\vec{x}} d^3\vec{x}. \tag{4.38}$$

Replacing $\vec{x} - \vec{x}_i = \vec{r}$, the above expression reads

$$\hat{\theta}_u(\vec{k}) = \sum_i N_i e^{-i\vec{k}\cdot\vec{x}_i} \int d^3\vec{r} \rho_{u,i}(|r|) e^{-i\vec{k}\cdot\vec{x}}. \tag{4.39}$$

The integral corresponds to the numerator of equation 4.32, which is the unnormalised FT of the density profile, $u(\vec{k}, m)$, so that

$$\hat{\theta}_u(\vec{k}) = \sum_i N_i e^{-i\vec{k}\cdot\vec{x}_i} u(\vec{k}, m). \tag{4.40}$$

In the same way, equation 4.37 in Fourier space reads

$$\langle \hat{\theta}_u(\vec{k})\hat{\theta}_v^*(\vec{k}')\rangle = (2\pi)^3\delta_D(\vec{k}-\vec{k}')P_{uv}(k), \tag{4.41}$$

where $P_{uv}(k)$ is the power spectrum. In order to find an HM expression for it, we replace the expression of the FT of a generic field from equation 4.40 in the above equation, obtaining

$$\langle \hat{\theta}_u(\vec{k})\hat{\theta}_v^*(\vec{k}')\rangle = \Big\langle \sum_{i,j} N_i N_j e^{-i\vec{k}\cdot\vec{x}_i} e^{+i\vec{k}'\cdot\vec{x}_j} u_u(k,m)u_v(k',m)\Big\rangle. \tag{4.42}$$

We split the sum into the contributions from $i = j$ and $i \neq j$. The first term describes the field correlation within a single halo and is therefore called the One-halo (1h) term. The second encodes the contribution of field correlation between distinct haloes and is the Two-halo (2h) term.

ONE-HALO TERM When considering $i = j$, equation 4.42 reads

$$\begin{aligned}\langle \hat{\theta}_u(\vec{k})\hat{\theta}_v^*(\vec{k}')\rangle &= \Big\langle \sum_i N_i^2 e^{-i\vec{k}\cdot\vec{x}_i} e^{+i\vec{k}'\cdot\vec{x}_i} u_u(k,m)u_v(k',m)\Big\rangle \\ &= \Big\langle \sum_i N_i e^{-i(\vec{k}-\vec{k}')\cdot\vec{x}_i} u_u(k,m)u_v(k',m)\Big\rangle,\end{aligned} \tag{4.43}$$

where in the last equality I used $N_i^2 = N_i$. Rewriting the ensemble average, with $N_i = \int dm \frac{dn}{dm}(m,z)\Delta V_i$ (with V_i denoting the volume element) and, consequently, writing the sum over the volume elements as an integral over the volume, then

$$\begin{aligned}\langle \hat{\theta}_u(\vec{k})\hat{\theta}_v^*(\vec{k}')\rangle &= \int dm \frac{dn}{dm}(m,z)u_u(k,m)u_v(k',m)\int d^3x e^{-i(\vec{k}-\vec{k}')\cdot\vec{x}_i} \\ &= \int dm \frac{dn}{dm}(m,z)u_u(k,m)u_v(k,m)(2\pi)^3\delta_D(\vec{k}-\vec{k}').\end{aligned} \tag{4.44}$$

The expression for the 1h term in an HM framework can be obtained by equating the above equation with equation 4.41, so that

$$(2\pi)^3\delta_D(\vec{k}-\vec{k}')P_{uv}^{1h}(k) = \int dm \frac{dn}{dm}(m,z)u_u(k,m)u_v(k,m)(2\pi)^3\delta_D(\vec{k}-\vec{k}'), \tag{4.45}$$

and, consequently, obtain

$$P_{uv}^{1h}(k) = \int dm \frac{dn}{dm}(m,z)u_u(k,m)u_v(k,m). \tag{4.46}$$

It is interesting to note that, as $k \to 0$, the dependence on the halo profile vanishes (see last part of the discussion in section 4.2.3); therefore, at large scales the 1h term only contributes as a constant $P(k)$. This behaviour can be visualised in figure 4.8 in the next section for the matter power spectrum.

TWO-HALO TERM Moving now to the $i \neq j$ case, equation 4.42 becomes

$$\langle \hat{\theta}_u(\vec{k})\hat{\theta}_v^*(\vec{k'})\rangle = \langle \sum_i \sum_j N_i N_j e^{-i\vec{k}\cdot\vec{x}_i} e^{+i\vec{k'}\cdot\vec{x}_j} u_u(k, m_1) u_v(k', m_2)\rangle. \tag{4.47}$$

Rewriting the ensemble average and the volume integral as in the 1h case, we obtain

$$\begin{aligned}\langle \hat{\theta}_u(\vec{k})\hat{\theta}_v^*(\vec{k'})\rangle = \int dm_1 \frac{dn}{dm}(m_1, z) \int dm_2 \frac{dn}{dm}(m_2, z) u_u(k, m_1) u_v(k', m_2) \\ \int d^3x \int d^3x' e^{-i\vec{k}\cdot\vec{x}} e^{+i\vec{k'}\cdot\vec{x'}} \langle N(m_1, \vec{x}) N(m_2, \vec{x'})\rangle,\end{aligned} \tag{4.48}$$

where in the angular brackets we also considered the correlation between the halo positions. Now, noting that

$$\begin{aligned}&\int d^3x \int d^3x' e^{-i\vec{k}\cdot\vec{x}} e^{+i\vec{k'}\cdot\vec{x'}} \langle N(m_1, \vec{x}) N(m_2, \vec{x'})\rangle \equiv \langle \hat{N}(m_1, k)\hat{N}^*(m_2, k)\rangle \\ &= (2\pi)^3 \delta_D(\vec{k} - \vec{k'}) P_{hh}(m_1, m_2, k),\end{aligned} \tag{4.49}$$

where in the last equivalence we used the definition of power spectrum in equation 4.41, and $P_{hh}(m_1, m_2, k)$ is therefore the power spectrum of the halo centres. Replacing equation 4.49 in equation 4.48, and equating the latter with 4.41, we obtain the expression for the 2h term evaluated in an HM framework between two generic fields, u and v, as

$$P_{uv}^{2h}(k) = \int\int dm_1 dm_2 \frac{dn}{dm}(m_1, z) \frac{dn}{dm}(m_2, z) u_u(k, m_1) u_v(k, m_2) P_{hh}(m_1, m_2, k). \tag{4.50}$$

Haloes are biased tracers of the underlying matter density field, so that

$$P_{hh}(m_1, m_2, k) = b(m_1) b(m_2) P_{mm}^{lin}(k), \tag{4.51}$$

with $P_{mm}^{lin}(k)$ being the linear matter power spectrum and $b(m)$ the halo bias introduced in section 4.2.2. Therefore the 2h term is

$$\begin{aligned}P_{uv}^{2h}(k) = P_{mm}^{lin}(k) \int dm_1 \frac{dn}{dm}(m_1, z) b(m_1) u_u(k, m_1) \times \\ \int dm_2 \frac{dn}{dm}(m_2, z) b(m_2) u_v(k, m_2).\end{aligned} \tag{4.52}$$

The final expression of the power spectrum between two generic fields u and v in an HM framework is therefore obtained by adding together the 1h and 2h terms as

$$P_{uv}(k) = P_{uv}^{1h}(k) + P_{uv}^{2h}(k). \tag{4.53}$$

In figure 4.7 I show an intuitive representation of the contributions arising from the 1h and the 2h terms when the two generic u and v fields are both represented by the galaxy field.

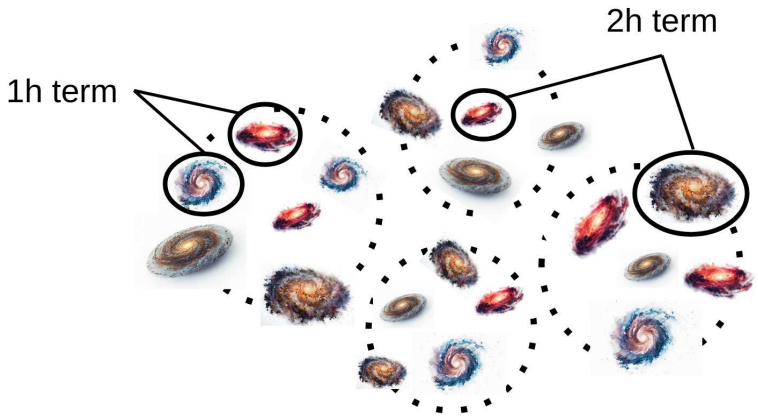

Figure 4.7: Schematic representation of the 1h and 2h terms contributions in case the u and v fields are the both the galaxy field.

In the following, I present the HM predictions for two significant cases. In section 4.3.1, I present the calculations when $u, v = m, m$, *i.e.* the case where both fields are the continuous matter density field, leading to the formulation of the matter power spectrum. The other case worth noticing is presented in section 4.3.2, where the two fields involved in the HM framework are the discrete galaxy field, *i.e.* $u, v = g, g$.

4.3.1 Matter Power Spectrum

Matter is a continuous field and, in order to have a prediction of its 1h and 2h terms starting from the most general expressions in equations 4.46 and 4.52, we only have to set the proper definition of the FT of the NFW profile in the case where $u_u(k, m) = u_v(k, m) \equiv u_m(k, m)$. Recalling that $u_m(k, m)$ represents the unnormalised FT of the NFW profile, then it can be written as an amplitude multiplying the normalised profile, so that

$$u_m(k, m) = \frac{m}{\bar{\rho}} \tilde{u}_m(k, m). \tag{4.54}$$

Making the above expression explicit in equations 4.46 and 4.52, we obtain

$$P^{1h}_{mm}(k) = \frac{1}{\bar{\rho}^2} \int dm \frac{dn}{dm}(m, z) m^2 \tilde{u}^2_m(k, m), \tag{4.55}$$

and

$$P^{2h}_{mm}(k) = P^{lin}_{mm}(k) \left[\frac{1}{\bar{\rho}} \int dm \frac{dn}{dm}(m, z) b(m) m \tilde{u}_m(k, m) \right]^2. \tag{4.56}$$

In the case of the matter distribution, the HMF and the halo bias must satisfy two conditions for the HM predictions to hold.
The first one makes explicit the HM assumption, ensuring that all matter in the Universe is contained in dark matter haloes:

$$\int dm \frac{dn}{dm}(m, z) m = \bar{\rho}. \tag{4.57}$$

The second condition is that matter is an unbiased tracer with respect to itself; therefore,

$$\int dm \frac{dn}{dm}(m,z)b(m)m = \bar{\rho}. \tag{4.58}$$

By making use of this last property, it is possible to clearly see that for $k \to 0$, $P^{2h}_{mm}(k) \to P^{lin}_{mm}(k)$[5]. As we will see in the next sections, this property does not apply to the other cases, since galaxies are biased tracers of the underlying matter density field.

Figure 4.8 shows the HM predictions for $P^{1h}_{mm}(k)$ and $P^{2h}_{mm}(k)$. At large scales, *i.e.* when $k \to 0$, it is possible to appreciate the flat behaviour in Fourier space of the 1h term, and the 2h term approaching the linear matter power spectrum.

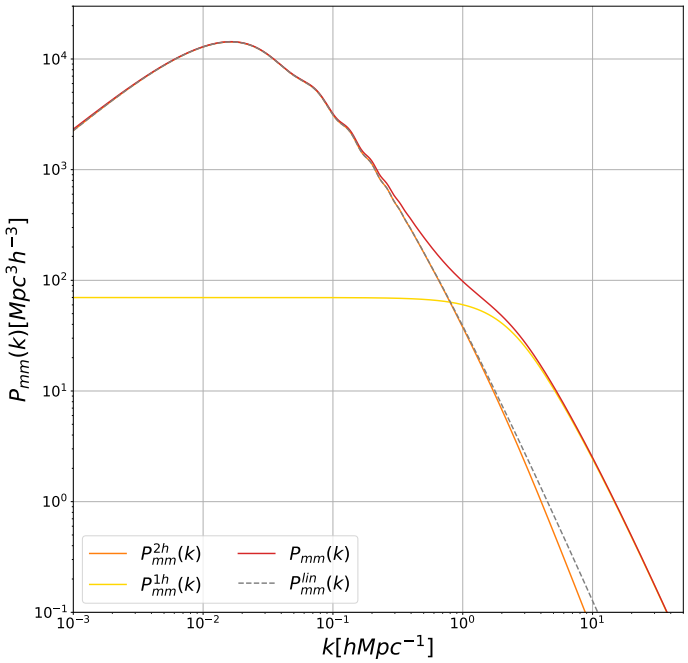

Figure 4.8: Prediction of the matter power spectrum evaluated in an HM framework at redshift $z = 0.5$. The 1h and 2h terms are displayed in yellow and orange, respectively; the total matter power spectrum given by the sum of the two is in red, while the grey dashed line represents the linear matter power spectrum computed using the `CAMB` Boltzmann solver.

4.3.2 Galaxy Power Spectrum

The case of galaxies requires a bit more attention. Galaxies are a discrete tracer of the matter density field; therefore, instead of just defining a proper FT of their density distribution, we have to account for the auto-correlation that a discrete field has with itself at zero separation $r = 0$. In Fourier space,

5 The normalised FT of the NFW profile appears inside the integral, therefore $\tilde{u}_m(k \to 0, m) = 1$.

this translates into a constant contribution over all wavenumbers k, which acts as a shot noise term given by

$$P_{gg}^{SN} = \frac{1}{\bar{n}}, \tag{4.59}$$

with $\bar{n}$ the average number density of galaxies. Given the galaxy field in real space defined as

$$\theta_g(\vec{x}) = \sum_i N_{g,i}(m)\rho_g(|\vec{x} - \vec{x}'|, m), \tag{4.60}$$

with $N_g(m)$ the number of galaxies in a halo with mass m, the galaxy number density $\bar{n}_g$ can be obtained by averaging the galaxy field:

$$\begin{aligned}\langle\theta_g(\vec{x})\rangle &= \langle\sum_i N_{g,i}(m)\rho_g(|\vec{x} - \vec{x}'|, m)\rangle \\ &= \int dm \frac{dn}{dm}(m,z) \int d^3x \rho_g(x,m).\end{aligned} \tag{4.61}$$

Rewriting the density profile in terms of its normalised version, $\rho_g(x,m) = N_g(m)\tilde{\rho}_g(x,m)$, we obtain

$$\begin{aligned}\langle\theta_g(\vec{x})\rangle &= \int dm \frac{dn}{dm}(m,z) N_g(m) \int d^3x \tilde{\rho}_g(x,m) \\ &= \int dm \frac{dn}{dm}(m,z) N_g(m) \equiv \bar{n}_g.\end{aligned} \tag{4.62}$$

In order to properly write the 1h term, we have to separate the cross-galaxy term from the auto-galaxy term by removing an $N_g(m)\tilde{u}_g(k,m)\frac{dn}{dm}$ contribution, and replace it with a shot noise $N_g(m)\frac{dn}{dm}$, so as to obtain

$$\begin{aligned}P_{gg}^{1h}(k) &= \frac{1}{\bar{n}_g^2} \int dm \frac{dn}{dm} [N_g^2(m)\tilde{u}_g^2(k,m) - N_g(m)\tilde{u}_g^2(k,m) + N_g(m)] \\ &= \frac{1}{\bar{n}_g^2} \int dm \frac{dn}{dm} N_g(m)(N_g(m) - 1)\tilde{u}_g^2(k,m) + \frac{1}{\bar{n}_g^2} \int dm \frac{dn}{dm} N_g(m) \\ &= \frac{1}{\bar{n}_g^2} \int dm \frac{dn}{dm} N_g(m)(N_g(m) - 1)\tilde{u}_g^2(k,m) + \frac{1}{\bar{n}_g},\end{aligned} \tag{4.63}$$

where in the last equivalence we make use of the definition of the average galaxy number density provided in equation 4.62. The last term is the shot noise. As explained in Asgari et al., 2023, this term is usually removed from datasets. Therefore, in the following expressions of the 1h terms, it will not be included.

In galaxy population modelling, it is a usual choice to separate galaxies into central, N_c, and satellite, N_s, galaxies, assuming that the first galaxy populating a dark matter halo is a central one, and that all the other galaxies are satellite galaxies[6]. This implies that the number of central galaxies is

6 Details about the prescription to model the number of both central and satellite galaxies based on the halo mass is encapsulated in the Halo Occupation Distribution (HOD) and is fully described in section 4.3.2.1

$N_c = 0, 1$. I provide details about the modelling of galaxy clustering in dark matter haloes later in section 4.3.2.1. For now, let us focus on equation 4.63 and explicitly write the number of galaxies as $N_g(m) = N_c(m) + N_s(m)$.

$$\begin{aligned} N_g(m)(N_g(m)-1)\tilde{u}_g^2(k,m) \equiv & [N_c(m)\tilde{u}_c(k,m) + N_s(m)\tilde{u}_s(k,m)] \\ & [N_c(m)\tilde{u}_c(k,m) + N_s(m)\tilde{u}_s(k,m) - 1] \\ = & N_c(m)\tilde{u}_c(k,m)[N_c(m)\tilde{u}_c(k,m) - 1] + \\ & N_s(m)\tilde{u}_s(k,m)[N_s(m)\tilde{u}_s(k,m) - 1] + \\ & 2N_c(m)N_s(m)\tilde{u}_c(k,m)\tilde{u}_s(k,m). \end{aligned} \tag{4.64}$$

Now assuming that

- $N_c(m)(N_c(m) - 1) \equiv 0$, because the number of central galaxies is either 0 or 1,
- the satellite galaxies follow Poisson statistics so that $N_s(m)(N_s(m) - 1) \approx N_s^2(m)$,
- the FT of the density profile of a central galaxy is just unity, $\tilde{u}_c(k,m) = 1$,
- we can equate the density profile of satellite galaxies to that of matter as $\tilde{u}_s(k,m) \equiv \tilde{u}(k,m)$,

then the above equation becomes

$$N_g(m)(N_g(m)-1)\tilde{u}_g^2(k,m) = 2N_c(m)N_s(m)\tilde{u}(k,m) + N_s^2(m)\tilde{u}^2(k,m). \tag{4.65}$$

Replacing this expression in equation 4.63, we finally obtain the expression of the 1h term for the galaxy field as

$$P_{gg}^{1h}(k) = \frac{1}{\bar{n}_g^2}\int dm \frac{dn}{dm}[2N_c(m)N_s(m)\tilde{u}(k,m) + N_s^2(m)\tilde{u}^2(k,m)]. \tag{4.66}$$

The computation of the 2h term for the galaxy field in an HM framework is more straightforward because there is no auto-correlation term to be considered. Therefore, the final expression can be obtained from the most general equation 4.52 by replacing $u_u(k,m) \to \dfrac{N_g(m)}{\bar{n}_g}\tilde{u}(k,m)$:

$$P_{gg}^{2h}(k) = P_{mm}^{lin}(k)\left[\frac{1}{\bar{n}_g}\int dm \frac{dn}{dm} b(m) N_g(m)\tilde{u}(k,m)\right]^2. \tag{4.67}$$

The final galaxy power spectrum in this scenario is then obtained by the sum of the 1h and 2h terms as

$$P_{gg}(k) = P_{gg}^{1h}(k) + P_{gg}^{2h}(k). \tag{4.68}$$

The HM predictions for the galaxy power spectrum are shown in figure 4.9. The dashed line serves as a reference and represents the linear matter power spectrum. It is possible to notice that, in this case, the 2h term does not overlap with the linear matter power spectrum as $k \to 0$ as it does for the case of matter (see figure 4.8). This happens because galaxies are a biased tracer of the underlying matter density field, while the matter power spectrum in the HM framework has been computed by imposing that matter is unbiased with respect to itself (see equation 4.58).

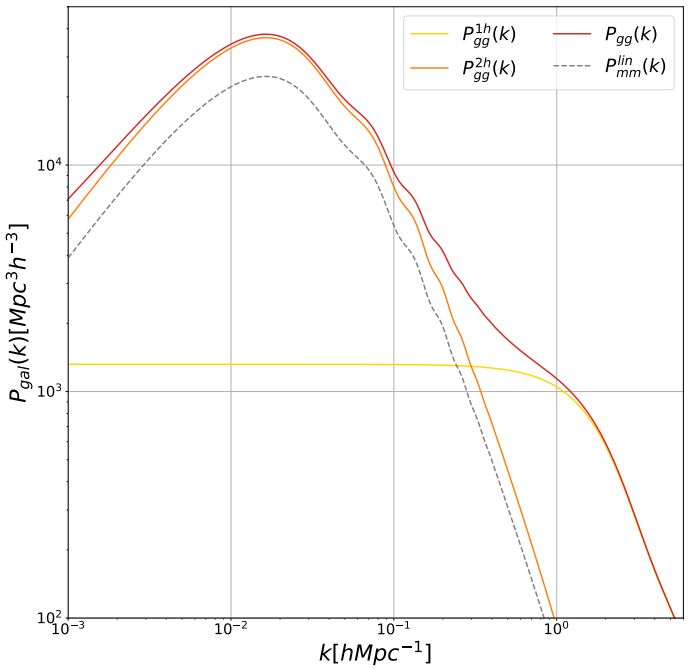

Figure 4.9: Prediction of the galaxy power spectrum evaluated in an HM framework at redshift $z = 0.001$. The 1h and 2h terms are displayed in yellow and orange, respectively, and the total galaxy power spectrum given by the sum of the two is shown in red. The grey dashed curve represents the CAMB output of the linear matter power spectrum and serves as a reference to better visualise the fact that galaxies are a biased tracer of the underlying matter density field.

An interesting case which will be relevant for the work presented in this thesis is the one encoding for two different discrete fields – *e.g.*, two distinct galaxy populations. In this case, also the computation of the 1h term naturally comes from the most general HM predictions of equation 4.46, because the auto-correlation term is no more present. When evaluating the 1h and 2h terms in an HM framework between two discrete tracers, g and g', we end up with

$$P_{gg'}^{1h}(k) = \frac{1}{\bar{n}_g}\frac{1}{\bar{n}_{g'}}\int dm \frac{dn}{dm}[(N_c^g(m)N_s^{g'}(m) + N_s^g(m)N_c^{g'}(m))\tilde{u}(k,m) + N_s^g(m)N_s^{g'}(m)\tilde{u}^2(k,m)], \tag{4.69}$$

for the 1h term, and

$$P^{2h}_{gg'}(k) = P^{lin}_{mm}(k)\left[\frac{1}{\bar{n}_g}\int dm \frac{dn}{dm} b(m) N_g(m) \tilde{u}(k,m)\right] \times \left[\frac{1}{\bar{n}_{g'}}\int dm \frac{dn}{dm} b(m) N_{g'}(m) \tilde{u}(k,m)\right], \tag{4.70}$$

for the 2h term.

4.3.2.1 Halo Occupation Distribution

One of the predictions of the HM is the fact that baryonic gas cools and, consequently, forms stars only in virialised dark matter haloes. Thus, the number of galaxies within a dark matter halo is related to the mass of the host halo itself. Furthermore, the HM framework assumes that the first galaxy forms at the centre of the hosting halo and it is what I previously called the central galaxy. All the other galaxies are satellite galaxies.

The information on how central and satellite galaxies populate dark matter haloes of a given mass m is provided by the HOD. The proposed models for the HOD in the literature are numerous. In the following, I list the main features of the several models, but I am going to focus on the prescription provided in Tinker and Wetzel, 2010 since it is the functional form most used in the literature.

- Zehavi et al., 2005 model encodes a simple central step and a Poisson distribution of satellite galaxies.
- Scoccimarro et al., 2001 does not explicitly model a central-satellite split of the galaxies populating dark matter haloes.
- In Berlind et al., 2003 there is the central-satellite split, but their results depend on simulations.

Tinker and Wetzel, 2010 model the number of central galaxies as

$$\langle N_{\rm cent}|M\rangle = \frac{1}{2}\left[1 + {\rm erf}\left(\frac{\log M - \log M_{\rm min}}{\sigma_{\log M}}\right)\right], \tag{4.71}$$

and the number of satellite galaxies as

$$\langle N_{\rm sat}|M\rangle = \frac{1}{2}\left[1 + {\rm erf}\left(\frac{\log M - \log 2M_{\rm min}}{\sigma_{\log M}}\right)\right]\left(\frac{M}{M_{\rm sat}}\right)^{\alpha}. \tag{4.72}$$

In the equations above, $M_{\rm min}$ represents the mass of a halo with 50% probability of having a central galaxy. The factor of 2 before $M_{\rm min}$ in equation 4.72 is to avoid dark matter haloes having a larger probability of hosting satellites than central galaxies; $\sigma_{\log M}$ regulates the transition in the dark matter halo from having no galaxies to having one central galaxy; $M_{\rm sat}$ is the minimum

mass for a dark matter halo in order to host a satellite galaxy, and α is the power-law index regulating the number of satellite galaxies.
Figure 4.10 shows the prediction for the number of central and satellite galaxies as functions of the halo masses.

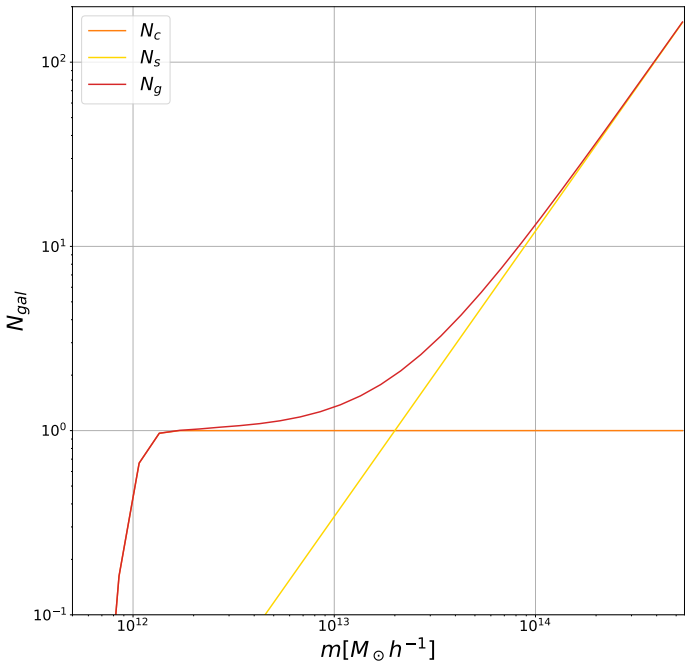

Figure 4.10: HOD predictions for the number of central (orange) and satellite (yellow) galaxies. The red curve shows the total number of galaxies as a function of the halo mass. The parameters used to obtain these predictions are: $\sigma_{\log M} = 0.1$, $M_{sat} = 20 M_{min}$, $\alpha = 1.55$, and $\log M_{min} = 12.0$.

This chapter provides an introduction to the basics of the HM. One of the most evident aspects is that there are many different prescriptions for the basic elements of the HM – *i.e.* the HMF, the halo bias, the NFW profile, and even the HOD. The wide range of possibilities reflects a fundamental uncertainty in the modelling of these elements. The choice of one prescription over another depends on several factors, according to the requirements of the specific analysis being performed.

To conclude, this chapter has introduced many ingredients which form the starting point for the modelling of numerous matter tracers, as well as one of the main concepts that I employ during the development of my original work presented throughout this thesis. Therefore, to better guide the reader into the following chapters, figure 4.11 schematically illustrates the idea behind the HM formalism.

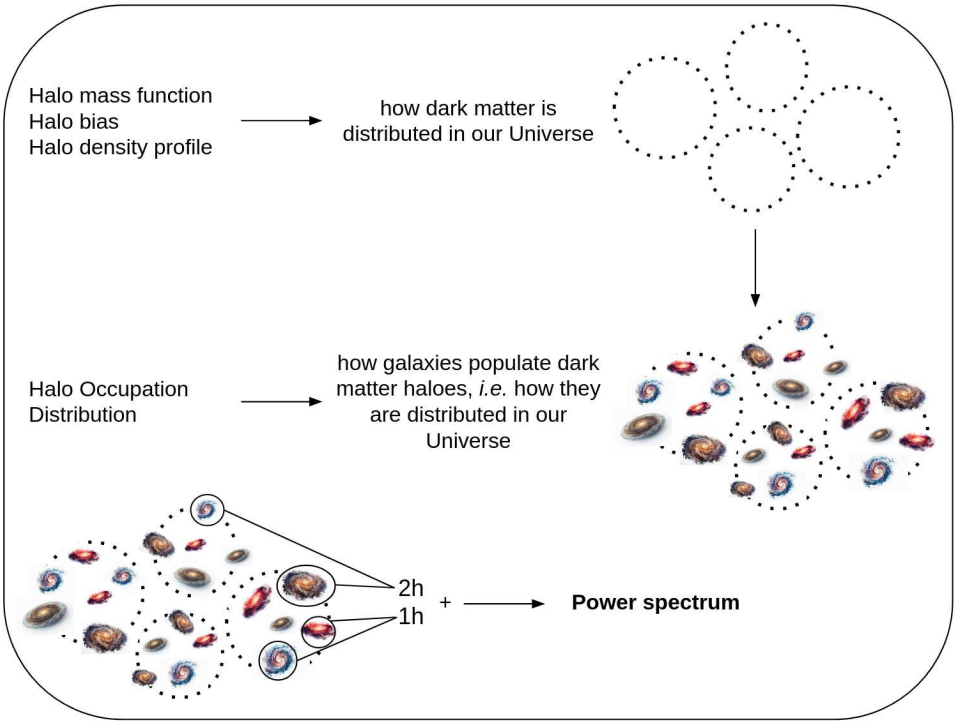

Figure 4.11: Schematic of the HM framework, showing the main ingredients and the steps leading to predictions for non-linear power spectra.

5 EXTRAGALACTIC EMISSIONS

This chapter is devoted to the description of extragalactic emissions, different from the CMB, relevant in the small-scale and millimetre/sub-millimetre regimes. Specifically, I describe the contributions from the Cosmic Infrared Background (CIB), the thermal Sunyaev-Zel'dovich (tSZ), and their cross-correlation, CIBxtSZ, the kinetic Sunyaev-Zel'dovich (kSZ), and radio sources. In addition, Galactic cirrus is included in the modelling, as its emission contributes to the same observables and angular scales probed by extragalactic foregrounds and must therefore be accounted for in order to avoid biases in the inference of extragalactic components.

Review of the main experiments

***COBE*/FIRAS** Space mission measuring the absolute CMB spectrum; FIRAS established its near–perfect blackbody and provided the first robust CIB detection and spectrum. [Fixsen et al., 1996; Puget et al., 1996]

PLANCK All–sky microwave/sub–mm survey (30–857 GHz); delivered multi–frequency CIB anisotropy maps, tSZ catalogues and power spectra with full–sky coverage. [Planck Collaboration, 2014b; 2016e]

ACT Ground–based arcminute–resolution CMB experiment operating at 98, 150, and 220 GHz; measured small–scale power including clustered/Poisson CIB, radio sources, and tSZ/kSZ. [Dunkley et al., 2013]

SPT 10 m South Pole Telescope (95/150/220 GHz); precise small–scale band powers used to constrain tSZ, kSZ, radio and CIB components; provides the kSZ template adopted in this work. [George et al., 2015]

***HERSCHEL*–SPIRE** ESA space telescope's sub–mm camera (250/350/500 μm); mapped CIB anisotropies and number counts with high sensitivity over degree scales. [Viero et al., 2013]

Before moving to the detailed description of each component, it is important to emphasise that millimetre–wave measurements are reported in different units depending on the experiment and frequency band:

- some experiments (*e.g.*, Planck, Atacama Cosmology Telescope (ACT), South Pole Telescope (SPT)) often provide power spectra in CMB temperature units (μK^2),
- others, or some external source catalogues, report results in flux density units ($Jy^2 sr^{-1}$).

In order to obtain a self-consistent description of the various emissions, it is crucial to understand how to convert between CMB and flux units, and vice versa. This is also essential when combining data from multiple instruments or comparing theoretical predictions (usually in CMB temperature units) with templates provided in flux units.

The conversion factors can be computed analytically. Specifically, the frequency–dependent conversion factor relating a change in temperature, δT_ν, to a change in specific intensity, δB_ν, is

$$\delta T_\nu = \left(\frac{\partial B_\nu}{\partial T}\right)^{-1}\Bigg|_{T_{CMB}} \delta B_\nu, \tag{5.1}$$

where $\left(\frac{\partial B_\nu}{\partial T}\right)$ is

$$\frac{\partial B_\nu}{\partial T}\Bigg|_{T_{CMB}} = \frac{2k}{c^2}\left(\frac{kT_{CMB}}{h}\right)^2 \frac{x^4 e^x}{(e^x-1)^2}, \qquad x \equiv \frac{h\nu}{kT_{CMB}}. \tag{5.2}$$

This expression is simply the derivative of the Planck function

$$B_\nu(T) = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/kT}-1}$$

with respect to temperature, evaluated at the CMB temperature $T_{CMB} = 2.7255$ K. The factor $\left(\frac{\partial B_\nu}{\partial T}\right)\Big|_{T_{CMB}}$ converts temperature fluctuations in CMB units to brightness fluctuations in physical flux units.
To summarise:

- to convert from CMB units to flux units, assuming that the same frequency channel is involved, the spectrum in μK^2 has to be multiplied by $\left(\frac{\partial B_\nu}{\partial T}\right)^2\Big|_{T_{CMB}}$ to obtain $Jy^2 sr^{-1}$. If the power spectrum is computed across two different frequency channels, ν and ν', because of the frequency dependence of the conversion factor in equation 5.2, two different conversion factors will be involved, so that the power spectrum will be multiplied by $\left(\frac{\partial B_\nu}{\partial T}\right)\Big|_{T_{CMB}} \times \left(\frac{\partial B_{\nu'}}{\partial T}\right)\Big|_{T_{CMB}}$;
- to convert from flux units to CMB units, the spectrum must be divided by the same factor.

These conversion factors will be used throughout this work in section 5.4 and section 5.5, to consistently have all the results presented in this introductory part in CMB units, and in chapter 6 and in chapter 7, to switch from CMB to flux units, depending on the datasets involved in the particular analysis.

Table 5.1 shows the conversion factors computed from equation 5.2 for the most used frequency channels throughout my work.

Table 5.1: Conversion factors (CF) for the most used frequency channels throughout my work.

Band [GHz]	CF $[Jy \cdot sr^{-1}\mu K^{-1}]$
27	21.980
39	44.936
93	213.515
145	385.327
217	483.469
225	482.525
280	428.903
353	296.423
545	57.038
857	1.432

5.1 COSMIC INFRARED BACKGROUND

The Cosmic Infrared Background (CIB) represents the emission from Dusty Star-Forming (DSF) galaxies across cosmic history, appearing as a diffuse, cosmological background. The CIB was first detected by Puget et al., 1996 with FIRAS on the COBE satellite [Fixsen et al., 1996], and partially resolved into light from individual galaxies using SCUBA on the *James Clerk Maxwell Telescope* [Eales et al., 1999; Hughes et al., 1998; Smail et al., 1997].

Detailed studies of CIB anisotropies have been conducted using, for example, power spectra measured from the SPT [Hall et al., 2010], *Planck* [Planck Collaboration, 2014b], and *Herschel*–SPIRE [Viero et al., 2013] data. These observations provide valuable cosmological and astrophysical information that both complements and extends beyond observations of individual sources, since at high redshifts most of the light originates from sources at or below the confusion limit. Moreover, there is a relation, which can be modelled through a HM formalism, between the DSF galaxies and the dark matter haloes in which they reside. In particular, anisotropies in this cosmological background trace the underlying dark matter halo distribution, thus probing the clustering properties of galaxies. Starting from the perturbations in the initial density field and through the spherical collapse model, the HM predicts how dark matter haloes are distributed in the Universe. Integrating this

framework with information on how galaxies populate dark matter haloes (see section 4.3.2.1) links observations of the CIB with the LSS of the Universe.

In section 4.3.2, I have already introduced how to obtain a prediction in the HM framework for the galaxy distribution across the Universe. At this point, in order to obtain a model of the CIB power spectrum, once we have defined their spatial distribution through a proper choice of the HOD (see section 4.3.2.1), we have to describe how these galaxies emit.

The information on how galaxies emit is encapsulated in the emissivity functions, describing the redshift distribution of sources falling below the detection limit of the observing experiment, S_ν^{lim}, and is defined as

$$j_\nu(z) = \int_0^{S_\nu^{lim}} \frac{d^3N(S_\nu, z)}{dS_\nu dz d\Omega} S_\nu dS_\nu. \tag{5.3}$$

In the above definition enter the following elements:

- S_ν, representing the flux density at a given frequency ν of a source at redshift z, which Hogg, 1999 defines as

  $$S_\nu = \frac{(1+z)L_{\nu'}}{4\pi D_L^2(z)},$$

  where $\nu' = \nu(1+z)$, and $D_L(z)$ the luminosity distance.

- $\frac{d^3N(S_\nu, z)}{dS_\nu dz d\Omega}$, defining the surface density of sources per unit flux density and redshift interval. This quantity is defined as

  $$\frac{d^3N(S_\nu, z)}{dS_\nu dz d\Omega} = \frac{\Phi(\log L_{\nu'}, z)}{L_{\nu'} \ln 10} \frac{dL_{\nu'}}{dS_\nu} \frac{d^2V}{dz d\Omega},$$

  with $\Phi(\log L_{\nu'}, z)$ the luminosity function (LF), and $\frac{d^2V}{dz d\Omega}$ the comoving volume corresponding to a redshift slice between z and $z + dz$ per unit solid angle. The specific modelling of the LF depends on the galaxy population—or populations—one wants to describe. For this reason, a detailed discussion is postponed in chapter 6, where I present my specific model of the CIB emission.

Figure 5.1 shows a prediction of the emissivity function for three different frequencies, chosen among *Planck* channels, *i.e.* 217, 353, and 545 GHz, and it serves only for visualisation purposes since, as already introduced, a more detailed modelling and presentation will be provided in chapter 6. From the figure it is possible to have an intuitive hint on how DSF galaxies emit at different redshifts.

At this point, we have all the elements to compute the CIB power spectrum, given the fact that we have a way to model both the distribution and the emission of DSF galaxies. Specifically, the CIB power spectrum is modelled as the sum of two different contributions:

$$C_{\ell,\nu\times\nu'}^{CIB} = C_{\ell,\nu\times\nu'}^{CIB,clus} + C_{\ell,\nu\times\nu'}^{CIB,SN}, \tag{5.4}$$

where

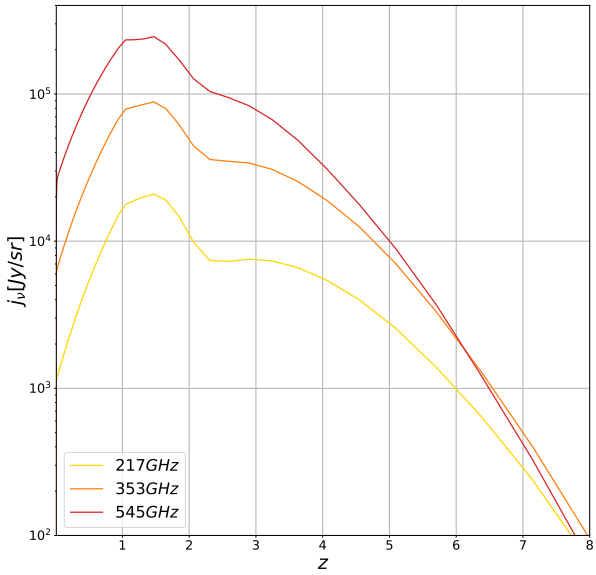

Figure 5.1: Emissivity functions as a function of redshift computed for a single-galaxy population model for three different frequency channels: 217 GHz in yellow, 353 GHz in orange, and 545 GHz in red. The flux cuts applied in the computation of the integral of equation 5.3 are $S_\nu^{lim} = 225, 315, 350$ mJy, respectively, according to *Planck* prescriptions.

- the clustering term $C_{\ell,\nu\times\nu'}^{CIB,clus}$ is defined as

$$C_{\ell,\nu\times\nu'}^{CIB,clus} = \int \frac{dz}{\chi^2}\frac{dz}{d\chi} j_\nu(z) j_{\nu'}(z) P_{gg}(k = \ell/\chi, z), \tag{5.5}$$

$j_\nu(z)$ are the emissivity functions describing how galaxies emit, $P_{gg}(k,z)$ has been defined in equation 4.68 and uses the Limber approximation [Limber, 1953; Limber, 1954] to project the three-dimensional galaxy power spectrum to obtain the angular one, and χ represents the comoving distance. Since from chapter 4 and, specifically, from equation 4.68, we know that the power spectra evaluated within a HM framework encode both the 1h and 2h terms, also the clustering term of the CIB emission, given its dependence on the galaxy power spectrum, shows 1h and 2h terms so that we have

$$C_{\ell,\nu\times\nu'}^{CIB} = C_{\ell,\nu\times\nu'}^{CIB,1h} + C_{\ell,\nu\times\nu'}^{CIB,2h}, \tag{5.6}$$

where

$$C_{\ell,\nu\times\nu'}^{CIB,1h} = \int \frac{dz}{\chi^2}\frac{dz}{d\chi} j_\nu(z) j_{\nu'}(z) \frac{1}{\bar{n}_g^2} \int dm \frac{dn}{dm} [2N_c(m)N_s(m)\tilde{u}(k,m) + N_s^2(m)\tilde{u}^2(k,m)], \tag{5.7}$$

for the 1h term, and

$$C_{\ell,\nu\times\nu'}^{CIB,2h} = \int \frac{dz}{\chi^2}\frac{dz}{d\chi} j_\nu(z) j_{\nu'}(z) P_{mm}^{lin}(k) \left[\frac{1}{\bar{n}_g}\int dm \frac{dn}{dm} b(m) N_g(m) \tilde{u}(k,m)\right]^2, \tag{5.8}$$

for the 2h term. Figure 5.2 shows these contributions in orange and yellow, respectively. The prescriptions for the HMF, dn/dm, and the halo bias, $b(m)$, are the same presented in chapter 4 in equations 4.15 and 4.25.

- The Shot Noise (SN) term, $C_{\ell,\nu\times\nu'}^{CIB,SN}$, accounts for the contribution of sources falling below a certain detection limit—specific for the considered experiment and frequency channel—, modelled as a Poisson term independent of the scale. Its computation is based on the galaxy number counts and reads as

$$C_{\ell,\nu\times\nu'}^{CIB,SN} = \int_0^{S_\nu^{lim}} S_\nu \frac{dN}{dS_\nu} dS_\nu. \tag{5.9}$$

Its flat contribution in C_ℓs is depicted in brown in figure 5.2.

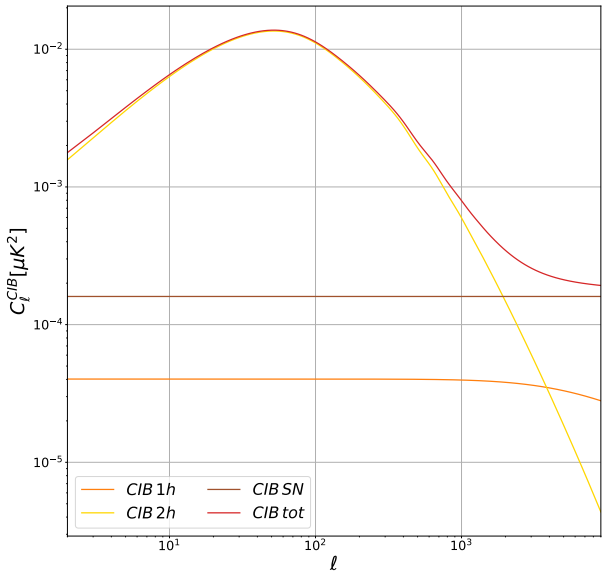

Figure 5.2: Predictions of the different components contributing to the CIB power spectrum evaluated in an HM framework. For illustrative purposes, I show the auto-frequency power spectrum evaluated at 280 GHz. The orange and yellow curves show the 1h and 2h terms contributing to the clustering part, and the brown flat contribution represents the SN term.

Figure 5.2 illustrates the different contributions to the CIB power spectrum. On the one hand, the 1h and 2h terms—shown in orange and yellow— contribute to the clustering part of the CIB power spectrum, which is the leading

contribution at large angular scales. On the other hand, the flat contribution of the unresolved sources falling below the detection limit, described by the SN term and shown in brown, starts to be relevant on smaller angular scales. The information on how galaxies are distributed across the Universe and their emission is encapsulated in the clustering term; therefore, large-scale observations are more effective in constraining parameters and discriminating among different modelling of galaxy clustering. On the contrary, small-scale observations provide additional information on the SN term. Chapter 6 is devoted to a wide discussion on a physically-motivated CIB modelling, exploring observations covering different scales and frequencies.

5.2 SUNYAEV-ZEL'DOVICH EFFECT

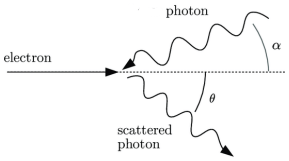

Figure 5.3: Scheme of the inverse Compton scattering.

This results in an energy transfer from the electrons to the photons, distorting the blackbody spectrum of the CMB: photons are shifted from the Rayleigh–Jeans part of the spectrum to the Wien tail. Two distinct contributions arise: the thermal (tSZ) and the kinetic (kSZ) Sunyaev-Zel'dovich effects. tSZ affects the intensity of the CMB blackbody spectrum only, and arises when the CMB photons are scattered off by the high-energy electrons in the SZ clusters. The kSZ effect happens when the free high-energy electrons present in the SZ clusters also have bulk peculiar velocities along the line of sight. A sketch of the two mechanisms explaining the tSZ and kSZ effects is shown in figure 5.5, while the spectral distortions of the CMB radiation induced by these effects are shown in figure 5.4.

The intensity variation depends on the physical properties of galaxy clusters:

$$\Delta T(x) = \Delta T_{thermal}(x, y) + \Delta T_{kinetic}(x, \tau, v_p), \qquad x \equiv \frac{h\nu}{kT_{CMB}}. \tag{5.10}$$

The physical quantities defining the intensity variation are the Compton-y parameter, y, the Thomson scattering depth, τ, and the peculiar velocity of the intracluster medium (ICM), v_p. Details about these effects and their physical properties are provided in the following.

5.2.1 Thermal Sunyaev-Zel'dovich Effect

The thermal Sunyaev-Zel'dovich (tSZ) effect corresponds to the spectral distortion of the CMB intensity spectrum produced by the thermal energy of the ICM. The relative CMB temperature fluctuation due to the tSZ is

$$\frac{\delta T_{tSZ}(\nu, y)}{T_{CMB}} = y\, g(\nu), \tag{5.11}$$

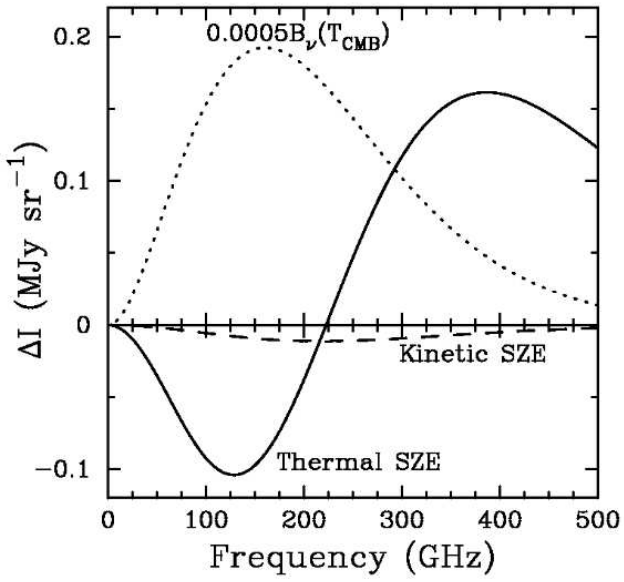

Figure 5.4: Spectral distortion of the CMB spectrum (dotted line) due to the tSZ (solid line) and kSZ (dashed line) effects. (From figure 2 of Carlstrom et al., 2002.)

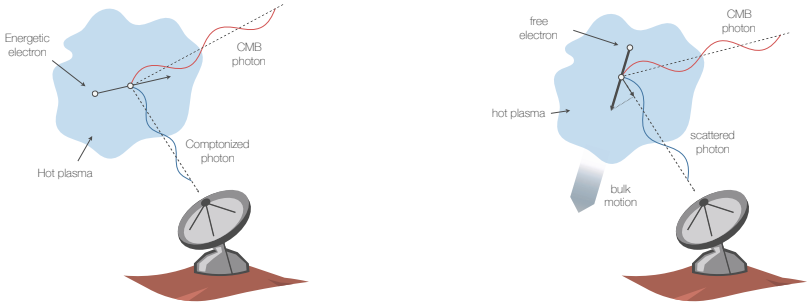

Figure 5.5: Illustration showing the mechanism originating the tSZ and the kSZ effects, from figures 1 and 3 of Mroczkowski et al., 2019

where y is the Compton-y parameter and $g(\nu)$ encodes the frequency dependence of the distortion,

$$g(\nu) = \frac{h\nu}{kT_{CMB}} \frac{e^{h\nu/kT_{CMB}} + 1}{e^{h\nu/kT_{CMB}} - 1} - 4. \tag{5.12}$$

The above equation holds in the non-relativistic limit only. For the sake of simplicity, the relativistic corrections are neglected throughout this thesis. This choice is supported by the fact that Planck Collaboration, 2016e do not report highly significant relativistic distortions to the tSZ power spectrum. For a recent synthesis of the cosmological information content of the tSZ

power spectrum and its scale dependence, see Douspis et al., 2022.
The Compton-y parameter is

$$y = \frac{\sigma_T}{m_e c^2} \int d\ell \, n_e k T_e = \frac{\sigma_T}{m_e c^2} \int d\ell \, P_e, \tag{5.13}$$

where σ_T is the Thomson cross-section, m_e the electron mass, n_e the electron number density, and P_e the electron pressure profile (modelled here with an empirical template).

The spatial power spectrum of the SZ clusters generating the tSZ can be computed employing a HM formalism, specifically from equations 4.46 and 4.52, where the FT of the density profile is taken to be the FT of the Compton-y parameter, $\tilde{y}$, so that

$$P_{yy}^{1h}(k) = \int dm \frac{dn}{dm}\bigg|_{500} \tilde{y}(k,m)^2, \tag{5.14}$$

for the 1h term, and

$$P_{yy}^{2h}(k) = P_{mm}^{lin} \left[\int dm \frac{dn}{dm}\bigg|_{500} b_{tSZ}(m) \tilde{y}(k,m) \right]^2, \tag{5.15}$$

for the 2h term. The total power spectrum is then given by the sum of these terms

$$P_{yy}(k) = P_{yy}^{1h}(k) + P_{yy}^{2h}(k). \tag{5.16}$$

A couple of elements require a deeper analysis. In both equations 5.14 and 5.15, the HMF shows a '500' subscript. The reason is that SZ clusters are expected to populate more massive and more clustered dark matter haloes than those hosting DSF galaxies; therefore the HMF has been computed using a density contrast $\Delta = 500$ rather than the $\Delta = 200$ employed in the CIB calculations. Figure 5.6 compares the predictions in terms of comoving number density of haloes capable of hosting DSF galaxies in red, and of haloes hosting SZ clusters in orange, for $z = 1.5$ and $z = 4.0$ in the left and right panels, respectively. The comoving number density of haloes hosting SZ clusters is always lower than the one for the DSF galaxies, and this difference is more pronounced for higher redshift values.

In equation 5.15 the bias is denoted as b_{tSZ} for the same reason: the SZ clusters populate more massive and more clustered haloes. Therefore, to describe the bias entering the 2h term, I employed the same functional form presented in equation 4.25, but using different values for the coefficients, evaluated in table 2 of Tinker et al., 2010a and reported below

$$\begin{aligned} A &= 1.0 + 0.24 y e^{-(4/y)^4}, \\ a &= 0.44 y - 0.88, \\ B &= 0.183, \\ b &= 1.5, \\ C &= 0.019 + 0.107 y + 0.19 e^{-(4/y)^4}, \\ c &= 2.4, \end{aligned} \tag{5.17}$$

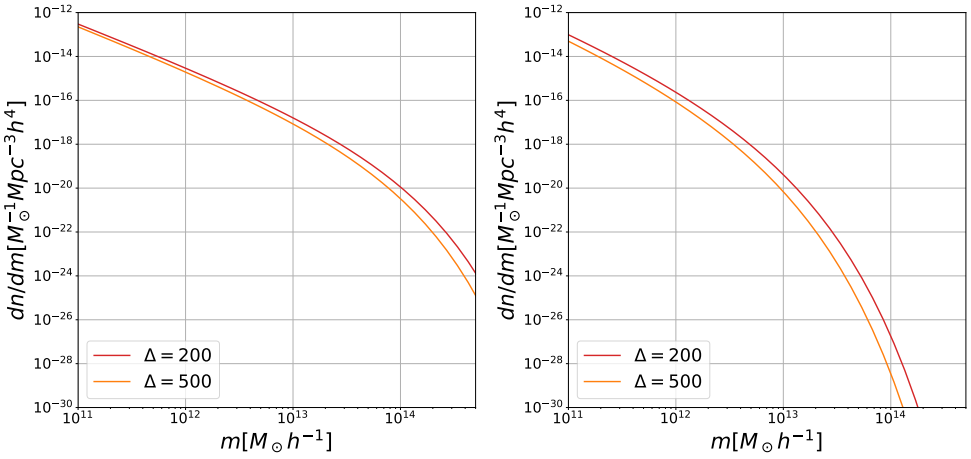

Figure 5.6: HMF predictions for $\Delta = 200$, in red, and $\Delta = 500$, in orange, dark matter haloes, at redshift $z = 1.5$ (left panel), and $z = 4.0$ (right panel).

with $y \equiv \log_{10} \Delta$.

The FT of the 3D Compton-y profile is

$$\tilde{y}(m,z) = \frac{\sigma_T}{m_e c^2} \frac{4\pi r_{500}}{\ell_{500}^2} C\, P_0 \int dx\, x^2 \frac{\sin(\ell x/\ell_{500})}{\ell x/\ell_{500}} P_e(x), \tag{5.18}$$

with

- $x = r/r_{500}$,
- $\ell_{500} = D_A/r_{500}$,
- $P_e(x)$ the electron pressure profile, modelled as a generalised NFW profile[1]:

  $$P_e(x) = (c_{500}x)^{-\gamma}[1 + (c_{500}x)^{\alpha}]^{(\gamma-\beta)/\alpha}.$$

  I use the best-fit values obtained by Planck Collaboration, 2014a analysis for the parameters defining the electron pressure profile. Specifically

  $$c_{500} = 1.81,\ \gamma = 0.31,\ \alpha = 1.33,\ \beta = 4.13.$$

- C a mass–dependent normalisation defined as

  $$C = 1.65 \left(\frac{h}{0.7}\right)^2 \left(\frac{H}{H_0}\right)^{8/3} \left[\frac{(h/0.7)\tilde{M}_{500}}{3 \times 10^{14} M_\odot}\right]^{\frac{2}{3}+0.12},$$

  which is proportional to the hydrostatic mass $\tilde{M}_{500}$. The hydrostatic mass differs from the true mass because of non–thermal pressure support, observational systematics, and model uncertainties[2]. The relation

1 Hydrodynamical simulations and X-ray/SZ observations have shown that cluster pressure profiles exhibit an approximately universal shape when expressed in scaled radial coordinates, which is well described by a generalised NFW functional form. For this reason, and because of its empirical success in reproducing both resolved cluster profiles and tSZ power spectrum measurements, I adopt this parametrisation calibrated on observations.

2 At a very first level of approximation, part of the bias induced by relativistic distortions to the tSZ spectrum is already encoded in the hydrostatic mass bias.

is parametrised by the hydrostatic mass bias B, such that $M_{500} = B\tilde{M}_{500}$. Following Planck Collaboration, 2016c, I use the notation b with $B = (1-b)^{-1}$. As $\ell_{500} \propto 1/r_{500}$, and $r_{500} \propto M_{500}^{1/3}$, this bias affects not only the normalisation but also the angular dependence of the tSZ power spectrum. Observational constraints on the possible mass/redshift evolution of the hydrostatic mass bias from gas–mass fraction measurements are presented in Salvati et al., 2019; Wicker et al., 2022.

The starting point in order to obtain the final prediction of the tSZ power spectrum is equation 5.11; specifically, the temperature variation caused by tSZ is

$$\delta T_\nu = T_{CMB} g(\nu) y,$$

therefore the tSZ power spectrum is

$$C^{tSZ}_{\ell,\nu\times\nu'} = T^2_{CMB} \int \frac{dV}{dz} dz\, g(\nu) g(\nu') P_{yy}(k = \ell/\chi), \tag{5.19}$$

where $P_{yy}(k)$ is computed using the Limber approximation, and its expression has been provided in equation 5.16. Therefore, also in this case, the tSZ power spectrum can be split into the two contributions coming from the 1h and 2h terms:

$$C^{tSZ}_{\ell,\nu\times\nu'} = C^{tSZ,1h}_{\ell,\nu\times\nu'} + C^{tSZ,2h}_{\ell,\nu\times\nu'}, \tag{5.20}$$

where

$$C^{tSZ,1h}_{\ell,\nu\times\nu'} = T^2_{CMB} \int \frac{dV}{dz} dz\, g(\nu) g(\nu') \int dm \left.\frac{dn}{dm}\right|_{500} \tilde{y}(k,m)^2, \tag{5.21}$$

for the 1h term, and

$$C^{tSZ,2h}_{\ell,\nu\times\nu'} = T^2_{CMB} \int \frac{dV}{dz} dz\, g(\nu) g(\nu') P^{lin}_{mm}(k) \left[\int dm \left.\frac{dn}{dm}\right|_{500} b_{tSZ}(m) \tilde{y}(k,m) \right]^2, \tag{5.22}$$

for the 2h term.

Figure 5.7 shows the prediction for the tSZ power spectrum. The 1h and 2h terms are shown in orange and yellow, respectively, while the total tSZ signal is depicted in red. It is possible to notice that, because of the scarcity of high-mass SZ clusters, the 1h term dominates across all scales.

5.2.2 Kinetic Sunyaev-Zel'dovich Effect

The kinetic Sunyaev-Zel'dovich (kSZ) effect originates from the Doppler shift induced by electrons with bulk velocities along the line of sight. The corresponding temperature fluctuation is

$$\delta T_{CMB}(x, \tau, v_p) = -\beta\tau, \tag{5.23}$$

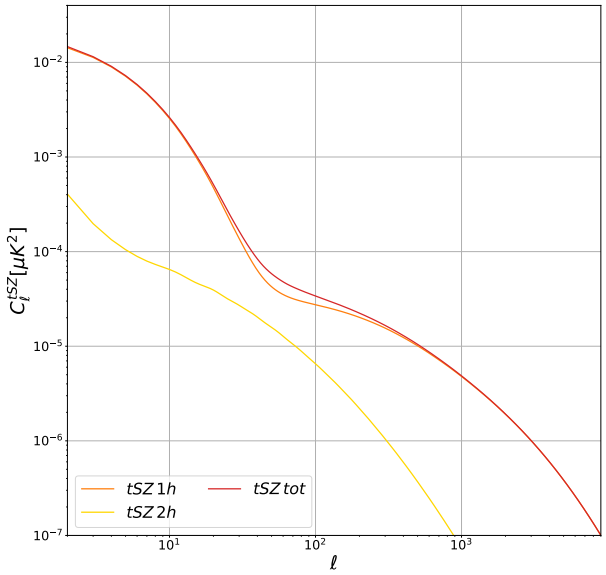

Figure 5.7: Predictions of the different components contributing to the tSZ power spectrum evaluated in an HM framework. For illustrative purposes, I show the auto-frequency power spectrum evaluated at 280 GHz. The orange and yellow curves show the 1h and 2h terms, while the red one represents the total tSZ signal.

where β and τ are the two parameters regulating the amplitude of this effect. The first one is the line–of–sight peculiar velocity (positive if approaching the observer), defined as $\beta = v_p/c$. The second parameter represents the Thomson scattering depth, and it is defined as $\tau = \sigma_T \int d\ell n_e$.

The kSZ signal is not well modelled by an HM formalism; instead, a template-based approach is used. The reason for this choice is that an HM description cannot fully capture the dominant contributions to the kSZ[3]. More precisely, the kSZ signal depends on large-scale, coherent velocities, and on the free-electron density, much of which lies outside haloes. Specifically, in my work I use the angular template $D_{0,\ell}^{kSZ}$ from George et al., 2015, which includes:

- a homogeneous contribution arising after reionization ($z < 8$) in a fully ionized Universe, calibrated from hydrodynamical simulations [Shaw et al., 2012],
- a patchy contribution sourced by ionized regions during reionization ($8 < z < 11$) [Zahn et al., 2012].

3 For a consolidated review of the homogeneous and patchy kSZ components and their cosmological/reionization dependences, see Gorce et al., 2022

The scaling of this template with cosmological parameters is given by

$$D_0^{kSZ} \propto \left(\frac{h}{0.71}\right)^{1.7} \left(\frac{\sigma_8}{0.80}\right)^{4.7} \left(\frac{\Omega_b}{0.044}\right)^{2.1} \left(\frac{\Omega_m}{0.264}\right)^{-0.4} \left(\frac{n_s}{0.96}\right)^{-0.2}, \quad (5.24)$$

where the amplitude of the template is normalised at $1\mu K^2$ at $\ell = 3000$, and assumes that reionization starts at $z = 11$ and ends at $z = 8$. However, more recent CMB polarisation measurements indicate a lower redshift of reionization, with $z_{\rm reio} \simeq 7$–8 [Planck Collaboration, 2020b]. As a consequence, the assumed reionization history underlying the template does not fully reflect current constraints, which constitutes one of the intrinsic limitations of a fixed-template description of the kSZ signal.
The final kSZ power spectrum is then given as the template rescaled by an amplitude often left as a free parameter to be constrained with data:

$$C_\ell^{kSZ} = \frac{2\pi}{\ell(\ell+1)} \mathcal{A}_{kSZ} D_{0,\ell}^{kSZ}. \quad (5.25)$$

In the above equation, the $\frac{2\pi}{\ell(\ell+1)}$ prefactor is used to have consistency across this chapter and report the final power spectra of the various extra-galactic emissions in units of C_ℓs despite the modelled template being given in band-power units—*i.e.* D_ℓs.

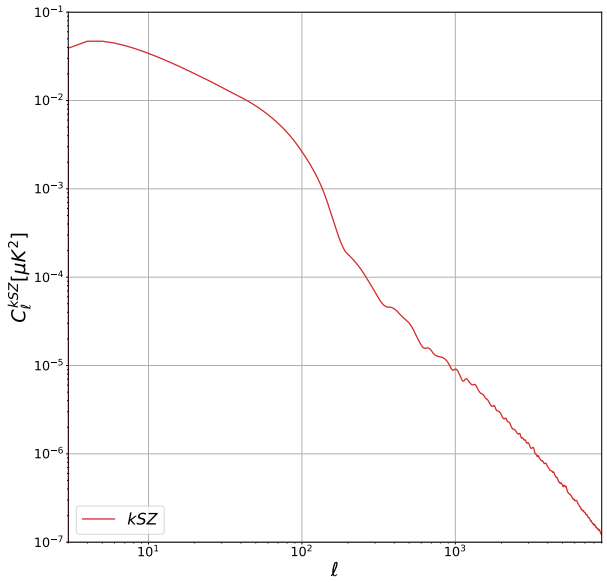

Figure 5.8: Prediction of the kSZ signal in C_ℓs units using the George et al., 2015 template and setting the free amplitude to their best-fit value, $\mathcal{A}_{kSZ} = 1.6\mu K^2$. This power spectrum is the same independent of the frequency channels involved.

Figure 5.8 shows the prediction of the kSZ power spectrum employing the angular template previously described and setting an amplitude $\mathcal{A}_{kSZ} =$

$1.6\mu K^2$ for the sake of visualisation. Given the fact that the kSZ signal consists of a Doppler shift of the CMB black-body spectrum, it does not show any frequency dependence. Therefore, the contribution shown in figure 5.8 is the same across all frequency channels considered.

5.3 CIB AND TSZ CROSS CORRELATION

The CIB, originating from the unresolved DSF galaxies, and the tSZ effect, resulting in a distortion of the CMB photons caused by the high-energy electrons in the ICM in galaxy clusters, are tracers of the same underlying matter density field, leading to a natural correlation between these two fields, as argued in Planck Collaboration, 2016d.
The HM framework provides a way to predict the signal arising from the CIBxtSZ correlation without the need to introduce extra parameters as done in Dunkley et al., 2013 and George et al., 2015. Specifically, Dunkley et al., 2013 and George et al., 2015 used a template-based approach, introducing a free parameter regulating the degree of correlation, assumed to be the same at all angular scales. Describing this signal within a HM framework, apart from the fact that it does not require the introduction of an additional parameter, is more physically motivated.

From the most general point of view, the power spectrum of the CIBxtSZ correlation is given in terms of an emission kernel, $\xi(\nu, z)$—describing how the two tracers emit as a function of frequency—and the spatial power spectrum $P(k)$ computed in a HM framework as

$$C_{\ell,\nu\times\nu'}^{CIB\times tSZ} = \int \frac{dV}{dz} dz\, \xi(\nu, z)\xi(\nu', z) P_{gy}(k = \ell/\chi). \tag{5.26}$$

The computation of the $P_{gy}(k)$–where again the Limber approximation has been used in its computation– in a HM framework can be done from the most general equations for the 1h and 2h terms presented in 4.46 and 4.52, where the two fields are the DSF galaxies and SZ clusters, and the FT of the NFW profiles are the ones used in section 5.1 and section 5.2.1. Therefore, the 1h term reads as

$$P_{gy}^{1h}(k) = \int dm \frac{dn}{dm}\bigg|_{500} \frac{N_g}{\bar{n}_g} \tilde{u}_g(k, m)\tilde{y}(k, m), \tag{5.27}$$

the 2h term as

$$\begin{aligned} P_{gy}^{2h}(k) =& P_{mm}^{lin}(k) \int dm \frac{dn}{dm} b_{CIB}(m) \frac{N_g}{\bar{n}_g} \tilde{u}_g(k, m) \times \\ & \int dm \frac{dn}{dm}\bigg|_{500} b_{tSZ}(m)\tilde{y}(k, m), \end{aligned} \tag{5.28}$$

and the final spectrum is simply the sum of 5.27 and 5.28.
The emission kernel has to encode the emission of both tracers and can be modelled as

$$\xi(\nu, z) = \frac{j_\nu(z)}{dV/dz} + T_{CMB} g(\nu), \tag{5.29}$$

where the first term contains the information about the frequency dependence of the DSF galaxies' emission, and the second term describes the emission dependence of the tSZ effect. Note that the denominator in the term describing the CIB emission appears in order to have consistency among the integration variables.
The computation of the final power spectrum requires a multiplication of the emission kernels at the two frequency channels involved. The multiplication leads to four different terms:

$$\begin{aligned}\xi(\nu, z)\xi(\nu', z) =& \frac{j_\nu(z)}{dV/dz}\frac{j_{\nu'}(z)}{dV/dz} + T_{CMB}\frac{j_\nu(z)}{dV/dz} g(\nu') + \\ & T_{CMB}\frac{j_{\nu'}(z)}{dV/dz} g(\nu) + T_{CMB}^2 g(\nu) g(\nu').\end{aligned} \tag{5.30}$$

However, the first and the last terms represent the CIB and tSZ only contributions, respectively. Therefore, the emission kernel describing the CIBxtSZ cross-correlation is simply

$$\xi(\nu, z)\xi(\nu', z) = T_{CMB}\frac{j_\nu(z)}{dV/dz} g(\nu') + T_{CMB}\frac{j_{\nu'}(z)}{dV/dz} g(\nu). \tag{5.31}$$

The final prediction within a HM framework of the CIBxtSZ signal from equation 5.26 is given by the sum of the 1h and 2h terms, reflecting the structure of the spatial power spectrum, $P_{gy}(k)$:

$$\begin{aligned}C_{\ell,\nu\times\nu'}^{CIB\times tSZ,1h} =& T_{CMB}\int \frac{dV}{dz} dz \left[\frac{j_\nu(z)}{dV/dz} g(\nu') + \frac{j_{\nu'}(z)}{dV/dz} g(\nu)\right] \\ & \int dm \frac{dn}{dm}\bigg|_{500} \frac{N_g}{\bar{n}_g} \tilde{u}_g(k, m)\tilde{y}(k, m)\end{aligned} \tag{5.32}$$

$$\begin{aligned}C_{\ell,\nu\times\nu'}^{CIB\times tSZ,2h} =& T_{CMB}\int \frac{dV}{dz} dz \left[\frac{j_\nu(z)}{dV/dz} g(\nu') + \frac{j_{\nu'}(z)}{dV/dz} g(\nu)\right] P_{mm}^{lin}(k) \\ & \int dm \frac{dn}{dm} b_{CIB}(m) \frac{N_g}{\bar{n}_g} \tilde{u}_g(k, m) \times \\ & \int dm \frac{dn}{dm}\bigg|_{500} b_{tSZ}(m)\tilde{y}(k, m).\end{aligned} \tag{5.33}$$

From a practical point of view, the 1h term describes the signal arising from CIB sources and SZ clusters belonging to the same dark matter halo, while the 2h term arises from the correlation between a DSF galaxy in one dark matter halo and a SZ cluster present in another dark matter halo.

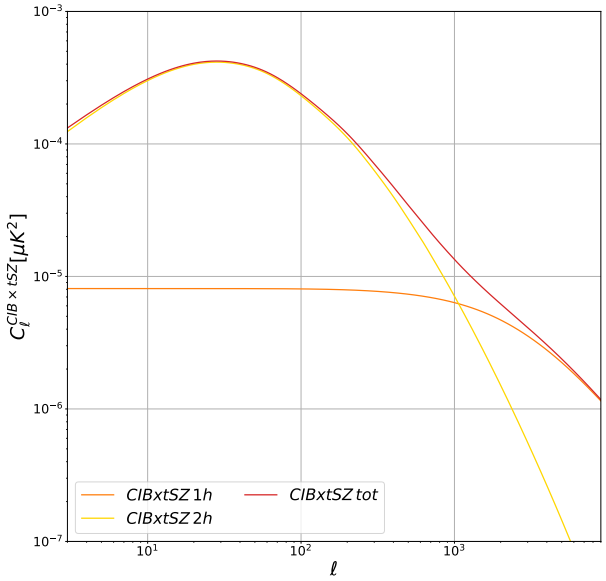

Figure 5.9: Predictions of the different components contributing to the CIBxtSZ power spectrum evaluated in a HM framework. For illustrative purposes, I show the auto-frequency power spectrum evaluated at 280 GHz. The orange and yellow curves show the 1h and 2h terms, while the red one represents the total CIBxtSZ signal.

Figure 5.9 shows the prediction for the CIBxtSZ power spectrum. The 1h and 2h terms are depicted in orange and yellow, respectively, while the total signal is the red curve. It is possible to notice a similar behaviour of the power spectrum to the one presented in figure 5.2 for the CIB-only case, but with lower power. This can be understood by noting that the leading contribution comes from the DSF galaxies, but is suppressed by the condition of looking at more massive dark matter haloes—meaning those that can host SZ clusters, too. Furthermore, also in this case the 2h term is subdominant, but not negligible as for the tSZ case. This can be explained by the case of a massive halo with no DSF galaxy in it: in this scenario, there is no contribution from the 1h term, but there can still be a contribution coming from the 2h term, arising from some overlap in the redshift distribution between DSF galaxies and SZ clusters. Another thing worth noticing is that, because the frequency dependence of the tSZ for low frequency channels can be negative (see figure 5.4 as a reference), for low frequencies involved in the analysis, the CIBxtSZ signal could also be negative.

5.4 RADIO SOURCES

The contribution from radio sources can be modelled as a Poisson term [Mauch et al., 2003; Mocanu et al., 2013], since their clustering is subdominant compared to the shot–noise component at the scales of interest (*i.e.* $\ell > 1000$). The angular power spectrum is described by

$$C^{radio}_{\ell,\nu\times\nu'} = \frac{2\pi}{\ell(\ell+1)}\mathcal{A}^{radio}\epsilon_{\nu\nu'}\left(\frac{\nu\nu'}{\nu_0^2}\right)^{\alpha_{radio}}\left(\frac{\ell}{3000}\right)^2, \tag{5.34}$$

where:

- the $\frac{2\pi}{\ell(\ell+1)}$ prefactor is used to have consistency throughout this chapter and obtain all power spectra in C_ℓs units,
- ν_0 is the normalisation frequency set to $\nu_0 = 150.2$ GHz from George et al., 2015,
- $\mathcal{A}^{radio}$ is the amplitude of the radio sources contribution measured at $\ell = 3000$ in CMB units, at the normalisation frequency ν_0. From George et al., 2015 I fix it at $\mathcal{A}^{radio} = 1.28\mu K^2$,
- $\epsilon_{\nu\nu'}$ is defined as

$$\epsilon_{\nu\nu'} = \frac{\frac{dB}{dT}\Big|_{\nu_0}\frac{dB}{dT}\Big|_{\nu_0}}{\frac{dB}{dT}\Big|_{\nu}\frac{dB}{dT}\Big|_{\nu'}} \tag{5.35}$$

  to obtain the final spectrum in CMB units,
- α_{radio} is the effective spectral index of the radio–source population, fixed at the best-fit obtained by George et al., 2015, $\alpha_{radio} = -0.945$,
- the last factor provides the angular template, normalised at $\ell = 3000$.

This phenomenological model reflects the fact that unresolved radio galaxies, dominated by Active Galactic Nuclei (AGN), produce approximately scale–independent shot noise in C_ℓ (hence $\propto \ell^2$ in D_ℓ).
The effective spectral index α_{radio} characterises the frequency dependence of the emission and is typically negative (*e.g.* $\alpha_{radio} \sim -0.7$ for synchrotron–dominated sources), though the precise value depends on the relative mix of flat–spectrum and steep–spectrum radio populations.
The amplitude $\mathcal{A}^{radio}$ is calibrated at $\ell = 3000$, a scale chosen to match previous analyses (*e.g.*, ACT, SPT, Planck) where the radio contribution is most clearly measured. For lower multipoles the radio term rapidly becomes subdominant compared to clustered CIB or tSZ.

Figure 5.10 shows the prediction for the power spectrum coming from the emission of radio sources, modelled as a Poisson contribution. It is important to notice that, although Equation 5.34 is given in C_ℓ units, figure 5.10 is in units of band powers, D_ℓ, for visualisation purposes.

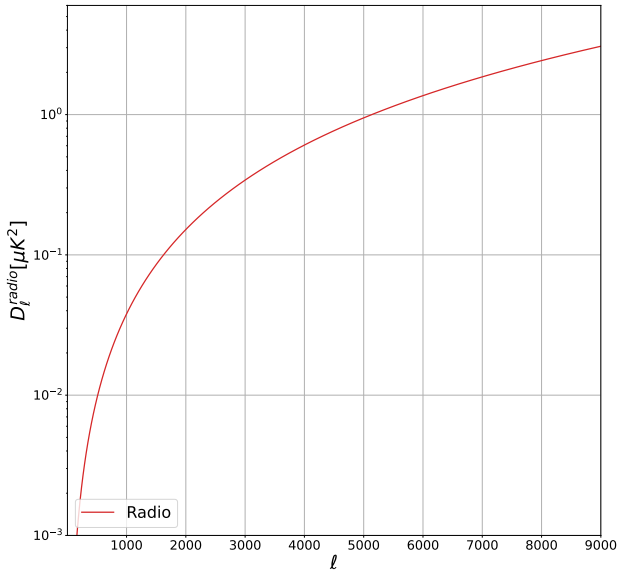

Figure 5.10: Predictions of the radio sources power spectrum evaluated using a template–based approach with $\nu_0 = 150.2$ GHz, $\mathcal{A}^{radio} = 1.28\mu K^2$, and $\alpha_{radio} = -0.945$. For illustrative purposes, I show the power spectrum evaluated at 280 GHz, in band–power units, *i.e.* $D_\ell = \dfrac{\ell(\ell+1)}{2\pi} C_\ell$.

5.5 GALACTIC CIRRUS

The final component entering this section is the emission from Galactic cirrus, *i.e.* diffuse thermal emission from interstellar dust grains within the Milky Way. Although of Galactic origin, this foreground contaminates extragalactic measurements at sub–millimetre frequencies and must be included in the modelling.

As for radio sources, I adopt a phenomenological template–based description in terms of an amplitude and separable angular and frequency dependencies [Fixsen et al., 1998]. The cirrus power spectrum is modelled as

$$C^{cirrus}_{\ell,\nu\times\nu'} = \frac{2\pi}{\ell(\ell+1)} \mathcal{A}^{cirrus}\, \epsilon_{\nu\nu'}\, \frac{\eta_\nu\, \eta_{\nu'}}{\eta^2_{\nu_0}} \left(\frac{\ell}{3000}\right)^{\alpha_{cirrus}}, \tag{5.36}$$

where:

- the $\dfrac{2\pi}{\ell(\ell+1)}$ prefactor is again used to have consistency throughout this chapter and obtain all power spectra in C_ℓ units;
- ν_0 is the normalisation frequency for Galactic cirrus set to $\nu_0 = 220.0$ GHz;
- $\mathcal{A}^{cirrus}$ is the amplitude of the cirrus power spectrum at $\ell = 3000$ in flux units;

- $\epsilon_{\nu\nu'}$ is the same factor defined in Equation 5.35 to obtain the final spectrum in CMB units;
- η_ν encodes the frequency dependence of the cirrus emission, modelled as a modified blackbody;
- α_{cirrus} is the effective angular spectral index, describing the ℓ dependence of the cirrus fluctuations.

The modified blackbody spectrum η_ν is typically parametrised as

$$\eta_\nu \propto \nu^\beta B_\nu(T_{cirrus}), \tag{5.37}$$

where $B_\nu(T_{cirrus})$ is the Planck function at a reference cirrus temperature of $T_{cirrus} = 17.6$K, and β is the cirrus emissivity index.
The angular dependence of cirrus emission is well described by a power law in multipole, $C_\ell \propto \ell^{-\gamma}$ (thus $D_\ell \propto \ell^{2-\gamma}$). Observations generally find $\gamma \simeq 2.5$–3, corresponding to $\alpha_{cirrus} \simeq -0.5$ to -1 in D_ℓ.
In summary, this template captures the essential properties of cirrus contamination: *(i)* a strong modified blackbody frequency dependence, and *(ii)* a power–law angular scaling. The amplitude $\mathcal{A}^{cirrus}$ is empirically calibrated at $\ell = 3000$.

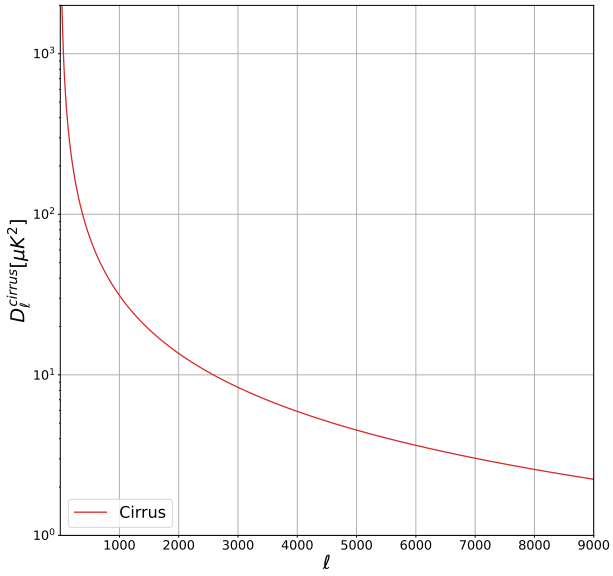

Figure 5.11: Predictions of the Galactic cirrus power spectrum evaluated using a template–based approach with $\mathcal{A}^{cirrus} = 0.65\mu K^2$, $\nu_0 = 220.0$ GHz, $T_{cirrus} = 17.6$K, $\beta = 1.8$, and $\alpha_{cirrus} = -1.2$. For illustrative purposes, I show the power spectrum evaluated at 280 GHz, represented in band–power units.

Figure 5.11 shows the prediction for the power spectrum coming from the emission of Galactic cirrus, modelled using a template–based approach. It is

important to notice that, as in the previous case for radio sources, although Equation 5.36 is given in C_ℓ units, figure 5.11 is in units of band powers, D_ℓ, for visualisation purposes.

Part III

A PHYSICALLY MOTIVATED MODELLING OF THE EXTRAGALACTIC EMISSIONS

This last part presents my original work on modelling the extragalactic emissions. I first develop and test a physically motivated model of the Cosmic Infrared Background (CIB) and fit it to three key CIB data sets. I then extend the framework to build a unified, self-consistent description of all major extragalactic components discussed in this thesis, and use it to forecast the constraining power of future high-precision Cosmic Microwave Background (CMB) measurements.

6 THE COSMIC INFRARED BACKGROUND MODEL

In this chapter, I present the study of a physically motivated model of the Cosmic Infrared Background (CIB) emission. This deep focus on this signal serves as a first step in my wider study and modelling of all other extragalactic emissions in the Universe, exploiting the strong connection with the underlying matter density field.

In section 6.1, I summarise the features of the model I implemented to provide a physically motivated description of the CIB emission. Because some of the model parameters cannot yet be robustly predicted from first principles, and are only loosely constrained by theory or simulations, I determine them by fitting the model to available CIB data sets, described in section 6.2. The results and the conclusions are presented in section 6.3 and section 6.4.

6.1 THE MODEL

As already introduced in the previous chapter, the CIB is the emission from Dusty Star-Forming (DSF) galaxies across cosmic history. Anisotropies in this radiation background trace the underlying matter distribution. Working under the very first assumption of the HM – *i.e.* the fact that all matter in the Universe is contained in dark matter haloes – I modelled the relation between DSF galaxies and the dark matter haloes in which they reside using the HM formalism.

Previous works in the literature followed this approach, and the crucial and most important aspect is the choice of the specific model for the distribution of galaxies in the dark matter haloes. Most previous studies used single-population models; however, in my work I followed the approach of Cai et al., 2013, hereafter C13, which distinguished DSF galaxies into two different populations: late-type (LT) galaxies and the progenitors of early-type (ET) galaxies.

The reason for this distinction mainly comes from observations. Specifically, by drawing different galaxies in a colour-magnitude diagram, galaxies distribute in the red sequence and in the blue cloud presented in figure 6.1, without populating the so-called green valley. These observations suggest a bimodal distribution of galaxies, motivating therefore C13 – and this work – following this approach.

In the following, I review the model of the two galaxy populations before plugging it into a full-shape HM formalism – see section 4.3 of chapter 4 and, specifically, section 4.3.2 for the full computation of the galaxy power

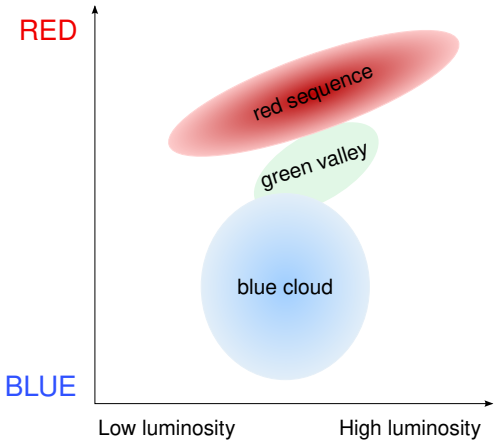

Figure 6.1: Galaxy distribution in a colour-magnitude diagram. Actual data do not observe galaxies populating the green valley, rather they appreciate a bimodal distribution between the red sequence and the blue cloud.

spectrum in the case of two different discrete tracers of the matter density field.

6.1.1 Modeling of the two galaxy populations

I populate dark matter haloes with two different galaxy populations. LT galaxies are characterised by young stellar populations ($\lesssim 7-8\,\mathrm{Gyr}$), corresponding to formation redshifts $z \lesssim 1-1.5$; they have low-to-moderate star formation rates ($\lesssim 10\,M_\odot\,\mathrm{yr}^{-1}$), and reside in less massive dark matter haloes. ET galaxies exhibit an old stellar population ($\gtrsim 8-9\,\mathrm{Gyr}$), indicative of formation redshifts $z \gtrsim 1.5$; they have low-to-null star formation rates, and reside in more massive haloes. However, during their formation phase, ET galaxies experience intense star formation activity (star formation rate up to 100-1000 $\mathrm{M}_\odot\,\mathrm{yr}^{-1}$) and are characterised by a high dust content, thus showing up as dust-obscured star-forming objects, which dominate the peak of the cosmic star formation activity in the Universe at $z \sim 2$. It is this early, highly star-forming, dust-obscured phase that I consider here for ET galaxies. Although, for simplicity, I keep using the acronym ET to refer to the proto-spheroidal galaxies.
The choice of populating dark matter haloes with ET and LT galaxies, together with the fact that the HM formalism naturally predicts the correlation between the two galaxy populations at the level of the galaxy power spectrum–and, consequently, of the CIB–is crucial to achieving a more realistic and coherent description of the CIB emission over a wide range of frequencies and angular scales.

The main outcome of the HM is the spatial distribution of dark matter haloes and, after a proper choice of the HOD, of the galaxies across the Universe. In the following, I describe how I arrive at a prediction of the

CIB power spectrum for a more specific modelling with respect to the most general approach described in section 5.1.
In chapter 5 section 5.1 I detailed the general approach and described the main quantities required to obtain the CIB power spectrum. However, when dealing with two distinct galaxy populations, there are three contributions to be encoded in the formalism and computed in the HM framework:

1. the contribution coming from ET galaxies only, to be computed analogously to the case of the galaxy power spectrum computation for a single galaxy population;
2. the contribution coming from LT galaxies only, to be computed as the previous one, but using different HOD parameters;
3. the contribution coming from the cross-term between the two galaxy populations. For this case, the galaxy power spectrum comes from Equations 4.69 and 4.70 since I am dealing with two distinct discrete tracers.

It is crucial to stress the fact that, aside from the different emission properties that will be detailed in the following, since I am dealing with two different populations, I also expect them to show different clustering properties. This is reflected in a different set of HOD parameters for the two galaxy populations. For the sake of clarity, with reference to section 4.3.2.1, when dealing with ET and LT galaxies I expect to have

$$\langle N^{ET}_{\rm cent}|M\rangle = \frac{1}{2}\left[1+{\rm erf}\left(\frac{\log M-\log M^{ET}_{\rm min}}{\sigma_{\log M}}\right)\right], \tag{6.1}$$

$$\langle N^{ET}_{\rm sat}|M\rangle = \frac{1}{2}\left[1+{\rm erf}\left(\frac{\log M-\log 2M^{ET}_{\rm min}}{\sigma_{\log M}}\right)\right]\left(\frac{M}{M^{ET}_{\rm sat}}\right)^{\alpha_{ET}}, \tag{6.2}$$

for the ET galaxies, and

$$\langle N^{LT}_{\rm cent}|M\rangle = \frac{1}{2}\left[1+{\rm erf}\left(\frac{\log M-\log M^{LT}_{\rm min}}{\sigma_{\log M}}\right)\right], \tag{6.3}$$

$$\langle N^{LT}_{\rm sat}|M\rangle = \frac{1}{2}\left[1+{\rm erf}\left(\frac{\log M-\log 2M^{LT}_{\rm min}}{\sigma_{\log M}}\right)\right]\left(\frac{M}{M^{LT}_{\rm sat}}\right)^{\alpha_{LT}}, \tag{6.4}$$

for the LT.

When plugging this information into equations 4.66 and 4.67, for the evaluation of the ET and LT contributions only, and into equations 4.69 and 4.70 for the mixing between the two populations, I obtain

$$P^{1h}_{\rm ET}(k) = \frac{1}{\left(\overline{n}^{\rm ET}_{\rm g}\right)^2}\int_0^\infty dm\frac{dn}{dm}[2N^{\rm ET}_c N^{\rm ET}_s\tilde{u}(m,k)+\left(N^{\rm ET}_s\right)^2\tilde{u}(k,m)^2], \tag{6.5}$$

$$P^{2h}_{\rm ET}(k) = P^{lin}_{mm}(k)\left[\int_0^\infty dm \frac{dn}{dm} b(m) \frac{N^{\rm ET}_{\rm g}}{\overline{n}^{\rm ET}_{\rm g}} \tilde{u}(k,m)\right]^2, \tag{6.6}$$

for the ET galaxies.

$$P^{1h}_{\rm LT}(k) = \frac{1}{\left(\overline{n}^{\rm LT}_{\rm gal}\right)^2} \int_0^\infty dm \frac{dn}{dm} [2N^{\rm LT}_c N^{\rm LT}_s \tilde{u}(m,k) + \left(N^{\rm LT}_s\right)^2 \tilde{u}(k,m)^2], \tag{6.7}$$

$$P^{2h}_{\rm LT}(k) = P^{lin}_{mm}(k)\left[\int_0^\infty dm \frac{dn}{dm} b(m) \frac{N^{\rm LT}_{\rm g}}{\overline{n}^{\rm LT}_{\rm g}} \tilde{u}(k,m)\right]^2, \tag{6.8}$$

for the LT galaxies only. And

$$P^{1h}_{\rm mix}(k) = \frac{1}{\overline{n}^{\rm ET}_{\rm g}\overline{n}^{\rm LT}_{\rm g}} \int_0^\infty dm \frac{dn}{dm} [(N^{\rm ET}_c N^{\rm LT}_s + N^{\rm ET}_s N^{\rm LT}_c)\tilde{u}(k,m) + N^{\rm ET}_s N^{\rm LT}_s \tilde{u}(k,m)^2], \tag{6.9}$$

$$P^{2h}_{\rm mix}(k) = P^{lin}_{mm}\left[\int_0^\infty dm \frac{dn}{dm} b(m) \frac{N^{\rm ET}_{\rm g}}{\overline{n}^{\rm ET}_{\rm g}} \tilde{u}(k,m)\right] \times \left[\int_0^\infty dm \frac{dn}{dm} b(m) \frac{N^{\rm LT}_{\rm g}}{\overline{n}^{\rm LT}_{\rm g}} \tilde{u}(k,m)\right], \tag{6.10}$$

for the mixing between the two galaxy populations. Therefore, the galaxy power spectrum computed within a HM framework with two different galaxy populations reads as

$$P_{gal}(k) = P_{ET}(k) + P_{LT}(k) + P_{mix}(k), \tag{6.11}$$

where each term is given by the sum of its 1h and 2h terms.

Values of the clustering parameters of the two galaxy populations estimated in different analyses from Ade et al., 2011b (hereafter P11), Xia et al., 2012 (hereafter X12), and C13, are reported in table 6.1. The main difference between these previous works is in the modelling of the galaxy populations. P11 fits for one single galaxy population, allowing the clustering parameters to vary depending on the frequency channel (this frequency dependence explains the reason why in table 6.1, for P11, I report a range of values for the clustering parameters rather than a single estimate). X12 and C13 consider two galaxy populations, but use different frequency ranges to constrain the clustering parameters. Specifically, X12 works in the 150–1200 GHz frequency range, while C13 in the 600–3000 GHz frequency range. Pushing to higher frequencies, C13 was able to constrain the minimum mass for the LT galaxies. Both works keep the $\alpha_{\rm LT}$ clustering parameter fixed to unity to reduce the

Table 6.1: Clustering parameters from previous analyses.

	X12	P11	C13
$\log(M_{\rm min}^{\rm ET}/M_\odot h^{-1})$	12.09 ± 0.06	$11.95 \pm 2.10-$ 12.21 ± 0.51	12.00 ± 0.04
$\alpha_{\rm ET}$	1.81 ± 0.04	$1.02 \pm 0.87-$ 1.30 ± 1.16	1.55 ± 0.05
$\log(M_{\rm min}^{\rm LT}/M_\odot h^{-1})$	$\equiv 10.85$	-	10.85 ± 0.06
$\alpha_{\rm LT}$	$\equiv 1$	-	$\equiv 1$

Notes. Constraints on key clustering parameters from previous analyses in X12, P11 and C13.

number of free parameters in the fit. As stated in P11, the other clustering parameters which I do not report in table 6.1–*i.e.*, $\sigma_{\log M}$ and $M_{\rm sat}$–are not critical parameters for the fit. I tested this assumption by taking a wide range of possible values for $\sigma_{\log M}$ and $M_{\rm sat}$ [Ade et al., 2011b; Tinker et al., 2010b] and calculating the CIB power spectrum for different frequencies and found no visible impact from these two clustering parameters. For this reason, I also fix them to $\sigma_{\log M} = 0.1$ and $M_{\rm sat} = 20 M_{\rm min}^{\rm ET/LT}$, respectively, for both galaxy populations.

In figure 6.2 I show the different components contributing to the galaxy power spectrum computed within my modelling at redshift $z = 0.001$. Contributions from ET, LT, and the mixing between the two populations are shown in red, yellow, and orange, respectively. The dashed lines are the 1h terms, the dotted ones show the 2h terms, while the sum of the two is represented by solid lines. The black curve is the total galaxy power spectrum, as written in equation 6.11.

With the full expression of the galaxy power spectrum in hand, in order to arrive at the prediction of the CIB power spectrum based on this two-population modelling, as already presented in section 5.1 and, more specifically, in equation 5.5, one has to incorporate the information contained in the emissivity functions, $j_\nu(z)$ (see equation 5.3).

The emissivity functions encapsulate the information on how galaxies emit as a function of redshift. They depend on the LF and, because of the implemented two-galaxy-populations modelling, one has to model differently the LF depending on whether I am dealing with ET or LT galaxies. In the following I review the two different modellings implemented in this framework.

6.1.1.1 Modelling of the Luminosity Function for Early-type galaxies

From N-body simulations, ET galaxies are expected to populate more massive and more static dark matter haloes, for which it is easier to model the evolution. In fact, it is possible to extract an analytic model for the bolometric

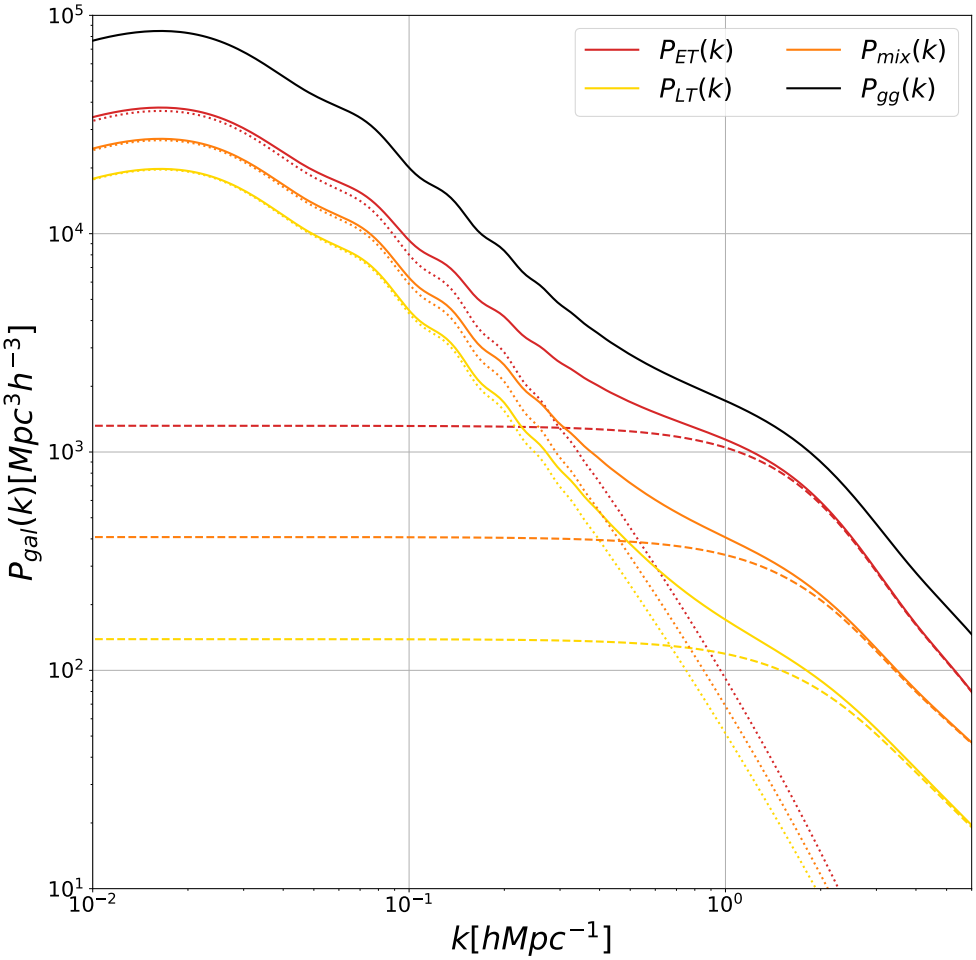

Figure 6.2: Predictions of the different components contributing to the galaxy power spectrum in a two-galaxy-populations HM framework. Specifically, the black curve is the galaxy power spectrum computed at redshift $z = 0.001$ as the sum of the contributions coming from ET and LT galaxies only, in red and yellow, and from the mixing between the two, in orange. Each single contribution is given by the sum of its own 1h and 2h terms, represented with the same colour legend, but with dashed and dotted curves, respectively.

LF of ET galaxies, obtained as a convolution of the halo formation rate, $\mathrm{d}n/\mathrm{d}t_{\mathrm{vir}}$, with the galaxy luminosity distribution, $P(\log L, z)$. To obtain the expression for the halo formation rate I follow the approach described in C13 and, in particular, a rewriting of the analytical expression of the HMF to explicitly encode the redshift evolution of the critical density for the collapse, $\delta_c(z)$, is required. To do that, I start from the Seth & Tormen (ST) parametrization of the HMF. I remind the reader of chapter a for a detailed description of how to move from a σ to a ν parametrization for the HMF. The starting point is equation 4.13

$$\frac{dn(m_{vir}, z)}{dm_{vir}} = \nu f(\nu) \frac{\bar{\rho}}{m_{vir}} \frac{d \ln \nu}{d \ln m_{vir}}. \tag{6.12}$$

The redshift evolution of the critical density for the collapse is given by

$$\delta_c(z) = \frac{\delta_0(z)D(0)}{D(z)}, \tag{6.13}$$

where $\delta_0(z)$ and $D(z)$, which are the extrapolated initial density and the linear growth factor, respectively, are approximated following the prescriptions of Lahav et al., 1991; Nakamura and Suto, 1997 and Carroll, 2001 as

$$\delta_0(z) \approx \frac{3}{20}(12\pi)^{2/3}[1 + 0.0123 \log \Omega_m(z)], \tag{6.14}$$

$$D(z) \approx \frac{5\Omega_m(z)}{2(1+z)}\left[\frac{1}{70} + \frac{209}{140}\Omega_m(z) - \frac{1}{140}\Omega_m^2(z) + \Omega_m^{4/7}(z)\right]^{-1}. \tag{6.15}$$

The redshift evolution of the critical density at the collapse enters the definition of the peak height, $\nu = (\delta_c(z)/\sigma(m_{vir}))^2$. Bardeen et al., 1986 and Sugiyama, 1995 also provide an analytical approximation for $\sigma(m_{vir})$ as

$$\begin{aligned}\sigma(m_{vir}) \approx \frac{0.8}{0.84}[&14.110393 - 1.1605397x - 0.0022104939x^2 \\ &+ 0.001331746x^3 - 2.1049631 \times 10^{-6}x^4],\end{aligned} \tag{6.16}$$

with $x \equiv \log m_{vir}/M_\odot$. This information about the redshift evolution enters the function describing the number density of peaks which, following ST formalism, is

$$f_{ST} \equiv \nu f(\nu) = A(1 + (a\nu)^{-p})\left(\frac{a\nu}{2\pi}\right)^{1/2} e^{-a\nu/2}, \tag{6.17}$$

with $A = 0.322$, $p = 0.3$, and $a = 0.707$.

The halo formation rate, dn/dt_{vir}, is then computed as the time derivative of the halo mass function:

$$\begin{aligned}\frac{dn}{dt_{vir}} &= \frac{dn(m_{vir}, z)}{dm_{vir}}\frac{d\ln f_{ST(\nu)}}{dt} \\ &\simeq \frac{dn(m_{vir}, z)}{dm_{vir}}\left[\frac{a\nu}{2} + \frac{p}{1 + (a\nu)^p}\right]\frac{d\ln\nu}{dz}\left|\frac{dz}{dt}\right|,\end{aligned} \tag{6.18}$$

with $dz/dt = -H_0(1+z)\sqrt{\Omega_{\Lambda,0} + \Omega_{m,0}(1+z)^3}$.

The second term I implemented for this modelling is the luminosity distribution computed at redshift z, inside a halo of mass m_{vir} which virialised at redshift z_{vir}, *i.e.* $P(\log L, z, m_{vir}, z_{vir})$. The total luminosity of a galaxy encodes the contributions arising from the stellar component and the active nucleus. For both components, the model I follow assumes a log-normal luminosity distribution

$$P(\log L|\log \bar{L})d\log L = \frac{e^{(-\log^2(L/\bar{L})/2\sigma^2)}}{\sqrt{2\pi\sigma^2}}d\log L, \tag{6.19}$$

where σ is the dispersion around the mean luminosity L, and it is different depending on whether I am considering the stellar component, $L_\star$, or the active nucleus, L_{AGN}, contributions. The distribution of the total luminosity, $L_{tot} = L_\star + L_{AGN}$, is therefore given by

$$\begin{aligned} P(\log L_{tot}|\log \bar{L}_\star, \log \bar{L}_{AGN}) d\log L = d\log L_{tot} \\ \times \int_{-\infty}^{\log L_{tot}} \frac{dx}{2\pi\sigma_\star\sigma_{AGN}} \frac{L_{tot}}{L_{tot}-10^x} e^{-(x-\log \bar{L}_\star)^2/2\sigma_\star^2} \\ \times e^{-[\log(L_{tot}-10^x)-\log \bar{L}_{AGN}]^2/2\sigma_{AGN}}. \end{aligned} \tag{6.20}$$

At this point, I have both elements required to obtain a prediction of the LF of ET galaxies, obtained as a convolution of the halo formation rate, in equation 6.18, and the galaxy luminosity distribution, in equation 6.20. The comoving differential LF, meaning the number density of galaxies per unit luminosity interval at a given redshift, is given by

$$\Phi(\log L, z) = \int_{m_{vir}^{min}}^{m_{vir}^{max}} \int_z^{z_{vir}^{max}} dz_{vir} \left|\frac{dt_{vir}}{dz_{vir}}\right| \frac{dn}{dt_{vir}} P(\log L, z|m_{vir}, z_{vir}). \tag{6.21}$$

6.1.1.2 Modelling of the Luminosity Function for Late-type galaxies

LT galaxies are expected to populate less massive dark matter haloes, which more likely merge with other low-mass dark matter haloes. This makes the modelling more challenging [Lapi and Cavaliere, 2011], and for this reason LT galaxies require an empirical parameterisation of their bolometric LF. Specifically, for the empirical parametrization I follow the C13 prescription, where LT galaxies are further divided into two sub-populations: warm star-burst galaxies and cold LT galaxies. The functional form adopted for both sub-populations is the one suggested by Saunders et al., 1990:

$$\Phi(\log L_{IR}, z) d\log L_{IR} = \Phi^* \left(\frac{L_{IR}}{L^*}\right)^{1-\alpha} \exp\left\{-\frac{\log^2(1+L_{IR}/L^*)}{2\sigma^2}\right\} d\log L_{IR}, \tag{6.22}$$

where Φ^* and L^* are the characteristic density and luminosity, respectively, α is the slope in the low-luminosity regime, and σ is the dispersion of the galaxy population (see table 1 of C13 for the values of these parameters).

6.1.1.3 Emissivity functions for the two galaxy populations model

Here I expand on what was already introduced in section 5.1 about the emissivity functions. Specifically, the emissivity function, defined in equation 5.3, depends on the LF involved in the calculations. Therefore, in this two galaxy populations modelling, I have two different predictions of the luminosity functions depending on whether I am dealing with ET or LT galaxies. Before moving to a more detailed description of the predictions in terms of the

emissivity functions for this modelling, it is important to underline the fact that the LFs reported in equation 6.21 and 6.22 are the bolometric expressions, meaning no frequency dependence is encoded. In order to move from the bolometric to the frequency-dependent expression of the LF I use templates Spectral Energy Distribution (SED)s for the two galaxy populations so that

$$L_\nu = \nu f(\nu) L_{bol}. \tag{6.23}$$

> The Spectral Energy Distribution (SED) represents a reconstruction of the spectrum of a galaxy as a superposition of different points obtained through photometric measurements. Specifically, it is a plot of energy versus frequency or wavelength of electromagnetic radiation. It is used in astronomy to characterise sources of galaxy populations.

I report figure 4 of C13 in figure 6.3 to show the template SEDs used for both ET galaxies–orange curve–and LT galaxies–given by both warm and cold sub-populations in blue and red, respectively.

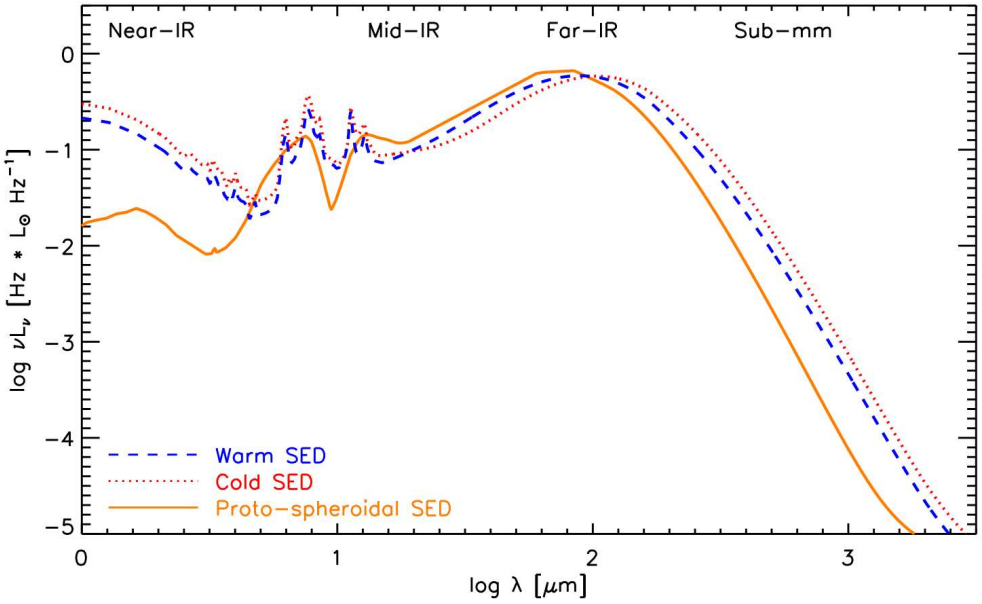

Figure 6.3: SED templates adopted for the two galaxy populations, as reported in C13. The orange curve represents the SED for ET galaxies, the blue and red curves describe the SEDs for the warm and cold galaxies, contributing to the LT population. Plot of the SEDs from figure 4 of C13.

At this point, I have all the elements needed in order to compute the LF for a specific frequency channel for both galaxy populations, and therefore to obtain predictions for the emissivity functions. For visualisation purposes, in figure 6.4 I show the emissivity functions computed for the 217, 353, 545, and 857 GHz frequency channels of the *Planck* satellite, computed for both galaxy populations. From figure 6.4 it is possible to appreciate how the different galaxy populations peak and contribute at different redshifts. Specifically,

LT galaxies have their emission peak at redshifts smaller than 2 and then decay fast, while ET galaxies do not contribute at redshifts below 1, but have a much broader extension at higher redshifts.

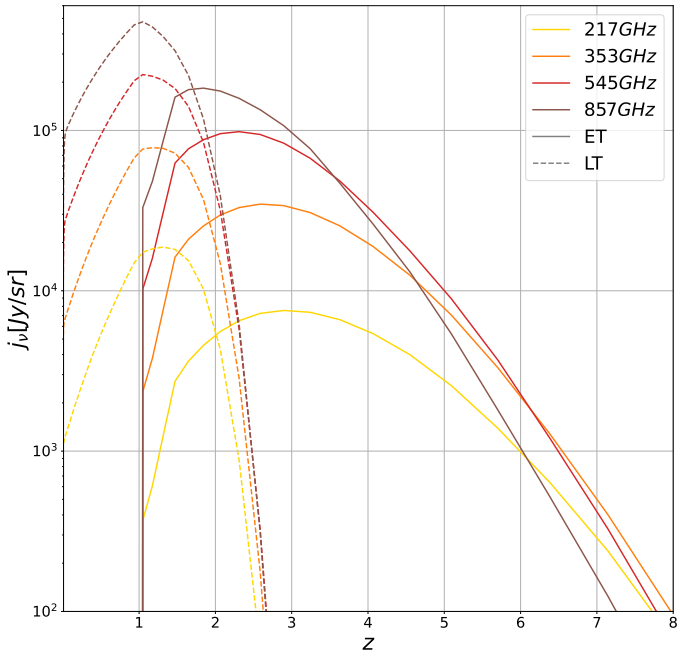

Figure 6.4: Emissivity functions for ET (*solid*) and LT (*dashed*) galaxy populations computed at 217 (*yellow*), 353 (*orange*), 545 (*red*), and 857 (*brown*) GHz frequency channels.

6.1.2 Modelling of the shot noise term

Even if from a theoretical point of view the SN component of the CIB power spectrum is obtained by integrating the contribution of sources falling below the flux cut (see equation 5.9 in section 5.1), at this level of my study I do not rely on computations based on the number counts. This is because number counts tend to be overestimated for the higher frequency channels due to resolution effects (see Bethermin et al., 2012), leading to much higher values of the SN with respect to those expected or measured.

Thus, I obtain the prediction for the SN power spectrum between two frequency channels by imposing a certain SN level for the single frequency channel, SN_ν, as

$$C^{SN}_{\ell,\nu\times\nu'} = \mathcal{C}_{\nu\times\nu'}\sqrt{SN_\nu \times SN_{\nu'}}, \tag{6.24}$$

where $\mathcal{C}_{\nu_1\times\nu_2}$ accounts for the possible correlation between emissions at different frequencies.

Within this first part of my work, I treat the SN level in each frequency channel as a free parameter of the model, to be constrained by comparison with the data. Specifically, the datasets described in the next section provide measured SN amplitudes for the different (auto- and cross-) spectra. I use these measurements to define informative priors for the SN parameters and as a benchmark for assessing the consistency of the inferred values with observations.

With the emissivity functions in hand, the complete two galaxy populations description of the galaxy power spectrum (see Equations 6.5–6.11), and the empirical parametrisation for the SN term employed in this work, I have all the elements to arrive at a prediction of the CIB power spectrum (see equations 5.6–5.9).

6.2 AVAILABLE CIB DATA

To constrain the parameters describing the clustering properties of the two galaxy populations, I fit my model to three datasets: the CIB power spectrum measurements from Planck Collaboration, 2014b (denoted as P14 in this work), the CIB power spectrum obtained from a re-analysis of *Planck* data by Lenz et al., 2019 (hereafter L19), and the angular power spectra from *Herschel*-SPIRE data presented by Viero et al., 2019 (hereafter V19).

6.2.1 Planck data

The *Planck* satellite, operated by the European Space Agency (ESA) in collaboration with National Aeronautics and Space Administration (NASA), was designed to map the entire sky at microwave and sub-millimetre wavelengths with unprecedented precision [Ade et al., 2011a; Planck Collaboration, 2011]. It observed in nine frequency channels, spanning 30 to 857 GHz, using two instruments: the Low Frequency Instrument (LFI) (30–70 GHz) based on radiometers, and the High Frequency Instrument (HFI) (100–857 GHz) employing bolometric detectors. Launched in May 2009, *Planck* conducted multiple full-sky surveys between 2009 and 2013, providing the most accurate measurements of the CMB temperature and polarisation anisotropies to date [Planck Collaboration, 2020a; b].

Figure 6.5: Planck satellite.

Beyond its primary cosmological target, *Planck* also delivered a wealth of information on Galactic and extragalactic emissions. The wide frequency coverage of the HFI channels enabled the separation of the CMB from foreground components such as thermal dust, synchrotron, and the CIB, while the highest-frequency bands (353–857 GHz) are dominated by the latter once the CMB is removed. These data therefore provide a unique opportunity to

study the statistical properties of the unresolved dusty galaxy population responsible for the CIB [Ade et al., 2011b; Planck Collaboration, 2014b].

In particular, P14 presented measurements of the CIB angular power spectra over five HFI frequencies (143, 217, 353, 545, and 857 GHz), using a total sky area of 2240deg^2. Flux-density cuts were applied at the map level to remove bright sources, with thresholds of $S_\nu^{\rm lim} = 350$, 225, 315, 350, and 710 mJy for the 143, 217, 353, 545, and 857 GHz channels, respectively (see table 1 of P14). In this study, I use ten CIB power spectra from P14, selecting only four frequency channels (217, 353, 545, and 857 GHz). Following P14, the 143 GHz data are excluded from the model fits, since this channel remains heavily contaminated by the CMB–its power spectrum at $\ell = 100$ being orders of magnitude higher than that of the CIB. Although one could apply significant corrections to mitigate this contamination, neither P14 nor the present analysis attempts such an approach.

The P14 dataset covers the multipole range from $\ell = 150$ to $\ell = 2500$, with logarithmic binning, $\Delta\ell/\ell = 0.3$. Across this range, P14 provides 8 data points for each CIB frequency auto- and cross-spectrum, resulting in a total of 80 data points in the dataset available for the model fit that is described in section 6.3. The error bars associated with each point account for both cosmic variance and instrumental noise. Since P14 data come without a covariance matrix, I conduct my analysis by building a diagonal covariance matrix from the spectra error bars and therefore neglecting possible correlations between different multipoles, as well as different frequency channels.

Data from *Planck* have been calibrated in two different ways, depending on the frequency channel. Specifically, the calibrator for the 217 and 353 GHz channels is the CMB orbital dipole, which boasts superior accuracy, while the 545 and 857 GHz channels are calibrated using planetary measurements. The calibration method generates a frequency-dependent uncertainty, with all values reported in table 6 of Planck Collaboration, 2016a and used in later sections when fitting the spectra. Additionally, updates to the calibrations were introduced between the first and the second *Planck* Public Releases, *i.e.* PR1 and PR2. The CIB spectra of P14 come from PR1 and to account for these calibration updates here I correct them with

$$C^{CIB}_{\ell,\nu_1\times\nu_2} = \frac{C^{CIB,PR1}_{\ell,\nu_1\times\nu_2}}{corr_{\nu_1}\times corr_{\nu_2}}. \tag{6.25}$$

In the above equation $C^{CIB}_{\ell,\nu_1\times\nu_2}$ represents the calibration-corrected data fed as input to the fit, $C^{CIB,PR1}_{\ell,\nu_1\times\nu_2}$ are the original CIB power spectra presented in P14, and $corr_{\nu_i}$ are the calibration corrections for each frequency, as detailed in Planck Collaboration, 2016a, equal to 0.991, 0.997, 1.018, 1.033 for the 217, 353, 545 and 857 GHz frequency channels, respectively.

As detailed in the next section, to compare the model predictions with the data I also need to apply colour corrections.

Colour correction coefficients account for the mismatch between the assumed reference spectrum used to calibrate an instrument and the actual SED of the CIB. Instruments such as *Planck* or *Herschel*-SPIRE are calibrated assuming a reference spectrum (*e.g.* $I_\nu \propto \nu^{-1}$, $\nu I_\nu = \text{const}$). The CIB, however, is produced by DSF galaxies with a modified blackbody SED, so the effective signal differs from what would be measured for the reference spectrum. Colour correction coefficients cc_ν are therefore introduced to convert the measured band-averaged intensity into a monochromatic intensity at the nominal band frequency consistent with the true CIB SED.

Specifically, P14 compute the colour corrections for each frequency channel using the CIB SED from Bethermin et al., 2012, ending up with colour-correction factors of 1.119, 1.097, 1.068 and 0.995 for the 217, 353, 545 and 857 GHz frequency channels, respectively.

6.2.2 Lenz data

The second dataset explored in my study is L19 (from Lenz et al., 2019), which provides six CIB power spectra using the 353, 545 and 857 GHz *Planck* maps–L19 excludes both 143 and 217 GHz channels on the basis of CMB contamination. As described in their paper, the authors of L19 employed a method based on neutral atomic hydrogen data to remove Galactic dust from the *Planck* intensity maps. This procedure allows one to extract the CIB power spectra down to lower multipoles compared to P14, reaching $\ell \sim 75$ (extending the P14 CIB power spectra starting from $\ell \sim 150$). L19 CIB data are also presented with a different and less aggressive binning scheme than P14's. Specifically, the authors apply a linear binning, including 64 multipoles in each bin, which results in a dataset with a total of 186 data points.

From the L19 dataset, I decide to exclude the first three bandpowers of each spectrum. I do this for two reasons: to better allow for a straightforward comparison with P14 data, and noting that the P14 data exhibit more power at lower multipoles compared to the L19 bandpowers (see figure 11 of L19). This could be due to a more aggressive dust cleaning procedure adopted in L19, but the consistency of the two datasets is beyond the scope of this work and not fully addressed in other literature for me to fully trust those additional bandpowers. I explore a few assumptions, including an investigation of the impact of the dust contamination in the L19 analysis in section 6.3.2.1.

As for P14, I apply to the L19 dataset colour corrections and the calibration factors, using the same values quoted in the previous section.

Similarly to P14, L19 spectra have been released without a covariance matrix. Once again, here I build a diagonal covariance of the bandpow-

ers, neglecting possible correlations between bands. Again, I explore this assumption in section 6.3.2.1.

6.2.3 Hershel-SPIRE data

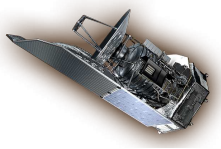

Figure 6.6: Hershel spacecraft.

The third dataset used in my work is the V19 CIB measurement from *Herschel*-SPIRE observations. The V19 dataset also includes auto- and cross-spectra involving SPT frequencies (95, 150 and 220 GHz) which could be beneficial for constraining the clustering parameters of ET galaxies. However, incorporating SPT data would have necessitated extending the modelling to additional extragalactic foregrounds which is outside the scope of this first part of the work. The comprehensive modelling of the extragalactic emission is the subject of the next chapter. The *Herschel Space Observatory* [Pilbratt et al., 2010], launched by the ESA in 2009, was equipped with a 3.5 m diameter mirror and operated over the far-infrared to sub-millimetre range (55–672 μm, corresponding to $\sim$ 450–5400 GHz). Among its three instruments, the *Spectral and Photometric Imaging Receiver* (SPIRE; Griffin et al., 2010) provided imaging and photometry in three broadband channels centred at 250, 350 and 500 μm (or 1200, 857, and 600 GHz, respectively), with beam Full Width Half Maximum (FWHM) of 18.1″, 24.9″, and 36.6″.[1] SPIRE's high sensitivity and angular resolution make it particularly suited to studies of the CIB, as it probes fainter dusty galaxies and smaller angular scales where the signal transitions from clustered to SN-dominated. Building upon these capabilities, several analyses–including V19– have exploited SPIRE maps to measure the CIB anisotropies at higher frequencies and multipoles than those accessible to *Planck*, providing complementary constraints on the emission and clustering properties of dusty star-forming galaxies.

More specifically, SPIRE observations approached the THz regime, covering 600, 857 and 1200 GHz from 90 deg^2. V19 applied a common flux density cut to the three frequency bands of $S_\nu^{lim} = 300$ mJy. The CIB spectra span multipoles from $\ell = 600$ to $\ell = 11000$. Following the binning scheme of Amblard et al., 2011, the bandpowers are linearly binned for $\ell < 1950$, and for higher multipoles a logarithmic binning with $\Delta\ell/\ell = 0.72$ is applied. In total, this dataset consists of 132 data points.

The maps are calibrated with an uncertainty of 5% for each frequency channel. The colour corrections needed at 600, 857 and 1200 GHz are 0.9988, 0.9929 and 0.9957, respectively [Lagache et al., 2003].

SPIRE spectra are accompanied by a covariance matrix which is, however, not well characterised[2]. In analogy with other studies (see *e.g.*, Maniyar et al.,

1 These correspond to effective angular resolutions of roughly 0.3–0.6 arcmin, higher than those of the *Planck* HFI channels at similar frequencies.

2 I started this work using the full, published covariance matrix and encountered a number of problems. After performing some tests, I concluded that this matrix exhibits a pathological behaviour. I found it to be occasionally negative-definite, thus not invertible. This was also noted by Maniyar et al., 2021.

2021) to overcome issues with the available covariance, I revert to assuming uncorrelated error bars also for V19 data, using a simple diagonal covariance matrix.

6.3 MODEL FIT TO DATA

Here I bring together the full model described in chapter 4, chapter 5, and more specifically in section 6.1, with the three datasets presented in section 6.2. I explore the model against the multi-frequency, multi-scale observations from *Planck* and *Herschel*-SPIRE individually. I discuss the results in the context of previous literature findings, and show that even this extended, two-population model–built to capture an extended frequency range–is not able to solve a tension between P14 and V19 CIB measurements which was reported in Maniyar et al., 2021 with a single-population model. This prevents me from achieving a joint fit using both datasets.

To compare the predictions of the model to the observed data, I correct the theory to account for instrument-specific corrections as follows:

$$C_{\ell,\nu_1\times\nu_2}^{CIB,\mathrm{data}} = \mathcal{A}_{\nu_1\times\nu_2} \times cc_{\nu_1} \times cc_{\nu_2} \times C_{\ell,\nu_1\times\nu_2}^{CIB,\mathrm{model}}. \tag{6.26}$$

Here, $C_{\ell,\nu_1\times\nu_2}^{CIB,\mathrm{model}}$ is the CIB power spectrum as predicted by the model; $\mathcal{A}_{\nu_1\times\nu_2} = \sqrt{f_{\mathrm{cal}}^{\nu_1} \times f_{\mathrm{cal}}^{\nu_2}}$ accounts for the absolute calibration uncertainties, *i.e.* $f_{\mathrm{cal}}^{\nu_i}$ for each different frequency channel; cc_{ν_i} is the colour correction per frequency channel.

To explore the parameters of the model, I perform a Markov Chain Monte Carlo (MCMC) analysis using the `emcee` ensemble sampler presented in Foreman-Mackey et al., 2013. The multi-dimensional parameter space is explored by sampling a Gaussian likelihood:

$$\log \mathcal{L}(C_\ell^{\mathrm{data}} | C_\ell^{\mathrm{model}}) \propto -\frac{1}{2} \sum_\ell \frac{\left(C_\ell^{\mathrm{data}} - C_\ell^{\mathrm{model}}\right)^2}{\sigma_\ell^2}, \tag{6.27}$$

where σ_ℓ^2 represents the error bars associated with the points of each dataset (*i.e.*, representing a diagonal covariance matrix for all three datasets).

The analysis considers different scenarios and the numbers of fixed or varied model parameters are set accordingly in each case as I describe below and in the individual subsections.

I perform the analyses covering two scenarios for the SN term modeled as in equation 6.24:

- In one case, I assume maximal correlation between the SN contributions, with the correlation coefficients in equation 6.24 all set to 1.
- In another case, the correlation coefficients are treated as free parameters in the MCMC analysis.

The data considered in this part of my study are able to constrain, with different degrees of sensitivity, the clustering parameters for both the early and the late galaxy populations, *i.e.*, $M_{\rm min}^{\rm ET}$, $\alpha_{\rm ET}$, $M_{\rm min}^{\rm LT}$ and $\alpha_{\rm LT}$; the SN levels for the considered frequency channels, SN_{ν_i}; and the calibration factors, $f_{\rm cal}^{\nu_i}$. However, the datasets covering *Planck* frequencies and multipole ranges are not able to constrain the $\alpha_{\rm ET}$ clustering parameter (see section 6.3.1 and figure 6.11), which I then fix to 1 for P14 and L19. On the other hand, as I show later, all datasets considered in this work can constrain $\alpha_{\rm LT}$ (either unconstrained or only weakly constrained in previous works), and this parameter is therefore always varied in the runs.

The choice of the prior distributions assigned to the free parameters, and the discussion and implications of the results from the individual model fits, are discussed in the following sub-sections.

6.3.1 P14 data analysis

For P14 I sample 10 and 16 free parameters for the two scenarios, respectively, with prior ranges listed in the second column of table 6.2. As mentioned above, since I expect no constraining power from *Planck*, in the baseline case I fix $\alpha_{\rm ET} = 1$. I vary the calibration factors with a Gaussian prior centered on 1 and with errors taken as twice the uncertainty in table 6 of Planck Collaboration, 2016a, except for the calibration factor at 217 GHz, which I fix to $f_{\rm cal}^{217} = 1$. This choice is driven by the very small uncertainty of the *Planck* measurement at 217 GHz, and it has been verified by extracting results treating $f_{\rm cal}^{217}$ as a free parameter and finding perfect agreement with the case in which it is fixed to unity. The clustering parameters and the SN levels are sampled from uniform distributions. When correlation coefficients are treated as free parameters, these are also sampled from uniform prior distributions.

Results (1σ confidence level) for the two model scenarios with fixed/varied correlations are reported in the third and the fourth columns of table 6.2. I obtain $\chi^2 = 81$ with 70 degrees of freedom when fixing the correlation coefficients, and $\chi^2 = 45$ when the correlation coefficients are free to vary (when the number of degrees of freedom is 64), with Probability To Exceed (PTE) of 17.9% and 96.9% for the two cases, respectively. I show how the best-fit model compares with the ten P14 CIB spectra in figure 6.7.

In figure 6.8, I report the full triangle plot showing the 1-dimensional posterior distributions and the 2-dimensional correlations (at 68 and 95% confidence) of all the parameters sampled in the MCMC analysis. Left panel of figure 6.9 shows again the 1-dimensional posterior distributions and the 2-dimensional correlations, but of the clustering parameters only, for the two scenarios considered in this work. Looking at both figures, I find that the two cases are in good agreement, and as expected I recover tighter constraints for some parameters due to the higher number of degrees of freedom in the MCMC when leaving correlations free to vary.

Table 6.2: Best-fit parameters from the model fit to P14 data.

Parameter	Prior	Results $\mathcal{C}_{\nu_1\times\nu_2}=1$	Results $\mathcal{C}_{\nu_1\times\nu_2}$ open
$\log(M^{\rm ET}_{\rm min}/M_\odot h^{-1})$	[10.7,12.8]	$11.45^{+0.16}_{-0.13}$	11.12 ± 0.19
$\log(M^{\rm LT}_{\rm min}/M_\odot h^{-1})$	[10.5,12.8]	$11.18^{+0.31}_{-0.27}$	$11.56^{+0.21}_{-0.17}$
$\alpha_{\rm LT}$	[0.2,3.5]	$1.337^{+0.063}_{-0.072}$	$1.436^{+0.075}_{-0.067}$
$SN_{217}[Jy^2/Sr]$	[0,50]	6.72 ± 0.76	22 ± 4
$SN_{353}[Jy^2/Sr]$	[50,500]	273 ± 16	276 ± 20
$SN_{545}[Jy^2/Sr]$	[400,4000]	1296 ± 120	1247 ± 100
$SN_{857}[Jy^2/Sr]$	[200,8000]	1827 ± 300	1899^{+400}_{-600}
$f^{353}_{\rm cal}$	1 ± 0.0156	0.999 ± 0.014	1.010 ± 0.015
$f^{545}_{\rm cal}$	1 ± 0.122	1.096 ± 0.032	1.094 ± 0.032
$f^{857}_{\rm cal}$	1 ± 0.128	1.289 ± 0.076	1.200 ± 0.072
$\mathcal{C}_{217\times353}$	[-1, 1]	-	$0.648^{+0.068}_{-0.075}$
$\mathcal{C}_{217\times545}$	[-1, 1]	-	0.469 ± 0.054
$\mathcal{C}_{217\times857}$	[-1, 1]	-	0.562 ± 0.070
$\mathcal{C}_{353\times545}$	[-1, 1]	-	$0.946^{+0.035}_{-0.031}$
$\mathcal{C}_{353\times857}$	[-1, 1]	-	$0.937^{+0.056}_{-0.042}$
$\mathcal{C}_{545\times857}$	[-1, 1]	-	$0.934^{+0.046}_{-0.039}$

Notes. First column: model parameters sampled for the fit to P14 data. Second column: priors and ranges of variation imposed on the model parameters. I adopt uniform priors for all parameters (ranges are within square brackets), except for the calibration factors for the 353, 545 and 857 GHz frequency channels, which are varied with Gaussian priors centered on 1. Third and fourth columns: mean and standard deviation of the model parameters for the case in which the correlations for the SN are fixed to one and the case in which they are free to vary, respectively. In the former case, I obtain $\chi^2=81$ with 80 points and 10 free parameters. The latter case has the same number of points, 16 free parameters, and $\chi^2=45$.

The relative impact seen here on the best-fit values for the minimum mass of the ET and LT galaxies is likely due to the fact that, within the multipole range probed by P14 data, the contribution to the CIB power spectrum from clustering dominates, while the SN is expected to dominate at smaller angular scales. Therefore, the different treatment of the SN levels does not have a significant impact on the clustering parameters.
Motivated by the results found on the L19 dataset (see next section), I re-run the analysis removing the spectra involving 217 GHz and note much closer agreement between the two scenarios as shown in the right panel of figure 6.9. To understand this, it is possible to look at the full parameter space explored by the fit. I find a tension in the two scenarios for the value of the SN level of the 217 GHz frequency channel, shown in figure 6.10. Specifically, I obtain a value for the SN level which is higher in the case with free correlations than in the case with fixed correlations at $\sim 4\sigma$ level. I also see that the correlation coefficients involving the 217 GHz frequency channel, *i.e.*, $\mathcal{C}_{217\times i}$ with $i=353,545,857$ (see fourth column of table 6.2), are significantly lower than unity. I explain this behaviour by noting the degeneracy between the

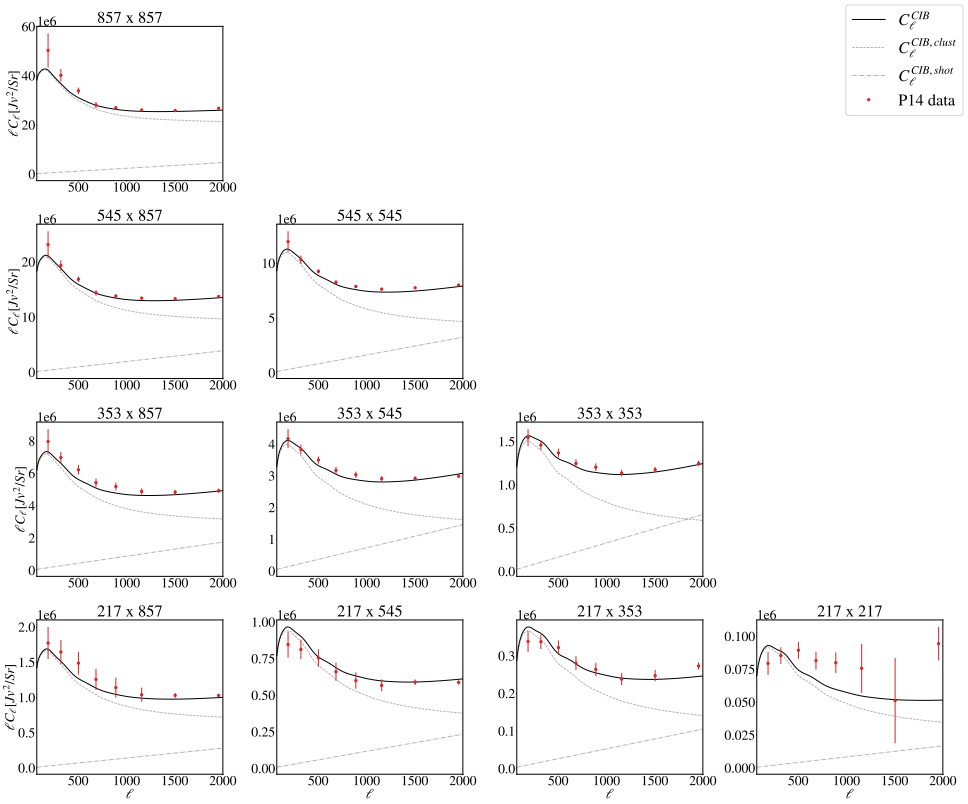

Figure 6.7: Comparison between P14 data and the model predictions obtained with the best-fit parameters in Table 6.2. The CIB power spectrum, the clustering term, and the SN term predicted by the model are represented with the solid black, the dashed grey, and the dot-dashed grey curves, respectively. P14 data with their error bars are represented in red.

SN level and the correlation coefficients at 217 GHz (see figure 6.10), which P14 data are not able to break. Specifically, they are anti-correlated, meaning that a shift of the SN level toward a lower value, closer to the one obtained in the first scenario, leads to higher values of the $\mathcal{C}_{217\times i}$.

When comparing the results with other literature findings, I note significant deviations between the values for the minimum masses of both galaxy populations extracted here and those found in the literature, more evident for the ET galaxies. Specifically, I find the minimum mass, $\log\left(M_{\rm min}^{\rm ET}/M_\odot h^{-1}\right)$, to be lower than previously reported values (see table 6.1). In particular, when comparing the value I obtain when fixing the correlation coefficients for the SN to unity, $\log\left(M_{\rm min}^{\rm ET}/M_\odot h^{-1}\right) = 11.45^{+0.16}_{-0.13}$, with the ones of C13, $\log\left(M_{\rm min}^{\rm ET}/M_\odot h^{-1}\right) = 12.00^{+0.04}_{-0.04}$, and X12, $\log\left(M_{\rm min}^{\rm ET}/M_\odot h^{-1}\right) = 12.09^{+0.06}_{-0.06}$, I find a tension at the level of $\sim 4\sigma$. In contrast, my analysis shows a higher value of the minimum mass for LT galaxies than the expected value found in C13 of $\log\left(M_{\rm min}^{\rm LT}/M_\odot h^{-1}\right) = 10.85 \pm 0.06$, but it is still compatible with the current constraint within $\sim 1\sigma$. This discrepancy raises a critical point as far as the model is concerned: my results suggest that the minimum mass for

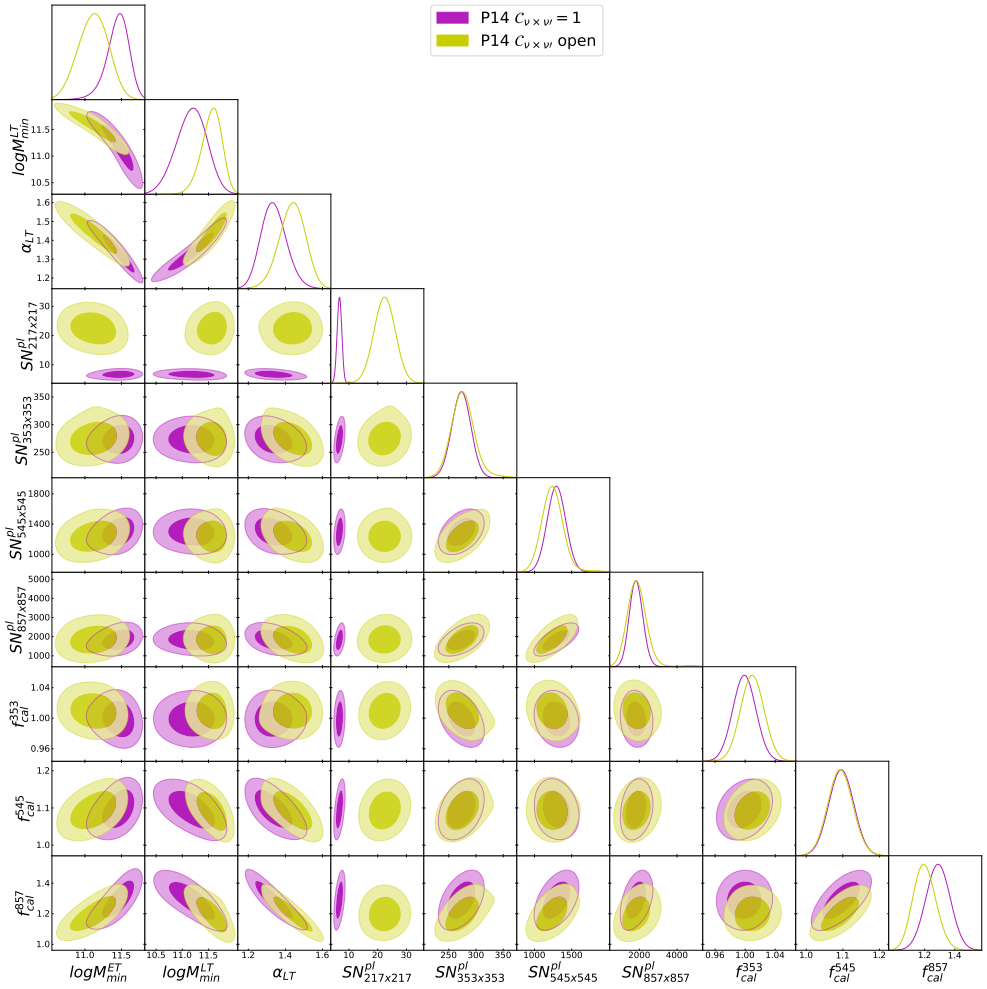

Figure 6.8: Posterior probability distributions of the free parameters of the model fit to P14 data for the two scenarios where the correlations for the SN are fixed to unity (in purple) and where they are free parameters in the fit (in green). I fix $\alpha_{\rm ET} = 1$.

ET galaxies is not higher than that for LT, contrary to prior assumptions. I speculate that this inconsistency may stem from the limited frequency and multipole ranges covered by P14 data. These ranges may not be sufficiently wide to effectively distinguish between the contributions of the two galaxy populations to the CIB spectrum, and, consequently, break the degeneracies among all three clustering parameters (see left panel of figure 6.9). Concerning $\alpha_{\rm LT}$, as already mentioned, previous analyses only considered it fixed to unity. Here, $\alpha_{\rm LT}$ is a free parameter of the model fit and is constrained at the best-fit values of $\alpha_{\rm LT} = 1.337^{+0.063}_{-0.072}$ and $\alpha_{\rm LT} = 1.436^{+0.075}_{-0.067}$ for the case where the correlation coefficients are fixed and free to vary, respectively.

I now comment on the SN correlation coefficients, in the scenario where they are free to vary (fourth column of table 6.2). To ease the comparison with

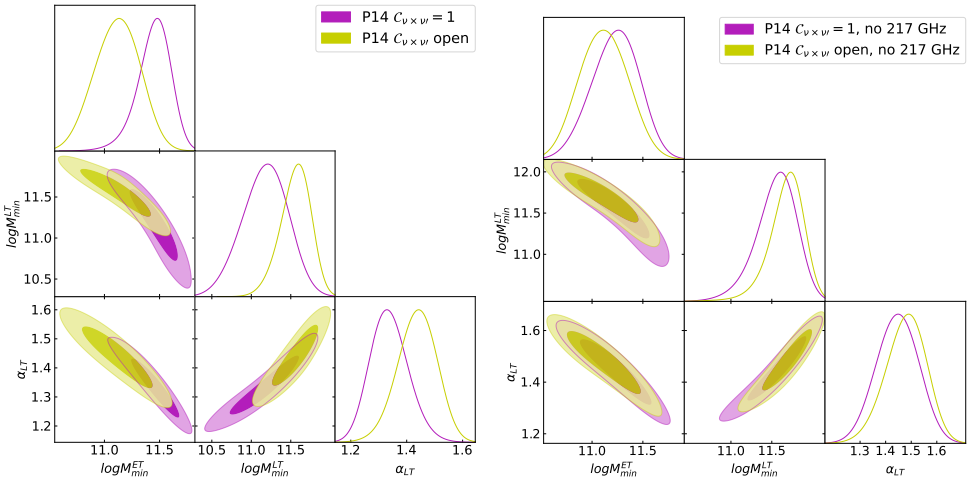

Figure 6.9: Same as figure 6.8 for the clustering parameters only. Left panel shows the 2-dimensional posteriors for the baseline case, while right panel explores the case when the 217 GHz frequency channel is excluded in the analysis. The posterior distributions for the two cases are in agreement, exploring the same regions of the parameter space. The agreement is even more pronounced when excluding the 217 GHz frequency channel.

previous *Planck* analyses, I also report in table 6.3 the values of the correlation coefficients extrapolated from table 6 of P14, as

$$\mathcal{C}_{\nu_1\times\nu_2} = \frac{C^{\rm Planck}_{\ell,\nu_1\times\nu_2}}{\sqrt{C^{\rm Planck}_{\ell,\nu_1\times\nu_1}\,C^{\rm Planck}_{\ell,\nu_2\times\nu_2}}}, \text{with}\,\nu_1\neq\nu_2, \tag{6.28}$$

where $C^{\rm Planck}_{\ell,\nu_1\times\nu_2}$ represent the entries of table 6 of P14. I observe that the correlation coefficients involving the 217 GHz frequency channel, $\mathcal{C}_{217\times i}$ with $i = 353, 545, 857$, are always lower than the values reported in table 6.3, even though they are still compatible within $\sim 2\sigma$. On the other hand, the remaining correlation coefficients are compatible with the values in table 6.3 within $\sim 1\sigma$. The observed low correlation coefficients are likely due to non-trivial degeneracies between the free parameters of the model, which the specific data are not able to efficiently break. In other words, the number of free parameters for this second scenario is too high and the model overfits the data, and this is confirmed by the low χ^2 obtained in this scenario ($\chi^2 = 45$).

I also report the results of the fit to P14 in an extended model where $\alpha_{\rm ET}$ is free to vary. I sample $\alpha_{\rm ET}$ from a uniform distribution in the range [0.2, 3.5]. The priors on the remaining parameters are the same as those reported in table 6.2. Figure 6.11 shows the posterior distributions of the clustering parameters.

P14 does not fully constrain $\alpha_{\rm ET}$ beyond setting an upper bound of $\alpha_{\rm ET} < 1.14$ at 95% CL. In addition, this parameter does not show significant correlation with the other clustering parameters. These considerations justify my choice of fixing $\alpha_{\rm ET} = 1$ in the baseline results.

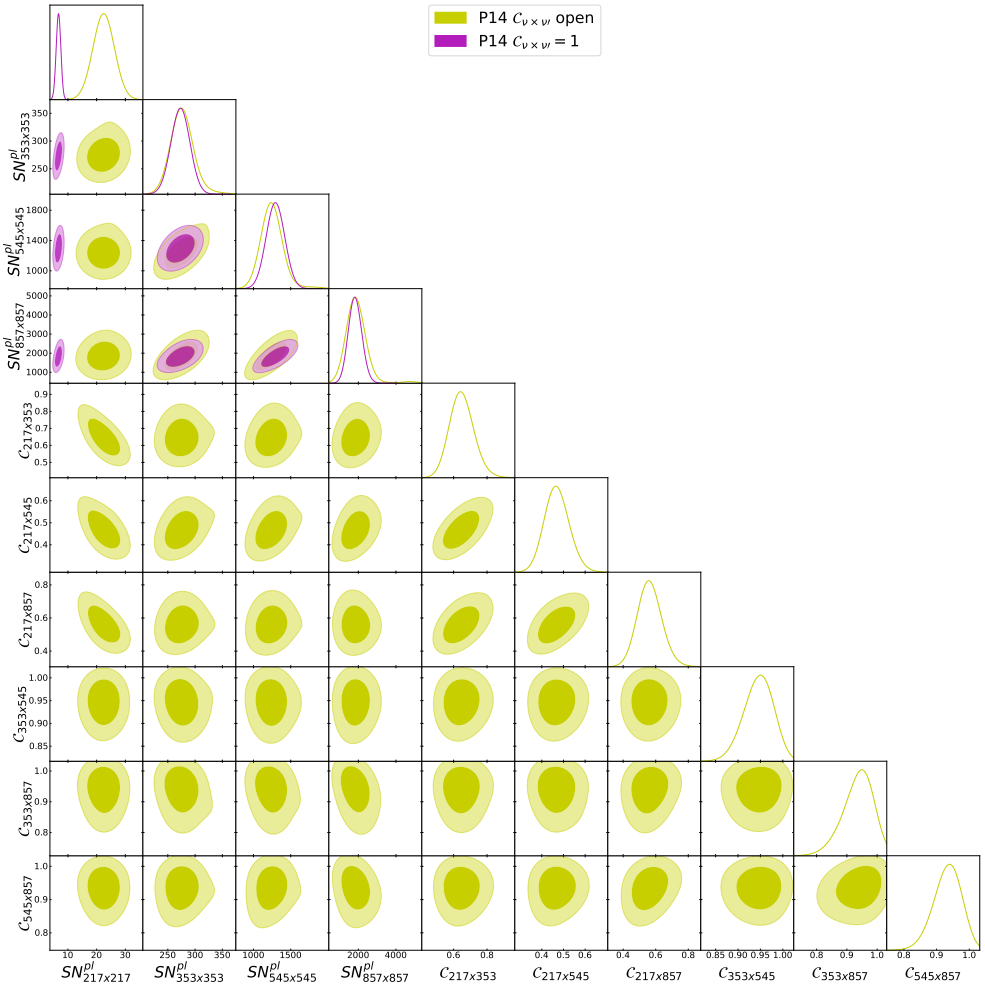

Figure 6.10: Posterior probability distributions of the SN levels for the scenario where the correlation coefficients are fixed to unity (in purple), and posterior probability distributions for the SN levels and the correlation coefficients for the scenario where the latter are free to vary (in green). This figure highlights the movement in the 217 GHz SN and shows the degeneracies with parameters characterising the spectra including at least one leg of the 217 GHz channel.

Table 6.3: P14 correlation coefficients.

	217	353	545	857
217	1	0.98 ± 0.24	0.88 ± 0.21	0.72 ± 0.17
353	-	1	0.95 ± 0.23	0.81 ± 0.20
545	-	-	1	0.93 ± 0.23
857	-	-	-	1

Notes. Correlation coefficients for the SN cross-spectra evaluated from equation 6.28 using the SN levels presented in table 6 of P14.

6.3.2 L19 data analysis

In this section, I present the analysis of the L19 dataset. I note that, as explained in more detail later, the model fits to L19 for both scenarios of

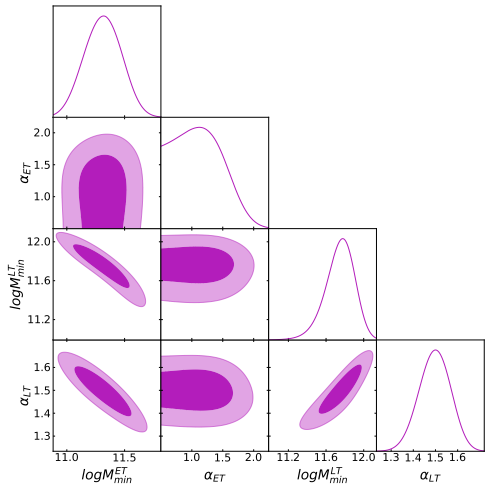

Figure 6.11: Posterior distributions of the clustering parameters from the analysis of the P14 dataset when allowing $\alpha_{\rm ET}$ to vary freely.

Table 6.4: Best-fit parameters from the model fit to L19 data.

Parameter	Prior	Results $\mathcal{C}_{\nu_1\times\nu_2}=1$	Results $\mathcal{C}_{\nu_1\times\nu_2}$ open
$\log(M^{\rm ET}_{\rm min}/M_\odot h^{-1})$	[10.7,12.8]	$11.505^{+0.057}_{-0.049}$	$11.488^{+0.063}_{-0.055}$
$\log(M^{\rm LT}_{\rm min}/M_\odot h^{-1})$	[10.5,12.8]	$10.75^{+0.15}_{-0.17}$	10.80 ± 0.16
$\alpha_{\rm LT}$	[0.2,3.5]	$1.254^{+0.023}_{-0.028}$	$1.261^{+0.027}_{-0.032}$
$SN_{353}[Jy^2/Sr]$	[50,500]	261.8 ± 6.9	264.1 ± 7.1
$SN_{545}[Jy^2/Sr]$	[400,4000]	1396 ± 45	1396 ± 46
$SN_{857}[Jy^2/Sr]$	[200,8000]	2391 ± 120	2425 ± 130
$f^{353}_{\rm cal}$	1 ± 0.0156	0.965 ± 0.015	0.966 ± 0.015
$f^{545}_{\rm cal}$	1 ± 0.122	1.174 ± 0.020	1.175 ± 0.021
$f^{857}_{\rm cal}$	1 ± 0.128	$1.468^{+0.042}_{-0.038}$	$1.460^{+0.048}_{-0.043}$
$\mathcal{C}_{353\times545}$	[-1, 1]	-	0.9838 ± 0.0076
$\mathcal{C}_{353\times857}$	[-1, 1]	-	$0.9970^{+0.0036}_{-0.0022}$
$\mathcal{C}_{545\times857}$	[-1, 1]	-	$0.9851^{+0.0093}_{-0.0082}$

Notes. Same as table 6.2 but for the L19 dataset.

correlations result in a poor χ^2. I, therefore, do not attempt to interpret the goodness of fit to this dataset and present results in this section only as a qualitative outcome.

The choice of the free parameters in the model fit and their prior distributions mirror what was described in the previous section and are listed in the first and second column of table 6.4, respectively. As done for the analysis of P14 data, $\alpha_{\rm ET}$ is fixed to unity, and the priors for the calibration factors of the 353, 545, and 857 GHz channels are based on *Planck* uncertainties.

In table 6.4, I also report the 1σ confidence interval results in both the scenarios of fixed (third column) and variable (fourth column) SN correlation coefficients. In figures 6.12 and 6.13, I show the posterior distributions of all

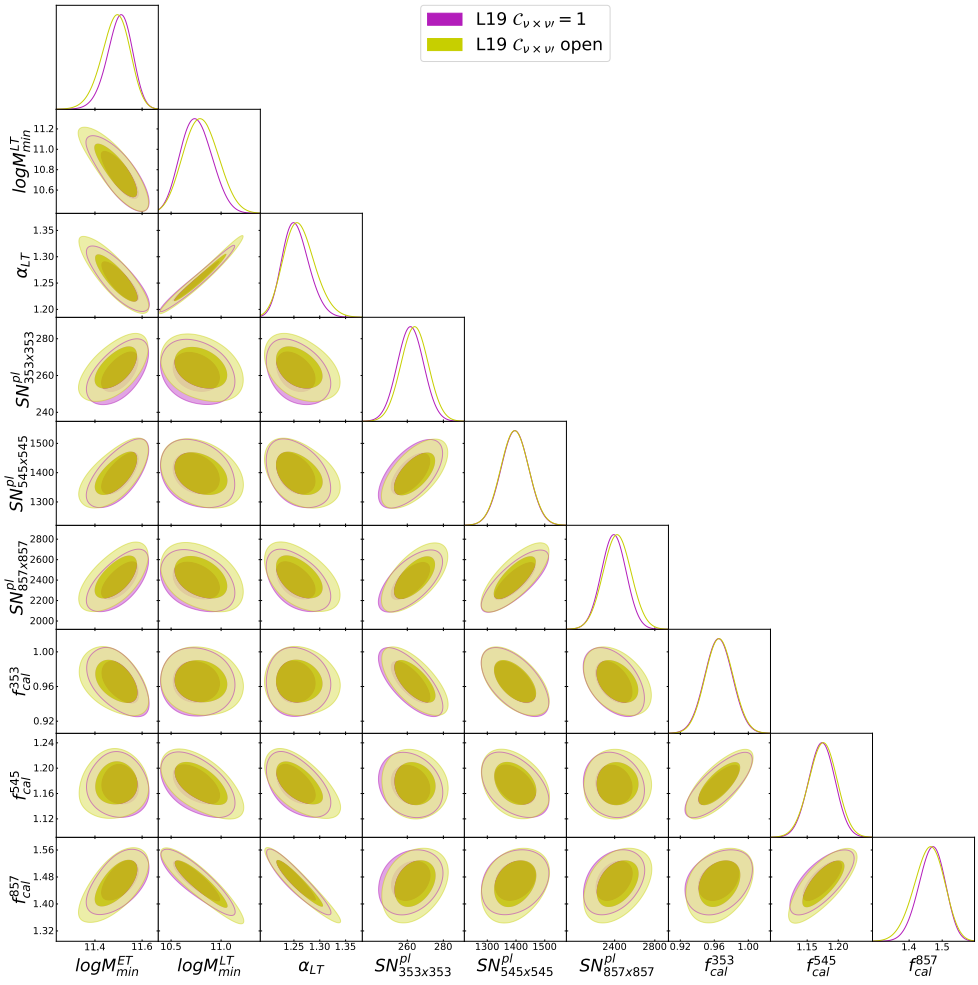

Figure 6.12: Same as figure 6.8 for L19 data.

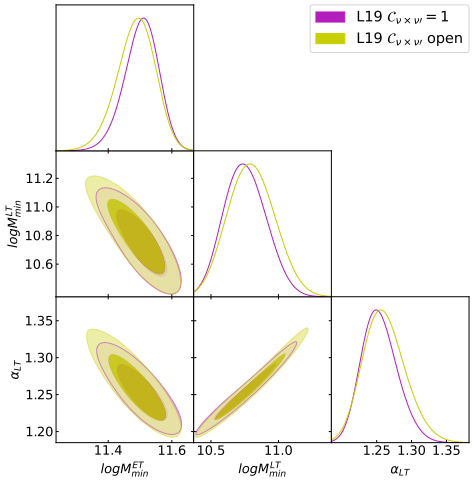

Figure 6.13: Same as figure 6.9 but for the L19 CIB dataset.

the parameters, and the clustering parameters only, for the two scenarios of SN correlation coefficients fixed to unity (in purple), and free-to-vary

correlation (in green). Focusing on the clustering parameters, the results show perfect agreement between the two scenarios, with only a minor broadening of the distributions in the case of increased degrees of freedom in the fit. Specifically, I note that the posterior distributions deviate much less than those of P14 seen in the left panel of the same figure, confirming that the exclusion of the 217 GHz frequency channel–not included at all in the L19 dataset–makes the results of the fits more stable among the two scenarios.

The model fits to the L19 data result in poor goodness-of-fit metrics, in particular exceedingly high χ^2 values (specifically I obtained a χ^2 of 655 and 650 for 159 and 156 degrees of freedom, respectively). However, I cannot definitively conclude that this is caused by deficits in the model. Section 6.3.2.1 is devoted to a detailed and closer examination of the L19 dataset; I show there how a smooth theory cannot meet the bandpowers as published, and that changing assumptions in the dataset is also not enough to improve the χ^2, and that the qualitative results presented here remain valid.

Without focusing on the exact results from L19, in the following I do a qualitative exploration of how the posteriors compare to those of P14. The clustering parameters for the two scenarios explored in this study are shown in the left and right panels of figure 6.14. The overlap between the two datasets is good (the largest shift is noted in the 1D posterior distribution of the minimum masses for the LT galaxies in the right panel of figure 6.14, approaching a $\sim 3\sigma$ difference).

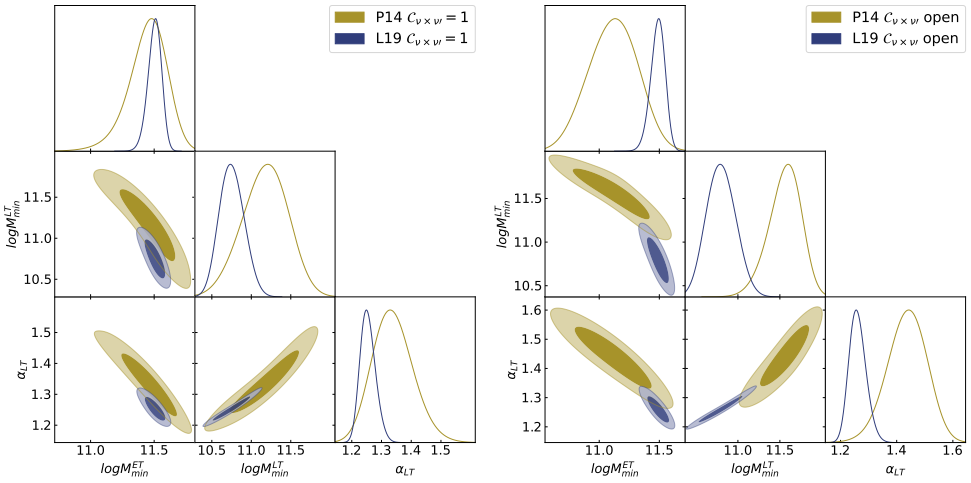

Figure 6.14: Left panel: Comparison of the posterior distributions of the clustering parameters for P14 (dark yellow) and L19 (dark purple) data in the case in which the correlation coefficients of the SN are fixed to unity. Right panel: Same as left panel, but for the case in which the correlation coefficients of the SN are free to vary.

In figure 6.15, I show the qualitative comparison between L19 bandpowers and the theoretical predictions for the CIB spectra obtained using the best-fit values reported in table 6.4 for the scenario where the SN correlation coefficients are fixed to unity. From this figure it is possible to immediately infer that, in the absence of strong correlations between points, the fit will

not have a good χ^2. A striking feature of this figure is the oscillating pattern of the L19 dataset which cannot be easily captured by the smooth predictions of the model.

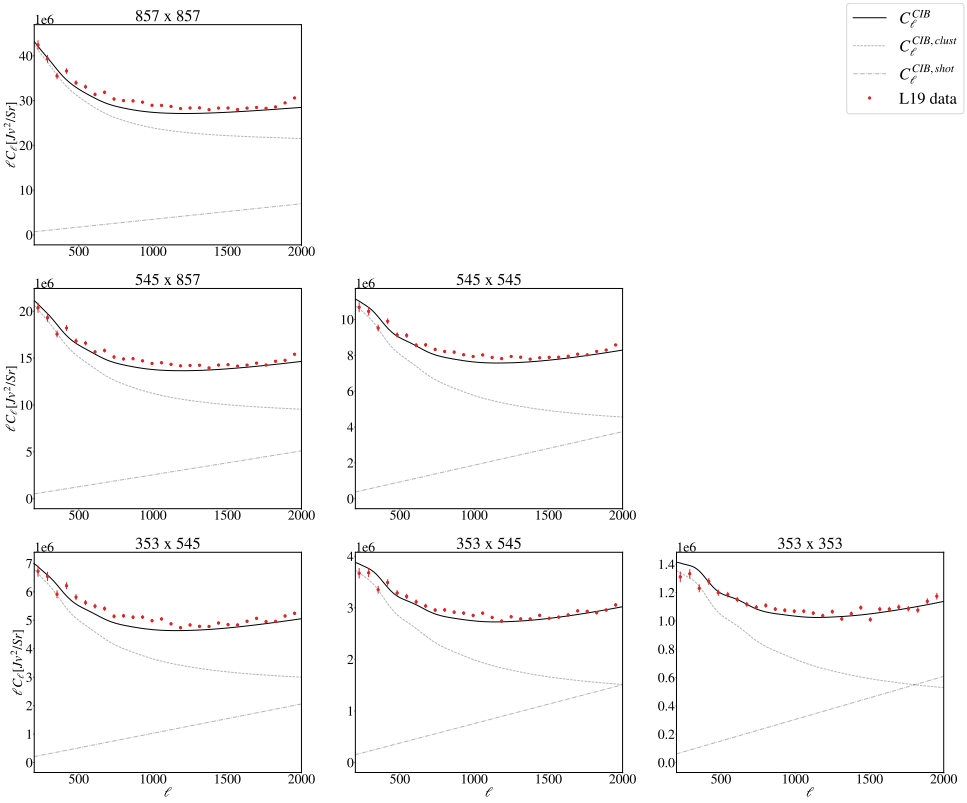

Figure 6.15: Same as figure 6.7, for the L19 dataset (red points) and using the best-fit parameters from table 6.4.

6.3.2.1 *Investigating the high χ^2*

To understand the L19 results further, I ran various tests on the L19 dataset taking the CIB intensity maps that they use and deriving from scratch CIB spectra. In the following I detail all the tests and further investigations I carried on, specifically

- I build a covariance matrix from scratch and run few MCMC analyses;
- I investigated the response in terms of model fit for single-frequency analyses.

COVARIANCE MATRIX CONSTRUCTION As an initial step, I implement L19's pipeline to compute the CIB power spectra and extend it to build a Gaussian covariance matrix. I do this to explore the impact of possible bin-to-bin correlations in the fit and in the χ^2. The covariance matrix that I obtain is shown in figure 6.16, while figure 6.17 reports the correlation matrix. From

these, it is possible to note that there is no significant correlation among different multipoles, but there is very significant correlation between the same bin at different frequencies. However, as I detail below, re-running the analysis including this new matrix showed no impact on the χ^2 of the fit.

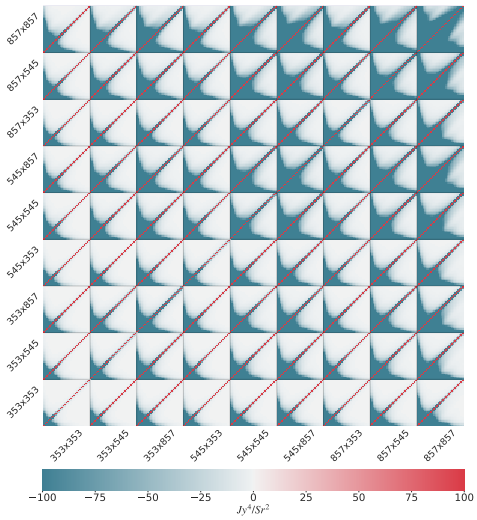

Figure 6.16: Covariance matrix of the L19 CIB measurements.

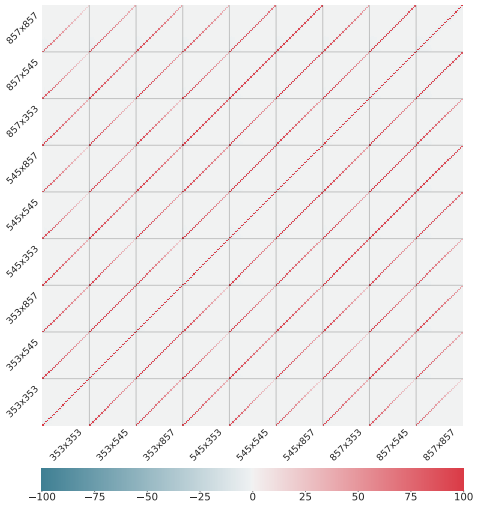

Figure 6.17: Correlation matrix of the L19 CIB measurements.

This covariance matrix accompanied three different configurations of the spectra:

- The first version of the dataset was obtained applying the same neutral hydrogen column density (N_{HI}) threshold of $2.5 \times 10^{20}\mathrm{cm}^{-2}$ across all three frequency channels (blue curve in figure 6.18, named *L19 v1*).
- The second dataset applied a lower N_{HI} threshold of $1.5 \times 10^{20}\mathrm{cm}^{-2}$ across all channels to explore the impact of dust and in particular of dust residuals in the analysis (green curve in figure 6.18, named *L19 v2*).
- The third dataset has been obtained following the prescription of L19 and imposing a different threshold for each frequency channel. Specifically I set $N_{HI} = 2.5 \times 10^{20}\mathrm{cm}^{-2}$, $2.0 \times 10^{20}\mathrm{cm}^{-2}$ and $1.8 \times 10^{20}\mathrm{cm}^{-2}$ for 353, 545 and 857 GHz respectively (orange curve in figure 6.18, named *L19 v3*).

Figure 6.18 compares these three datasets with the publicly available L19 CIB power spectra (shown as black data points with error bars).

I start the tests by using the first dataset to investigate the impact of different covariance choices on the goodness of fit. To do this I go through:

1. Running a MCMC with the full covariance matrix;
2. Using the full covariance matrix but limiting the analysis to the information from the frequency auto-spectra only;

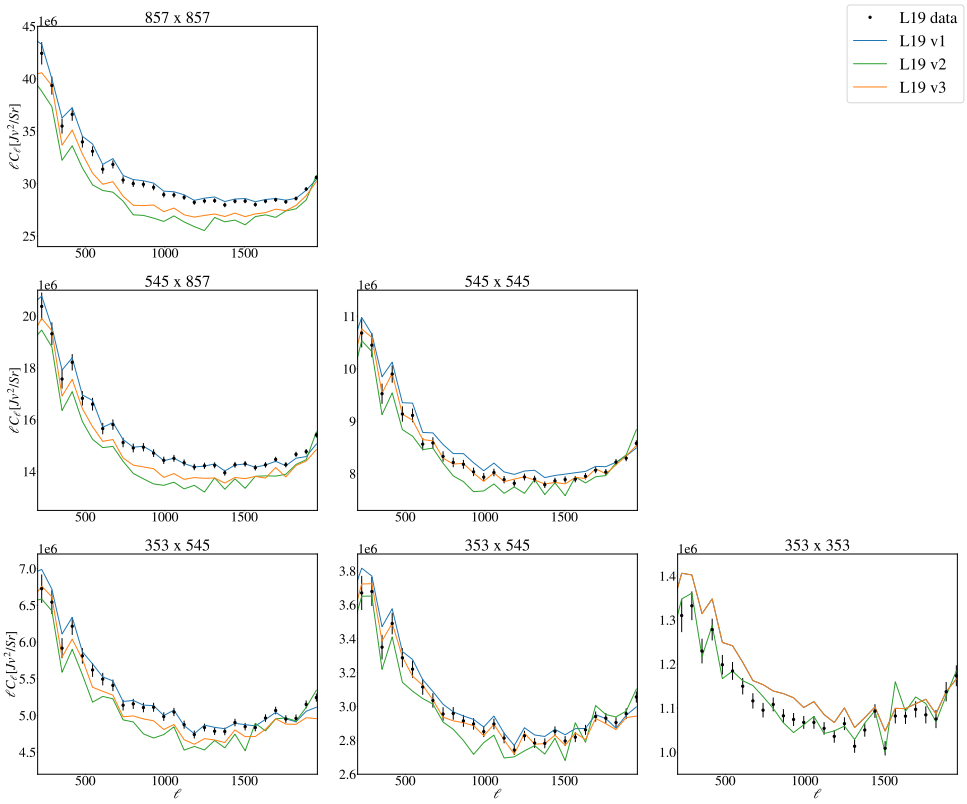

Figure 6.18: CIB power spectra from L19 (black points) compared to the spectra obtained after a re-analysis of the maps and applying different neutral hydrogen column density thresholds. Blue: threshold $2.5 \times 10^{20}\mathrm{cm}^{-2}$ across all three frequency channels; green: $1.5 \times 10^{20}\mathrm{cm}^{-2}$ for all channels; orange: different thresholds per channel following L19 ($N_{HI} = 2.5, 2.0, 1.8 \times 10^{20}\mathrm{cm}^{-2}$ for 353, 545 and 857 GHz respectively).

3. Employing only the diagonal of the covariance matrix;

Full covariance, full spectra The first run concluded that even when using the full covariance matrix, the model does not attain a good χ^2, continuing to find values of ~ 700. Furthermore, I also observed very different behaviours in the posteriors and best-fit values compared to the fiducial run presented above. Specifically, the run using the full covariance matrix yielded very similar values for the minimum masses of ET and LT galaxies, compared to differences of $\sim 3\sigma$ in the nominal case. Comparing the parameter posteriors with those obtained in the fiducial case presented previously (see figure 6.13 for the clustering parameters only and figure 6.12 for the full triangle plot), I noted that the minimum mass for the ET galaxies is no longer degenerate with the other clustering parameters. In contrast, the minimum mass for the LT galaxies and the α_{LT}, and all the remaining free parameters of the fit, show the same degeneracies as the fiducial case.

Full covariance, auto-frequency spectra only The runs that employed the full covariance with the frequency auto-spectra only and that using only

the diagonal of the covariance matrix produced results that aligned more closely with the original L19 dataset findings, both in terms of best-fit values and parameter space degeneracies. This better alignment with the main findings, especially when considering the run employing only the diagonal of the covariance matrix, was expected as it closely resembles the scenario presented previously. However, none of these tests led to better values of the χ^2.

Diagonal of the full covariance matrix, auto-frequency spectra only When using the full covariance matrix but limiting the analysis to the information from the frequency auto-spectra only and employing only the diagonal of the covariance matrix, I observe that the agreement with the fiducial case is stronger for the clustering parameters and slightly worse for the SN levels and the calibration factors, especially for the highest frequency channel. This led me to explore a possible dust residual which would be more important at higher frequencies. For this reason, I further investigated the model fit using single-frequency data.

SINGLE-FREQUENCY ANALYSES None of the single-frequency MCMC analyses could effectively constrain the minimum mass of the LT galaxies or the SN levels. The upper panel of figure 6.19 compares the recovered clustering parameters and shows a strong tension in the minimum mass of the ET galaxies between the 353 GHz frequency channel and the 545 and 857 GHz frequency channels. Specifically, the two higher frequency channels result in much higher best-fit values for the minimum mass of ET galaxies, both when compared to the best-fit value for the single-frequency run involving the 353 GHz frequency channel, and when compared to the best-fit value of the minimum mass found in my fiducial case and reported in table 6.4. Recalling that this clustering parameter acts as a re-scaling of the power spectrum, the high values could possibly be hinting at an excess of power coming from dust residuals.

To better explore this hypothesis, I evaluate the model fit to single-frequency data from the second version of the dataset (see lower left panel of figure 6.19) which employs a more aggressive mask and therefore should have smaller contamination. I do not find a difference in the constraining power for the minimum mass of LT galaxies and for the SN levels. However, the behaviour of the posterior distributions of the clustering parameters changes significantly. In particular, there is no tension between the minimum masses of ET galaxies, supporting the hypothesis of a dust residual impacting the results from higher frequencies in the previous dataset.

Finally, the third dataset, which used different thresholds for the three frequency channels, showed no significant differences from the second dataset (see lower right panel of figure 6.19), except for the 353 GHz frequency channel which provided more constraining power due to more data retained.

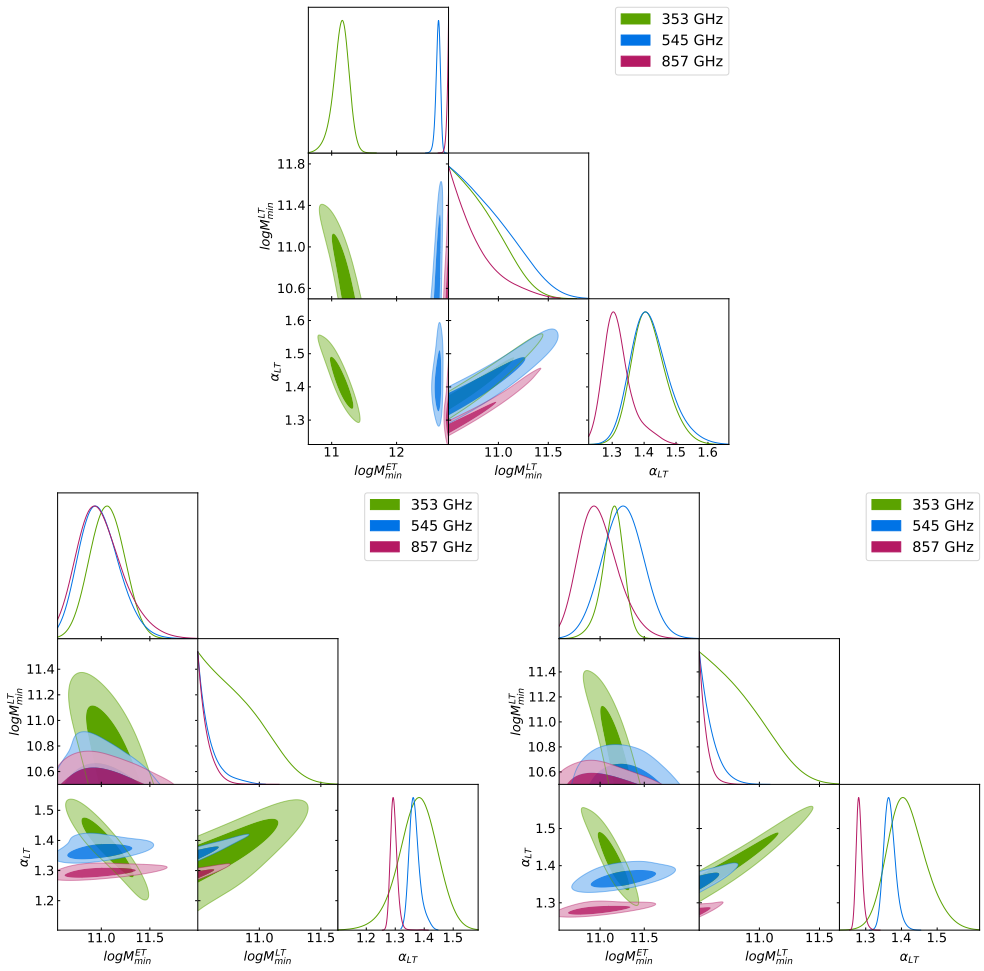

Figure 6.19: Upper: Comparison of the posterior distributions of clustering parameters from the single-frequency MCMC using the first L19 dataset. 353, 545 and 857 GHz frequency channels are represented in green, blue and pink, respectively. Lower left: Same as the upper panel for the second L19 dataset. Lower right: Same as the upper panel for the third L19 dataset.

This exploration highlights that the performance of the model in the multipole region where clustering terms dominate is highly sensitive to dust presence. Ensuring a dataset free from foreground contamination is crucial for testing the model and achieving reliable predictions.

6.3.3 V19 data analysis

I now move to discuss the results from the V19 *Herschel*-SPIRE dataset.

The model parameters and the prior distributions adopted in the analysis are reported in the first and second column of table 6.5, respectively. Unlike the previous datasets, the SPIRE data allow to constrain the $\alpha_{\rm ET}$ clustering parameter. I therefore let it be free to vary along with the other clustering parameters. The calibration factors for the three frequency channels, $f^i_{\rm cal}$ with

$i = 600, 857, 1200$, are varied with a Gaussian prior centered on 1 and a 1σ width determined as twice the calibration accuracy reported in Valtchanov, 2017. As done for the P14 and L19 analyses, I consider two scenarios of fixed and free-to-vary SN correlation coefficients.

The results of the model fit (68% confidence level) for the two cases under study are reported in the third and the fourth columns of table 6.5. I obtain $\chi^2 = 101$ for 122 degrees of freedom, when fixing the correlation coefficients for the SN, and $\chi^2 = 80$ for the scenario where the correlation coefficients are free to vary and the number of degrees of freedom is 119. The spectra–model comparison is shown in figure 6.20, for the case where the correlation coefficients for the SN are fixed to unity across all the V19 spectra. Specifically, V19 data start to be sensitive not only to the information coming from the clustering part, but also to the SN contribution. I find good agreement both in the multipole region where the clustering dominates and in the region where the SN dominates, at all frequencies.

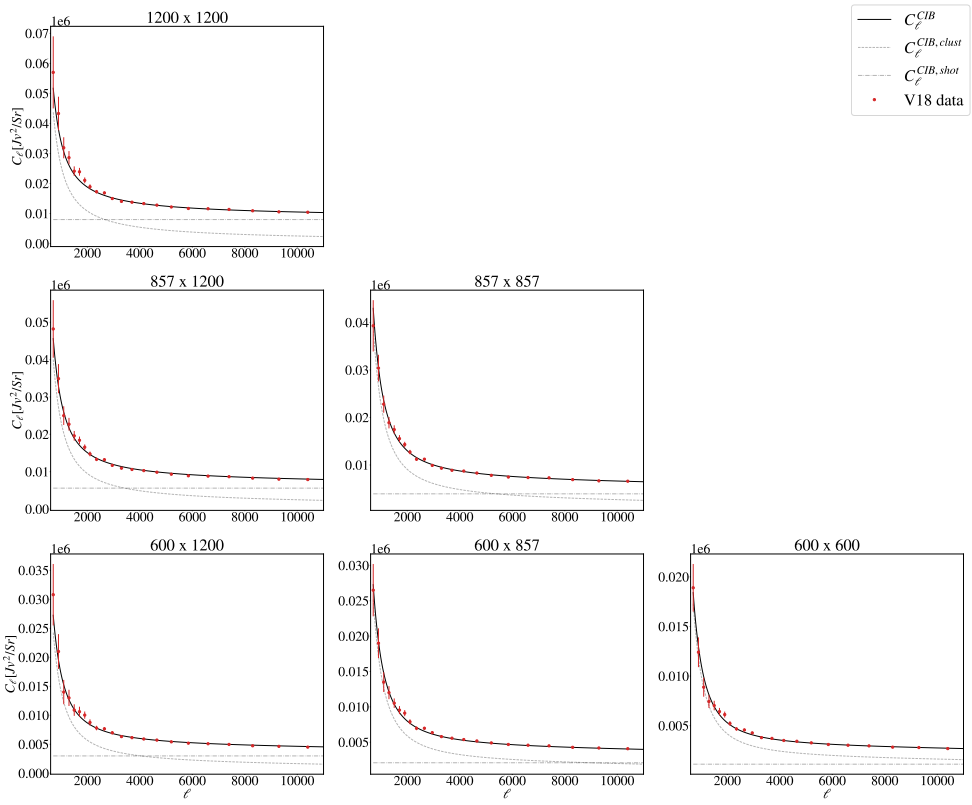

Figure 6.20: Same as figure 6.7 but for the V19 dataset and using the best-fit parameters from Table 6.5.

Figures 6.21 and 6.22 show the posterior distributions of the free model parameters and the clustering parameters only when fixing the correlation coefficients for the SN to unity (purple) and when letting them free to vary (green).

Table 6.5: Best-fit parameters from the model fit to V19 data.

Parameter	Prior	Results $\mathcal{C}_{\nu_1\times\nu_2}=1$	Results $\mathcal{C}_{\nu_1\times\nu_2}$ open
$\log(M^{\rm ET}_{\rm min}/M_\odot h^{-1})$	[10.7,12.8]	12.40 ± 0.11	$12.38^{+0.22}_{-0.18}$
$\alpha_{\rm ET}$	[0.2,3.5]	0.67 ± 0.23	1.12 ± 0.31
$\log(M^{\rm LT}_{\rm min}/M_\odot h^{-1})$	[10.5,12.8]	$11.79^{+0.25}_{-0.19}$	$11.91^{+0.34}_{-0.25}$
$\alpha_{\rm LT}$	[0.2,3.5]	1.253 ± 0.044	1.311 ± 0.071
$SN_{600}[Jy^2/Sr]$	[300,3000]	1117^{+120}_{-140}	1788 ± 300
$SN_{857}[Jy^2/Sr]$	[2000,7000]	3739 ± 340	5153 ± 530
$SN_{1200}[Jy^2/Sr]$	[5000,13000]	8506 ± 710	9688 ± 900
$f^{600}_{\rm cal}$	1 ± 0.11	1.055 ± 0.071	1.046 ± 0.084
$f^{857}_{\rm cal}$	1 ± 0.11	1.072 ± 0.064	0.973 ± 0.066
$f^{1200}_{\rm cal}$	1 ± 0.11	0.962 ± 0.066	0.907 ± 0.069
$\mathcal{C}_{600\times857}$	[-1, 1]	-	$0.983^{+0.014}_{-0.012}$
$\mathcal{C}_{600\times1200}$	[-1, 1]	-	0.885 ± 0.032
$\mathcal{C}_{857\times1200}$	[-1, 1]	-	0.9558 ± 0.0092

Notes. Same as Table 6.2 but for the V19 dataset. When the correlation coefficients are fixed to unity, I obtain $\chi^2=101$ with 132 points and 10 free parameters. In the case where the correlation coefficients are free to vary, I have 13 free parameters and obtain $\chi^2=80$.

The best-fit values which I obtain for the minimum mass of the ET galaxies are higher than those obtained in the analysis conducted on P14 data. They are also higher than the ones previously found in C13 and X12 (and reported in table 6.1). Similarly, the minimum mass for LT galaxies in both scenarios explored in this work appears to be higher than the value found in C13 ($\log(M^{\rm LT}_{\rm min}/M_\odot h^{-1})=10.85\pm0.06$), amounting to a $\sim 4\sigma$ tension between this work and previous literature. This difference could be explained by the fact that I am exploring an extended model compared to C13, who fixed $\alpha_{\rm LT}=1$. In fact, from figure 6.22 it is possible to note that, due to the positive correlation between $\log(M^{\rm LT}_{\rm min})$ and $\alpha_{\rm LT}$, values of $\alpha_{\rm LT}$ closer to unity will pull $\log(M^{\rm LT}_{\rm min})$ toward lower values.

These results also showcase the ability of V19 data to distinguish among the two galaxy populations and therefore their sensitivity to $\alpha_{\rm ET}$, which is unconstrained by *Planck*. The constraints on $\alpha_{\rm ET}$ are in agreement with semi-analytical models [Gao et al., 2004; Hansen et al., 2009], predicting $\alpha_{\rm ET}\leq1.0$. However, these results are not in agreement with the estimated values of $\alpha_{\rm ET}=1.55\pm0.05$ and $\alpha_{\rm ET}=1.81\pm0.04$ found in C13 and X12, corresponding to a tension of $\sim 4\sigma$ and 5σ respectively. The discrepancy between the findings of my work and those of C13 and X12 reduces when considering the correlation coefficients of the SN as free parameters. Specifically, the $\alpha_{\rm ET}$ clustering parameter in that scenario is compatible with the values of C13 and X12 within 1σ and 2σ, respectively. I can explain this better agreement looking at how the model parameters act on the spectra. The $\alpha_{\rm ET}$ clustering parameter (as well as $\alpha_{\rm LT}$) re-scale the overall power, similarly to calibration factors. From table 6.5, I see that in the case of free correlations, the SN

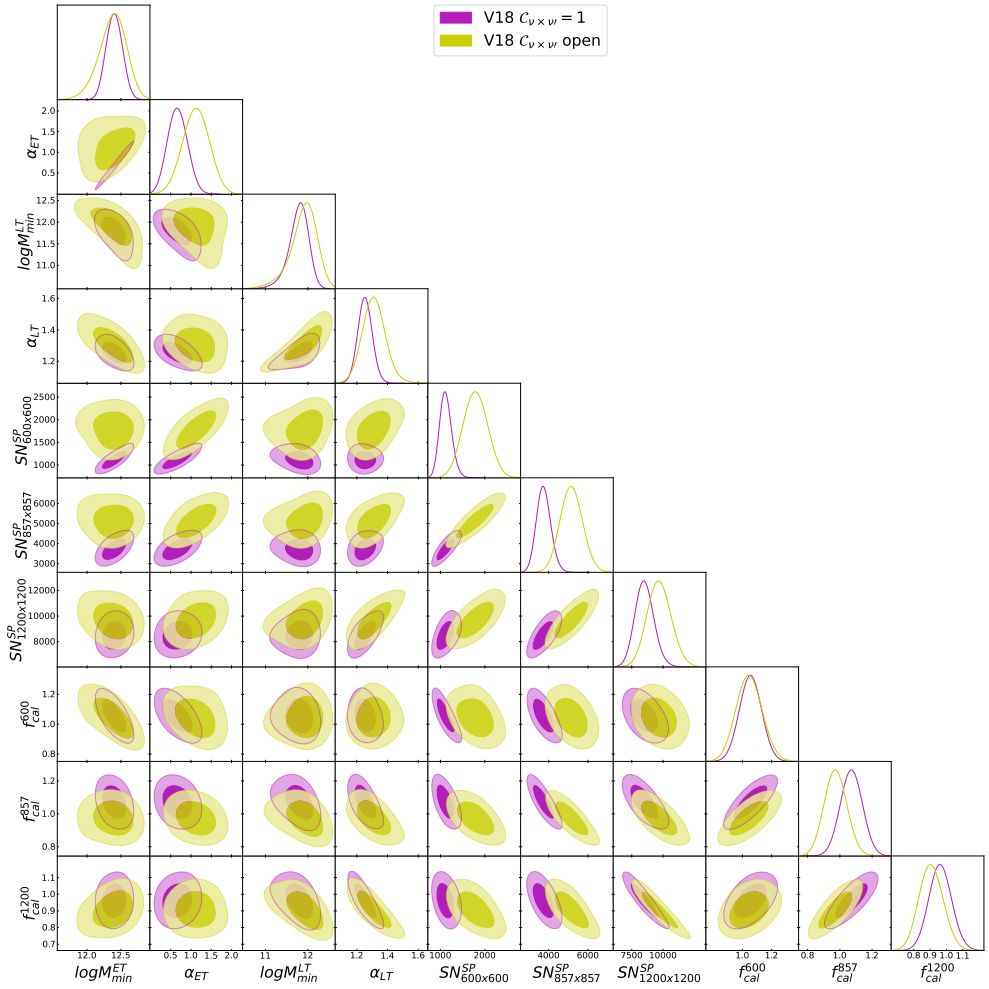

Figure 6.21: Same as figure 6.8 the V19 dataset.

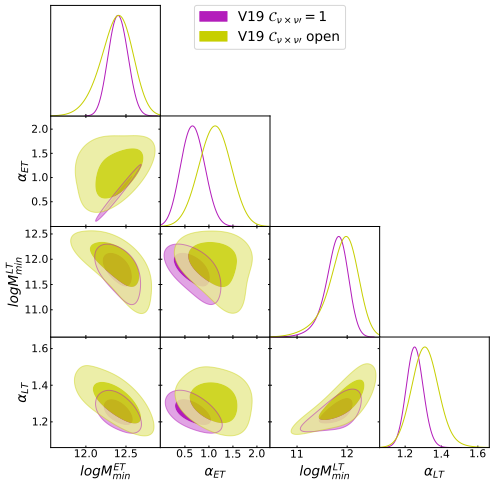

Figure 6.22: Same as figure 6.9 but for the V19 CIB dataset.

values change and carry with them the correlation factors, which go low and reduce power in the spectra. This is then compensated by power added

back in by higher clustering parameters. Concerning the spectral index of LT galaxies, a previous study conducted by C13 shows that $\alpha_{\rm LT}$ was only weakly constrained and therefore kept fixed at unity. Here I show that the V19 data are capable of constraining this parameter at $\alpha_{\rm LT} = 1.253 \pm 0.044$ and $\alpha_{\rm LT} = 1.311 \pm 0.071$ for the two scenarios.

SN levels depend on the scenario under scrutiny. When assuming maximal correlation between frequencies (correlation coefficients fixed to unity), the SN values are a factor of ~ 2 smaller than the levels predicted by the model of Bethermin et al., 2012 and reported in table 4 of Lagache et al., 2020. Allowing SN correlation coefficients to vary, *i.e.* giving more freedom to the model fit, leads the parameters to move toward values that are closer to the ones reported in the literature.

Concerning the calibration factors, they are generally consistent with 1 within 1σ regardless of the treatment of the SN correlation coefficients. I only report a low value of the 1200 GHz factor $f_{\rm cal}^{1200}$, which deviates from 1 at 2σ in the case of free-to-vary correlations between SN levels. This may compensate for a lower value of the SN level at 1200 GHz with respect to the value found in the case of fixed SN correlations.

6.3.4 P14 and V19 comparison

Given the general agreement between the posterior distributions of P14 and L19 shown in section 6.3.2 and given the qualitative nature of the L19 results, in this section I compare the results between *Planck* and *Herschel*-SPIRE data taking the P14 and V19 results.

Figure 6.23 show the posterior distributions of the clustering parameters for V19 (in dark blue) and P14 (in dark yellow) for the case in which the correlation coefficients of the SN are fixed to unity and the case in which they are free to vary, in the left and right panel respectively. From the inspection of these figures, it is possible to note a tension between the two datasets in the best-fit values of the minimum mass of the ET galaxies, at the level of $\sim 5\sigma$ for both scenarios. I find better agreement for the $\log\left(M_{\rm min}^{\rm LT}/M_{\odot}h^{-1}\right)$ and $\alpha_{\rm LT}$ clustering parameters. Specifically, the best-fit values for the minimum mass for LT galaxies from the two datasets are in agreement within $\sim 2\sigma$ when fixing the correlation coefficients, and $\sim 1\sigma$ for the case where the correlation coefficients are free to vary. Lastly, the best-fit values for the $\alpha_{\rm LT}$ clustering parameter are in agreement within $\sim 1\sigma$ for the two datasets and for both the scenarios explored in this study. I remind the reader that the P14 data do not allow me to provide significant constraining power for $\alpha_{\rm ET}$, which is thus fixed to 1.

I performed a number of tests to verify this comparison and to understand the impact of method and model assumptions on the agreement/disagreement. In particular, for P14 I tried excluding the 217 GHz frequency channel and fixing the $\alpha_{\rm ET}$ clustering parameter to values different from 1 (*e.g.*

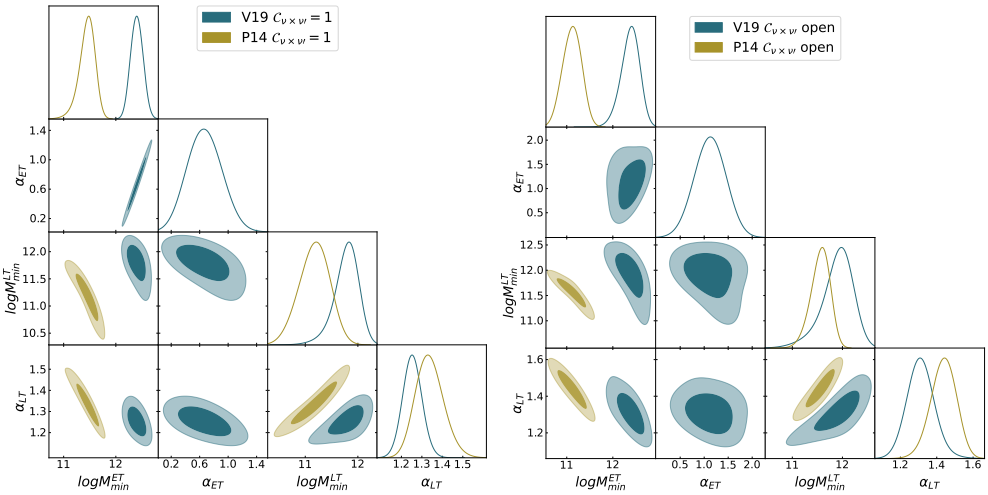

Figure 6.23: Comparison of the posterior distributions of the clustering parameters from P14 (dark yellow) and V19 (dark blue) data for the case in which the correlation coefficients of the SN are fixed to unity (left panel), and the case in which they are free to vary (right panel). For both cases, the posterior distributions of the two datasets are in tension, not exploring the same regions of the parameter space.

$\alpha_{ET} = 0.5$ and $\alpha_{ET} = 1.5$). For V19, I performed the MCMC fixing the α_{ET} clustering parameter in order to recover the same setup used for P14. None of these tests improved the agreement between the two datasets.

Given the tension between the two datasets, I do not perform a joint analysis. The tests I have done and the fact that the common scales of the two datasets are not sufficiently wide to allow for a comparison of the results in a region observed by both of them, do not allow me to clearly identify the source of the discrepancy between the datasets. The tension may equally point at inherent differences in the datasets or a failure in the ability of this model to capture the complexity of the physics under scrutiny. For example, I am not including the luminosity dependence of the HOD, as done in Viero et al., 2013; an investigation of the role this luminosity dependence could have on the results is left to a future work. The analysis of independent data from other surveys may help shed light on this tension.

6.4 CONCLUSIONS ABOUT THE CIB MODELLING

In this first part of my work, I applied a HM formalism to derive a physically motivated model of the CIB emission. The model includes contributions to the CIB emission generated by two different types of galaxy populations, namely ET and LT galaxies, which show different halo distribution and emission characteristics. For the first time, the mixing terms between these populations have also been fully integrated into the calculations. In the model, I distinguish between a contribution from the clustering of matter

to the CIB power spectrum (dominating at larger angular scales) and a SN contribution (dominating at smaller scales). I populate dark matter halos with two galaxy populations to explain two phenomena. Firstly, galaxy number counts exhibit a bimodal distribution, suggesting different behaviours for different galaxies. Secondly, previous attempts to model the CIB power spectrum using a single galaxy population have failed to fit all data accurately. I compare the model against three state-of-the-art datasets–data from the official *Planck* release (P14), a reanalysis of *Planck* data by Lenz (L19), and data from *Herschel*-SPIRE (V19). The datasets have been chosen so as to cover two observational regimes: lower frequency and lower multipole ranges (*Planck*); high-frequency, high-multipole ranges (*Herschel*-SPIRE). The two regimes map onto different sensitivities to the model parameters and contributions from galaxy populations. As I summarise below, despite this extension, the model still faces challenges in describing all these datasets. I tried to better explore issues related to my modelling by confronting it with simulations, but the simulations currently available cannot help in this kind of analysis. I provide a more detailed review of tests conducted using simulations in chapter b.

I found that P14 data provide information on the clustering parameters. However, the frequency range probed with P14 is not wide enough to effectively distinguish between contributions from the two galaxy populations. Qualitatively, L19 data tell the same story. The same does not apply to V19. In this case, the analysis allows me to constrain the clustering parameters of both ET and LT galaxies.

When comparing the results from P14 and V19, I report–in agreement with previous analyses (see *e.g.*, Maniyar et al., 2021)–a tension between the two datasets. This discrepancy is severe enough to prevent me from performing a joint fit. At this stage, I am also unable to clearly pinpoint the source of this discrepancy, which might be due to either the data themselves or to model assumptions (or both). If I assume that the two-population model is a correct and complete description of all CIB emission, the difficulty in fitting the data may be due to, for example, other signals contaminating what I consider a CIB-only dataset or a mistreatment of other experimental factors (such as beams or calibrations). In section 6.3.2.1, I demonstrate how possible dust residuals affect the clustering parameters for the L19 dataset. Alternatively, if I take current data at face value and assume that they are a clean measurement of the CIB emission, the tension might indicate that the model is still incomplete. However, at present, there are no additional astrophysical signatures guiding me in refining the model further.

This work also highlights that the results are highly dependent on the choices made in the analysis and on which priors and external information are imposed on the parameters. I performed a long list of tests during this work, but found limitations in the constraining power of the datasets at hand; I need new observations to be able to lock in further the model.

The tension observed between the low- and high-frequency/multipole range experiments suggests that surveys spanning wider frequency/angular-scale coverage could help shed light on the current discrepancies, as well as advance our understanding of the physics of CIB emission.

The use of data spanning a wider range of angular scales and frequencies would improve constraints on the model parameters, but would also require a consistent modelling of the additional extragalactic emission components. In particular, the approach used here for the CIB can be extended to provide a physical description of the SZ effect. Furthermore, being CIB and SZ tracers of the same underlying matter distribution, the extension of the HM approach to the SZ effect will also allow one to properly predict the correlation between the two tracers (building on other existing literature). The next chapter is entirely devoted to the extension of this modelling to a comprehensive representation of the extragalactic emissions.

COMPREHENSIVE MODELLING OF THE EXTRAGALACTIC EMISSIONS

This chapter extends the description of the Cosmic Infrared Background (CIB) emission to other extragalactic signals, such as Sunyaev-Zel'dovich (SZ) – both thermal Sunyaev-Zel'dovich (tSZ) and kinetic Sunyaev-Zel'dovich (kSZ) –, the correlation between CIB and tSZ, radio sources, and Galactic cirrus.

Extragalactic signals arise from the emission of astrophysical sources tracing the underlying matter density field, contributing across cosmic history at different redshifts and angular scales. In the context of CMB studies, these emissions are typically treated as contaminants – foregrounds – to the primary CMB signal, and substantial effort has been devoted to remove them to maximise the cosmological information encoded in the CMB power spectrum [Dunkley et al., 2009; Planck Collaboration, 2020b; Reichardt et al., 2021]. Yet, these same signals are also a rich source of information about galaxy formation, feedback processes, and the LSS of the Universe [Cooray et al., 2010; Lagache et al., 2005]. Secondary CMB anisotropies – such as the thermal and kinetic SZ effects or gravitational lensing – have been shown to contain unique information about cosmological parameters, astrophysical processes, and, in particular, the kSZ effect provides a direct handle on the epoch of reionization [Aghanim et al., 2008; Battaglia et al., 2012; 2013; George et al., 2015]. Understanding these secondary signals and their correlations with LSS tracers can provide constraints on the sum of the neutrino masses [Carbone et al., 2016; Louis et al., 2017], dark energy evolution [Bolliet et al., 2018], and feedback in galaxy clusters [Hill, 2018; Planck Collaboration, 2016b].

In this part of my work, I implement a unified framework that describes the extragalactic emissions with both a Halo Model (HM)-based and a template-based approach. Building on Zagatti et al., 2024 and specifically on what was presented in the previous chapter, I extend the HM description to the CIB, tSZ, and their cross-correlation, while treating the kSZ effect, radio sources, and Galactic cirrus with a template-based approach. The overall framework is briefly described in section 7.1.

Of particular relevance for this study is the new-generation ground-based CMB experiment, the Simons Obervatory (SO) [Ade et al., 2019], located in the Atacama Desert in Chile. SO will observe the sky with two complementary instruments: a 6-metre Large Aperture Telescope (LAT), optimised for small-scale temperature and polarisation anisotropies over a wide range of multipoles, and an array of Small Aperture Telescope (SAT)s, designed to target large angular scales with high sensitivity to primordial B-modes. Together, LAT and SATs cover multiple frequency bands between 27 and 280

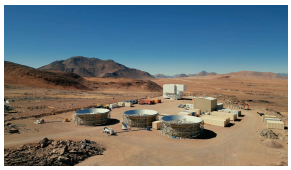

Figure 7.1: Simons Observatory site in the Atacama Desert, *Chile*.

GHz, enabling efficient separation of Galactic and extragalactic components and cross-checks of systematic effects. Operating over multipoles roughly in the range $30 \lesssim \ell \lesssim 9000$, SO will deliver high-precision measurements in a foreground-dominated regime, providing an ideal data set to test the physical modelling of the extragalactic emissions developed in this work. For this reason, while waiting for actual observed data to come, I generate the theoretical data set mimicking SO LAT observations described in section 7.2, spanning $\ell \in [1500, 8000]$ at the six LAT frequency channels. Using this data set, I first perform Fisher forecasts in section 7.3.1, to explore parameter degeneracies and identify which quantities are well constrained, so as to focus on them and, consequently, reduce the parameter space. In section 7.3.2 I carry out a MCMC analysis to evaluate the constraining power of an SO-like experiment on foreground parameters.

Although I employ publicly available information to reproduce the expected performance of a SO-like experiment, the results of this chapter should not be considered an official SO result.

7.1 MODELLING OF THE EXTRAGALACTIC COMPONENTS

This section provides a very brief description of the modelling of the different emissions, since it follows what was already introduced in chapter 4, chapter 5, and chapter 6. Therefore, it serves as a summary and guide throughout this chapter.

CIB modelling

For the CIB modelling, I do not change the HM-based approach presented in chapter 6, which encodes two galaxy populations, *i.e.* early-type (ET) and late-type (LT) galaxies. Therefore, the clustering term contributing to the total CIB power spectrum is given by equation 5.5 and reported below:

$$C_{\ell,\nu\times\nu'}^{CIB,clus} = \int \frac{dz}{\chi^2}\frac{dz}{d\chi} j_\nu(z) j_{\nu'}(z) P_{gg}(k = \ell/\chi, z), \tag{7.1}$$

with P_{gg} evaluated following equation 6.11, which encodes the contributions from ET, LT, and the mixing between the two galaxy populations.

A recap of the different formulations for the SN adopted in this thesis.

- In chapter 5 I presented the expression for the SN based on the number counts;

$$C_{\ell,\nu}^{CIB,SN} = \int_0^{S_\nu^{lim}} s_\nu \frac{dN}{dS_\nu} dS_\nu.$$

  However, when dealing with actual observed data, this prescription leads to higher SN levels than those observed.

- When dealing with the model fit to CIB data, in order to avoid overestimating the contribution of the SN, I evaluate the SN contribution between two frequency channels as

$$C_{\ell,\nu\times\nu'}^{CIB,SN} = \mathcal{C}_{\nu\times\nu'} \sqrt{SN_\nu \times SN_{\nu'}},$$

  where SN_ν is a certain SN level for the single frequency channel, treated as a free parameter.

- This chapter introduces a slightly more physical description of the SN term, presenting it as a template rescaled by an amplitude. The main reason for this third modelling is to reduce the number of free parameters.

A different approach is adopted for the modelling of the Shot Noise (SN). However, because of the lack of data or precision in the number counts across all frequencies of interest, and in order to avoid adding a free parameter for each frequency channel – as done in the previous chapter for the P14 and V19 analyses – I adopt a simplified modelling that follows a template-based approach, so that:

$$C_{\ell,\nu\times\nu'}^{SN} = \frac{2\pi}{\ell(\ell+1)} \mathcal{A}_{SN} \epsilon_{\nu\nu'} \frac{\eta_\nu \eta_{\nu'}}{\eta_0 \eta_0} \left(\frac{\ell}{3000}\right)^2, \tag{7.2}$$

where:

- the prefactor is to express the final result in C_ℓ's;
- $\mathcal{A}_{SN}$ is the amplitude of the CIB SN signal;
- $\epsilon_{\nu\nu'}$ allows for the conversion from flux to CMB units;
- η_ν represents a modified blackbody emission, given by $\eta_\nu = B_\nu(T_d)\nu^{\beta_{SN}}$[1], and provides the information on the frequency dependence of the signal;

1 Here T_d and ν_0 represent the dust temperature and frequency, which are set to $T_d = 20K$ and $\nu_0^d = 153.8GHz$.

- the term in parentheses sets the angular template of the CIB SN emission.

The above expression for the SN describes a template which rapidly approaches a scale-independent (flat) behaviour.

tSZ modelling

The modelling of the tSZ signal follows precisely the prescription provided in section 5.2.1 of chapter 5.

$$C^{tSZ}_{\ell,\nu\times\nu'} = T^2_{CMB}\int \frac{dV}{dz}dz g(\nu)g(\nu')\int dm \frac{dn}{dm}\Big|_{500}\tilde{y}(k,m)^2 + T^2_{CMB}\int \frac{dV}{dz}dz g(\nu)g(\nu')P^{lin}_{mm}(k)\left(\int dm\frac{dn}{dm}\Big|_{500} b_{tSZ}(m)\tilde{y}(k,m)\right)^2. \tag{7.3}$$

CIB-tSZ cross correlation

As already detailed in chapter 5, the correlation between the CIB and tSZ, being tracers of the same underlying matter distribution, can be described within a full HM framework. However, the description of this correlation must now take into account the fact that there are two different galaxy populations in the overall picture. In the following, I provide the final expressions for the 1h and 2h terms for the sake of simplicity, whereas in section a.2 of chapter a, I present the full calculations, making the contributions from the two galaxy populations explicit.

$$\begin{aligned} C^{tSZ\times CIB,1h}_{\ell,\nu\times\nu'} &= T_{CMB}\int dz \frac{dV}{dz}\left[\, g(\nu)S_{\nu'}(z) + g(\nu')S_{\nu}(z)\,\right] \\ &\times \int dm \frac{dn}{dm}\Big|_{500}\frac{N_{gal}(m)}{\bar{n}_{gal}}\, u_{gal}(m,k)\,\tilde{y}_\ell(m,z), \end{aligned} \tag{7.4}$$

while the two–halo term is

$$\begin{aligned} C^{tSZ\times CIB,2h}_{\ell,\nu\times\nu'} &= T_{CMB}\int dz \frac{dV}{dz}\left[\, g(\nu)S_{\nu'}(z) + g(\nu')S_{\nu}(z)\,\right] P^{lin}_{mm}(k) \\ &\times \left[\int dm\,\frac{dn}{dm}\Big|_{500} b_{tSZ}(m)\,\tilde{y}_\ell(m,z)\right] \\ &\times \left[\int dm\,\frac{dn}{dm}\, b(m)\,\frac{N_{gal}(m)}{\bar{n}_{gal}}\, u_{gal}(m,k)\right], \end{aligned} \tag{7.5}$$

where in both equations N_{gal} and $\bar{n}_{gal}$ account for the sum of the contributions from the ET and LT galaxy populations.

kSZ, radio sources, Galactic cirrus modelling

The kSZ signal, and the emissions from radio sources and Galactic cirrus are all modelled exactly using the template-based approach described in chapter 5. For completeness, I report the equations used to compute their power spectra.

$$
\begin{aligned}
C_\ell^{kSZ} &= \frac{2\pi}{\ell(\ell+1)} \mathcal{A}_{kSZ}\, D_{0,\ell}^{kSZ}, \\
C_{\ell,\nu\times\nu'}^{radio} &= \frac{2\pi}{\ell(\ell+1)} \mathcal{A}^{radio}\, \epsilon_{\nu\nu'} \left(\frac{\nu\nu'}{\nu_0^2}\right)^{\alpha_{radio}} \left(\frac{\ell}{3000}\right)^2, \\
C_{\ell,\nu\times\nu'}^{cirrus} &= \frac{2\pi}{\ell(\ell+1)} \mathcal{A}^{cirrus}\, \epsilon_{\nu\nu'} \frac{\eta_\nu\, \eta_{\nu'}}{\eta_{\nu_0}^2} \left(\frac{\ell}{3000}\right)^{\alpha_{cirrus}}.
\end{aligned}
\tag{7.6}
$$

7.2 CREATION OF A SMOOTH DATASET

In this section, I describe how I create the smooth theory dataset used for this work, meaning I create a dataset which mimics the upcoming SO LAT data – theoretical C_ℓs generated from my model, together with a power-spectrum-based covariance matrix built from the SO LAT noise curves (including beams and sky fraction).

7.2.1 Smooth power spectra

The aim is to build a dataset that mimics the instrumental characteristics of the Simons Obervatory (SO). The dataset therefore spans the SO LAT frequency channels, namely 27, 39, 93, 145, 225, and 280 GHz.

As a first step, I fix fiducial values for the model parameters (see table 7.1) and compute the 21 auto– and cross–frequency spectra over the multipole range $1500 \leq \ell \leq 8000$. The theoretical model introduced in the previous section is largely independent of observational inputs, with the exception of the emissivity functions, which enter the clustering part of the CIB. These functions rely on galaxy number counts measured at specific frequency bands. Direct measurements of the differential number counts for the individual galaxy populations at the SO LAT frequencies are not yet available. As a temporary but practical solution, I extrapolate the counts at SPT frequencies to the SO channels. This approach is justified because SPT operates in overlapping millimetre/sub–millimetre bands (95, 150, 220 GHz).
During this extrapolation, I encounter unphysical negative values of the differential number counts for the two lowest SO channels (27 and 39 GHz). These arise because the extrapolation extends far outside the frequency range directly probed by SPT, where the fitted spectral models are poorly constrained. To maintain physical consistency, I impose a physical prior that number counts must be non-negative, setting them to zero where extrapo-

lations fail. Lastly, I fix the adopted flux cut at $S_\nu^{lim} = 7\,mJy$, consistent with the SO forecast paper [Ade et al., 2019]. The final emissivity functions used in this work for the computation of the CIB clustering term are shown in figure 7.2. The figure illustrates the frequency dependence of $j_\nu(z)$ for all SO channels after applying the flux cut and extrapolation procedure.

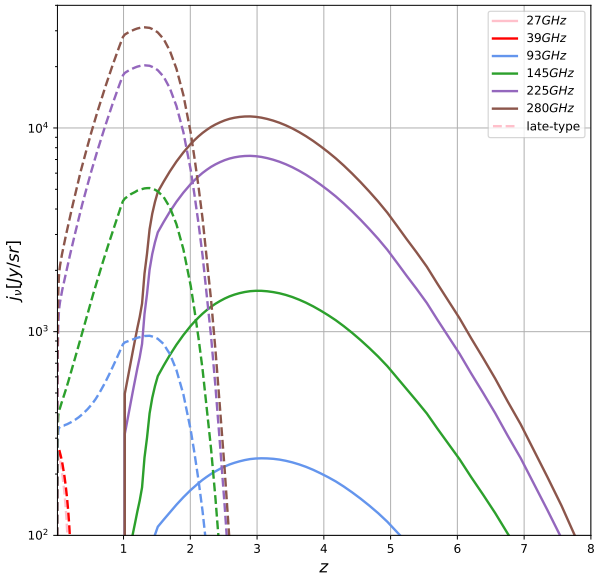

Figure 7.2: Emissivity functions for the SO LAT frequency channels, computed using differential number counts extrapolated from SPT measurements and applying a 7 mJy flux cut.

A similar approach is adopted for the SN component of the CIB power spectrum. The SN arises from the Poisson distribution of unresolved galaxies below the flux cut. In principle, it can be computed directly from measured dN/dS. However, due to the same lack of data at SO frequencies, I extrapolate the expected amplitudes from SPT measurements. In this case, rather than fixing the SN levels directly from the extrapolation, I use the extrapolated values as a reference to define physically motivated parameter ranges for the SN parameters in equation 7.2, *i.e.* $\mathcal{A}_{SN}$ and β_{SN}, which are then treated as free parameters in the MCMC analysis.

To produce the final dataset used in the analysis, I bin the theoretical ℓ–by–ℓ power spectra following the SO binning scheme implemented in the multifrequency likelihood for the LAT, `MFLike`[2]. Given the binning matrix $B_{\ell b}$, the binned power spectra C_b are obtained from the theoretical spectra C_ℓ as

$$C_b = \sum_\ell B_{\ell b} C_\ell. \tag{7.7}$$

2 https://github.com/simonsobs/LAT_MFLike.git

Table 7.1: Values of the model parameters employed for the creation of the smooth power spectra.

Parameter	Value
$\mathcal{A}_{radio}$	1.28
α_{radio}	-0.945
$\mathcal{A}_{cirrus}$	0.65
α_{cirrus}	-1.2
β_{cirrus}	1.8
$\mathcal{A}_{kSZ}$	1.9
b	0.53
$\log(M^{\rm ET}_{\rm min}/M_\odot h^{-1})$	12.8
$\log(M^{\rm LT}_{\rm min}/M_\odot h^{-1})$	10.85
$\alpha_{\rm ET}$	1.55
$\alpha_{\rm LT}$	1.0
$\mathcal{A}_{SN}$	9.0
β_{SN}	1.1

Notes. Values of the parameters for constructing the smooth theory dataset.

I then restrict the dataset to the range $1500 \leq \ell \leq 8000$. For visualisation, figure 7.3 shows both the unbinned ℓ–by–ℓ spectra and the binned auto–frequency spectra, with the associated error bars computed from the diagonal of the covariance matrix described in the following section.

7.2.2 Smooth covariance matrix

I start with the computation of a Gaussian (power–spectrum–based) covariance matrix of the form

$$\Sigma\left(C_\ell^{\nu_1\nu_2}, C_\ell^{\nu_3\nu_4}\right) = \frac{1}{(2\ell+1)f_{sky}}\left(C_\ell^{\nu_1\nu_4}C_\ell^{\nu_2\nu_3} + C_\ell^{\nu_1\nu_3}C_\ell^{\nu_2\nu_4}\right), \tag{7.8}$$

with $f_{sky} = 0.4$ as the assumed sky fraction. The total spectra entering the covariance are

$$C_\ell^{XY} = C_\ell^{XY,\,theo} + \frac{N_\ell^{XY}}{B_\ell^X B_\ell^Y}, \tag{7.9}$$

where N_ℓ^{XY} are the noise spectra and B_ℓ are the beam transfer functions. Since I aim to construct an SO LAT–like dataset, the noise curves N_ℓ for each frequency channel are computed using the publicly available `so_noise_models` package[3] provided by the SO collaboration.

This covariance is Gaussian in the sense that it neglects contributions from higher–order statistics (*e.g.* trispectra). However, I note that, as analysed in Louis et al., 2025, extragalactic foregrounds are expected to be non-Gaussian. Louis et al., 2025 state that the impact of including non-Gaussian contributions in cosmological analyses for ACT frequencies is at the sub–10% level. For

3 https://github.com/simonsobs/so_noise_models.git

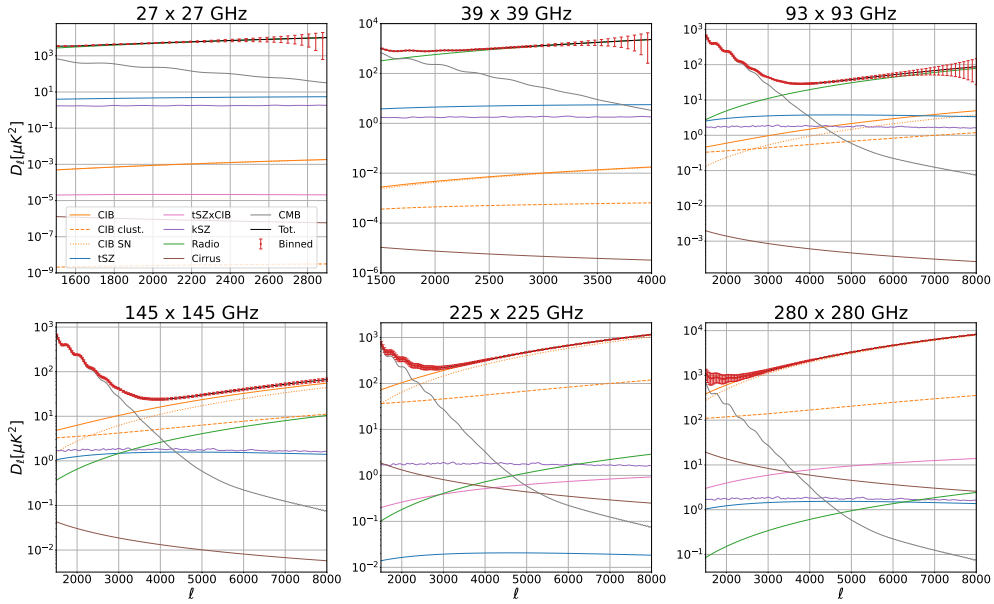

Figure 7.3: Auto–frequency spectra of the various components, represented in units of D_ℓs, entering the model. CIB contributions are shown in orange: total CIB (solid), CIB clustering (dashed), and CIB SN (dotted). tSZ, tSZ×CIB, and kSZ are plotted in blue, pink, and purple, respectively. Radio sources and Galactic cirrus are displayed in green and brown, while the CMB signal is in gray. The total signal is shown in black, and its binned version (using the `MFLike` scheme) along with the error bars inferred from the covariance matrix are in red. Note that the 27×27 and 39×39 GHz power spectra are truncated to lower multipoles for visualisation purposes because of very large error bars.

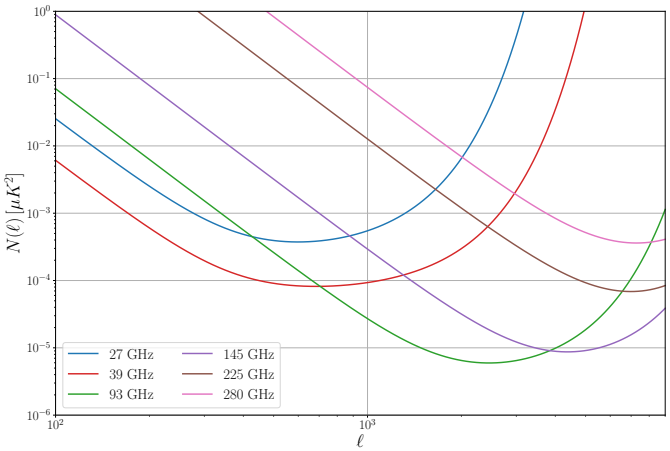

Figure 7.4: Noise curves for the auto-frequency channels computed with the notebook and specifications provided by the SO collaboration.

the SO LAT frequency channels I expect a slightly larger impact, but the following results are presented using this Gaussian covariance matrix.

As in the previous section, both the theoretical spectra and the covariance matrix are binned using the `MFLike` scheme. To summarise, denoting the binning matrix by $B_{\ell b}$, the binned power spectra and covariance are

$$C_b = \sum_\ell B_{\ell b} C_\ell, \qquad \Sigma_{bb'} = B^T_{\ell b}\, \Sigma_{\ell\ell'}\, B_{\ell' b'}. \tag{7.10}$$

7.3 RESULTS

The goal of this section is to explore the parameter space of the comprehensive model presented above, identifying which parameters can be constrained, which remain unconstrained, and where degeneracies appear. As a first step, I perform a Fisher forecast, which provides an approximate but fast assessment of parameter sensitivities. A more detailed exploration with MCMC methods is left to section 7.3.2.

7.3.1 Fisher Forecasts

The Fisher formalism is a standard tool for forecasting how well a given dataset can constrain model parameters, without requiring simulated data realisations.

For a set of N model parameters $\{\theta_i\}$, the Fisher matrix $\mathbf{F}$ is an $N \times N$ symmetric matrix defined as

$$\mathbf{F} = \mathbf{D}\, \mathbf{\Sigma}^{-1}\, \mathbf{D}^T, \tag{7.11}$$

where $\mathbf{\Sigma}$ is the covariance matrix defined in equation (7.8), and $\mathbf{D}$ is the derivative matrix of dimension $N \times m$, with m the length of the data vector. Each row of $\mathbf{D}$ contains the derivatives of the power spectra with respect to a single model parameter:

$$\mathbf{D} = \begin{pmatrix} \partial C_{\ell_1}/\partial\theta_1 & \cdots & \partial C_{\ell_m}/\partial\theta_1 \\ \vdots & & \vdots \\ \partial C_{\ell_1}/\partial\theta_N & \cdots & \partial C_{\ell_m}/\partial\theta_N \end{pmatrix}. \tag{7.12}$$

I evaluate the derivatives numerically using a finite–difference scheme,

$$\frac{\partial C_\ell}{\partial \theta_i} \simeq \frac{C_\ell(\theta_{i,0} + \Delta\theta) - C_\ell(\theta_{i,0} - \Delta\theta)}{2\Delta\theta}, \tag{7.13}$$

where $\theta_{i,0}$ is a fiducial starting point and $\Delta\theta = \epsilon\, \theta_{i,0}$ with $\epsilon = 0.07$. I verified that the results are stable against variations of the step size, testing $\epsilon = 0.1$ and $\epsilon = 0.5$. In practice, this means that for each parameter I recompute the theoretical spectra at slightly shifted values around the fiducial point and use the difference to approximate the derivative.

In the following I present both the Fisher correlation matrix and the Fisher Signal–to–Noise Ratio (SNR) evaluated as $SNR(\theta_i) \equiv \sqrt{F_{ii}}$ for three progressively more complex scenarios:

1. *Foreground parameters only*: I first vary the 17 parameters describing the foreground model, keeping the cosmology fixed. Figures 7.5 and 7.6 show the Fisher correlation matrix and the Fisher SNRs for this setup.

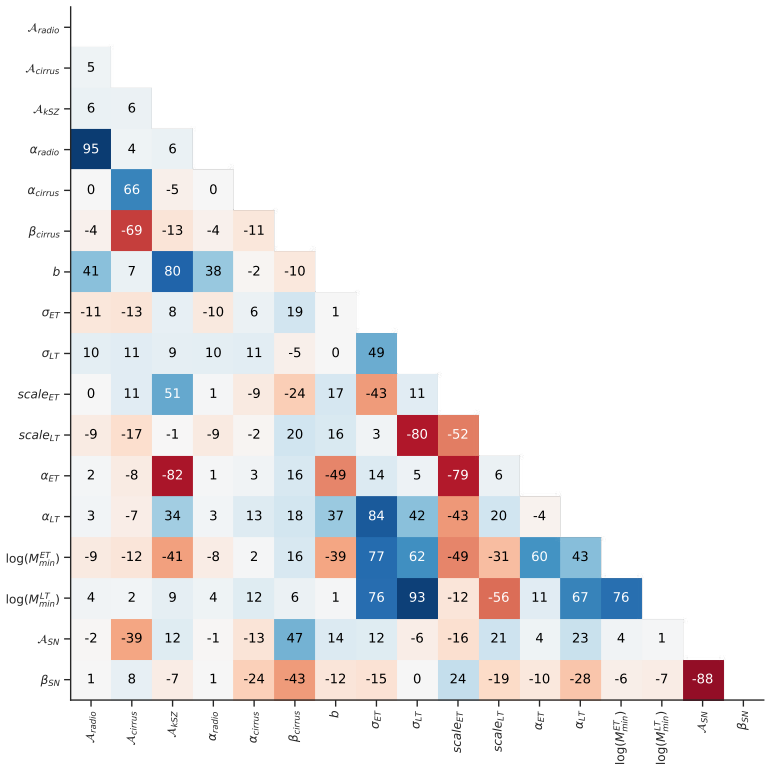

Figure 7.5: Fisher correlation matrix for the 17 foreground parameters only. Strong correlations are visible among radio parameters, cirrus parameters, and among those regulating the CIB emission.

As expected, radio parameters are highly correlated, as are cirrus parameters, and the subset of CIB parameters entering the two – galaxy – population modelling, both SN and clustering. Turning to the SNRs, I find that the minimum halo mass of ET galaxies is well constrained. In general, ET parameters show higher SNR than LT ones, consistent with the fact that SO LAT frequencies are more sensitive to ET contributions. I remind the reader that higher frequencies observe the nearby Universe, therefore a region where the dominant contribution comes from the LT galaxies; SO LAT frequencies are dominated by the ET contribution. In this work I am modelling within the same HM framework the CIB, the tSZ, and their cross-correlation. The latter naturally comes without the need to introduce any additional free parameters, leading to a good constraining power on the hydrostatic mass bias b. Radio parameters are also well constrained, thanks to the dominance of radio emission in

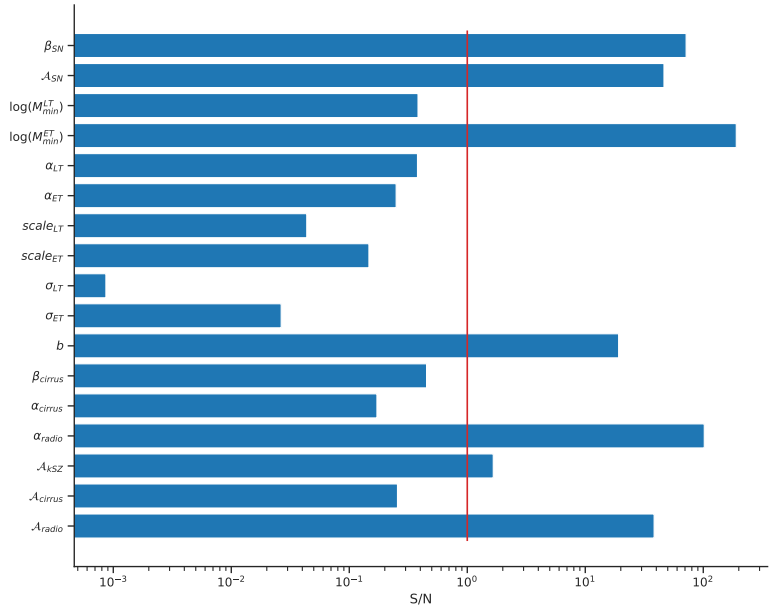

Figure 7.6: Fisher SNR for the foreground parameters. The vertical red line marks SNR= 1, used as a reference threshold.

the two lowest SO frequencies (27 and 39 GHz). Finally, the amplitude of the kSZ effect is measurable with SNR> 1 even in this conservative setup, where all parameters are allowed to vary freely without external priors.

2. *Adding cosmological parameters, fixed cosmology in foregrounds*: In the second setup, I open the cosmological parameters affecting the theoretical predictions at the level of CMB spectra, while keeping them fixed in the foregrounds modelling. Upper and lower panels of figure 7.7 show the results for this scenario. In this setup, the correlations among foreground parameters remain unchanged relative to the previous case. While noticing correlations among the cosmological parameters themselves, I see essentially no coupling between foreground and cosmological parameters. Foreground constraints remain unchanged compared to the previous case.

3. *Cosmological parameters with varying cosmology in foregrounds*: Finally, I allow the cosmology to vary consistently both in the CMB and in the foreground model. Upper and lower panels of figure 7.8 show the resulting correlations and Fisher SNRs. In this more realistic scenario, overall it is possible to appreciate slight correlations appearing between cosmological and foreground parameters. Among all these correlations, it is worth noticing the high degree of correlation the hydrostatic mass bias, b, shows with the background cosmology – which was not present in the previous case. The fact that b becomes correlated with cosmology

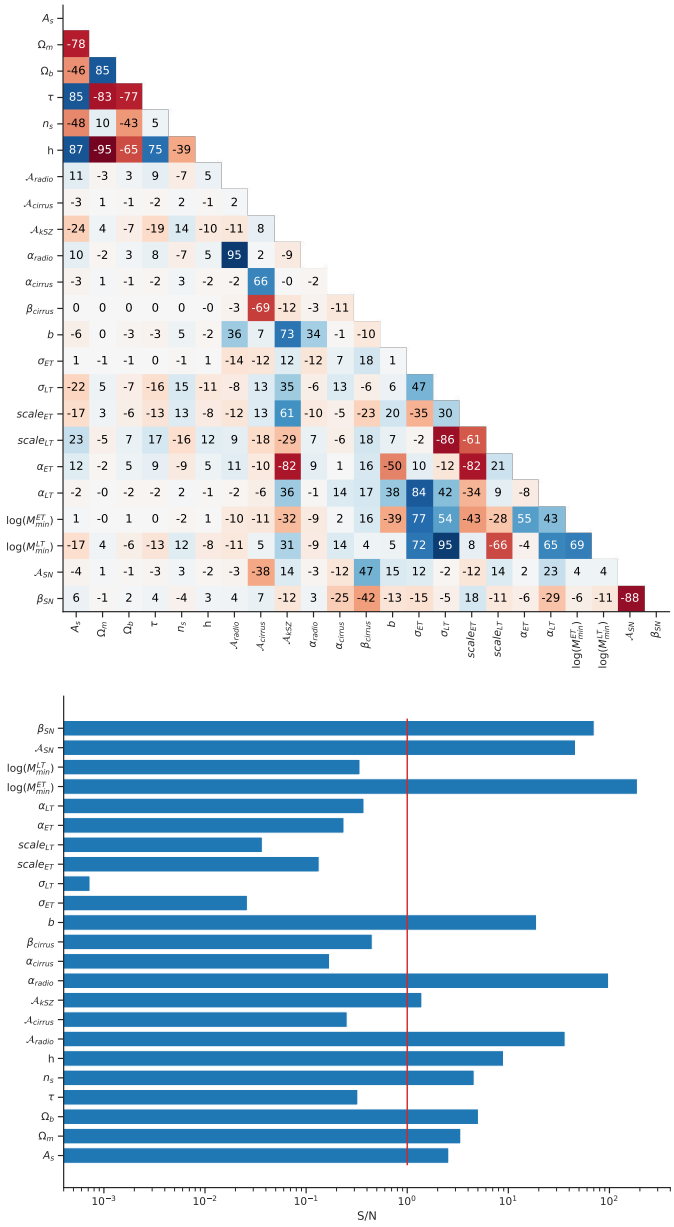

Figure 7.7: Upper panel: Fisher correlation matrix with cosmological parameters varied, but with a fixed cosmology in the foreground model. Lower panel: Fisher SNR for the case with cosmological parameters varied but fixed in the foreground model.

is because it effectively rescales the halo masses in a way that mimics the impact cosmological parameters have on the LSS. Specifically, an increase of b reduces the effective halo masses and hence the overall foregrounds amplitude. This effect can be compensated by increasing all the cosmological parameters that act on the amplitude of the spectrum such as $\mathcal{A}_s$, τ, and h – this is confirmed by the positive correlation observed. Conversely, the negative degree of correlation observed between b and Ω_m–Ω_b is easily explained by the fact that higher values of these two

cosmological parameters enhance the halo abundance and gas pressure, thereby increasing the foreground signal, requiring a decrease in b to compensate the effect. The Fisher SNRs for cosmological parameters decrease compared to the previous case, reflecting the degeneracies introduced by allowing cosmology to affect both the primary CMB and the foreground model. The foreground parameters themselves remain broadly consistent with the earlier scenarios.

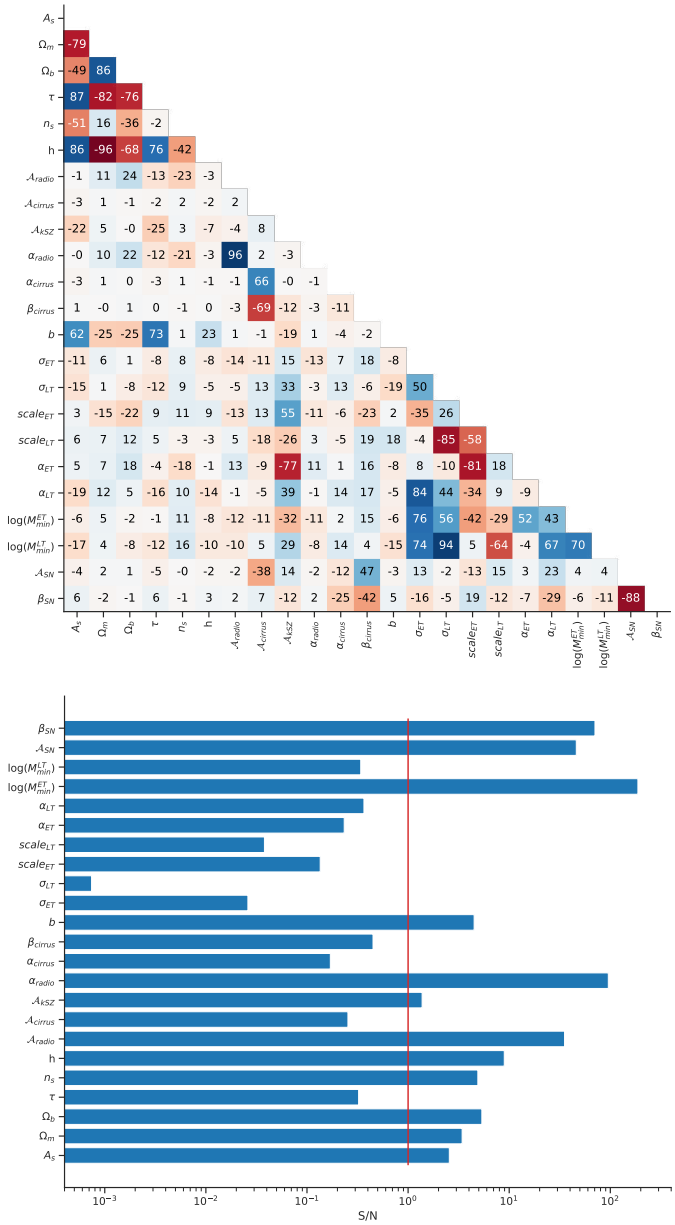

Figure 7.8: Upper panel: Fisher correlation matrix with both cosmology and foregrounds varied consistently. Lower panel: Fisher SNR for the case with cosmology and foregrounds varied consistently.

To highlight the distinction between the second and the third case, figure 7.9 directly compares the Fisher SNRs of the cosmological parameters in cases *(2)* and *(3)*. When all parameters are varied simultaneously, the inclusion of cosmology–dependent foregrounds introduces additional degeneracies that reduce the marginalised Fisher SNRs of cosmological parameters relative to case (2). However, if one restricts the Fisher analysis to the cosmological subspace alone, the effective constraining power improves in case (3), since the information content of the foreground model is now coherently tied to the underlying cosmology. In other words, although degeneracies with foreground parameters degrade the fully marginalised errors, a cosmology–dependent foreground treatment captures additional information that can tighten constraints within the cosmological sector itself. This underlines the importance of a consistent physical modelling of extragalactic foregrounds when deriving cosmological parameter forecasts.

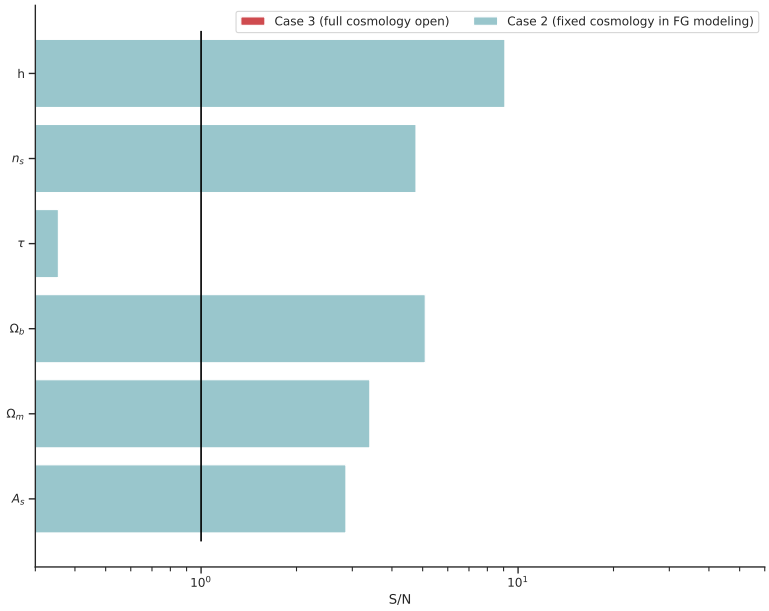

Figure 7.9: Comparison of the Fisher SNR of the cosmological parameters in the case of fixed cosmology in the foreground model (blue) and the case with varying cosmology in the foreground model (red). Allowing cosmology to propagate consistently into the foregrounds enhances the constraining power.

It is important to stress that this constitutes only a preliminary analysis. A more detailed assessment of the quantitative impact on cosmological parameter inference will require going beyond Fisher forecasts and incorporating more realistic data–driven tests, which I leave to future work.

7.3.2 MCMC analysis

I now present the results of a MCMC analysis performed with the `Cobaya` ensemble sampler [Torrado and Lewis, 2021]. The goal of this step is to

explore the multidimensional parameter space of this foreground model and to assess the constraining power of an SO–like experiment on these parameters.

The parameter space is explored using a Gaussian likelihood built from the power–spectrum covariance matrix of equation (7.8):

$$\log \mathcal{L}(C_\ell^{data} | C_\ell^{model}) \propto -\frac{1}{2} \Delta_{C_\ell}^T \Sigma^{-1} \Delta_{C_\ell}, \tag{7.14}$$

where $\Delta_{C_\ell} = C_\ell^{data} - C_\ell^{model}$ is the difference between the simulated SO–like data vector (see section 7.2) and the theoretical predictions, considering all 21 auto- and cross-frequency spectra, and Σ is the covariance matrix. Both data and the covariance matrix are binned with the binning scheme introduced in section 7.2.
This likelihood treats the bandpowers as Gaussian–distributed and neglects non–Gaussian covariance terms from higher–order correlations.

To reduce computational cost and avoid exploring poorly constrained directions, I use the Fisher–forecast results of section 7.3.1 as a guide. Parameters that (i) are strongly degenerate with others or (ii) exhibit very low Fisher SNR are fixed to their fiducial values. This reduces the dimensionality of the parameter space while retaining the most informative degrees of freedom. As a consequence, the resulting forecasts should be interpreted as conditional constraints, reflecting the information content of the data on the retained parameters rather than fully marginalised constraints over the complete model parameter space.

Specifically, for the baseline case:

- because of strong degeneracies, I fix the power–law slope of the radio sources (α_{radio});
- because of the low Fisher SNR and the strong mutual correlations, I fix all cirrus parameters ($\mathcal{A}_{cirrus}, \alpha_{cirrus}, \beta_{cirrus}$);
- I also fix four CIB HOD parameters ($\sigma_{ET}, \sigma_{LT}, M_{sat}^{ET}, M_{sat}^{LT}$) because of their low Fisher SNR;
- because of the high degree of correlation in the two parameters regulating the SN, I keep the power-law coefficient fixed during the MCMC analysis (β_{SN}).

After these reductions, the baseline run explores eight free parameters. Their priors – motivated by previous analyses (SPT, ACT, *Planck*) or physically reasonable ranges – are listed in the second column of tables 7.2 and 7.3, along with the resulting posterior constraints.

To test the stability of the constraints and study parameter degeneracies, I also run three alternative configurations:

- Case A: Fixing the hydrostatic mass bias b.

Table 7.2: Best-fit parameters from the model fit to the smooth data set for the baseline and case A presented in this study.

Parameter	Prior	Results	
		Baseline	Case A
$\mathcal{A}_{radio}$	1.28 ± 0.19	1.2796 ± 0.0090	1.2799 ± 0.0087
$\mathcal{A}_{kSZ}$	[0,4]	$1.80^{+0.59}_{-0.47}$	1.92 ± 0.33
b	[0.2,0.7]	0.527 ± 0.015	–
$\alpha_{\rm ET}$	[0.0,3.0]	$1.62^{+0.41}_{-0.32}$	$1.59^{+0.43}_{-0.33}$
$\alpha_{\rm LT}$	[0.0,2.1]	$0.85^{+0.38}_{-0.14}$	$0.92^{+0.28}_{-0.11}$
$\log(M^{\rm ET}_{\rm min}/M_\odot h^{-1})$	[11,12.8]	$11.86^{+0.31}_{-0.12}$	$11.83^{+0.31}_{-0.14}$
$\log(M^{\rm LT}_{\rm min}/M_\odot h^{-1})$	[10.0,12.0]	$11.18^{+0.58}_{-0.21}$	$11.22^{+0.56}_{-0.17}$
$\mathcal{A}_{SN}$	[7.0, 11.0]	8.995 ± 0.051	8.994 ± 0.051

Notes. First column: model parameters sampled for the model fit to SO-like smooth data. Second column: priors and ranges of variation imposed on the model parameter. I impose a Gaussian prior only on the amplitude of the radio sources emission motivated by previous studies; I let all the other parameters free to vary within a flat prior. Third and fourth columns: best fit values within 1σ confidence levels for the model free parameters of the baseline and case A presented in this analysis.

Table 7.3: Best-fit parameters from the model fit to the smooth data set for cases B and C presented in this study.

Parameter	Prior	Results	
		Case B	Case C
$\mathcal{A}_{radio}$	1.28 ± 0.19	1.2798 ± 0.0090	1.2798 ± 0.0086
$\mathcal{A}_{kSZ}$	[0,4]	1.97 ± 0.25	1.96 ± 0.19
b	[0.2,0.7]	0.532 ± 0.013	–
$\alpha_{\rm ET}$	[0.0,3.0]	1.60 ± 0.27	1.58 ± 0.24
$\alpha_{\rm LT}$	[0.0,2.1]	–	–
$\log(M^{\rm ET}_{\rm min}/M_\odot h^{-1})$	[11,12.8]	$11.87^{+0.23}_{-0.10}$	$11.88^{+0.22}_{-0.098}$
$\log(M^{\rm LT}_{\rm min}/M_\odot h^{-1})$	[10.0,12.0]	$11.14^{+0.64}_{-0.26}$	$11.11^{+0.67}_{-0.43}$
$\mathcal{A}_{SN}$	[7.0, 11.0]	8.992 ± 0.051	8.991 ± 0.051

Notes. Same as table 7.2 for cases B and C.

- Case B: Fixing the power–law slope for the LT satellite population, α_{LT}.
- Case C: Fixing both b and α_{LT}.

These tests are motivated by the significant correlations observed between b, α_{LT}, the kSZ amplitude, and the radio foreground parameters (see section 7.3.1). Since SO frequencies are expected to be less sensitive to LT galaxies, cases B and C probe whether freeing α_{LT} meaningfully degrades the constraints. All other priors remain identical to those of the baseline run.

Sampling convergence is assessed with the Gelman–Rubin statistic, requiring $R - 1 < 0.01$. A burn–in fraction of 30% is applied to all chains before computing the posteriors.

Tables 7.2 and 7.3 list the best–fit values and 1σ uncertainties for the four scenarios considered, while figure 7.10 presents the corresponding one–dimensional marginalised posteriors and two–dimensional parameter correlations at the 68% and 95% confidence levels.
Across all configurations, the results are mutually consistent, confirming that the Fisher–based parameter selection was robust. As expected, progressively fixing additional parameters in cases A–C narrows the posterior widths relative to the baseline run, demonstrating the anticipated gain in constraining power.
The amplitude of the radio–source signal, $\mathcal{A}_{radio}$, is very well constrained and matches its fiducial value in all cases, reflecting the dominance of radio sources in the lowest SO frequency channels (27 and 39 GHz) and their significant contribution even at 93 GHz on small angular scales.
The kSZ amplitude, $\mathcal{A}_{kSZ}$, is likewise well constrained and consistent with its fiducial value, but its precision noticeably improves when either b or α_{LT} (or both) are fixed, as visible in the triangle plot. This behaviour agrees with the positive correlations between $\mathcal{A}_{kSZ}$ and these parameters already evident in the Fisher forecasts (*i.e.*, see figure 7.5).
The hydrostatic mass bias, b, is tightly constrained in the baseline and case B runs, again matching the fiducial value and consistent with its high Fisher SNR; fixing $\alpha_{\rm LT}$ alone produces no significant additional gain in b's precision.
Parameters describing ET–galaxy clustering are moderately well constrained, and their posterior widths decrease substantially when α_{LT} is fixed (cases B and C), which reduces degeneracy between ET and LT populations. By contrast, LT–galaxy clustering parameters remain poorly constrained in all configurations. Even when α_{LT} is fixed, there is little improvement in the constraints on $\log M^{LT}_{min}$, and its posterior peak shifts toward higher values compared with the fiducial, suggesting that SO–like frequency coverage provides only limited sensitivity to the LT population, which is subdominant at these frequencies. Finally, the amplitude of the CIB SN contribution, $\mathcal{A}_{SN}$, remains stable and in excellent agreement with its fiducial value across all four scenarios, consistent with its weak correlations observed in both the Fisher analysis and the two–dimensional posteriors.

7.4 CONCLUSIONS AND FUTURE PROSPECTS

In this work I have extended the modelling of extragalactic foregrounds developed in Zagatti et al., 2024 and presented in chapter 6, to include additional components: the thermal and kinetic Sunyaev–Zel'dovich (tSZ and kSZ) effects, the cross–correlation between the CIB and tSZ (CIB×tSZ), and the contributions from radio sources and Galactic cirrus. I provide a modular and publicly available Python code that implements these components in a unified framework. In particular, the CIB, tSZ, and CIB×tSZ power spectra

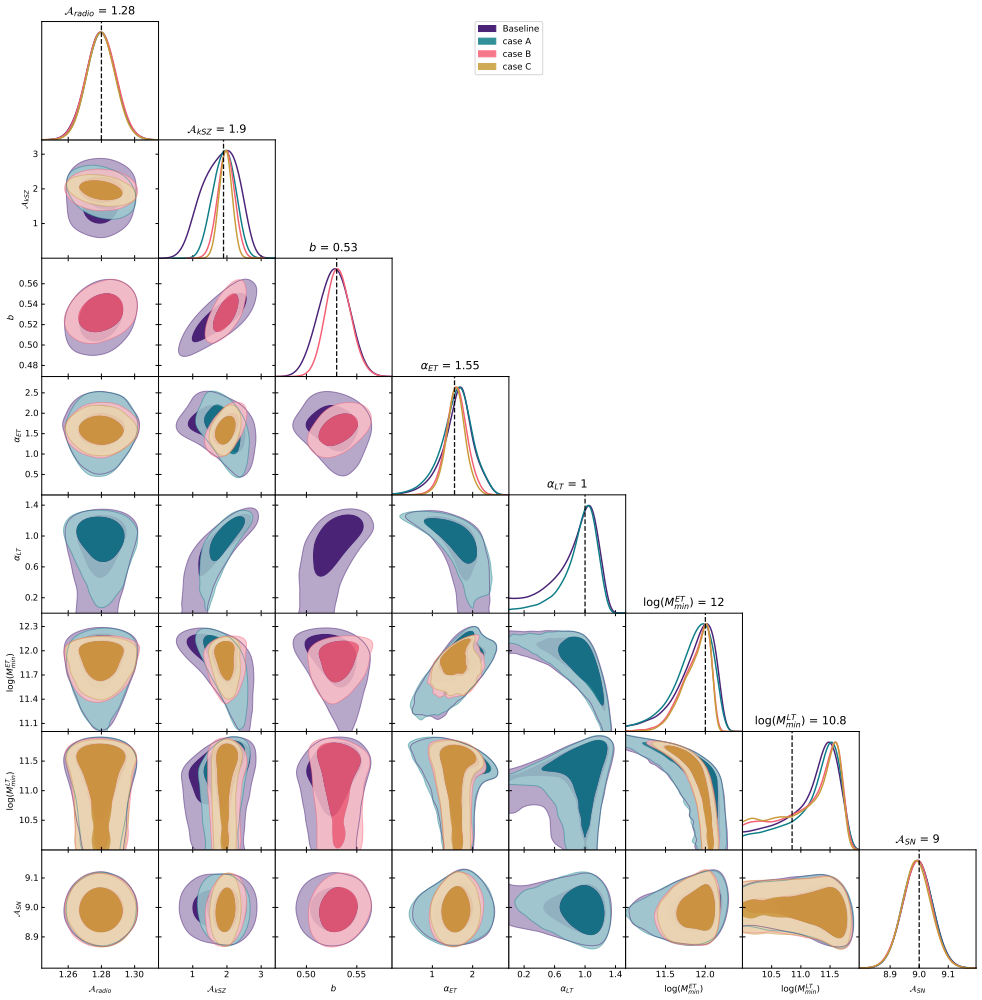

Figure 7.10: Posterior probability distributions for the four MCMC configurations. The baseline case is shown in purple, and cases A, B, and C in teal, pink, and yellow, respectively. Diagonal panels show one–dimensional marginalised distributions with fiducial values indicated by dashed vertical lines. Off–diagonal panels show two–dimensional contours at 68% and 95% confidence levels, highlighting correlations between parameters.

are described using a full HM formalism, while radio sources, Galactic cirrus, and the kSZ effect are treated with a template–based approach.

Using these ingredients, I generated a smooth theoretical dataset designed to mimic the instrumental features of the SO LAT. The dataset covers six SO frequency channels and spans multipoles $\ell \in [1500, 8000]$. Noise curves were computed following the official SO prescriptions, and the resulting power spectra and covariance matrix were binned using the MF_Like scheme. This smooth dataset was then employed to forecast the constraining power of an SO–like experiment on foreground parameters and to explore potential degeneracies between foregrounds and cosmology.

The Fisher forecasts suggest that a physically motivated, cosmology–dependent foreground model will be crucial once SO data probe small an-

gular scales with unprecedented sensitivity. Although a full exploration of cosmological parameter impacts is beyond the scope of this work, the results in section 7.3.1 indicate that neglecting cosmology–foreground correlations may degrade cosmological constraints in the low–variance, high–foreground regime. A more detailed study of these effects – including the joint estimation of cosmological parameters – will be a key direction for future work.

Regarding foreground parameters, SO frequency coverage and sensitivity will not be sufficient to tightly constrain the clustering properties of both galaxy populations. In particular, LT galaxies – being subdominant at SO frequencies – remain poorly constrained. By contrast, the strong contribution of radio sources at low frequencies leads to precise constraints on their amplitude, motivating future efforts to develop a HM description for the radio population. Such a framework would not only be more physically motivated than a purely phenomenological template, but would also allow for a natural prediction of correlations between radio sources and DSF galaxies (responsible for the CIB), as well as between radio sources and tSZ clusters, in a manner analogous to our treatment of the CIB×tSZ correlation.

Finally, the forecasts indicate that the amplitude of the kSZ effect should be well constrained by an SO–like experiment, especially when degeneracies with the hydrostatic mass bias and LT galaxy parameters are mitigated. These findings reinforce the importance of self–consistent modelling of all major extragalactic foregrounds when interpreting forthcoming high–resolution CMB observations. Future analyses combining SO data with measurements from other surveys (*e.g.*, ACT, SPT, Vera Rubin Observatory (VRO) [Ivezić et al., 2019], or CCAT-Prime [Aravena et al., 2023]) could further improve parameter constraints and provide independent tests of the physical models presented here.

CONCLUSIONS

This thesis presented a physically motivated model of the extragalactic emissions relevant for current and forthcoming high-precision Cosmic Microwave Background (CMB) experiments [Ade et al., 2019; Aiola et al., 2020; Balkenhol et al., 2023]. In particular, their increasing sensitivity pushes the exploration of the CMB power spectrum towards foreground-dominated regions. Therefore, a consistent, physically motivated description of these emissions is required in order to maximise the cosmological information that can be extracted from CMB analyses, avoid biasing cosmological inference, and exploit the information these signals can provide in terms of galaxy evolution and Large–scale structure (LSS) formation.

The first main part of this work focused on the Cosmic Infrared Background (CIB), whose modelling was developed in a Halo Model (HM) formalism. Within this framework, I populated dark matter haloes with two distinct galaxy populations, early-type (ET) and late-type (LT) galaxies, showing different emission and clustering properties. For the first time, the contribution from the correlation between the two galaxy populations has been coherently modelled in this framework. The HM allows us to obtain a prediction of the clustering term contributing to the CIB power spectrum. Furthermore, I also accounted for the contribution of the Shot Noise (SN) signal to obtain the full prediction of the power spectrum of the first extragalactic signal studied in this work.

Once the theoretical setup was built, I tested the model predictions against available datasets; *Planck* CIB power spectra (P14, from Planck Collaboration, 2014b), a re-analysis of Planck data from Lenz et al., 2019 (L19), and *Herschel*-SPIRE data (V19, from Viero et al., 2019).
P14 data explore angular scales dominated by the contribution from the clustering of the CIB. However, the frequency range covered by this experiment turns out to be too narrow to disentangle the contribution from the two different galaxy populations. By contrast, V19 extends to higher frequencies and smaller angular scales; this allows for a better constraining power on the clustering parameters of both ET and LT galaxies, as well as a better exploration of the SN signal. However, when comparing the results for the two datasets, a significant tension emerges which prevented me from performing a joint model fit to the data, and, consequently, from ending up with a unique model capable of describing the CIB signal over wide frequency and multipole ranges. The observed tension between the P14 and V19 results may originate from limitations in the adopted modelling assumptions, from residual systematics in the data sets, or from a combination of both. On the modelling side, the tension could reflect an overly restrictive parame-

terisation of the galaxy populations, simplified prescriptions for the halo occupation or Spectral Energy Distribution (SED)s, or residual degeneracies between clustering and SN contributions that are treated differently across angular scales. On the data side, differences in frequency coverage, angular resolution, foreground treatment, and calibration between *Planck* and *Herschel*-SPIRE could lead to effective inconsistencies when the same physical model is applied to both data sets. Disentangling these effects would require targeted consistency tests, such as varying specific modelling assumptions, performing scale- or frequency-restricted fits, or jointly analysing CIB data with complementary tracers, in order to isolate whether the tension is driven primarily by modelling choices or by data-related systematics.

For this reason, in the second part of the thesis, I presented the extension of the modelling to a comprehensive description of the major extragalactic components. Within a unified framework, I modelled CIB, thermal Sunyaev-Zel'dovich (tSZ), and their correlation using a common HM description, while employing a template-based approach for kinetic Sunyaev-Zel'dovich (kSZ), radio sources and Galactic cirrus. I then constructed a dataset based on my theoretical predictions for these emissions which mimics upcoming Simons Obervatory (SO) Large Aperture Telescope (LAT) observations. Employing these data, I carried out Fisher forecasts and Markov Chain Monte Carlo (MCMC) to assess how well a new-generation experiment like SO can constrain these model parameters and to understand how a proper modelling of the interplay between extragalactic emissions and background cosmology affects cosmological inference. From this analysis I found that, even if the two galaxy populations are not equally accessible, for upcoming SO data I expect a high constraining power in the detection of the parameters regulating radio-source emission and, more interestingly in terms of the physics that can be explored, the detection of the amplitude of the kSZ signal. Additionally, preliminary results suggest that when foreground modelling is properly propagated in cosmological inference, there is an improvement in the constraining power on cosmological parameters. This confirms the central role extragalactic emissions have in future cosmological analyses.

The work developed within this thesis confirms that extragalactic foregrounds are not only contaminants to be removed from the primary CMB, but also signals that must be modelled coherently with cosmology. In particular, neglecting or inaccurately modelling these components can lead to biased cosmological parameter estimates, due to degeneracies between foreground amplitudes and cosmological parameters that project onto similar angular and frequency dependences. Moreover, jointly modelling foregrounds and cosmology allows the full information content of the data to be exploited, leading to more robust and, in some cases, tighter constraints than approaches that treat foregrounds as fixed or purely nuisance contributions.

This thesis provides:

- a two-galaxy-population, HM CIB framework;
- a joint HM description of CIB, tSZ, and their cross-correlation;
- a modular implementation of additional extragalactic components;
- forecasts and constraints for SO-like data.

The work done during my PhD is just a first step into a much wider investigation and opened several possibilities. While detailed studies and re-analyses related to the CIB modelling will be carried out, extending the HM description to other extragalactic components will lead to a unified and coherent framework which is currently missing. Furthermore, the HM also naturally predicts bispectra and trispectra, therefore an extension of the HM predictions to higher order statistics will make future analyses more robust. Additionally, this type of approach offers a unique window for the exploration of cross-correlation science.

Part IV

APPENDICES

HALO MODEL CALCULATIONS

This appendix provides useful calculations not fully described within the main text of this thesis.

A.1 DIFFERENT HALO MASS FUNCTION PARAMETRISATION

There is no unified choice in terms of the parametrisation of the Halo Mass Function (HMF); either one may parametrise it in terms of peak height ν or variance of the matter density field σ is acceptable. In the following, I provide the full calculations to move from the To8 HMF expressed in terms of σ to the same expression as a function of ν, and vice versa.
Equation 4.13 provides the relation between the comoving number density of dark matter haloes and the function describing the number density of peaks, $\nu f(\nu)$. I report the same equation below for the sake of simplicity,

$$\frac{m^2 dn/dm(m,z)}{\bar{\rho}} \frac{dm}{m} = \nu f(\nu) \frac{d\nu}{\nu}. \tag{a.1}$$

Simulation results from Jenkins et al., 2001 show that, when moving from mass parametrisation to $\sigma(m)$ parametrisation, the following relation holds:

$$f(\sigma) = \frac{m}{\bar{\rho}} \frac{dn}{d\ln\sigma^{-1}}. \tag{a.2}$$

In order to understand how $\nu f(\nu)$ relates to $f(\sigma)$, starting from equation a.1 I obtain

$$\frac{m}{\bar{\rho}} dn = \nu f(\nu) \frac{d\nu}{\nu}. \tag{a.3}$$

At this point, recalling that $\nu = \delta_c^2/\sigma^2$, I rewrite the $d\nu/\nu$ term as

$$\frac{d\nu}{\nu} = -2d\ln\sigma = 2d\ln\sigma^{-1}, \tag{a.4}$$

so that

$$\frac{m}{\bar{\rho}} dn = 2\nu f(\nu) d\ln\sigma^{-1}. \tag{a.5}$$

Moving the last term on the right-hand side to the left, I obtain

$$\frac{m}{\bar{\rho}} \frac{dn}{d\ln\sigma^{-1}} = 2\nu f(\nu), \tag{a.6}$$

and by equating the above equation with a.2 I end up with the relation between $\nu f(\nu)$ and $f(\sigma)$ as

$$f(\sigma) = 2\nu f(\nu). \tag{a.7}$$

Having this relation in hand, moving from equation 4.13 to equation 4.15 is simply a matter of changing variables.

A.2 CIB–TSZ POWER SPECTRUM: FULL CALCULATIONS

Here I detail all the calculations needed to obtain the power spectrum of the cross-correlation between CIB and tSZ when working with the two-galaxy-populations model. In particular, Sunyaev-Zel'dovich (SZ) clusters can correlate with both ET and LT galaxies.

The power spectrum of the two tracers evaluated in a HM framework is

$$P_{g\times y}(k) = P^{1h}_{g\times y} + P^{2h}_{g\times y}. \tag{a.8}$$

However, because of the two-galaxy-populations model, both the One-halo (1h) and Two-halo (2h) terms are given by the contributions of both galaxy populations so as to have

$$P^{1h}_{g\times y}(k) = P^{1h}_{ET\times y}(k) + P^{1h}_{LT\times y}(k), \tag{a.9}$$

and

$$P^{2h}_{g\times y}(k) = P^{2h}_{ET\times y}(k) + P^{2h}_{LT\times y}(k). \tag{a.10}$$

The expressions of the 1h and 2h terms in both cases of ET and LT galaxies naturally follow the generalised expressions provided in chapter 5 in equations 5.27 and 5.28. For completeness, below I report the specific cases when the different galaxy populations are involved:

$$P^{1h}_{ET\times y}(k) = \int dm \frac{dn}{dm}\bigg|_{500} \frac{N^{ET}_g}{\bar{n}^{ET}_g} \tilde{u}_g(k,m)\tilde{y}(k,m), \tag{a.11}$$

$$P^{1h}_{LT\times y}(k) = \int dm \frac{dn}{dm}\bigg|_{500} \frac{N^{LT}_g}{\bar{n}^{LT}_g} \tilde{u}_g(k,m)\tilde{y}(k,m), \tag{a.12}$$

$$\begin{aligned} P^{2h}_{ET\times y}(k) =& P^{lin}_{mm}(k) \int dm \frac{dn}{dm} b_{CIB}(m) \frac{N^{ET}_g}{\bar{n}^{ET}_g} \tilde{u}_g(k,m)\times \\ & \int dm \frac{dn}{dm}\bigg|_{500} b_{tSZ}(m)\tilde{y}(k,m), \end{aligned} \tag{a.13}$$

$$P^{2h}_{LT\times y}(k) = P^{lin}_{mm}(k) \int dm \frac{dn}{dm} b_{CIB}(m) \frac{N^{LT}_g}{\bar{n}^{LT}_g} \tilde{u}_g(k,m) \times \int dm \frac{dn}{dm}\bigg|_{500} b_{tSZ}(m) \tilde{y}(k,m). \tag{a.14}$$

When adding together the two terms contributing to the 1h part, and the two contributing to the 2h, we obtain,

$$P^{1h}_{CIB\times y}(k) = \int dm \frac{dn}{dm}\bigg|_{500} \tilde{u}_g(k,m)\tilde{y}(k,m) \left[\frac{N^{ET}_g}{\bar{n}^{ET}_g} + \frac{N^{LT}_g}{\bar{n}^{LT}_g}\right], \tag{a.15}$$

for the 1h term, and

$$P^{2h}_{CIB\times y}(k) = P^{lin}_{mm}(k) \int dm \frac{dn}{dm}\bigg|_{500} b_{tSZ}(m) \tilde{y}(k,m) \times \int dm \frac{dn}{dm} b_{CIB}(m) \tilde{u}_g(k,m) \left[\frac{N^{ET}_g}{\bar{n}^{ET}_g} + \frac{N^{LT}_g}{\bar{n}^{LT}_g}\right]. \tag{a.16}$$

At this point, to obtain the final expression of the power spectrum describing the CIB×tSZ signal, we have to properly define the emission kernel described from the most general point of view in equation 5.29 of chapter 5. Specifically, let me define the emission associated with the CIB as

$$\mathcal{E}_\nu(z) = \frac{j_\nu(z)}{dV/dz}. \tag{a.17}$$

Because of the two galaxy populations and the different predictions in terms of the emissivity functions, $j_\nu(z)$, we expect to have different expressions for the emission kernel $\xi(\nu,z)$, depending on whether we are dealing with the cross-correlation SZ clusters have with ET or LT galaxies. Accordingly, the two emission kernels to describe this emission are

$$\xi^{ET}(\nu,z) = \mathcal{E}^{ET}_\nu(z) + T_{CMB} g(\nu), \tag{a.18}$$

and

$$\xi^{LT}(\nu,z) = \mathcal{E}^{LT}_\nu(z) + T_{CMB} g(\nu). \tag{a.19}$$

The general computation of the power spectrum for two tracers u and v requires

$$C^{u\times v}_{\ell,\nu\times\nu'} = \int \frac{dV}{dz} dz \xi(\nu,z)\xi(\nu',z) P_{u\times v}(k). \tag{a.20}$$

As already explained in chapter 5, when considering the product between two emission kernels describing the CIB–tSZ cross-correlation, only two terms among the total four of the product survive, so that

$$\xi(\nu,z)\xi(\nu',z) = T_{CMB}[\mathcal{E}_\nu(z) g(\nu') + \mathcal{E}_{\nu'}(z) g(\nu)], \tag{a.21}$$

for both galaxy populations. However, when summing the two contributions from ET and LT galaxies, we can further simplify the integral defining the product between the emission kernels as

$$\xi(\nu,z)\xi(\nu',z) = T_{CMB}[\mathcal{E}_{\nu}^{gal}(z)g(\nu') + \mathcal{E}_{\nu'}^{gal}(z)g(\nu)], \tag{a.22}$$

with $\mathcal{E}_{\nu}^{gal}(z) = \mathcal{E}_{\nu}^{ET}(z) + \mathcal{E}_{\nu}^{LT}(z)$. At this point, the entire structure to obtain the power spectrum describing the CIB×tSZ signal is set and we can write

$$\begin{aligned} C_{\ell,\nu\times\nu'}^{CIB\times tSZ,1h} =& T_{CMB}\int \frac{dV}{dz}dz[\mathcal{E}_{\nu}^{gal}(z)g(\nu') + \mathcal{E}_{\nu'}^{gal}(z)g(\nu)] \\ & \int dm \frac{dn}{dm}\bigg|_{500} \tilde{u}_g(m,z)\tilde{y}(m,z)\left[\frac{N_g^{ET}}{\bar{n}_g^{ET}} + \frac{N_g^{LT}}{\bar{n}_g^{LT}}\right], \end{aligned} \tag{a.23}$$

$$\begin{aligned} C_{\ell,\nu\times\nu'}^{CIB\times tSZ,2h} =& T_{CMB}\int \frac{dV}{dz}dz[\mathcal{E}_{\nu}^{gal}(z)g(\nu') + \mathcal{E}_{\nu'}^{gal}(z)g(\nu)]P_{mm}^{lin}(k) \\ & \int dm \frac{dn}{dm}\bigg|_{500} b_{tSZ}(m)\tilde{y}(m,z) \\ & \int dm \frac{dn}{dm} b_{CIB}(m)\tilde{u}_g(m,z)\left[\frac{N_g^{ET}}{\bar{n}_g^{ET}} + \frac{N_g^{LT}}{\bar{n}_g^{LT}}\right]. \end{aligned} \tag{a.24}$$

WEBSKY SIMULATIONS FOR CIB MODELLING

After the results presented in chapter 6 pointing to a tension between the datasets employed in the analysis, and before moving to the comprehensive description of chapter 7, I focus on simulations of the CIB.

Since current datasets cannot help to understand whether the issue comes from the data, the modelling, or a combination of both, the idea is to focus as a first step on simulations, to better investigate the modelling. This is because, from a theoretical point of view, simulation products should describe the background signal without the complications coming from the modelling of experimental features such as bandpasses, calibration factors, and colour corrections. For this purpose, I use the simulation products of `Websky` [Stein et al., 2020].

In order to better study the model of the CIB I implemented, `Websky` simulations provide maps of the CIB only for the *Planck* frequency channels. The aim is to create a set of simulated data to use as a reference for the model fit. The logical steps to follow in order to end up with the dataset to use in the MCMC analysis are

1. downgrade the CIB maps[1] to end up with products at a resolution of `NSIDE=2048`;

2. from the downgraded maps of the CIB at the different *Planck* frequency channels, obtain the associated masks by applying the flux cut of 7mJy (left and right panels of figure b.1 show the CIB map and the associated mask for the 353, GHz frequency channel only);

> **Mask apodization**
> I do not apodize the masks employed in this analysis. After numerous tests, I find that the results are very insensitive to whether or not the mask is apodized. The same behaviour has been found by Ade et al., 2011b.
> **The order matters**
> In order to obtain the correct results at the desired resolution, one has to follow in the same order the steps I just described. Evaluating the mask over the full-resolution map and then downgrading the mask and the map to compute power spectra does not work.

1 `Websky` products come at a resolution of `NSIDE=4096`, and I want to reproduce *Planck* observations with a resolution of `NSIDE=2048`.

Table b.1: Temperature sensitivity for *Planck* HFI frequency channels.

Frequency [GHz]	σ [deg]	σ [rad]	σ [Jy/Sr]	$N_\ell^{auto}\,[Jy/Sr]^2$
143	$0.55\ \mu K \cdot deg$	$0.00960 \mu K \cdot rad$	3.021×10^{-5}	9.126×10^{-10}
217	$0.78\ \mu K \cdot deg$	$0.01361 \mu K \cdot rad$	2.814×10^{-5}	7.896×10^{-10}
353	$2.56\ \mu K \cdot deg$	$0.4468 \mu K \cdot rad$	15.5×10^{-5}	2.402×10^{-8}
545	$0.78\ kJy \cdot deg/Sr$	$0.01361 kJy \cdot rad/Sr$	13.61	185.23
857	$0.72\ kJy \cdot deg/Sr$	$0.01257 kJy \cdot rad/Sr$	12.57	158.00

Notes. Second column; values of the temperature sensitivities for *Planck* High Frequency Instrument (HFI) frequency channels. For the three lowest frequency channels they are provided in CMB units, while the sensitivities for the 353, and the 545 GHz channels are in flux units. Third and fourth columns show the same sensitivities converted to radians and in flux units, respectively. Last column lists the instrumental noise levels for auto-frequency only.

3. having both the maps and the masks, use the `NaMaster` Python module [Alonso et al., 2019] to compute the associated power spectra;

4. compute the noise curves for the different auto- and cross-frequency spectra using the instrumental noise levels presented in table 6 of Planck Collaboration, 2016a and reported in the second column of table b.1 for the sake of simplicity. In order to end up with the noise curves associated with these sensitivities, I first convert all the values reported in the *Planck* paper from degrees to radians (third column of table b.1). Because the final spectra are in flux units, I employ the conversion factors described in the introduction of chapter 5 to move from CMB to flux units (fourth column of table b.1). The noise curves follow

$$N_{\ell,\nu\times\nu'} = \sigma_\nu \cdot \sigma_{\nu'}.$$

Values for the auto-frequency spectra are reported in the last column of table b.1 and are in agreement with the findings of McCarthy, 2024;

5. compute the covariance matrix, encoding the *Planck* noise curves and the effect of the mask. This last step is carried out employing the `NaMaster` Python module.

After step 5, I end up with the simulated data needed to perform an MCMC analysis.

Figure b.2 shows the auto-frequency spectra only obtained from `Websky` simulations, compared with the available datasets.

Unfortunately, this analysis does not lead to useful results. As can be seen from the figure, results from simulations are never in agreement with the observed data. Even if I do not expect a perfect agreement, this behaviour suggests that simulations are not actually reproducing the CIB signal.

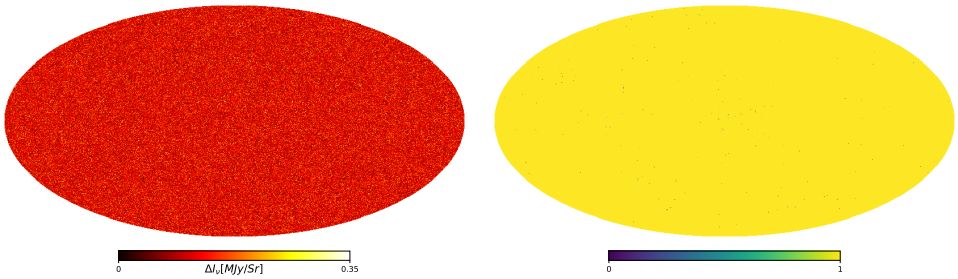

Figure b.1: Left panel: map of the CIB from `Websky` simulations. Right panel: mask associated with the CIB evaluated after imposing a flux cut of $7\,mJy$.

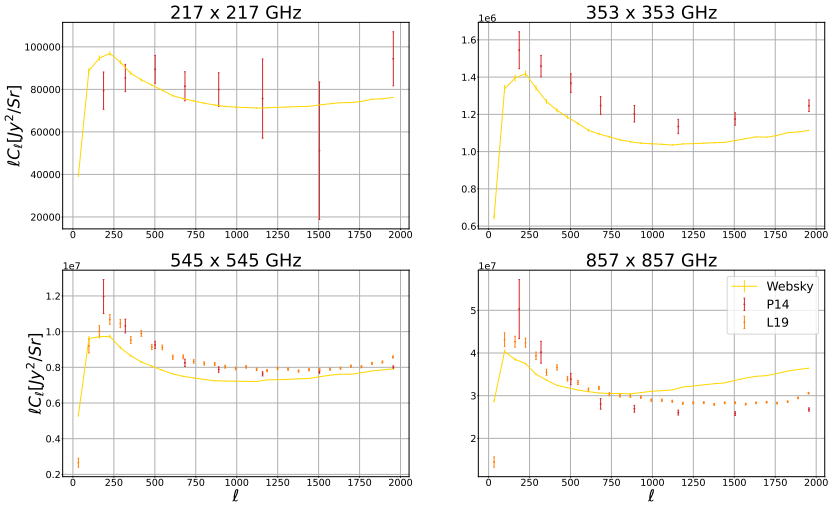

Figure b.2: Auto-frequency spectra only evaluated starting from `Websky` simulations in yellow, compared with actual CIB observations from P14 and L19 in red and orange, respectively.

This hypothesis is confirmed by a very poor MCMC analysis. Preliminary results show a complete inability of the simulated data to capture the model; many parameters are unconstrained, and many others fall well below their expected values.

The results of the MCMC analysis, along with the comparison of the power spectra from `Websky` against the observed CIB data, are sufficient to lead me to abandon this path of analysis. When more reliable simulations, such as the ones under development presented by Bayer et al., 2025, are available, I will resume this analysis.

ANISOTROPIC COSMIC BIREFRINGENCE

Although this thesis focuses on extragalactic foregrounds, here I summarise separate work on anisotropic Cosmic Birefringence (CB) based on *Planck* PR3 and PR4 (NPIPE) data. The goal is to provide a concise record of methodology and results; full details, derivations, validation, and extended figures are in the published paper and its appendices.

C.1 INTRODUCTION

The CMB is linearly polarised at the 1–10% level due to Thomson scattering. In the standard scenario, Maxwell electromagnetism preserves parity,

$$\mathcal{L}_{em}^{SM} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{C.1}$$

implying vanishing TB and EB cross-correlations. Recent *Planck*-based analyses [Diego-Palazuelos et al., 2022; Eskilt, 2022; Eskilt and Komatsu, 2022; Minami and Komatsu, 2020a], using techniques that disentangle instrumental angle from CB [Minami, 2020; Minami and Komatsu, 2020b; Minami et al., 2019], hint at an isotropic CB angle $\beta \sim 0.3^\circ$ at $\sim$2.5–3σ, while anisotropic CB remains compatible with null [Bianchini et al., 2020; Bortolami et al., 2022; Contreras et al., 2017; Gruppuso et al., 2020; Namikawa et al., 2020] (see [Komatsu, 2022] for a review).

A parity-violating Chern–Simons extension,

$$\mathcal{L}_{em} = \mathcal{L}_{em}^{SM} - \frac{\lambda}{4f} \phi F_{\mu\nu} \tilde{F}^{\mu\nu}, \tag{C.2}$$

with a (pseudo-)scalar field ϕ, produces an isotropic CB if ϕ is homogeneous and anisotropic CB from its fluctuations [Carroll et al., 1990; Gubitosi et al., 2011; Kamionkowski, 2009; Li and Zhang, 2008]. This motivates updated constraints on anisotropic CB and its cross-correlations with CMB fields (see also [Caldwell et al., 2011]).

C.2 HARMONIC ESTIMATOR

Observed polarisation is the sum of primordial modes and rotation-induced modes [Kamionkowski, 2009]:

$$\begin{aligned} a_{\ell m}^{E,\mathrm{tot}} &= a_{\ell m}^{E} + \delta a_{\ell m}^{E}, \\ a_{\ell m}^{B,\mathrm{tot}} &= a_{\ell m}^{B} + \delta a_{\ell m}^{B} \simeq \delta a_{\ell m}^{B}, \end{aligned} \tag{C.3}$$

assuming negligible primordial B at last scattering. For the CB field $\alpha(\hat{n}) = \sum_{LM} \alpha_{LM} Y_{LM}$,

$$\begin{aligned} \delta a^B_{\ell m} &= 2 \sum_{LM} \sum_{\ell' m'} \alpha_{LM} a^E_{\ell' m'} \xi^{LM}_{\ell m \ell' m'} H^L_{\ell \ell'}, \\ \delta a^E_{\ell m} &= 2i \sum_{LM} \sum_{\ell' m'} \alpha_{LM} a^E_{\ell' m'} \xi^{LM}_{\ell m \ell' m'} H^L_{\ell \ell'}, \end{aligned} \tag{c.4}$$

with the Wigner-3j kernels $\xi^{LM}_{\ell m \ell' m'}$ and $H^L_{\ell \ell'}$ as in Kamionkowski, 2009. The resulting (parity-violating) cross-correlations can be written via rotational invariants

$$\begin{aligned} \langle a^{X,\mathrm{tot}}_{\ell m} a^{X',\mathrm{tot}*}_{\ell' m'} \rangle &= \sum_{LM} D^{LM,XX'}_{\ell \ell'} \xi^{LM}_{\ell m \ell' m'}, \\ D^{LM,XX'}_{\ell \ell'} &= 2 \alpha_{LM} Z^{XX'}_{\ell \ell'} H^L_{\ell \ell'}, \end{aligned} \tag{c.5}$$

where $Z^{XX'}_{\ell \ell'}$ depends on the primordial spectra [Gluscevic et al., 2009]. Including a Gaussian window function W_ℓ,

$$D^{LM,XX',\mathrm{map}}_{\ell \ell'} = 2 \alpha_{LM} Z^{XX'}_{\ell \ell'} H^L_{\ell \ell'} W_\ell W_{\ell'}. \tag{c.6}$$

Two equivalent estimators follow [Gluscevic et al., 2009; Pullen and Kamionkowski, 2007]:

$$\begin{aligned} \hat{D}^{LM,XX',\mathrm{map}}_{\ell \ell'} &= 2 \hat{\alpha}_{LM} Z^{XX'}_{\ell \ell'} H^L_{\ell \ell'} W_\ell W_{\ell'}, \\ \hat{D}^{LM,XX',\mathrm{map}}_{\ell \ell'} &= G^{-1}_{\ell \ell'} \sum_{m m'} a^{X,\mathrm{map}}_{\ell m} a^{X',\mathrm{map}*}_{\ell' m'} \xi^{LM}_{\ell m \ell' m'}, \end{aligned} \tag{c.7}$$

with $G_{\ell \ell'} = \frac{(2\ell+1)(2\ell'+1)}{4\pi}$. For each (ℓ, ℓ'),

$$(\overline{\alpha}_{LM})^{XX'}_{\ell \ell'} = \frac{\hat{D}^{LM,XX',\mathrm{map}}_{\ell \ell'}}{F^{L,XX'}_{\ell \ell'}}, \qquad F^{L,XX'}_{\ell \ell'} \equiv 2 Z^{XX'}_{\ell \ell'} H^L_{\ell \ell'} W_\ell W_{\ell'}. \tag{c.8}$$

An inverse-variance combination yields

$$\overline{\alpha}^{XX'}_{LM} = \frac{\sum_{\ell \ell'} (\overline{\alpha}_{LM})^{XX'}_{\ell \ell'} / (\sigma^2_L)^{XX'}_{\ell \ell'}}{\sum_{\ell \ell'} 1 / (\sigma^2_L)^{XX'}_{\ell \ell'}}, \quad (\sigma^2_L)^{A}_{\ell \ell'} = \frac{C^{\ell \ell'}_{AA'}}{G_{\ell \ell'} F^{L,A}_{\ell \ell'} F^{L,A'}_{\ell \ell'}}. \tag{c.9}$$

c.2.1 EB-only implementation

Since EB has higher Signal–to–Noise Ratio (SNR) than TB for *Planck*, I use EB only. Using

$$\langle a^{E,\mathrm{tot}}_{\ell m} a^{B,\mathrm{tot}}_{\ell' m'} \rangle = 2 \sum_{LM} \alpha_{LM} C^{EE}_\ell \xi^{LM}_{\ell m \ell' m'} H^L_{\ell \ell'}, \tag{c.10}$$

the fast, un-normalised form is

$$\overline{\alpha}^{\mathrm{UN}}_{LM} = \int d\hat{n} Y_{LM}(\hat{n}) 2 \left(Q^E U^B - U^E Q^B \right), \tag{c.11}$$

with map-like objects

$$
\begin{aligned}
Q^E \pm iU^E &= \sum_{\ell m} \left(C_\ell^{EE} \bar{a}^E_{\ell m} \right)_{\pm 2} Y_{\ell m}, \quad \bar{a}^{E*}_{\ell m} = \frac{a^{E,\mathrm{map}}_{\ell m}}{C_\ell^{EE,\mathrm{map}}} W_\ell, \\
Q^B \pm iU^B &= \sum_{\ell m} \left(\pm i \bar{a}^B_{\ell m} \right)_{\pm 2} Y_{\ell m}, \quad \bar{a}^{B*}_{\ell m} = \frac{a^{B,\mathrm{map}}_{\ell m}}{C_\ell^{BB,\mathrm{map}}} W_\ell,
\end{aligned}
\tag{C.12}
$$

and $C_\ell^{XX,\mathrm{map}} = C_\ell^{XX} W_\ell^2 + N_\ell^{XX}$. The normalised estimates and mean-field subtraction are

$$
\bar{\alpha}_{LM} = \frac{(\bar{\alpha}^{\mathrm{UN},*}_{LM})^*}{\sigma_L^{-2}}, \qquad \hat{\alpha}_{LM} = \bar{\alpha}_{LM} - \langle \bar{\alpha}_{LM} \rangle_{\mathrm{sims}}. \tag{C.13}
$$

C.3 POWER SPECTRUM AND DE-BIASING

From the de-biased coefficients,

$$
C_L^{\hat{\alpha}\hat{\alpha}} = \frac{1}{f_{sky}} \frac{1}{2L+1} \sum_M \hat{\alpha}_{LM} \hat{\alpha}^*_{LM}. \tag{C.14}
$$

The final estimate removes an isotropic and a Monte Carlo bias:

$$
\hat{C}_L^{\alpha\alpha} = C_L^{\hat{\alpha}\hat{\alpha}} - \left(C_L^{\mathrm{bias,iso}} + C_L^{\mathrm{bias,MC}} \right), \tag{C.15}
$$

where $C_L^{\mathrm{bias,iso}}$ selects the diagonal $\ell = \ell'$ part analytically [Gluscevic et al., 2012] and $C_L^{\mathrm{bias,MC}}$ captures off-diagonal contributions from mask, inhomogeneous noise, and lensing using simulations:

$$
C_L^{\mathrm{bias,MC}} = \left\langle C_L^{\hat{\alpha}\hat{\alpha}} - C_L^{\mathrm{bias,iso}} \right\rangle_{\mathrm{sims}}. \tag{C.16}
$$

In data, $C_L^{\hat{\alpha}\hat{\alpha}}$ is large pre-debias because it scales like $\sim C_\ell^{EE,\mathrm{map}} C_{\ell'}^{BB,\mathrm{map}}$, often noise-dominated.

C.4 DATA AND SIMULATIONS

I use *Planck* PR3 [Planck Collaboration, 2020a] and PR4/NPIPE [Planck Collaboration, 2020c] CMB polarisation maps at `HEALPix Nside=2048`, cleaned with `commander` for the main results and compared with `SMICA`, `NILC`, and `SEVEM`. NPIPE provides 400 E/B simulations for `commander` and 600 for `SEVEM`; PR3 supplies 1000 CMB MC plus noise simulations (splits and full mission). Masks correspond to $f_{sky} \simeq 78\%$ (NPIPE) and 75% (PR3). Unless stated, I take $\ell^{CMB}_{\mathrm{min}} = 50$, $\ell^{CMB}_{\mathrm{max}} = 2000$, and $L \in [0, 1500]$.

C.5 RESULTS

In the following sections, I briefly list and describe the main results of this work, starting from the results of the NPIPE full-mission data, and moving to a short description about the different component separations tested in this analysis, the consistency checks of the estimation procedure, to end up with the cross-correlations of the CB field with the three CMB temperature and polarisation fields.

C.5.1 Full-mission NPIPE

The de-biased $\hat{C}_L^{\alpha\alpha}$ is consistent with zero (Probability To Exceed (PTE) = 84.35%) and yields a 95% upper limit on a scale-invariant spectrum in $D_L \equiv L(L+1)C_L/(2\pi)$ of $\mathcal{A}_{CB} < 0.09\,\mathrm{deg}^2$, compatible with Bortolami et al., 2022. Results agree with ACT-Pol and SPT-Pol null detections [Bianchini et al., 2020; Namikawa et al., 2020].

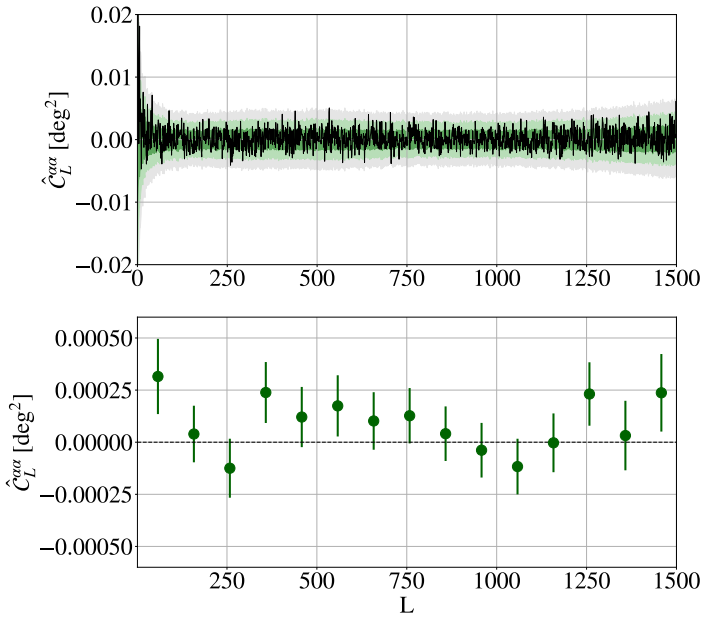

Figure C.1: De-biased CB power spectrum from *Planck* PR4/NPIPE full-mission data (black), with 1, 2, 3σ bands from fully de-biased simulations. Binning: 100 multipoles per bin (first 8 multipoles excluded).

Figure C.1 shows these results. Specifically, the upper panel represents the *L-by-L* power spectrum, and the shaded areas are the 1, 2, and 3σs confidence levels. While the lower panel shows the binned version of the CB power spectrum, after applying a linear binning scheme with 100 multipoles per bin.

C.5.2 Component-separation

PR3/NPIPE results agree across different component separation methods, as shown in figure C.2.

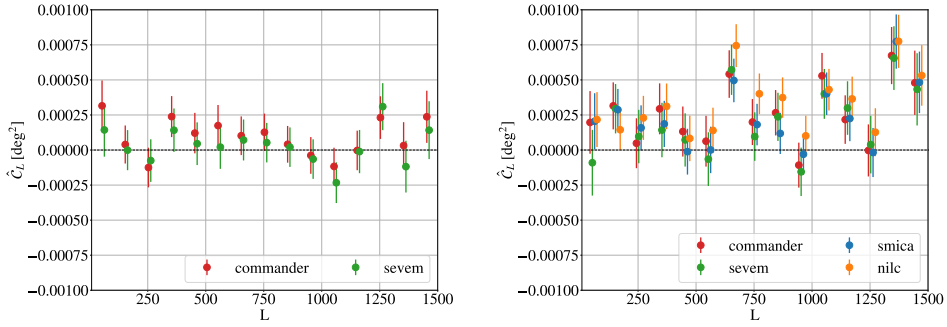

Figure C.2: Left panel: De-biased $\alpha\alpha$ power spectrum after binning with 100 multipoles per bin and excluding the first 8 multipoles using *Planck* NPIPE *full mission* data. We compare the different component separation methods; `commander` in red and `SEVEM` in green. Right panel Same as the left panel, but for *Planck* PR3 data and for the four component separation methods available for PR3; in blue we show the power spectrum obtained using `SMICA` component separation method, in orange `NILC`, in green `SEVEM` and in red `commander`.

I also evaluate the stability of the results against different combination of data splits. Specifically, PR3 is divided into *half-mission 1*, hereafter HM1, and *half-mission 2*, HM2, while NPIPE is composed of data split A and B.

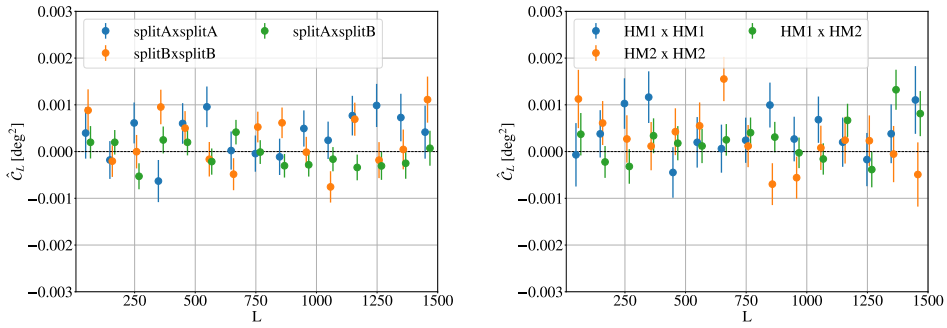

Figure C.3: Left panel: De-biased $\alpha\alpha$ power spectrum after binning with 100 multipoles per bin, excluding the first 8 multipoles from the analysis, obtained auto-correlating α_{LM} estimates from data split A (blue), data split B (orange) and cross-correlating estimates from the two data splits (green) of *Planck* NPIPE data products. Right panel Same as the left panel, for the auto-correlation of α_{LM} estimates from *half mission 1* (blue), *half mission 2* (orange) and for the cross-correlation from the two half missions (green) of *Planck* PR3 data products.

Specifically, left and right panels of figure C.3 show the results for the different combinations of data splits for NPIPE (left) and PR3 (right), confirming the CB power spectrum compatible with zero.

c.5.3 Consistency checks

In figures c.4 and c.5 I also show that varying $\ell_{\rm min}^{CMB} \in 10, 30, 50, 100$, $\ell_{\rm max}^{CMB} \in 1500, 2000, 2500$, and masks with $f_{sky} \in 78, 59, 40, 20\%$ yields stable results.

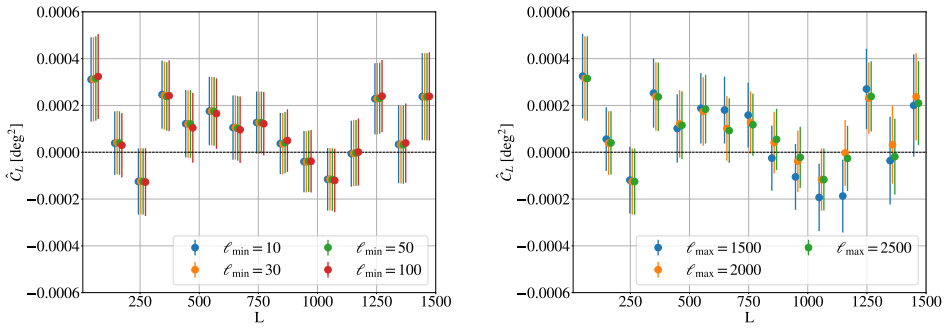

Figure c.4: CB power spectrum evaluated from *Planck* NPIPE *full mission* data products, binning with 100 multipoles per bin and excluding the first 8 multipoles. Left panel Power spectra for different values of ℓ_{min}^{CMB} included in the analysis and $\ell_{max}^{CMB} = 2000$ for all cases. We indicate in *blue* the power spectrum obtained with $\ell_{min}^{CMB} = 10$, in *orange* the one with $\ell_{min}^{CMB} = 30$, in *green* the one with $\ell_{min}^{CMB} = 50$ and in *red* the one with $\ell_{min}^{CMB} = 100$. Right panel Power spectra for different values of ℓ_{max}^{CMB} included in the analysis and $\ell_{min}^{CMB} = 50$ for all cases. We indicate in *blue*, *orange* and *green* the CB power spectra obtained including CMB multipoles up to $\ell_{max}^{CMB} = 1500$, $\ell_{max}^{CMB} = 2000$ and $\ell_{max}^{CMB} = 2500$, respectively.

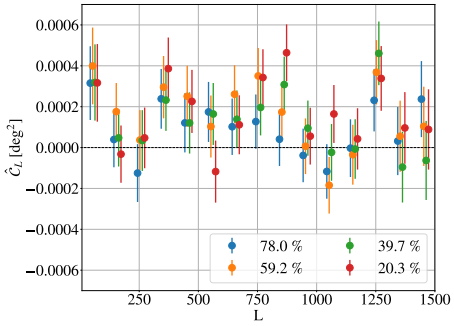

Figure c.5: CB power spectrum evaluated from *Planck* NPIPE *full mission* data products, binning over 100 multipoles per bin and excluding the first 8 multipoles, for different masks applied to *Planck* maps. In *blue* we show the power spectrum obtained with the standard mask, corresponding to a $f_{sky} = 78.0\%$; in *orange*, *green* and *red* we show the power spectrum obtained applying masks corresponding to $f_{sky} = 59.2\%, 39.7\%, 20.3\%$ respectively.

C.5.4 Cross-correlations

The last interesting result of this work is the evaluation up to small scales of the cross-spectra between CB and CMB fields. I build de-biased α maps (smoothed at 1° FWHM) and measure αT (in D_L), αE, and αB (in C_L).

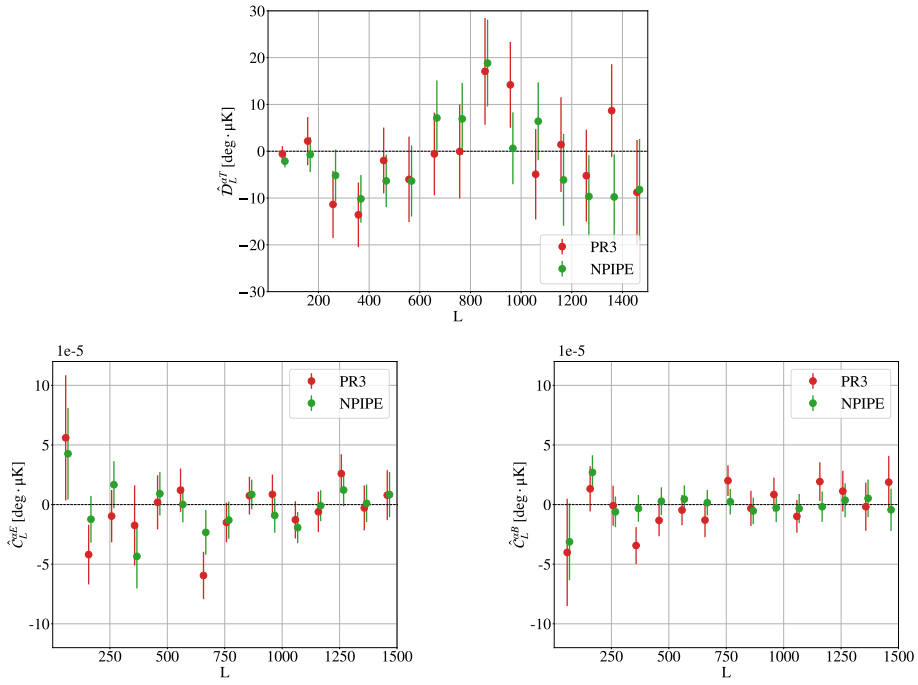

Figure C.6: Cross-spectra between the CMB fields (temperature and polarisation) and the CB field for NPIPE, in green, and PR3, in red, data products. Upper panel αT cross-spectrum in *band-powers*; Lower left panel αE cross-spectrum; Lower right panel αB cross-spectrum.

In figure C.6, the upper panel illustrates the cross-correlation between CB and the CMB temperature field (αT) displayed in band-powers, for both *Planck* NPIPE (green points) and PR3 (red points) datasets. The lower panels display the cross-correlation of the CB field with the CMB polarisation fields (αE and αB).

C.6 BIREFRINGENCE–GALAXY CROSS-CORRELATION AND CONSTRAINTS ON AXION-LIKE FIELDS

In complementary work [Arcari et al., 2025], I contributed to the first measurement of the cross-correlation between anisotropic CB and the large-scale distribution of galaxies, using *Planck* PR4/NPIPE data and the `Quaia` quasar catalogue [Dustin, 2014; Meisner et al., 2019; Vallenari et al., 2023]. The analysis builds directly on the EB-based CB reconstruction summarised in this

appendix, and extends it by correlating the reconstructed $\alpha(\hat{n})$ field with tracers of the underlying matter density.

The theoretical framework assumes a pseudo-scalar (axion-like) field ϕ coupled to electromagnetism via a Chern–Simons term,

$$\mathcal{L} \supset -\frac{1}{4} g_{\phi\gamma} \phi F_{\mu\nu} \tilde{F}^{\mu\nu}, \tag{c.17}$$

so that spatial fluctuations of ϕ induce anisotropic CB, while the same field, through its coupling to metric perturbations, is correlated with the late-time matter distribution. This naturally generates a non-zero cross-correlation between the CB field and galaxy number counts, $C_L^{\alpha g}$, whose shape and amplitude depend on the axion mass, initial misalignment angle, and axion–photon coupling $g_{\phi\gamma}$.

On the data side, the CB map is obtained by applying the harmonic EB estimator (with mean-field subtraction and debiasing) to *Planck* NPIPE polarisation maps, as described above. The galaxy field is constructed from the latest `Quaia` quasar sample, providing a high-redshift, wide-area tracer of the LSS. The cross-spectrum $C_L^{\alpha g}$ is then measured combining a Quadratic Maximum-Likelihood (QML) estimator on large scales with a pseudo-C_L estimator at higher multipoles, using consistent masks and a covariance matrix calibrated on simulations of both probes.

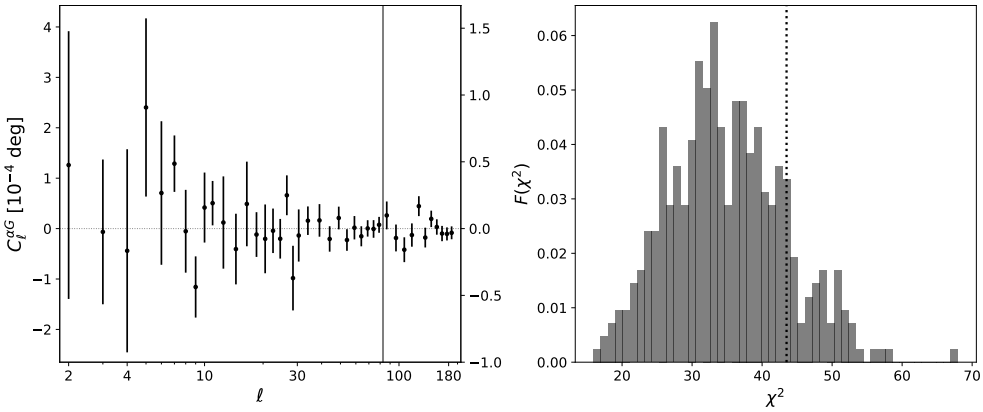

Figure c.7: Left panel: *Planck* NPIPE cross `Quaia` angular power spectrum between anisotropic birefringence and galaxy number counts. Right panel: Corresponding Gaussian χ^2 distribution from the 400 realisations. The dotted vertical line refers to the data χ^2 of the left panel.

The measured birefringence–galaxy cross-correlation is found to be statistically consistent with zero over the full multipole range, with PTE values in broad agreement with the null hypothesis for all robust binning choices (see left and right panels of figure c.7). This null result is then interpreted in terms of the axion-like scenario above. Using a Gaussian likelihood for $C_L^{\alpha g}$, the analysis explores the three-dimensional parameter space of axion mass, initial misalignment, and $g_{\phi\gamma}$. Because very small misalignment angles suppress the signal independently of the coupling, there is a strong

degeneracy that prevents a single, model-independent bound when fully marginalising over all parameters. Instead, the study derives an upper limit on $g_{\phi\gamma}$ at each point in the (m_ϕ, θ_i) plane, shown in figure c.8, identifying regions where the cross-correlation would be large enough to be excluded by the data. In the most favourable part of parameter space, this yields competitive and, in the targeted ultralight mass range, unprecedented constraints on the axion–photon coupling, significantly improving upon straightforward extrapolations of existing bounds at higher masses.

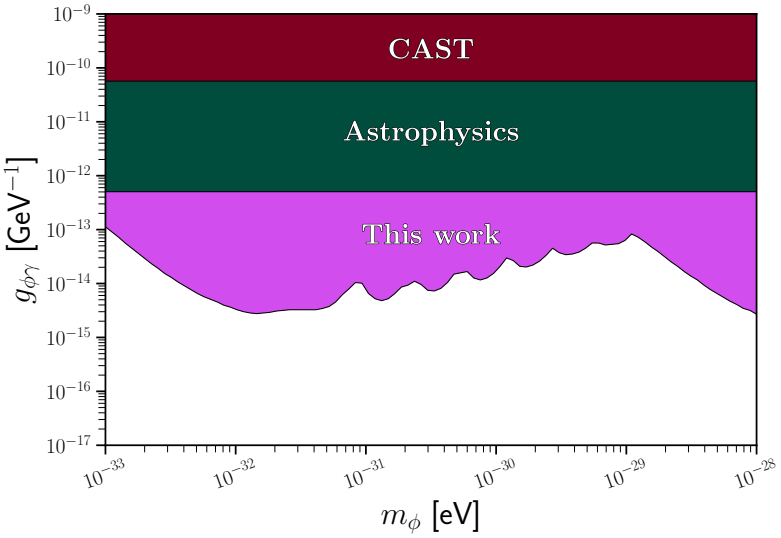

Figure c.8: Best achievable upper bound, at 95% confidence level, on the axion-photon coupling, $g_{\phi\gamma}$, as a function of the axion mass, m_ϕ, as derived from the cross-correlation analysis of *Planck* PR4 CB and `Quaia` galaxy number counts. The result is compared with the extension to lower masses of current bounds, derived in a higher mass range.

From the anisotropic CB perspective, this work demonstrates how birefringence maps, originally developed and validated in the analysis summarised in this appendix, can be turned into a powerful probe of parity-violating physics when combined with LSS tracers. While current data are noise-dominated, the methodology and first constraints highlight the potential of future high-sensitivity CMB polarisation experiments and deep galaxy surveys to either detect a birefringence–galaxy correlation or to further tighten constraints on axion-like fields in a mass range otherwise very difficult to access.

C.7 CONCLUSIONS

In this work I implemented a harmonic EB estimator for the CB spherical harmonic coefficients, including mean-field subtraction and a two-term debiasing at the power-spectrum level. Applied to *Planck* PR3 and PR4/NPIPE data (full mission and splits), I find $C_L^{\alpha\alpha}$ consistent with zero at $\sim 2\sigma$ across the

range, with an upper limit $\mathcal{A}_{CB} < 0.09, \mathrm{deg}^2$. Results are robust to component separation, multipole cuts, and masks, and I also report αT, αE, and αB cross-spectra up to $L = 1500$. Forecasts (not shown here) indicate substantial sensitivity gains for upcoming experiments, with `LiteBIRD` and `CMB-S4` expected to improve constraints by factors $\sim$25 and $\sim$1000, respectively, at the power-spectrum level.

In parallel, I contributed to a complementary analysis correlating the reconstructed anisotropic CB field with the `Quaia` quasar catalogue, targeting the cross-spectrum $C_L^{\alpha g}$ as a probe of axion-like fields coupled to photons. Also in this case the measurements are consistent with null, but they translate into new, competitive constraints on the axion–photon coupling in the ultralight mass regime, under well-defined model assumptions. Together, the *Planck*-only birefringence reconstruction and its cross-correlation with LSS demonstrate that current data already probe parity-violating physics beyond the standard model, and they establish a framework that future high-sensitivity CMB and galaxy surveys can exploit to either detect anisotropic CB or significantly tighten constraints on axion-like scenarios.

The estimator and pipeline, along with CB spectra, cross-spectra products, and the birefringence maps used in these analyses, are publicly available (see `https://github.com/paganol/alpha_lm` for the code, and `https://github.com/giorgiazagatti/CB_Planck_maps_spectra.git` for the products).

ACRONYMS

1h	One-halo
2h	Two-halo
ACT	Atacama Cosmology Telescope
AGN	Active Galactic Nuclei
BAO	Baryon Acoustic Oscillations
CB	Cosmic Birefringence
CDM	Cold Dark Matter
CIB	Cosmic Infrared Background
CMB	Cosmic Microwave Background
DSF	Dusty Star-Forming
EoS	Equation of State
ESA	European Space Agency
ET	early-type
FLRW	Friedmann–Lemaitre–Robertson–Walker
FWHM	Full Width Half Maximum
FT	Fourier Transform
GR	General Relativity
HFI	High Frequency Instrument
HM	Halo Model
HMF	Halo Mass Function
HOD	Halo Occupation Distribution
ICM	intracluster medium
ISW	Integrated Sach–Wolfe
kSZ	kinetic Sunyaev-Zel'dovich
LAT	Large Aperture Telescope
LF	luminosity function
LFI	Low Frequency Instrument
LSS	Large–scale structure
LT	late-type
MCMC	Markov Chain Monte Carlo
NASA	National Aeronautics and Space Administration
NFW	Navarro–Frenk–White
PTE	Probability To Exceed
QML	Quadratic Maximum-Likelihood
rms	root–mean–square
SAT	Small Aperture Telescope
SED	Spectral Energy Distribution
SN	Shot Noise
SNR	Signal–to–Noise Ratio
ST	Seth & Tormen
SO	Simons Obervatory
SPT	South Pole Telescope
SW	Sachs-Wolfe
SZ	Sunyaev-Zel'dovich
tSZ	thermal Sunyaev-Zel'dovich
VRO	Vera Rubin Observatory

BIBLIOGRAPHY

Abitbol, M. et al. (2025). "The Simons Observatory: science goals and forecasts for the enhanced Large Aperture Telescope." In: *JCAP* **08**, p. 034. DOI: 10.1088/1475-7516/2025/08/034. arXiv: 2503.00636 [astro-ph.IM] (cit. on pp. v, vi).

Ade, P. A. R. et al. (2011a). "Planck early results. VI. The High Frequency Instrument data processing." In: *Astron. Astrophys.* **536**, A6. DOI: 10.1051/0004-6361/201116462. arXiv: 1101.2048 [astro-ph.CO] (cit. on p. 105).

– (2011b). "Planck early results. XVIII. The power spectrum of cosmic infrared background anisotropies." In: *Astron. Astrophys.* **536**. DOI: 10.1051/0004-6361/201116461. arXiv: 1101.2028 [astro-ph.CO] (cit. on pp. 98, 99, 106, 161).

Ade, P. et al. (2019). "The Simons Observatory: Science goals and forecasts." In: *JCAP* **02**, p. 056. DOI: 10.1088/1475-7516/2019/02/056. arXiv: 1808.07445 [astro-ph.CO] (cit. on pp. v, vi, 131, 136, 151).

Aghanim, N., S. Majumdar, and J. Silk (2008). "Secondary anisotropies of the CMB." In: *Rept. Prog. Phys.* **71**, p. 066902. DOI: 10.1088/0034-4885/71/6/066902. arXiv: 0711.0518 [astro-ph] (cit. on p. 131).

Aiola, S. et al. (2020). "The Atacama Cosmology Telescope: DR4 Maps and Cosmological Parameters." In: *JCAP* **12**, p. 047. DOI: 10.1088/1475-7516/2020/12/047. arXiv: 2007.07288 [astro-ph.CO] (cit. on p. 151).

Alonso, D., J. Sanchez, and A. Slosar (2019). "A unified pseudo-C_ℓ framework." In: *Mon. Not. Roy. Astron. Soc.* **484**.3, pp. 4127–4151. DOI: 10.1093/mnras/stz093. arXiv: 1809.09603 [astro-ph.CO] (cit. on p. 162).

Alonso, D. et al. (Mar. 2021). "Combining information from multiple cosmological surveys: inference and modeling challenges." In: arXiv: 2103.05320 [astro-ph.CO] (cit. on pp. v, vi).

Amblard, A. et al. (2011). "Sub-millimetre galaxies reside in dark matter halos with masses greater than 3×10^{11} solar masses." In: *Nature* **470**, pp. 510–512. DOI: 10.1038/nature09771. arXiv: 1101.1080 [astro-ph.CO] (cit. on p. 108).

Aravena, M. et al. (2023). "CCAT-prime Collaboration: Science Goals and Forecasts with Prime-Cam on the Fred Young Submillimeter Telescope." In: *Astrophys. J. Suppl.* **264**.1, p. 7. DOI: 10.3847/1538-4365/ac9838. arXiv: 2107.10364 [astro-ph.CO] (cit. on p. 149).

Arcari, S., N. Bartolo, G. Fabbian, A. Greco, A. Gruppuso, M. Lattanzi, P. Natoli, L. Pagano, and G. Zagatti (Sept. 2025). "Stairway to Axions: the cross-correlation of birefringence and galaxies from NPIPE and Quaia." In: arXiv: 2509.22273 [astro-ph.CO] (cit. on p. 171).

Asgari, M., A. J. Mead, and C. Heymans (Mar. 2023). "The halo model for cosmology: a pedagogical review." In: DOI: 10.21105/astro.2303.08752. arXiv: 2303.08752 [astro-ph.CO] (cit. on pp. 54, 61, 66).

Balkenhol, L. et al. (2023). "Measurement of the CMB temperature power spectrum and constraints on cosmology from the SPT-3G 2018 TT, TE, and EE dataset." In: *Phys. Rev. D* **108**.2, p. 023510. DOI: 10.1103/PhysRevD.108.023510. arXiv: 2212.05642 [astro-ph.CO] (cit. on p. 151).

Bardeen, J. M., J. R. Bond, N. Kaiser, and A. S. Szalay (1986). "The Statistics of Peaks of Gaussian Random Fields." In: *Astrophys. J.* **304**, pp. 15–61. DOI: 10.1086/164143 (cit. on pp. 58, 101).

Battaglia, N., J. R. Bond, C. Pfrommer, and J. L. Sievers (2012). "On the Cluster Physics of Sunyaev-Zel'dovich and X-Ray Surveys. II. Deconstructing the Thermal SZ Power Spectrum." In: *ApJ* **758**.2, 75, p. 75. DOI: 10.1088/0004-637X/758/2/75. arXiv: 1109.3711 [astro-ph.CO] (cit. on p. 131).

Battaglia, N., A. Natarajan, H. Trac, R. Cen, and A. Loeb (Oct. 2013). "Reionization on Large Scales. III. Predictions for Low-l Cosmic Microwave Background Polarization and High-l Kinetic Sunyaev-Zel'dovich Observables." In: *ApJ* **776**.2, 83, p. 83. DOI: 10.1088/0004-637X/776/2/83. arXiv: 1211.2832 [astro-ph.CO] (cit. on p. 131).

Bayer, A. E., Y. Zhong, Z. Li, J. DeRose, Y. Feng, and J. Liu (2025). "The HalfDome multi-survey cosmological simulations: N-body simulations." In: *JCAP* **05**, p. 016. DOI: 10.1088/1475-7516/2025/05/016. arXiv: 2407.17462 [astro-ph.CO] (cit. on p. 163).

Berlind, A. A., D. H. Weinberg, A. J. Benson, C. M. Baugh, S. Cole, R. Dave, C. S. Frenk, A. Jenkins, N. Katz, and C. G. Lacey (2003). "The Halo occupation distribution and the physics of galaxy formation." In: *Astrophys. J.* **593**, pp. 1–25. DOI: 10.1086/376517. arXiv: astro-ph/0212357 (cit. on p. 69).

Bertschinger, E. (1985). "The self-similar evolution of holes in an Einstein-de Sitter universe." In: *Astrophys. J. Suppl.* **58**, p. 1. DOI: 10.1086/191027 (cit. on p. 58).

Bethermin, M. et al. (2012). "A unified empirical model for infrared galaxy counts based on observed physical evolution of distant galaxies." In: *Astrophys. J. Lett.* **757**, p. L23. DOI: 10.1088/2041-8205/757/2/L23. arXiv: 1208.6512 [astro-ph.CO] (cit. on pp. 104, 107, 127).

Bianchini, F. et al. (2020). "Searching for Anisotropic Cosmic Birefringence with Polarization Data from SPTpol." In: *Phys. Rev. D* **102**.8, p. 083504. DOI: 10.1103/PhysRevD.102.083504. arXiv: 2006.08061 [astro-ph.CO] (cit. on pp. 165, 168).

Blas, D., J. Lesgourgues, and T. Tram (2011). "The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes." In: *JCAP* **07**, p. 034. DOI: 10.1088/1475-7516/2011/07/034. arXiv: 1104.2933 [astro-ph.CO] (cit. on p. 22).

Bocquet, S., K. Heitmann, S. Habib, E. Lawrence, T. Uram, N. Frontiere, A. Pope, and H. Finkel (2020). "The Mira-Titan Universe. III. Emulation of

the Halo Mass Function." In: *Astrophys. J.* **901**.1, p. 5. DOI: 10.3847/1538-4357/abac5c. arXiv: 2003.12116 [astro-ph.CO] (cit. on p. 54).

Bolliet, B., B. Comis, E. Komatsu, and J. F. Macías-Pérez (2018). "Dark energy constraints from the thermal Sunyaev–Zeldovich power spectrum." In: *Mon. Not. Roy. Astron. Soc.* **477**.4, pp. 4957–4967. DOI: 10.1093/mnras/sty823. arXiv: 1712.00788 [astro-ph.CO] (cit. on p. 131).

Bortolami, M., M. Billi, A. Gruppuso, P. Natoli, and L. Pagano (2022). "Planck constraints on cross-correlations between anisotropic cosmic birefringence and CMB polarization." In: *JCAP* **09**, p. 075. DOI: 10.1088/1475-7516/2022/09/075. arXiv: 2206.01635 [astro-ph.CO] (cit. on pp. 165, 168).

Bullock, J. S., A. Dekel, T. S. Kolatt, A. V. Kravtsov, A. A. Klypin, C. Porciani, and J. R. Primack (2001a). "A Universal angular momentum profile for galactic halos." In: *Astrophys. J.* **555**, pp. 240–257. DOI: 10.1086/321477. arXiv: astro-ph/0011001 (cit. on p. 59).

Bullock, J. S., T. S. Kolatt, Y. Sigad, R. S. Somerville, A. V. Kravtsov, A. A. Klypin, J. R. Primack, and A. Dekel (2001b). "Profiles of dark haloes. Evolution, scatter, and environment." In: *Mon. Not. Roy. Astron. Soc.* **321**, pp. 559–575. DOI: 10.1046/j.1365-8711.2001.04068.x. arXiv: astro-ph/9908159 (cit. on pp. 59, 60).

Cai, Z.-Y., A. Lapi, J.-Q. Xia, G. De Zotti, M. Negrello, C. Gruppioni, E. Rigby, G. Castex, J. Delabrouille, and L. Danese (2013). "A hybrid model for the evolution of galaxies and Active Galactic Nuclei in the infrared." In: *Astrophys. J.* **768**, p. 21. DOI: 10.1088/0004-637X/768/1/21. arXiv: 1303.2335 [astro-ph.CO] (cit. on p. 95).

Caldwell, R. R., V. Gluscevic, and M. Kamionkowski (2011). "Cross-Correlation of Cosmological Birefringence with CMB Temperature." In: *Phys. Rev. D* **84**, p. 043504. DOI: 10.1103/PhysRevD.84.043504. arXiv: 1104.1634 [astro-ph.CO] (cit. on p. 165).

Carbone, C., M. Petkova, and K. Dolag (2016). "DEMNUni: ISW, Rees-Sciama, and weak-lensing in the presence of massive neutrinos." In: *JCAP* **07**, p. 034. DOI: 10.1088/1475-7516/2016/07/034. arXiv: 1605.02024 [astro-ph.CO] (cit. on p. 131).

Carlstrom, J. E., G. P. Holder, and E. D. Reese (2002). "Cosmology with the Sunyaev-Zel'dovich effect." In: *Ann. Rev. Astron. Astrophys.* **40**, pp. 643–680. DOI: 10.1146/annurev.astro.40.060401.093803. arXiv: astro-ph/0208192 (cit. on p. 80).

Carroll, S. M. (2001). "The Cosmological constant." In: *Living Rev. Rel.* **4**, p. 1. DOI: 10.12942/lrr-2001-1. arXiv: astro-ph/0004075 (cit. on p. 101).

Carroll, S. M., G. B. Field, and R. Jackiw (1990). "Limits on a Lorentz and Parity Violating Modification of Electrodynamics." In: *Phys. Rev. D* **41**, p. 1231. DOI: 10.1103/PhysRevD.41.1231 (cit. on p. 165).

Contreras, D., P. Boubel, and D. Scott (2017). "Constraints on direction-dependent cosmic birefringence from Planck polarization data." In: *JCAP*

12, p. 046. DOI: 10.1088/1475-7516/2017/12/046. arXiv: 1705.06387 [astro-ph.CO] (cit. on p. 165).

Cooray, A. R., D. E. Holz, and R. Caldwell (2010). "Measuring dark energy spatial inhomogeneity with supernova data." In: *JCAP* **11**, p. 015. DOI: 10.1088/1475-7516/2010/11/015. arXiv: 0812.0376 [astro-ph] (cit. on p. 131).

Cooray, A. and R. K. Sheth (2002). "Halo Models of Large Scale Structure." In: *Phys. Rept.* **372**, pp. 1–129. DOI: 10.1016/S0370-1573(02)00276-4. arXiv: astro-ph/0206508 (cit. on p. 49).

Crocce, M., P. Fosalba, F. J. Castander, and E. Gaztañaga (2010). "Simulating the Universe with MICE: the abundance of massive clusters." In: *MNRAS* **403**.3, pp. 1353–1367. DOI: 10.1111/j.1365-2966.2009.16194.x. arXiv: 0907.0019 [astro-ph.CO] (cit. on p. 53).

Del Popolo, A., F. Pace, and M. Le Delliou (2017). "A high precision semi-analytic mass function." In: *JCAP* **03**, p. 032. DOI: 10.1088/1475-7516/2017/03/032. arXiv: 1703.06918 [astro-ph.CO] (cit. on p. 54).

Despali, G., C. Giocoli, R. E. Angulo, G. Tormen, R. K. Sheth, G. Baso, and L. Moscardini (2016). "The universality of the virial halo mass function and models for non-universality of other halo definitions." In: *Mon. Not. Roy. Astron. Soc.* **456**.3, pp. 2486–2504. DOI: 10.1093/mnras/stv2842. arXiv: 1507.05627 [astro-ph.CO] (cit. on p. 53).

Diego-Palazuelos, P. et al. (Mar. 2022). "Cosmic Birefringence from Planck Public Release 4." In: *56th Rencontres de Moriond on Cosmology*. arXiv: 2203.04830 [astro-ph.CO] (cit. on p. 165).

Dodelson, S. and F. Schmidt (2020). *Modern Cosmology*. DOI: 10.1016/C2017-0-01943-2 (cit. on pp. 18, 26, 29, 30, 33, 34, 44).

Douspis, M., L. Salvati, A. Gorce, and N. Aghanim (2022). "Retrieving cosmological information from small-scale CMB foregrounds - I. The thermal Sunyaev Zel'dovich effect." In: *Astron. Astrophys.* **659**, A99. DOI: 10.1051/0004-6361/202142004. arXiv: 2109.03272 [astro-ph.CO] (cit. on p. 81).

Dunkley, J. et al. (Feb. 2009). "Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Likelihoods and Parameters from the WMAP data." In: *ApJS* **180**, pp. 306–376. arXiv: 0803.0586 (cit. on p. 131).

Dunkley, J. et al. (2013). "The Atacama Cosmology Telescope: likelihood for small-scale CMB data." In: *JCAP* **07**, p. 025. DOI: 10.1088/1475-7516/2013/07/025. arXiv: 1301.0776 [astro-ph.CO] (cit. on pp. 73, 86).

Dustin, L. (2014). "unWISE: Unblurred Coadds of the WISE Imaging." In: *AJ* **147**.5, 108, p. 108. DOI: 10.1088/0004-6256/147/5/108. arXiv: 1405.0308 [astro-ph.IM] (cit. on p. 171).

Eales, S., S. Lilly, W. Gear, L. Dunne, J. R. Bond, F. Hammer, O. Le Fevre, and D. Crampton (1999). "The canada-uk deep submillimetre survey: first submillimetre images, the source counts, and resolution of the background." In: *Astrophys. J.* **515**, p. 518. DOI: 10.1086/307069. arXiv: astro-ph/9808040 (cit. on p. 75).

Eskilt, J. R. (2022). "Frequency-dependent constraints on cosmic birefringence from the LFI and HFI Planck Data Release 4." In: *Astron. Astrophys.* **662**, A10. DOI: 10.1051/0004-6361/202243269. arXiv: 2201.13347 [astro-ph.CO] (cit. on p. 165).

Eskilt, J. R. and E. Komatsu (2022). "Improved constraints on cosmic birefringence from the WMAP and Planck cosmic microwave background polarization data." In: *Phys. Rev. D* **106**.6, p. 063503. DOI: 10.1103/PhysRevD.106.063503. arXiv: 2205.13962 [astro-ph.CO] (cit. on p. 165).

Fillmore, J. A. and P. Goldreich (1984). "Self-similiar gravitational collapse in an expanding universe." In: *Astrophys. J.* **281**, pp. 1–8. DOI: 10.1086/162070 (cit. on p. 58).

Fixsen, D. J. (Dec. 2009). "The Temperature of the Cosmic Microwave Background." In: *ApJ* **707**, pp. 916–920. DOI: 10.1088/0004-637X/707/2/916. arXiv: 0911.1955 [astro-ph.CO] (cit. on p. 11).

Fixsen, D. J., E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright (1996). "The Cosmic Microwave Background spectrum from the full COBE FIRAS data set." In: *Astrophys. J.* **473**, p. 576. DOI: 10.1086/178173. arXiv: astro-ph/9605054 (cit. on pp. 73, 75).

Fixsen, D. J., E. Dwek, J. C. Mather, C. L. Bennett, and R. A. Shafer (1998). "The Spectrum of the extragalactic far infrared background from the COBE FIRAS observations." In: *Astrophys. J.* **508**, p. 123. DOI: 10.1086/306383. arXiv: astro-ph/9803021 (cit. on p. 90).

Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman (2013). "emcee: The MCMC Hammer." In: *Publ. Astron. Soc. Pac.* **125**, pp. 306–312. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph.IM] (cit. on p. 109).

Freedman, W. L. and B. F. Madore (Aug. 2010). "The Hubble Constant." In: *Annual Review of Astronomy and Astrophysics* **48**.1. arXiv:1004.1856 [astro-ph, physics:gr-qc], pp. 673–710. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev-astro-082708-101829. URL: http://arxiv.org/abs/1004.1856 (visited on 09/10/2022) (cit. on p. 5).

Gao, L., S. D. M. White, A. Jenkins, F. Stoehr, and V. Springel (2004). "The Subhalo populations of lambda-CDM dark halos." In: *Mon. Not. Roy. Astron. Soc.* **355**, p. 819. DOI: 10.1111/j.1365-2966.2004.08360.x. arXiv: astro-ph/0404589 (cit. on p. 125).

George, E. M. et al. (2015). "A measurement of secondary cosmic microwave background anisotropies from the 2500-square-degree SPT-SZ survey." In: *Astrophys. J.* **799**.2, p. 177. DOI: 10.1088/0004-637X/799/2/177. arXiv: 1408.3161 [astro-ph.CO] (cit. on pp. 73, 84–86, 89, 131).

Gluscevic, V., D. Hanson, M. Kamionkowski, and C. M. Hirata (2012). "First CMB Constraints on Direction-Dependent Cosmological Birefringence from WMAP-7." In: *Phys. Rev. D* **86**, p. 103529. DOI: 10.1103/PhysRevD.86.103529. arXiv: 1206.5546 [astro-ph.CO] (cit. on p. 167).

Gluscevic, V., M. Kamionkowski, and A. Cooray (2009). "De-Rotation of the Cosmic Microwave Background Polarization: Full-Sky Formalism." In: *Phys.*

Rev. D **80**, p. 023510. DOI: 10.1103/PhysRevD.80.023510. arXiv: 0905.1687 [astro-ph.CO] (cit. on p. 166).

Gorce, A., M. Douspis, and L. Salvati (2022). "Retrieving cosmological information from small-scale CMB foregrounds - II. The kinetic Sunyaev Zel'dovich effect." In: *Astron. Astrophys.* **662**, A122. DOI: 10.1051/0004-6361/202243351. arXiv: 2202.08698 [astro-ph.CO] (cit. on p. 84).

Griffin, M. et al. (2010). "The Herschel-SPIRE instrument and its in-flight performance." In: *Astron. Astrophys.* **518**. DOI: 10.1051/0004-6361/201014519. arXiv: 1005.5123 [astro-ph.IM] (cit. on p. 108).

Gruppuso, A., D. Molinari, P. Natoli, and L. Pagano (2020). "Planck 2018 constraints on anisotropic birefringence and its cross-correlation with CMB anisotropy." In: *JCAP* **11**, p. 066. DOI: 10.1088/1475-7516/2020/11/066. arXiv: 2008.10334 [astro-ph.CO] (cit. on p. 165).

Gubitosi, G., M. Migliaccio, L. Pagano, G. Amelino-Camelia, A. Melchiorri, P. Natoli, and G. Polenta (2011). "Using CMB data to constrain non-isotropic Planck-scale modifications to Electrodynamics." In: *JCAP* **11**, p. 003. DOI: 10.1088/1475-7516/2011/11/003. arXiv: 1106.6049 [astro-ph.CO] (cit. on p. 165).

Gunn, J. E. and J. R. Gott III (1972). "On the Infall of Matter into Clusters of Galaxies and Some Effects on Their Evolution." In: *Astrophys. J.* **176**, pp. 1–19. DOI: 10.1086/151605 (cit. on p. 49).

Hall, N. R. et al. (2010). "Angular Power Spectra of the Millimeter Wavelength Background Light from Dusty Star-forming Galaxies with the South Pole Telescope." In: *Astrophys. J.* **718**, pp. 632–646. DOI: 10.1088/0004-637X/718/2/632. arXiv: 0912.4315 [astro-ph.CO] (cit. on p. 75).

Hansen, S. M., E. S. Sheldon, R. H. Wechsler, and B. P. Koester (2009). "The Galaxy Content of SDSS Clusters and Groups." In: *Astrophys. J.* **699**, pp. 1333–1353. DOI: 10.1088/0004-637X/699/2/1333. arXiv: 0710.3780 [astro-ph] (cit. on p. 125).

Hernquist, L. (1990). "An Analytical Model for Spherical Galaxies and Bulges." In: *Astrophys. J.* **356**, p. 359. DOI: 10.1086/168845 (cit. on p. 59).

Hill, J. C. (2018). "Foreground Biases on Primordial Non-Gaussianity Measurements from the CMB Temperature Bispectrum: Implications for Planck and Beyond." In: *Phys. Rev. D* **98**.8, p. 083542. DOI: 10.1103/PhysRevD.98.083542. arXiv: 1807.07324 [astro-ph.CO] (cit. on p. 131).

Hogg, D. W. (May 1999). "Distance measures in cosmology." In: arXiv: astro-ph/9905116 (cit. on pp. 12, 76).

Hu, W. and S. Dodelson (2002). "Cosmic Microwave Background Anisotropies." In: *Ann. Rev. Astron. Astrophys.* **40**, pp. 171–216. DOI: 10.1146/annurev.astro.40.060401.093926. arXiv: astro-ph/0110414 (cit. on p. 21).

Hughes, D. et al. (1998). "High - redshift star formation in the Hubble Deep Field revealed by a submillimeter - wavelength survey." In: *Nature* **394**, pp. 241–247. DOI: 10.1038/28328. arXiv: astro-ph/9806297 (cit. on p. 75).

Ivezić, Ž. et al. (2019). "LSST: from Science Drivers to Reference Design and Anticipated Data Products." In: *Astrophys. J.* **873**.2, p. 111. DOI: 10.3847/1538-4357/ab042c. arXiv: 0805.2366 [astro-ph] (cit. on p. 149).

Jenkins, A., C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida (2001). "The Mass function of dark matter halos." In: *Mon. Not. Roy. Astron. Soc.* **321**, p. 372. DOI: 10.1046/j.1365-8711.2001.04029.x. arXiv: astro-ph/0005260 (cit. on pp. 53, 54, 157).

Jing, Y. P. (1999). "The density profile of equilibrium and nonequilibrium dark matter halos." In: arXiv: astro-ph/9901340 (cit. on p. 59).

Kamionkowski, M. (2009). "How to De-Rotate the Cosmic Microwave Background Polarization." In: *Phys. Rev. Lett.* **102**, p. 111302. DOI: 10.1103/PhysRevLett.102.111302. arXiv: 0810.1286 [astro-ph] (cit. on pp. 165, 166).

Komatsu, E. (2022). "New physics from the polarized light of the cosmic microwave background." In: *Nature Rev. Phys.* **4**.7, pp. 452–469. DOI: 10.1038/s42254-022-00452-4. arXiv: 2202.13919 [astro-ph.CO] (cit. on p. 165).

Lagache, G., M. Béthermin, L. Montier, P. Serra, and M. Tucci (2020). "Impact of polarised extragalactic sources on the measurement of CMB B-mode anisotropies." In: *Astron. Astrophys.* **642**, A232. DOI: 10.1051/0004-6361/201937147. arXiv: 1911.09466 [astro-ph.CO] (cit. on p. 127).

Lagache, G., H. Dole, and J. L. Puget (2003). "Modeling the infrared galaxy evolution using a phenomenological approach." In: *Mon. Not. Roy. Astron. Soc.* **338**, p. 555. DOI: 10.1046/j.1365-8711.2003.05971.x. arXiv: astro-ph/0209115 (cit. on p. 108).

Lagache, G., J.-L. Puget, and H. Dole (2005). "Dusty infrared galaxies: Sources of the cosmic infrared background." In: *Ann. Rev. Astron. Astrophys.* **43**, pp. 727–768. DOI: 10.1146/annurev.astro.43.072103.150606. arXiv: astro-ph/0507298 (cit. on p. 131).

Lahav, O., P. B. Lilje, J. R. Primack, and M. J. Rees (1991). "Dynamical effects of the cosmological constant." In: *Mon. Not. Roy. Astron. Soc.* **251**, pp. 128–136 (cit. on p. 101).

Lapi, A. and A. Cavaliere (2011). "Dark Matter Halos: The Dynamical Basis of Effective Empirical Models." In: *Adv. Astron.* **2011**, p. 903429. DOI: 10.1155/2011/903429. arXiv: 1010.2602 [astro-ph.CO] (cit. on p. 102).

Lenz, D., O. Doré, and G. Lagache (2019). "Large-scale Maps of the Cosmic Infrared Background from Planck." In: *Astrophys. J.* **883**.1, p. 75. DOI: 10.3847/1538-4357/ab3c2b. arXiv: 1905.00426 [astro-ph.CO] (cit. on pp. 105, 107, 151).

Lewis, A., A. Challinor, and A. Lasenby (2000). "Efficient computation of CMB anisotropies in closed FRW models." In: *ApJ* **538**, pp. 473–476. DOI: 10.1086/309179. arXiv: astro-ph/9911177 [astro-ph] (cit. on p. 22).

Li, M. and X. Zhang (2008). "Cosmological CPT violating effect on CMB polarization." In: *Phys. Rev. D* **78**, p. 103516. DOI: 10.1103/PhysRevD.78.103516. arXiv: 0810.0403 [astro-ph] (cit. on p. 165).

Limber, D. N. (1953). "The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field." In: *ApJ* **117**, p. 134. DOI: 10.1086/145672 (cit. on p. 77).

Limber, D. N. (1954). "The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. II." In: *Astrophys. J.* **119**, p. 655. DOI: 10.1086/145870 (cit. on p. 77).

Louis, T. et al. (2017). "The Atacama Cosmology Telescope: Two-Season ACTPol Spectra and Parameters." In: *JCAP* **06**, p. 031. DOI: 10.1088/1475-7516/2017/06/031. arXiv: 1610.02360 [astro-ph.CO] (cit. on p. 131).

– (Mar. 2025). "The Atacama Cosmology Telescope: DR6 Power Spectra, Likelihoods and ΛCDM Parameters." In: arXiv: 2503.14452 [astro-ph.CO] (cit. on p. 137).

Ma, C.-P. and E. Bertschinger (1995). "Cosmological perturbation theory in the synchronous and conformal Newtonian gauges." In: *Astrophys. J.* **455**, pp. 7–25. DOI: 10.1086/176550. arXiv: astro-ph/9506072 (cit. on p. 21).

Maartens, R. (2011). "Is the Universe homogeneous?" In: *Phil. Trans. Roy. Soc. Lond. A* **369**, pp. 5115–5137. DOI: 10.1098/rsta.2011.0289. arXiv: 1104.1300 [astro-ph.CO] (cit. on p. 4).

Maniyar, A. S., M. Béthermin, and G. Lagache (2021). "Simple halo model formalism for the cosmic infrared background and its correlation with the thermal Sunyaev-Zel'dovich effect." In: *Astron. Astrophys.* **645**, A40. DOI: 10.1051/0004-6361/202038790. arXiv: 2006.16329 [astro-ph.CO] (cit. on pp. 108, 109, 129).

Mauch, T., T. Murphy, H. J. Buttery, J. Curran, R. W. Hunstead, B. Piestrzynski, J. G. Robertson, and E. M. Sadler (2003). "SUMSS: A wide-field radio imaging survey of the southern sky. 2. The Source catalogue." In: *Mon. Not. Roy. Astron. Soc.* **342**, p. 1117. DOI: 10.1046/j.1365-8711.2003.06605.x. arXiv: astro-ph/0303188 (cit. on p. 89).

McCarthy, F. (May 2024). "Large-scale galactic-dust-cleaned cosmic infrared background maps from \textitPlanck PR4 and HI4PI with \textttpyilc." In: arXiv: 2405.13470 [astro-ph.CO] (cit. on p. 162).

Meisner, A. M., D. Lang, E. F. Schlafly, and D. J. Schlegel (2019). "unWISE Coadds: The Five-year Data Set." In: *Publ. Astron. Soc. Pac.* **131**.1006, p. 124504. DOI: 10.1088/1538-3873/ab3df4. arXiv: 1909.05444 [astro-ph.IM] (cit. on p. 171).

Meszaros, P. (1974). "The behaviour of point masses in an expanding cosmological substratum." In: *A&A* **37**.2, pp. 225–228 (cit. on p. 42).

Minami, Y. (2020). "Determination of miscalibrated polarization angles from observed cosmic microwave background and foreground *EB* power spectra: Application to partial-sky observation." In: *PTEP* **2020**.6, 063E01. DOI: 10.1093/ptep/ptaa057. arXiv: 2002.03572 [astro-ph.CO] (cit. on p. 165).

Minami, Y. and E. Komatsu (2020a). "New Extraction of the Cosmic Birefringence from the Planck 2018 Polarization Data." In: *Phys. Rev. Lett.* **125**.22, p. 221301. DOI: 10.1103/PhysRevLett.125.221301. arXiv: 2011.11254 [astro-ph.CO] (cit. on p. 165).

– (2020b). "Simultaneous determination of the cosmic birefringence and miscalibrated polarization angles II: Including cross frequency spectra." In: *PTEP* **2020**.10, 103E02. DOI: 10.1093/ptep/ptaa130. arXiv: 2006.15982 [astro-ph.CO] (cit. on p. 165).

Minami, Y., H. Ochi, K. Ichiki, N. Katayama, E. Komatsu, and T. Matsumura (2019). "Simultaneous determination of the cosmic birefringence and miscalibrated polarization angles from CMB experiments." In: *PTEP* **2019**.8, 083E02. DOI: 10.1093/ptep/ptz079. arXiv: 1904.12440 [astro-ph.CO] (cit. on p. 165).

Mo, H. J. and S. D. M. White (1996). "An Analytic model for the spatial clustering of dark matter halos." In: *Mon. Not. Roy. Astron. Soc.* **282**, p. 347. DOI: 10.1093/mnras/282.2.347. arXiv: astro-ph/9512127 (cit. on p. 56).

Mocanu, L. M. et al. (2013). "Extragalactic Millimeter-wave Point-source Catalog, Number Counts and Statistics from 771 deg^2 of the SPT-SZ Survey." In: *Astrophys. J.* **779**, p. 61. DOI: 10.1088/0004-637X/779/1/61. arXiv: 1306.3470 [astro-ph.CO] (cit. on p. 89).

Moore, B., T. R. Quinn, F. Governato, J. Stadel, and G. Lake (1999). "Cold collapse and the core catastrophe." In: *Mon. Not. Roy. Astron. Soc.* **310**, pp. 1147–1152. DOI: 10.1046/j.1365-8711.1999.03039.x. arXiv: astro-ph/9903164 (cit. on p. 59).

Mroczkowski, T. et al. (2019). "Astrophysics with the Spatially and Spectrally Resolved Sunyaev-Zeldovich Effects: A Millimetre/Submillimetre Probe of the Warm and Hot Universe." In: *Space Sci. Rev.* **215**.1, p. 17. DOI: 10.1007/s11214-019-0581-2. arXiv: 1811.02310 [astro-ph.CO] (cit. on p. 80).

Nakamura, T. T. and Y. Suto (1997). "Strong gravitational lensing and velocity function as tools to probe cosmological parameters: Current constraints and future predictions." In: *Prog. Theor. Phys.* **97**, pp. 49–81. DOI: 10.1143/PTP.97.49. arXiv: astro-ph/9612074 (cit. on p. 101).

Namikawa, T. et al. (2020). "Atacama Cosmology Telescope: Constraints on cosmic birefringence." In: *Phys. Rev. D* **101**.8, p. 083527. DOI: 10.1103/PhysRevD.101.083527. arXiv: 2001.10465 [astro-ph.CO] (cit. on pp. 165, 168).

Navarro, J. F., C. S. Frenk, and S. D. M. White (1996). "The Structure of cold dark matter halos." In: *Astrophys. J.* **462**, pp. 563–575. DOI: 10.1086/177173. arXiv: astro-ph/9508025 (cit. on pp. 58, 59).

Ohanian, H. C. and R. Ruffini (2013). *Gravitation and Spacetime*. 3rd ed. Cambridge University Press. DOI: 10.1017/CBO9781139003391 (cit. on p. 7).

Peacock, J. A. (1999). *Cosmological physics* (cit. on p. 12).

– (2007). "Testing anthropic predictions for Lambda and the CMB temperature." In: *Mon. Not. Roy. Astron. Soc.* **379**, pp. 1067–1074. DOI: 10.1111/j.1365-2966.2007.11978.x. arXiv: 0705.0898 [astro-ph] (cit. on p. 53).

Pilbratt, G. et al. (2010). "Herschel Space Observatory. An ESA facility for far-infrared and submillimetre astronomy." In: *Astron. Astrophys.* **518**. DOI: 10.1051/0004-6361/201014759. arXiv: 1005.5331 [astro-ph.IM] (cit. on p. 108).

Planck Collaboration (2011). "Planck Early Results. I. The Planck mission." In: *Astron. Astrophys.* **536**, A1. DOI: 10.1051/0004-6361/201116464. arXiv: 1101.2022 [astro-ph.IM] (cit. on p. 105).

– (2014a). "Planck 2013 results. XV. CMB power spectra and likelihood." In: *Astron. Astrophys.* **571**, A15. DOI: 10.1051/0004-6361/201321573. arXiv: 1303.5075 [astro-ph.CO] (cit. on p. 82).

– (2014b). "Planck 2013 results. XXX. Cosmic infrared background measurements and implications for star formation." In: *Astron. Astrophys.* **571**, A30. DOI: 10.1051/0004-6361/201322093. arXiv: 1309.0382 [astro-ph.CO] (cit. on pp. v, vi, 73, 75, 105, 106, 151).

– (2016a). "Planck 2015 results. VIII. High Frequency Instrument data processing: Calibration and maps." In: *Astron. Astrophys.* **594**, A8. DOI: 10.1051/0004-6361/201525820. arXiv: 1502.01587 [astro-ph.CO] (cit. on pp. 106, 110, 162).

– (2016b). "Planck 2015 results. XXI. The integrated Sachs-Wolfe effect." In: *Astron. Astrophys.* **594**, A21. DOI: 10.1051/0004-6361/201525831. arXiv: 1502.01595 [astro-ph.CO] (cit. on p. 131).

– (2016c). "Planck 2015 results. XXII. A map of the thermal Sunyaev-Zeldovich effect." In: *Astron. Astrophys.* **594**, A22. DOI: 10.1051/0004-6361/201525826. arXiv: 1502.01596 [astro-ph.CO] (cit. on p. 83).

– (2016d). "Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts." In: *Astron. Astrophys.* **594**, A24. DOI: 10.1051/0004-6361/201525833. arXiv: 1502.01597 [astro-ph.CO] (cit. on p. 86).

– (2016e). "Planck intermediate results. XLVII. Planck constraints on reionization history." In: *Astron. Astrophys.* **596**, A108. DOI: 10.1051/0004-6361/201628897. arXiv: 1605.03507 [astro-ph.CO] (cit. on pp. 73, 80).

– (2020a). "Planck 2018 results. I. Overview and the cosmological legacy of Planck." In: *Astron. Astrophys.* **641**, A1. DOI: 10.1051/0004-6361/201833880. arXiv: 1807.06205 [astro-ph.CO] (cit. on pp. 36, 37, 105, 167).

– (2020b). "Planck 2018 results. VI. Cosmological parameters." In: *Astron. Astrophys.* **641**. [Erratum: Astron.Astrophys. 652, C4 (2021)], A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO] (cit. on pp. 85, 105, 131).

– (2020c). "*Planck* intermediate results. LVII. Joint Planck LFI and HFI data processing." In: *Astron. Astrophys.* **643**, A42. DOI: 10.1051/0004-6361/202038073. arXiv: 2007.04997 [astro-ph.CO] (cit. on p. 167).

Press, W. H. and P. Schechter (1974). "Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation." In: *Astrophys. J.* **187**, pp. 425–438. DOI: 10.1086/152650 (cit. on p. 52).

Puget, J. L., A. Abergel, J. P. Bernard, F. Boulanger, W. B. Burton, F. X. Desert, and D. Hartmann (1996). "Tentative detection of a cosmic far - infrared background with COBE." In: *Astron. Astrophys.* **308**, p. L5 (cit. on pp. 73, 75).

Pullen, A. R. and M. Kamionkowski (2007). "Cosmic Microwave Background Statistics for a Direction-Dependent Primordial Power Spectrum." In: *Phys. Rev. D* **76**, p. 103529. DOI: 10.1103/PhysRevD.76.103529. arXiv: 0709.1144 [astro-ph] (cit. on p. 166).

Reed, D., R. Bower, C. Frenk, A. Jenkins, and T. Theuns (2007). "The halo mass function from the dark ages through the present day." In: *Mon. Not. Roy. Astron. Soc.* **374**, pp. 2–15. DOI: 10.1111/j.1365-2966.2006.11204.x. arXiv: astro-ph/0607150 (cit. on p. 53).

Reichardt, C. L. et al. (2021). "An Improved Measurement of the Secondary Cosmic Microwave Background Anisotropies from the SPT-SZ + SPTpol Surveys." In: *Astrophys. J.* **908**.2, p. 199. DOI: 10.3847/1538-4357/abd407. arXiv: 2002.06197 [astro-ph.CO] (cit. on p. 131).

Salvati, L., M. Douspis, A. Ritz, N. Aghanim, and A. Babul (2019). "Mass bias evolution in tSZ cluster cosmology." In: *Astron. Astrophys.* **626**, A27. DOI: 10.1051/0004-6361/201935041. arXiv: 1901.03096 [astro-ph.CO] (cit. on p. 83).

Saunders, W., M. Rowan-Robinson, A. Lawrence, G. Efstathiou, N. Kaiser, R. S. Ellis, and C. S. Frenk (1990). "The 60-micron and far-infrared luminosity functions of IRAS galaxies." In: *Mon. Not. Roy. Astron. Soc.* **242**, pp. 318–337 (cit. on p. 102).

Scoccimarro, R., R. K. Sheth, L. Hui, and B. Jain (2001). "How many galaxies fit in a halo? Constraints on galaxy formation efficiency from spatial clustering." In: *Astrophys. J.* **546**, pp. 20–34. DOI: 10.1086/318261. arXiv: astro-ph/0006319 (cit. on p. 69).

Shaw, L. D., D. H. Rudd, and D. Nagai (2012). "Deconstructing the Kinetic SZ Power Spectrum." In: *ApJ* **756**.1, 15, p. 15. DOI: 10.1088/0004-637X/756/1/15. arXiv: 1109.0553 [astro-ph.CO] (cit. on p. 84).

Sheth, R. K. and G. Tormen (1999). "Large scale bias and the peak background split." In: *Mon. Not. Roy. Astron. Soc.* **308**, p. 119. DOI: 10.1046/j.1365-8711.1999.02692.x. arXiv: astro-ph/9901122 (cit. on p. 53).

Smail, I., R. J. Ivison, and A. W. Blain (1997). "A Deep submillimeter survey of lensing clusters: A New window on galaxy formation and evolution." In: *Astrophys. J. Lett.* **490**, p. L5. DOI: 10.1086/311017. arXiv: astro-ph/9708135 (cit. on p. 75).

Stein, G., M. A. Alvarez, J. R. Bond, A. van Engelen, and N. Battaglia (2020). "The Websky Extragalactic CMB Simulations." In: *JCAP* **10**, p. 012. DOI:

10.1088/1475-7516/2020/10/012. arXiv: 2001.08787 [astro-ph.CO] (cit. on p. 161).

Sugiyama, N. (1995). "Cosmic background anistropies in CDM cosmology." In: *Astrophys. J. Suppl.* **100**, p. 281. DOI: 10.1086/192220. arXiv: astro-ph/9412025 (cit. on p. 101).

Tinker, J. L., A. V. Kravtsov, A. Klypin, K. Abazajian, M. S. Warren, G. Yepes, S. Gottlober, and D. E. Holz (2008). "Toward a halo mass function for precision cosmology: The Limits of universality." In: *Astrophys. J.* **688**, pp. 709–728. DOI: 10.1086/591439. arXiv: 0803.2706 [astro-ph] (cit. on pp. 53, 54).

Tinker, J. L., B. E. Robertson, A. V. Kravtsov, A. Klypin, M. S. Warren, G. Yepes, and S. Gottlober (2010a). "The Large Scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests." In: *Astrophys. J.* **724**, pp. 878–886. DOI: 10.1088/0004-637X/724/2/878. arXiv: 1001.3162 [astro-ph.CO] (cit. on pp. 53, 81).

Tinker, J. L., R. H. Wechsler, and Z. Zheng (2010b). "Interpreting the Clustering of Distant Red Galaxies." In: *Astrophys. J.* **709**, pp. 67–76. DOI: 10.1088/0004-637X/709/1/67. arXiv: 0902.1748 [astro-ph.CO] (cit. on p. 99).

Tinker, J. L. and A. R. Wetzel (2010). "What Does Clustering Tell Us About the Buildup of the Red Sequence?" In: *Astrophys. J.* **719**, pp. 88–103. DOI: 10.1088/0004-637X/719/1/88. arXiv: 0909.1325 [astro-ph.CO] (cit. on p. 69).

Torrado, J. and A. Lewis (2021). "Cobaya: Code for Bayesian Analysis of hierarchical physical models." In: *JCAP* **05**, p. 057. DOI: 10.1088/1475-7516/2021/05/057. arXiv: 2005.05290 [astro-ph.IM] (cit. on p. 144).

Tristram, M. et al. (2022). "Improved limits on the tensor-to-scalar ratio using BICEP and Planck data." In: *Phys. Rev. D* **105**.8, p. 083524. DOI: 10.1103/PhysRevD.105.083524. arXiv: 2112.07961 [astro-ph.CO] (cit. on p. 35).

Vallenari, A. et al. (2023). "Gaia Data Release 3 - Summary of the content and survey properties." In: *Astron. Astrophys.* **674**, A1. DOI: 10.1051/0004-6361/202243940. arXiv: 2208.00211 (cit. on p. 171).

Valtchanov, I. E. (2017). *SPIRE Handbook v3.1, Herschel Explanatory Supplement vol. IV*. HERSCHEL-HSC-DOC-0798 (cit. on p. 124).

Viero, M. P. et al. (2013). "HerMES: Cosmic Infrared Background Anisotropies and the Clustering of Dusty Star-Forming Galaxies." In: *Astrophys. J.* **772**, p. 77. DOI: 10.1088/0004-637X/772/1/77. arXiv: 1208.5049 [astro-ph.CO] (cit. on pp. 73, 75, 128).

– (2019). "Measurements of the Cross Spectra of the Cosmic Infrared and Microwave Backgrounds from 95 to 1200 GHz." In: *Astrophys. J.* **881**.2, p. 96. DOI: 10.3847/1538-4357/ab2da0. arXiv: 1810.10643 [astro-ph.CO] (cit. on pp. v, vi, 105, 151).

Warren, M. S., K. Abazajian, D. E. Holz, and L. Teodoro (2006). "Precision determination of the mass function of dark matter halos." In: *Astrophys. J.*

646, pp. 881–885. DOI: 10.1086/504962. arXiv: astro-ph/0506395 (cit. on p. 53).

Watson, W. A., I. T. Iliev, A. D'Aloisio, A. Knebe, P. R. Shapiro, and G. Yepes (2013). "The halo mass function through the cosmic ages." In: *MNRAS* **433**.2, pp. 1230–1245. DOI: 10.1093/mnras/stt791. arXiv: 1212.0095 [astro-ph.CO] (cit. on p. 53).

Wicker, R., M. Douspis, L. Salvati, and N. Aghanim (2022). "Constraining hydrostatic mass bias and cosmological parameters with the gas mass fraction in galaxy clusters." In: *EPJ Web Conf.* **257**, p. 00046. DOI: 10.1051/epjconf/202225700046. arXiv: 2111.01490 [astro-ph.CO] (cit. on p. 83).

Xia, J.-Q., M. Negrello, A. Lapi, G. De Zotti, L. Danese, and M. Viel (2012). "Clustering of sub-millimeter galaxies in a self-regulated baryon collapse model." In: *Mon. Not. Roy. Astron. Soc.* **422**, pp. 1324–1331. DOI: 10.1111/j.1365-2966.2012.20705.x. arXiv: 1111.4212 [astro-ph.CO] (cit. on p. 98).

Zagatti, G., E. Calabrese, C. Chiocchetta, M. Gerbino, M. Negrello, and L. Pagano (2024). "A halo model approach to describe clustering and emission of the two main star-forming galaxy populations for cosmic infrared background studies." In: *Astron. Astrophys.* **692**, A190. DOI: 10.1051/0004-6361/202451424. arXiv: 2406.19069 [astro-ph.CO] (cit. on pp. v, vi, 131, 147).

Zahn, O. et al. (2012). "Cosmic Microwave Background Constraints on the Duration and Timing of Reionization from the South Pole Telescope." In: *ApJ* **756**.1, 65, p. 65. DOI: 10.1088/0004-637X/756/1/65. arXiv: 1111.6386 [astro-ph.CO] (cit. on p. 84).

Zehavi, I. et al. (2005). "The Luminosity and color dependence of the galaxy correlation function." In: *Astrophys. J.* **630**, pp. 1–27. DOI: 10.1086/431891. arXiv: astro-ph/0408569 (cit. on p. 69).

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Final Version as of January 16, 2026 (classicthesis v4.8).